UNCOVERING TRANSCRIPTOMICS RESPONSES AND REGULATIONS DURING FUSARIUM OXYSPORUM – PLANT INTERACTIONS

Houlin Yu
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Recommended Citation
https://doi.org/10.7275/35789052 https://scholarworks.umass.edu/dissertations_2/2926

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
UNCOVERING TRANSCRIPTOMICS RESPONSES AND REGULATIONS DURING *Fusarium oxysporum* – PLANT INTERACTIONS

A Dissertation Presented

by

Houlin Yu

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY

September 2023

Plant Biology
A Dissertation Presented

by

Houlin Yu

Approved as to style and content by:

__________________________________________________________________________________

Li-Jun Ma, Chair

__________________________________________________________________________________

Samuel P. Hazen, Member

__________________________________________________________________________________

Ana L. Caicedo, Member

__________________________________________________________________________________

Antonio Di Pietro, Member

__________________________________________________________________________________

Samuel P. Hazen, Director
Plant Biology Program
DEDICATION

To all the visionary individuals who have dedicated their lives
to scientific discoveries that uplift humanity

And to their beloved ones who have ceaselessly offered unwavering
support throughout their research careers
ACKNOWLEDGMENTS

I am deeply grateful to Dr. Li-Jun Ma for her exceptional guidance, mentorship, passion, confidence, and remarkable approaches to science. Her influence has played an instrumental role in shaping my career, and I am forever indebted to her. I vividly recall the stage I was when I first met her four years ago, and I cannot imagine where I would be today without her support.

I also want to express my heartfelt appreciation to my committee members, Dr. Antonio Di Pietro, Dr. Samuel Hazen, and Dr. Ana Caicedo, for their invaluable support in many fronts and insightful advice throughout my journey. Their expertise and guidance have significantly contributed to the development of my work.

I am sincerely grateful to all the past and present members of the Ma lab. Without their assistance and support, this endeavor would not have been possible. I would also like to thank the Interdepartmental Graduate Programs in Life Sciences and the Department of Biochemistry and Molecular Biology staff for their tireless efforts in ensuring the smooth progress of my research. I also appreciate all my collaborators’ support and assistance on my projects. I want to thank all my friends for your support and encouragement. I want to express my gratitude to my senior mentors who constantly advise me, including but not limited to my M.S. thesis advisor Dr. Elizabeth Schultz and my undergraduate thesis advisor Dr. Jun Chen.

I want to acknowledge the generous support of the Translational Data Science group at Biogen Inc., particularly Dr. Sarbottam Piya, Dr. Jing Zhu, Dr.
Fergal Casey, and Dr. Baohong Zhang, during my cooperative education experience internship. Their backing has been invaluable to my growth and development.

Finally, I want to express my deepest gratitude to my family for their constant love and support. Despite never having the opportunity for any formal education, my grandma, Li, was a continuous presence in my childhood and instilled an intense curiosity and a thirst for knowledge about the world in my heart. My dad, Jisheng, and my mom, Hong, have always embraced my new ideas and have supported every decision I made as I navigated through adulthood. My wife, Le, has become my significant and extremely supportive life partner on all fronts, including the scientific endeavor. I consider myself incredibly fortunate to have her by my side.

Thank you all for your immense contributions and unwavering support!
ABSTRACT

UNCOVERING TRANSCRIPTOMICS RESPONSES AND REGULATIONS DURING FUSARIUM OXYSPORUM – PLANT INTERACTIONS

SEPTEMBER 2023

HOULIN YU, B.S., NORTHWEST A&F UNIVERSITY, YANGLING, CHINA
M.S., THE UNIVERSITY OF LETHBRIDGE, LETHBRIDGE, CANADA
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor LI-JUN MA

_Fusarium oxysporum_ causes devastating wilt diseases in more than one hundred crops. To provide a foundation for developing technologies to enhance plant health, sustain a healthy ecosystem, and feed a continuously growing human population, my dissertation research has focused on uncovering transcriptional responses and regulations during _F. oxysporum_ and plant interactions from fungi and plant host aspects. Through a literature review (Chapter 1), we learned that accessory chromosomes (ACs) and transcriptional regulators are essential for the host-specific virulence of _F. oxysporum_.

Transcription factors, key regulatory elements in the sensory and response networks of these fungi, undoubtedly play a fundamental role in establishing the adaptability of this group. We compared the TFome (Chapter 2), which consists of all TFs found within a genome, to look at a species' evolutionary history of its regulatory mechanisms. The study revealed both the conservation and diversity of _F. oxysporum_ TFome.

The ACs from each _F. oxysporum_ genome likely dictate strain-specific interactions with a particular host. This allows a comparative study that minimizes
genetic differences between strains to address the underlying mechanism that results in distinct phenotypes (e.g., pathogenic vs. non-pathogenic outcomes). We performed a comparative transcriptomics study (Chapter 3) of infection by an endophytic (Fo47) and a pathogenic (Fo5176) strain of *F. oxysporum* in the context of the *F. oxysporum*-Arabidopsis pathosystem, which revealed the transcriptional plasticity of plant defense responses.

*F. oxysporum* penetrates the root epidermis, propagating and moving toward the vasculature. These pathogens' occupation of the vascular system blocks water and nutrient transport, further causing devastating wilt disease. Extensive efforts have been taken by bulk transcriptome profiling to probe the integral plant responses to the *F. oxysporum* stress. However, as the pathogen journeys through multiple layers to establish the infection, different root cell types likely respond differently. We performed a single-nucleus RNA sequencing analysis in the Arabidopsis-*F. oxysporum* pathosystem (Chapter 4). Our research revealed a unique pathogen-induced cell cluster enriched for defense-related functions that could aid in developing new strategies to improve plant defense mechanisms.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
<tr>
<td>CHAPTER 1. ACCESSORY CHROMOSOMES OF THE &lt;em&gt;Fusarium oxysporum&lt;/em&gt;</td>
<td>1</td>
</tr>
<tr>
<td>SPECIES COMPLEX AND THEIR CONTRIBUTION TO HOST Niche Adaptation</td>
<td></td>
</tr>
<tr>
<td>1.1 Fusarium oxysporum: a species complex occupying diverse ecological niches</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Accessory chromosomes: contributors of host-specific pathogenicity among the FOSC</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Determinants of host-specific pathogenicity</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 An expanded kinase family enhances environmental sensing</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Crosstalk coordinates the functions of core and accessory chromosomes</td>
<td>10</td>
</tr>
<tr>
<td>1.3.3 Effectors disarm host defense</td>
<td>12</td>
</tr>
<tr>
<td>1.3.4 Convergent points highlight adaptation to both abiotic and biotic stresses</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Mechanisms of niche adaptation: Genome evolution and maintenance</td>
<td>15</td>
</tr>
<tr>
<td>1.5 Conclusion and perspective</td>
<td>19</td>
</tr>
<tr>
<td>1.5.1 Improving the quality of AC assemblies</td>
<td>19</td>
</tr>
<tr>
<td>1.5.2 Exploring the origin of ACs</td>
<td>20</td>
</tr>
<tr>
<td>1.5.3 Effective regulation of ACs</td>
<td>22</td>
</tr>
<tr>
<td>1.5.4 Novel therapeutic and management strategies</td>
<td>23</td>
</tr>
<tr>
<td>1.6 References</td>
<td>25</td>
</tr>
<tr>
<td>CHAPTER 2. CONSERVATION AND EXPANSION OF TRANSCRIPTIONAL FACTOR REPERTOIRE IN THE &lt;em&gt;Fusarium oxysporum&lt;/em&gt; SPECIES COMPLEX</td>
<td>42</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>42</td>
</tr>
<tr>
<td>2.2 Materials and Methods</td>
<td>45</td>
</tr>
<tr>
<td>2.2.1 Generation of fungal TFomes</td>
<td>45</td>
</tr>
<tr>
<td>2.2.2 RNA-seq analysis</td>
<td>45</td>
</tr>
<tr>
<td>2.2.3 Genome partition</td>
<td>46</td>
</tr>
<tr>
<td>2.2.4 Phylogenetics analysis</td>
<td>46</td>
</tr>
<tr>
<td>2.2.5 Expansion index calculation</td>
<td>47</td>
</tr>
<tr>
<td>2.3.1 FOSC TFome expansion resulted from the acquisition of ACs</td>
<td>47</td>
</tr>
</tbody>
</table>
2.3.2 Conserved TF families that are primarily associated with general/global transcription factors ........................................50

2.3.2.1 Transcription/Translation regulation........................................50
2.3.2.2 Cell cycle control........................................................................51
2.3.3 Gene family contractions in FOSC partially caused by whole genome duplication in yeast ..................................................53

2.3.4 Significant FOSC TFome expansion driven by a few exceedingly expanded TF families ..............................................56
2.3.4.1 Gain-of-Function among Filamentous Ascomycete Fungi..........................................................................................56
2.3.4.2 Seven exceedingly expanded TF families ..................................57
2.3.4.3 Other families..............................................................................59

2.3.5 Orthologous survey of TF families that were manually curated......................................................................................61

2.3.6 Transcriptome analysis to probe the essential TFs during host colonization...............................................................64

2.4 Discussion..............................................................................................66
2.5 References .............................................................................................69

3. METATRANSCRIPTOMIC COMPARISON OF ENDOPHYTIC AND PATHOGENIC FUSARIUM–ARABIDOPSIS INTERACTIONS REVEALS PLANT TRANSCRIPTIONAL PLASTICITY ........................................92

3.1 Introduction.............................................................................................92
3.2 Results.....................................................................................................95

3.2.1 A pathosystem that reveals both endophytic and pathogenic interactions ..........................................................95

3.2.2 Reprogramming of the plant transcriptome in response to a fungal pathogen or endophyte ........................................97

3.2.3 Perturbation of plant immunity ..........................................................102
3.2.3.1 Conserved immune response toward an endophyte and a pathogen........................................................................102

3.2.3.2 Stronger induction of plant PTI responses by the pathogen.............................................................................104

3.2.3.3 Suppressed plant immunity in the presence of the endophyte..............................................................................105

3.2.4 PTI- and ETI-sensing genes induced by the endophyte and the pathogen ..............................................................106
3.2.4.1 RLP and RLK genes ...................................................................107
3.2.4.2 NLR genes ................................................................................108

3.2.5 Accessory chromosomes in two strains harbor genes induced during infection and with distinct biological functions........................................................................................................109

3.3 Discussion..............................................................................................112
3.4 Materials and Methods .........................................................................115
3.4.1 Plant and fungal growth..................................................................115
3.4.2 Plant infection assay .......................................................................115
3.4.3 RNA preparation, sequencing, and data analysis................................116
3.4.4 Functional analysis and visualization.................................118
3.4.5 Synteny and phylogenetic tree construction..............................119
3.5 References ...........................................................................119

4. UNVEILING THE TRANSCRIPTIONAL DYNAMICS OF
ARABIDOPSIS-FUSARIAUM OXYSPORUM INTERACTIONS
THROUGH SINGLE-NUCLEUS RNA SEQUENCING...........................137
4.1 Introduction ...........................................................................137
4.2 Results ..................................................................................141
  4.2.1 Capturing key transcriptomic changes in Arabidopsis
    response to F. oxysporum at bulk and single-nucleus
    resolutions ................................................................................141
  4.2.2 Infection pseudo-time inferred by examining the
    trajectory of nuclei from mature cortex clusters .......................147
  4.2.3 The regulatory networks coordinating development and
    defense .....................................................................................149
  4.2.4 Fungal nuclei displayed distinct stages during infection ......153
4.3 Discussion .............................................................................154
4.4 Materials and Methods ..........................................................157
  4.4.1 Sample collection and library preparation ............................157
  4.4.2 snRNA-seq data processing ...............................................158
  4.4.3 Differential gene expression analysis .................................159
  4.4.4 Functional enrichment analysis .........................................159
  4.4.5 Score functions and network construction .........................160
  4.4.6 Experimental validation ...................................................162
4.5 References .............................................................................169

APPENDIX I - Genome Sequence of Fusarium oxysporum f.
sp. matthiolae, a Brassicaceae Pathogen ......................................175
REFERENCES .............................................................................182
LIST OF TABLES

Table 2.1 Fungal genomes used in this study ........................................87
Table 2.2 Expansion index ($E_l$) of 48 TF families ..................................89
Table 2.2 Ortholog copy number and expansion index ($E_l$) of the characterized and expanded TFs in F. oxysporum .........................91

Table 3.1 Genes that play important roles in danger sensing and signaling suppressed by the endophytic interaction ..................136
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Morphology and plant colonization of <em>Fusarium oxysporum</em> strain Fo5176, phytopathogen of <em>Brassicaceae</em></td>
<td>39</td>
</tr>
<tr>
<td>1.2</td>
<td><em>Fusarium oxysporum</em> genome structure</td>
<td>40</td>
</tr>
<tr>
<td>1.3</td>
<td>Characteristics that differentiate accessory chromosomes from core chromosomes in <em>Fusarium oxysporum</em> (This figure was published in <a href="https://doi.org/10.1094/PHYTO-03-20-0069-IA">https://doi.org/10.1094/PHYTO-03-20-0069-IA</a>)</td>
<td>41</td>
</tr>
<tr>
<td>2.1</td>
<td>Phylogeny of fungal genomes included in this study</td>
<td>83</td>
</tr>
<tr>
<td>2.2</td>
<td>TFome conservation and variation among ascomycete fungi: baseline description</td>
<td>84</td>
</tr>
<tr>
<td>2.3</td>
<td>Evolutinal trajectory of heat shock factors (HSFs) suggesting genome expansion and adaptation</td>
<td>85</td>
</tr>
<tr>
<td>2.4</td>
<td>Unique expansion of some TFs, driven by ACs, may provide clues to host-specific adaptation</td>
<td>86</td>
</tr>
<tr>
<td>3.1</td>
<td>Compatible versus incompatible Arabidopsis interaction with an endophytic (Fo47) versus a pathogenic (Fo5176) <em>Fusarium oxysporum</em> strain</td>
<td>129</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparative genomics reveals unique sets of accessory chromosomes in <em>Fusarium oxysporum</em> Fo47 and Fo5176</td>
<td>130</td>
</tr>
<tr>
<td>3.3</td>
<td>Expression profiling of Arabidopsis roots inoculated with an endophytic versus a pathogenic <em>Fusarium oxysporum</em> strain</td>
<td>131</td>
</tr>
<tr>
<td>3.4</td>
<td>A summary of transcriptomic changes occurring at 12 h postinoculation</td>
<td>132</td>
</tr>
<tr>
<td>3.5</td>
<td>Expression and gene ontology (GO) enrichment of immunity gene clusters</td>
<td>133</td>
</tr>
<tr>
<td>3.6</td>
<td>Distinct biological functions for induced accessory chromosome (AC) genes in the endophyte Fo47 and the pathogen Fo5176</td>
<td>134</td>
</tr>
</tbody>
</table>
Figure 3.7 Model of transcriptomic plasticity in beneficial and antagonistic plant–fungal interactions .......................................................... 135

Figure 4.1 Capturing key transcriptomic changes in Arabidopsis response to *F. oxysporum* at bulk and single-nucleus resolutions ........................................................................................................ 164

Figure 4.2 Infection pseudo-time inferred by examining the trajectory of nuclei from mature cortex clusters ................................................................. 165

Figure 4.3 The regulatory networks coordinating development and defense .................................................................................................................. 166

Figure 4.4 Fungal nuclei displayed distinct stages during infection .......... 168
CHAPTER 1
ACCESSORY CHROMOSOMES OF THE FUSARIUM OXYSPORUM SPECIES COMPLEX AND THEIR CONTRIBUTION TO HOST NICHE ADAPTATION

1.1 Fusarium oxysporum: a species complex occupying diverse ecological niches

The genus Fusarium contains over 300 phylogenetically closely related species that inhabit different ecological niches (Aoki et al. 2014), including plants (Dean et al. 2012; Fones et al. 2020) and humans (O'Donnell et al. 2007). As a cross-kingdom pathogen that causes devastating vascular wilts in many crops and severe infectious diseases in humans, Fusarium oxysporum is one of the most important pathogenic Fusaria (Ma et al. 2013; Michielse and Rep 2009; Ploetz 2015; Edel-Hermann and Lecomte 2019; Pegg et al. 2019; Zhang et al. 2020). While a sexual stage (teleomorph) has been described for some Fusarium species, it has not been observed for F. oxysporum. In the absence of sexual reproduction, these organisms cannot strictly be defined as a biological species. The term F. oxysporum species complex (FOSC) is used to describe all F. oxysporum strains based on morphology and phylogeny using molecular markers (Ma et al. 2013). Members within the FOSC are classified as different forma speciales based on their specific adaptation to particular plant hosts (Armstrong and Armstrong 1981). For instance, strains within F. oxysporum f. sp. lycopersici

---

and *F. oxysporum* f. sp. *cubense* cause wilt disease on tomato (*Solanum lycopersicium*) and banana (*Musa* spp.) plants, respectively. Over 120 *forma speciales* have been described in the FOSC.

As *Fusarium* wilts cause substantial economic losses in many economically important crops, including potato (*Solanum tuberosum*), sugarcane (*Saccharum* spp.), bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*), banana, and date palms (*Phoenix dactylifera*), *F. oxysporum* is considered to be among the top five most economically important plant pathogens (Dean et al. 2012). The fungus impairs the vascular transport of nutrients and water in the infected plant, causing wilting and, eventually, death (Gordon 2017; Husaini et al. 2018).

A common reason for the devastation of *Fusarium* wilts is the lack of effective means to remove the pathogen from infected fields. The fungus produces three types of asexual spores: microconidia (Figure 1.1A-C), macroconidia, and chlamydospores. All of these spore types contribute to the persistence of the disease (Gordon 2017; Husaini et al. 2018). Microconidia or propagules are abundantly produced and can be found in soil. Macroconidia are usually formed on plant stems or leaves. Thick-walled resistant chlamydospores are produced upon the exhaustion of nutrients and can survive in the soil for many years (Nelson 1981).

One of the most severe epidemics of agricultural crops to date involved *Fusarium* wilt of banana (Viljoen et al. 2020), historically referred to as Panama disease. In the 1950s, millions of ‘Gros Michel’ banana plants were destroyed by
race 1 of *F. oxysporum* f. sp. *cubense* (*Foc*) in Central America (Stover 1962). Almost a century later, a new *Foc* race, tropical race 4 (TR4), was detected as the causal agent of *Fusarium* wilt in ‘Cavendish’ bananas, the *Foc* race 1-resistant cultivar that revived the Central American banana export industry and became the most economically important banana cultivar in global markets (Ploetz 2015). TR4 was first described in Southeast Asia (Boehm et al. 1994) and has spread rapidly in Australia (Shivas et al. 1995), Africa (Visser et al. 2010), the Middle East (Maymon et al. 2018), Europe (Özarslandan and Akgül, 2020), and Latin America (Garcia-Bastidas et al. 2020). The history and ongoing problems of this disease illustrate the ability of *F. oxysporum* to overcome host resistance and adapt quickly to changing environments (Martin et al. 2006; Seidl and Thomma 2014).

Even though there are more reports on plant-pathogenic FOSC strains, nonpathogenic *F. oxysporum* strains, including some biocontrol agents, are also prevalent in nature. For instance, a comprehensive study of *Fusarium* wilt suppressive soils from the Châteaurenard region in France established a novel approach to control *Fusarium* wilt (Alabouvette 1999). Similarly, the biocontrol *F. oxysporum* strain CS-20 was reported to induce host defense and reduce wilt symptoms in tomato plants (Shcherbakova et al. 2016).

In addition to being plant pathogens, some *F. oxysporum* strains are also human pathogens and can cause severe skin, nail, or disseminated infections, known as fusariosis (O'Donnell et al. 2007; Zhang et al. 2020). Much more invasive than other infectious fungal pathogens, the infective agent can be
detected in blood samples among approximately 50% of fusariosis patients (Nucci and Anaissie 2007). Because *Fusarium* spp. are broadly resistant to most of the clinically available antifungals, fusariosis has a high mortality rate (Scheel et al. 2013; Nucci et al. 2014; Prajna et al. 2016). Indeed, a 100% mortality rate was reported among persistently neutropenic fusariosis patients (Boutati and Anaissie 1997; Nucci and Anaissie 2002). *Fusarium* species were also reported as the most common causative agents of fungal keratitis in India (Lalitha et al. 2015; Hassan et al. 2016), China (Wang et al. 2009; He et al. 2011), South Africa (O'Sullivan et al. 1997), and Brazil (Ibrahim et al. 2009). Collectively, fusariosis has been listed as an emerging fungal disease and the second most common opportunistic mold infection after aspergillosis (Guarro 2013; Nucci et al. 2014).

1.2 Accessory chromosomes: contributors of host-specific pathogenicity among the FOSC

The broad host range and host-specific pathogenicity of the FOSC have been attributed to the existence of diverse and horizontally transmitted accessory chromosomes (ACs) that result in highly dynamic and compartmentalized genomes (Ma et al. 2010, 2013; Rep and Kistler 2010; Kistler 2013; van der Does et al. 2012) (Figure 1.2). In contrast to core chromosomes (CCs), which are conserved among all *F. oxysporum* genomes, ACs are highly variable (Ma et al. 2010, 2013). First reported in the early 1990s, fungal ACs were also referred to as supernumerary chromosomes, conditionally dispensable chromosomes, and lineage-specific chromosomes (Croll and McDonald 2012; Croll et al. 2013; Ma
2014; Bertazzoni et al. 2018; Yang et al. 2020) and were suggested to be horizontally transferred (Masel et al. 1996; Akagi et al. 2009), as opposed to the vertical transmission from parent to offspring.

The presence of ACs in *F. oxysporum* was solidified in a 2010 study examining the genome of the tomato pathogen *F. oxysporum* f. sp. *lycopersici* strain 4287 (Fol4287) (Ma et al. 2010). Four ACs lacking homologous counterparts in two closely related species, *F. graminearum* and *F. verticillioides*, were reported in Fol4287. This study also demonstrated the discordance in phylogenetic relationships of genes between CCs and ACs and further tested the hypothesis that ACs are transferred horizontally. The accessory sequences of *F. oxysporum* can exist as separate chromosomes (*i.e.*, ACs) or can be attached to CCs. The Fol4287 genome contains four separate ACs and additional accessory sequences are located at the ends of CCs, including chromosomes 1 and 2. In the genome of *F. oxysporum* f. sp. *cubense* TR4, all accessory sequences are attached to the ends of CCs (Zhang 2019).

The existence of ACs and their contribution to fusariosis were reported in the genomes of two human-pathogenic *F. oxysporum* isolates, NRRL 32931 and NRRL 47514, a clinical strain isolated from a leukemia patient, and a strain associated with the USA 2005/06 *Fusarium* keratitis outbreak, respectively (Zhang et al. 2020), 10 years after the initial report of ACs in a *F. oxysporum* plant-pathogenic strain (Ma et al. 2010).

ACs were reported to determine host-specific pathogenicity among plant-infecting *F. oxysporum* isolates, including other *F. oxysporum* f. sp. *lycopersici*
strains (Vlaardingerbroek et al. 2016), *F. oxysporum* f. sp. *melonis* (Schmidt et al. 2016), *F. oxysporum* f. sp. *cepae* (Armitage et al. 2018), *F. oxysporum* f. sp. *radicis-cucumerinum* (Van Dam et al. 2017), legume-infecting strains (Williams et al. 2016), and an Arabidopsis-infecting strain (Thatcher et al. 2012) (Figure 1.1D). Experimentally, horizontal transfer of whole or partial ACs was confirmed in Fol4287 and the *F. oxysporum* f. sp. *radicis-cucumerinum* isolate Forc016 (Ma et al. 2010; Van Dam et al. 2017; Li et al. 2020). A direct link between ACs and pathogenicity was documented in the lab by introducing pathogenicity into the non-pathogenic strain Fo47 after the migration of one of the ACs, the Chromosome 14, from Fol4287 (Ma et al. 2010; van der Does and Rep 2012). A later study demonstrated that only part of the small arm of Chromosome 14 was sufficient to induce virulence symptoms (Li et al. 2020). The observation that these ACs can be horizontally transferred explains the polyphyletic nature of *forma specialis* groups (O’Donnell et al. 1998; Katan 1999; Alves-Santos et al. 1999; Abo et al. 2005).

Characterized by a low gene density and high repeat content, ACs are enriched in diverse transposable elements (TEs). In the Fol4287 genome, 74% of all transposons occur in ACs, and Fol4287 ACs contain 95% of the DNA transposons (Ma et al. 2010). High repeat content may contribute to the plasticity of ACs and explain the high frequency of large chromosomal duplications and partial or complete chromosomal losses among ACs (Ma et al. 2010; Vlaardingerbroek et al. 2016; Li et al. 2020). For example, the genome of
Fol4287 has a large segmental duplication between Chromosomes 3 and 6 (Ma et al. 2010).

In addition, CCs and ACs show distinct epigenetic patterns (Fokkens et al. 2018). In the Fol4287 genome, CCs are enriched in H3 lysine 4 dimethylation (H3K4me2), a histone marker for euchromatin, in non-centromeric and non-subtelomeric sequences. By contrast, ACs are enriched in the heterochromatic signal H3 lysine 27 trimethylation (K3K27me3). Furthermore, the differences in histone modifications correspond to differences in transcriptional regulation. H3K27me3-associated regions show lower transcript levels, whereas H3K4me2-enriched regions are characterized by high gene density and exhibit higher transcript levels than the rest of the genome (Fokkens et al. 2018), as described in other filamentous fungi (Connolly et al. 2013; Soyer et al. 2014; Galazka and Freitag 2014).

The presence of ACs and CCs compartmentalizes the *F. oxysporum* genome. While CCs share homologous chromosomes in sister species, ACs may have size and sequence polymorphisms and disrupt the chromosomal collinearity even among isolates of the same forma speciales (Bertazzoni et al. 2018; Yang et al. 2020), introducing different evolutionary speeds within the same species genome. Fokkens et al. (2018) revealed an even more complex "multi-speed" genome structure by identifying AC patterns on the three smallest CCs, termed ‘fast-core chromosomes’. These chromosomes have similar gene densities and repeat contents as the other CCs but are on a faster evolutionary track (Fokkens
et al. 2018). These findings give rise to many intriguing questions about the evolvability and function of the epigenetic machinery in *F. oxysporum*.

### 1.3 Determinants of host-specific pathogenicity

To explore how members of the FOSC interact with a broad range of hosts, researchers have examined the virulence-related genes that contribute to pathogenicity and host specificity, the interplay between CCs and ACs, and the regulation across these two compartments within the same genome (Ma et al. 2013; Yang et al. 2020; Zuriegat et al. 2021). Below, we discuss the genomic and genetic characteristics of *F. oxysporum*, focusing on the kinases that mediate the signal transduction involved in host recognition, transcription factors (TFs) that are activated by the signal transduction pathways and then regulate other genes, and effectors produced in response to host signals (Figure 1.3).

#### 1.3.1 An expanded kinase family enhances environmental sensing

Kinases are essential components of host–fungal interactions, playing key roles in host sensing and signal transduction and mediating an effective response to the sensed stimuli. A comprehensive survey of the *F. oxysporum* kinome (*i.e.*, a complete set of kinases encoded in the genome) revealed an expansion of kinase proteins among members of the FOSC, facilitated in part by the presence of ACs (Delulio et al. 2018). This study also highlighted 99 “core” kinase families highly conserved among ascomycetes fungi.

The “core” kinase families include the mitogen-activated protein kinase
(MAPK) signaling pathway, one of the most evolutionarily conserved eukaryotic cellular mechanisms of extracellular information perception and transduction. In fungi, the MAPK pathway is involved in mating, sporulation, cell cycle, morphogenesis, cell wall integrity, autophagy, cell–cell interactions, fungus–host interactions (neutral or beneficial), the response to stress stimuli (such as the damage-associated molecular pattern (DAMP) response), the DNA damage response (Milo-Cochavi et al. 2019), and pathogenesis (reviewed by Martínez-Soto and Ruiz-Herrera 2017). Three *F. oxysporum* MAPKs, Mpk1, Hog1, and Fmk1, have been functionally characterized. Mpk1 and Hog1 are involved in the response to stress, colonization, and virulence, while Fmk1 is required for host adhesion and penetration (Segorbe et al. 2017). Mpk1 is the key element in the chemotropic response pathway of *F. oxysporum* (Turrà et al. 2015), as it senses nutrients of root exudates and responds rapidly by promoting germination and guiding the germ tubes and hyphae toward the stimulus (Turrà and Di Pietro 2015; Turrà et al. 2015).

The expanded *F. oxysporum* kinome includes atypical kinases and histidine kinases. One expanded kinase family is the target of rapamycin (TOR) kinase belonging to the atypical kinase family and is a key regulator involved in nutrient sensing, the cell cycle, survival, growth, and fungal development. The expansion of the histidine kinase family is equally interesting, as they function in signaling pathways that sense the condition of the external environment (Loewith and Hall 2011; Delulio et al. 2018). The observed kinome expansions among members of the FOSC appear to have equipped the pathogens to survive in
diverse hosts. The comparative kinome analysis among FOSC suggests a convergent evolution that shapes individual *F. oxysporum* isolates with an enhanced and unique capacity to perceive the environment and activate the associated downstream responses (Delulio et al. 2018).

### 1.3.2 Crosstalk coordinates the functions of core and accessory chromosomes

Signaling pathways are part of a complex and holistic network, involving the interconnected activation or deactivation of master regulators (*e.g.* TFs) and the subsequent induction or repression of genes (*e.g.* effectors) in response to the environment, such as plant stimuli (Husaini et al. 2018). Even in a compartmentalized genome, *F. oxysporum* ACs and CCs do not function as two autonomous entities. Rather, crosstalk between these two compartments is likely to be coordinated by mediators such as TFs (Michielse et al. 2009; Yang et al. 2020), with a key set of TFs functioning during host colonization in *F. oxysporum* (Guo et al. 2014; 2021).

One example documenting the importance of host-specific pathogenicity accomplished through such crosstalk is the expression of *SIX* (*Secreted In Xylem*) effector genes located in the ACs of *F. oxysporum* f. sp. *lycopersici*. The expression of *SIX* effector genes is directly regulated by the TF *Sge1* (*SIX gene expression 1*), which is located on a CC (Michielse et al. 2009). The expression of *Sge1* is, in turn, regulated by the AC-encoding TF *Ftf1* (*Fusarium TF1*) (van der Does et al. 2016; Niño-Sánchez et al. 2016).
Another expanded TF family is *Ebr* (*Enhanced Branching*), which regulates general metabolism and virulence (Jonkers et al. 2014; Zhang et al. 2020). Three paralogous copies of this gene, *Ebr2, Ebr3*, and *Ebr4*, are encoded in ACs and regulated by *Ebr1*, which is encoded in the core genome. Knocking out the core *Ebr1* gene reduced pathogenicity and resulted in growth defects (Jonkers et al. 2014), suggesting the importance of this ortholog. However, the AC encoding *Ebr2* rescues the *Ebr1* knockout mutation when controlled by the *Ebr1* promoter, indicating some functional redundancy among members of this family.

In addition to *Ftf* (Niño-Sánchez et al. 2016) and *Ebr* groups (Jonkers et al. 2014; Zhang et al. 2020), an expansion of the alkaline pH-responsive TF *PacC/Rim1p* was observed in the human-pathogenic strains NRRL 32931 (isolated from a leukemia patient) and NRRL 47514 (a strain associated with the *Fusarium* keratitis outbreak) (Zhang et al. 2020). These expansions are expected to play a role in the fungus’s response to high pH and other extreme environmental conditions of the human host (Zhang et al. 2020). Similar to the kinases, an expansion of TFs facilitated in part by the acquisition of ACs was also observed, which may suggest that regulatory networks involved in environmental adaptation are fine-tuned to support complex cross-kingdom interactions (Ma lab, unpublished).

Although roughly 5% of the genes in the *F. oxysporum* genome encode TFs (Ma et al. 2010), only 26 TFs have been functionally characterized in *F. oxysporum* (Zuriegat et al. 2021). Further research is much needed.
1.3.3 Effectors disarm host defense

Plants have sophisticated defense mechanisms against pathogen invasion (Glazebrook 2005; Ponce de León and Montesano 2013; Nishad et al. 2020). For successful colonization, a phytopathogen must break down the plant cell wall and suppress plant immunity. Secreted fungal effectors are major players at the plant–microbe interface (De Wit et al. 2009; Lo Presti et al. 2015), contributing to processes ranging from host recognition to intracellular invasion of the plant tissue (Stergiopoulos and de Wit 2009; Giraldo and Valent 2013).

Members of the FOSC employ effectors throughout the course of infection, including during penetration of the plant root epidermis, propagation through the cortical cell layers, and eventually during colonization of the root vasculature. Many effectors are encoded by ACs. For instance, 12 out of 14 SIX proteins, the first set of effectors described in *F. oxysporum* (Rep et al. 2004), are encoded by genes localized on Chromosome 14, an AC (Ma et al. 2010; Schmidt et al. 2013). Among them, SIX1, SIX3, SIX5, and SIX6 confer full virulence to Fol4287 (Rep et al. 2004; Ma et al. 2013; Gawehns et al. 2014), and SIX1 suppresses plant immunity (Tintor et al. 2020). Another important group of effectors in the FOSC are enzymes involved in the degradation of plant compounds (such as carbohydrate-active enzymes, *i.e.*, CAZymes) (Ma lab, unpublished). Widely dispersed in the fungal world, particularly in phytopathogens (Zhao et al. 2013), CAZymes play crucial roles in degrading plant compounds during appressorial penetration, nutrient uptake, and plant
tissue colonization (Stergiopoulos and de Wit 2009; Dodds and Rathjen 2010; Giraldo and Valent 2013).

Pan-genomic analyses of all predicted FOSC effector genes revealed tremendous effector diversity (Ma Lab, unpublished; van Dam et al. 2016; Constantin et al. 2021) and suggested that the effector gene repertoire in each formae speciales is directly related to the fungal lifestyle. For example, genomes of *F. oxysporum* plant pathogens such as Fol4287 and Fo5176 maintain genes encoding more SIX effectors and CAZymes than endophytic and human-pathogenic strains (Ma lab, unpublished).

Candidate effectors can be predicted based on their small size (typically <300 amino acids in length), the presence of a secretion signal peptide, and enrichment for cysteine residues. Based on this simple definition, *F. oxysporum* effector genes can be found in both CCs and ACs. Effectors encoded by CCs are conserved in both pathogenic and endophytic *Fusarium* strains. A recent study demonstrated that CC-encoded effectors determine endophytic growth and multi-host plant compatibility for both endophytic and pathogenic *F. oxysporum* (Redkar et al. 2021), further reinforcing the notion that AC effectors specifically determine host-specific pathogenicity. Additional experimental work is needed to decipher the functions of different effectors.

### 1.3.4 Convergent points highlight adaptation to both abiotic and biotic stresses

Studies of *F. oxysporum*, a cross-kingdom pathogen, are usually focused on fungal pathogenesis. However, as a cosmopolitan fungus, members of the
FOSC have been isolated from diverse ecological niches, including soil, air, plants, animals, and even the International Space Station (Urbaniak et al. 2019; Schuerger et al. 2021). Signaling pathways involved in the fungus’s response to biotic and abiotic stresses are distinct, with adaptation to abiotic stresses influencing how microbes interact with their hosts. Possessing a remarkable capacity to adapt to changing environmental stresses, members of the FOSC could be a model for exploring both biotic and abiotic stress signaling pathways and the intersection between them. Here, we summarize some research on temperature and pH adaptation, both of which are tightly controlled physiological parameters that can be considered to be abiotic stresses.

Significant shifts in phenotypic traits and expression profiles were observed when a plant-pathogenic strain of *F. oxysporum* was subjected to high temperature (37°C) (Segorbe et al. 2017). By contrast, this temperature did not affect the growth rate of the human pathogen *F. oxysporum* MRL8996 (NRRL 47514) (Zhang et al. 2020; Ayhan 2021). The expression profile shift of this human pathogen in response to high temperature is different from that of the plant-pathogenic strain (Ma lab, unpublished).

The genome of the human-pathogenic *F. oxysporum* strain NRRL 32931 contains four ACs with 812 genes that are significantly enriched in genes encoding metal ion and cation transporters (Zhang et al. 2020). These genes could be important for overcoming human nutritional immunity by sequestering trace minerals to limit microbial infection. The genome of this fungus contains additional copies of the TF encoding *PacC/Rim1*, which is involved in the response to pH
changes and other stresses (Peñalva and Arst, Jr. 2004; Zhang et al. 2020). Both phenotypic and genotypic observations support the importance of genetic weapons gained by the FOSC, perhaps via ACs, in the fungus’s response to different environmental stresses. We anticipate that this topic will be investigated further using diverse approaches. One such approach is experimental evolution, which enables adaptation trajectories to be observed in a controlled environment (Garland 2009). This approach has been successfully applied to study microbial adaptation under in vitro (Barrick and Lenski 2013) and in vivo (Lescat et al. 2017) conditions, and provides an opportunity to study the crosstalk between microbial responses to biotic and abiotic stresses. For instance, this approach could be used to analyze the adaptation of an isolate to elevated temperature and pH, which is thought to facilitate opportunistic infection through the bloodstream of humans.

1.4 Mechanisms of niche adaptation: Genome evolution and maintenance

Natural selection of genetic variations is the basis of evolution. By monitoring biological processes such as mutagenesis, selection, adaptation, and speciation, biologists predict patterns that give rise to diverse organisms and track changes in these patterns over time. Genome plasticity is a driving force in the "arms race" between a pathogen and its host. The capacity to colonize more than 100 different hosts highlights the genomic plasticity of FOSC members.

Meiotic recombination is a well-known process underlying genomic variation. With the increasing availability of sequenced genomes, it has become clear that asexual species, such as *F. oxysporum*, also show a high degree of
genetic variability (Takken and Rep 2010; Karasov et al. 2014; Perez-Nadales et al. 2014; Gladieux et al. 2018). In addition to the horizontally transmitted ACs, there is tremendous interest in studying the evolvability of members within the FOSC that operate under the model of predominant clonal evolution. Such studies mostly involve inspecting patterns of mutation contributed by single nucleotide variations (SNVs), insertions and deletions (INDELs), transposon insertion and excision variations, large segmental deletions, duplications, and translocations in the genome (Ayhan 2021).

Exposure to DNA-damaging agents triggers a range of stress-related responses across the tree of life (Fry et al. 2005). SNVs occur when damaged DNA is not completely repaired either before or after DNA replication. Several conserved DNA repair mechanisms are responsible for repairing the damage and maintaining genome stability. DNA damage repair in fungi has mostly been studied in *Saccharomyces cerevisiae* (Workman et al. 2006; Smolka et al. 2007; Shalem et al. 2008; Fu et al. 2008; Bandyopadhyay et al. 2010; Guénolé et al. 2013). Even though many pathways are shared among yeast and filamentous fungi, there are important differences. Several DNA-damaging agents, such as methyl methanesulfonate (MMS), UV radiation, and hydroxyurea were used to study the processes involved in DNA damage tolerance and repair capacity in the FOSC (Milo-Cochavi et al. 2019).

MMS exposure induces some strong and shared responses in *S. cerevisiae* and *F. oxysporum*, including the activation of the Chk1-Chk2 signal transduction pathway, proteasome components, and the Xbp1-Yap1 networks.
(Milo-Cochavi et al. 2019). However, some clear differences were reported in a recent study (Milo-Cochavi et al. 2019), including unique upregulation of genes encoding the splicing module, the basic transcription machinery, and several RNA pol II-associated proteins in *F. oxysporum* but not in *S. cerevisiae*. This study suggested that *F. oxysporum* responded to the chronic DNA damage by recycling the transcription machinery in response to stalled RNA polymerases and consequently reactivated the splicing machinery (Milo-Cochavi et al. 2019). The other observed difference was the regulation of ribonucleotide reductase (RNR), a key enzyme that mediates the synthesis of deoxyribonucleotides (dNTPs) in response to DNA damage. In contrast to *S. cerevisiae* and other organisms, *F. oxysporum* did not upregulate the expression of RNR. Accordingly, dNTP pools in *F. oxysporum*, but not in yeast, were decreased in response to the DNA damage agent hydroxyurea (Cohen et al. 2019).

UV radiation is probably the most common toxic environmental mutagen. Genes involved in repairing UV-induced DNA damage are important for reducing premutagenic lesions in DNA and transcription errors. A recent study (Milo-Cochavi et al. 2019) reported that *F. oxysporum* responded to UV radiation through a developmentally regulated oscillation in the expression of the UV repair genes photolyase (*Phr1*) and UV endonuclease (*Uvde*). Both gene products can specifically bind to lesions caused by UV radiation. At the early stages of germination, *Phr1* expression was induced, while *Uvde* expression was reduced. The trend was reversed at 14 hours post-inoculation when spores were fully germinated and filaments were established (Milo-Cochavi et al. 2019). This
observation led to the hypothesis that *F. oxysporum* operates a photolyase-based, transient, and precise UV repair machinery, and that the fungus only actives this machinery upon UV exposure to minimize the cost of transcription when it is not necessary (Milo-Cochavi et al. 2019). We should note that most of the current knowledge about DNA damage repair in *Fusarium* is based on patterns of gene expression. While induction of a certain gene by DNA damage may indicate a functional role in damage tolerance and survival, other processes, mostly post-translational modifications such as phosphorylation, neddylation, ubiquitination, SUMOylation, and PARylation, may also play important roles and should be comprehensively studied.

Chromatin state also has a crucial role in DNA damage and repair (Dabin et al. 2016; Stadler and Richly 2017; Allshire and Madhani 2018). The heterochromatic as well as transposon- and repeat-rich environment of *F. oxysporum* ACs may affect their mutation rate and repair efficiency and have either beneficial or adverse effects on the genomic plasticity required for genetic changes associated with plant–fungal interactions.

In addition to SNVs, large- and small-scale genome rearrangements and TEs can also promote rapid adaptation to new environments. Transposons enriched among ACs contribute significantly to fungal pathogenicity. For instance, the insertion of a Hornet-like DNA transposon into the *SIX3* effector gene enables the pathogenic *F. oxysporum* f. sp. *lycopersici* to evade plant immunity and evolve into a new disease-causing race (Inami et al. 2012). Members of another DNA transposon family, the miniature Impala elements or
MIMPs, are associated with promoter regions. Miniature Fot5 associates with the downstream SIX genes in an AC, Chromosome 14 of Fol4287 (Schmidt et al. 2013). Similarly, Helitrons are found upstream of SIX9 homologs in the Arabidopsis-pathogenic strain Fo5176 (Chellapan et al. 2016).

How these mechanisms impact on mutational patterns in an organism, particularly asexual fungi, needs to be investigated further. These data, which have been mostly obtained from observations of only a few F. oxysporum strains, should be expanded to other fungal relatives to reveal what aspects are genus-, species-, or even strain-specific.

1.5 Conclusion and perspective

The presence of fungal ACs and their contribution to fungal phytopathogenicity are widely accepted (Ma et al. 2010; Croll and McDonald 2012; Bertazzoni et al. 2018). A recent study provided clear evidence of the link between fungal ACs and adaptation to human host conditions (Zhang et al. 2020), extending the contribution of fungal ACs to adaptation beyond plant hosts. There is still much to learn about the genome dynamics of ACs. Here, we discuss a few unresolved questions.

1.5.1 Improving the quality of AC assemblies

Due to the high level of repetitive sequences, ACs are rarely assembled to the chromosomal level when sequenced. A strategy incorporating long-read sequencing, such as PacBio and Nanopore reads, with high-throughput chromosome conformation capture (Hi-C), has been successful in producing
chromosomal scale assemblies of the *F. oxysporum* strains Fo47 and Fo5176 (Wang et al. 2020; Fokkens et al. 2021). An alternative strategy is to sequence one chromosome at a time (Peng et al. 2019). Single ACs can be separated using contour-clamped homogeneous electric field electrophoresis and then extracted from the gel (O’Brien et al. 2006) or flow-sorted from the genomic DNA pool (Jain et al. 2016).

### 1.5.2 Exploring the origin of ACs

The evolutionary origin of ACs is an intriguing puzzle for the fungal community. Horizontal transfer is one widely accepted hypothesis for the acquisition of fungal ACs, and conidial anastomosis during germination is postulated to mediate conidial fusion and facilitate intra- and interspecies genetic exchange (Gabriela Roca et al. 2005; Mehrabi et al. 2011). Mobile genetic elements, such as plasmids, can be horizontally transferred. The mobility of ACs suggests that they have a plasmid-like nature. The known sizes of fungal ACs are mostly below 2 Mb, and plasmids with insert sizes larger than 1 Mb have also been reported (Finan et al. 1986; Harrison et al. 2010), which suggests that fungal ACs may be novel mobile genetic elements. However, ACs have a linear structure, with a centromere and telomeric sequences (Fokkens et al. 2021), rather than a circular structure as is typical of plasmids. To test this hypothesis, a detailed survey might be necessary.

ACs can also be generated from CC genome duplication, and partial loss (possibly different processes are involved back and forth). Active links are observed between ACs and the telomeres of CCs, as reflected by their shared
enrichment of effector genes and other genes involved in host invasion (Ma and Xu 2019; Peng et al. 2019), supporting the alternative hypothesis that ACs can originate from the nondisjunction of duplicated chromosomes (e.g., CCs) followed by mutations and/or degradation (Croll et al. 2013). Of course, we cannot exclude the possibility that different mechanisms may contribute to different ACs.

Many questions remain to be explored, such as: Why is the FOSC so versatile in acquiring extra genetic materials? Are there critical factors that determine the fate of an AC once it emerges in a new genomic landscape? Do evolutionary events (e.g., changes in the ability to maintain chromosome stability) underly the origin/acquisition of ACs? The distinct genomic nature, as well as differences in the rate of evolution between CCs and ACs, suggest that different mechanisms maintain these types of chromosome. Does the heterochromatic structure of ACs affect the recruitment and accessibility of repair factors to the damage site? Do the same repair factors function across the genome, or do they differ by type (e.g., CCs vs. ACs)? Do CCs and ACs show different DNA repair efficiency, and if so, how does this affect the mutation rate and evolution of the ACs?

Mutagenesis experiments coupled with experimental evolution, genetics, and genomic studies of DNA repair of different damage types will shed light on this intriguing aspect. In addition, constructing a F. oxysporum pan-genome that incorporates a substantial number of high-quality genome assemblies, or a comprehensive comparative analysis of ACs belonging to the same formae
speciales, will be necessary to track the evolutionary origin and trajectory of ACs.

1.5.3 Effective regulation of ACs

There seems to be a proximal association between the expansion of TFs and their target genes. For instance, *Ftf1* genes are close to clusters of *SIX* genes on the Fol4287 pathogenicity-conferring chromosome (Schmidt et al. 2013). de Vega-Bartol *et al.* (2011) postulated that the different number of TFs and effector genes may be related to the fine-tuning of the host-specific infection process, and this may be achieved through copy number and sequence variations of *Ftf1* binding sites on *SIX* gene promoters (Zhao *et al.* 2020). This possibility remains to be tested, preferably by chromatin immunoprecipitation sequencing (Park 2009) or DNA affinity purification sequencing (Baumgart *et al.* 2021), which would not only reveal the actual binding sites but also identify the genome-wide targets.

Crosstalk between core and accessory components of the genome is critical for the coordinated response to environmental changes and the switch between vegetative growth and pathogenicity. What is the relationship between the regulatory roles of CC and AC TFs? Do AC TFs regulate core genes, and do they compete for the same binding sites? What determines the expansion and contraction of gene regulators during evolution? Aside from TFs, protein kinases are also largely expanded in ACs of *F. oxysporum*. How do CC kinases regulate the biological activity of AC-encoding proteins, or vice versa? Undoubtedly, the relationship and crosstalk between TFs and protein kinases in ACs and CCs will impact the global regulatory networks of the fungus.
1.5.4 Novel therapeutic and management strategies

The severity of the problem caused by this cross-kingdom pathogen underscores the urgency in developing novel therapies and disease management strategies. It is of great interest to explore novel targets, as most marketed antifungal agents have the same mode of action, perturbing cell membrane integrity.

Loss of epigenetic marks can potentially change the transcriptional program and consequently affect many aspects of the life and disease cycle of pathogenic fungi, including resistance to stresses and pathogenicity. A link between histone acetylation, antifungal resistance, pathogenicity, and stress tolerance was reported in several important fungal pathogens, such as *Candida albicans* and *Cryptococcus neoformans* (Robbins et al. 2012; Lamoth et al. 2015; Li et al. 2015; Garnaud et al. 2016; Brandão et al. 2018). Several histone deacetylase (HDAC) inhibitors also showed antifungal activity, particularly when they were combined with existing fungicides (Mai et al. 2007; Pfaller et al. 2009; Lamoth et al. 2015; Brandão et al. 2018). A recent study showed a potential connection between epigenetics and metabolism in *F. oxysporum*, focusing on Sirtuin, a type of HDAC and one of the primary consumers of cellular NAD$^+$. Disrupting the NAD$^+$ biosynthesis salvage pathway inhibited growth and reduced the biomass of *F. oxysporum*, revealing a potential approach to develop novel antifungals (Anand et al. 2019).

The absence of a specific DNA repair enzyme from a specific phylogenetic group of fungi or individual species may indicate that members of this group are
more vulnerable to DNA damage agents. A comparative genomic approach, such as the one performed by Milo et al. (2019), can be a starting point for designing functional experiments to identify synthetic lethal interactions that can be further exploited for developing sophisticated new fungicides. The unorthodox behavior of RNR upon DNA damage in one strain of *F. oxysporum* may suggest that this fungus has increased sensitivity to RNR inhibition; therefore, RNR could be a promising candidate for the development of species-specific pesticides (Cohen et al. 2019).

Comparative systems, *e.g.*, of endophytic and pathogenic *Fusarium*, have illustrated the power of distinguishing the key events associated with fungal pathogenicity and plant host defense (Guo et al. 2021). Additional comprehensive omics approaches, such as proteomics and metabolomics, will provide valuable insight into host–pathogen interactions. Different post-translational modifications, such as phosphorylation, glycosylation, and ubiquitination, that are related to alterations in both the pathogen and host proteome can also be elucidated. In plants, pathogenic *F. oxysporum* travels through many cell layers, starting at the epidermis and moving through the cortex, endodermis, and root vasculature to establish the infection. Investigating gene transcription regulation at single-cell resolution in response to the fungal pathogen is of high research interest and is now feasible through single-cell RNA sequencing (Plant Cell Atlas Consortium et al. 2021; Cole et al. 2021). Collectively, a better understanding of fungal pathogenicity and host defense
mechanisms will be the key to controlling these infections. Technological advancements mentioned about will make this endeavor feasible.

1.6 References


Fokkens L, Guo L, Dora S, et al (2021) A chromosome-scale genome assembly for the *Fusarium oxysporum* Strain Fo5176 to establish a model *Arabidopsis*-fungal pathosystem. G3 (Bethesda) 10:3549-3555. https://doi.org/10.1534/g3.120.401375


Stover RH 1962. Fusarial wilt (Panama disease) of bananas and other *Musa* species. Commonwealth Mycological Institute, Kew, Surrey, UK.


Figure 1.1 Morphology and plant colonization of *Fusarium oxysporum* strain Fo5176, phytopathogen of *Brassicaceae*.

(A-C) Germination of microconidia. (A) Bright field; (B) Fluorescence of fungal cell wall stained by Calcofluor white (excitation at 380 nm and emission at 475 nm); and (C) Merge. (D) Fo5176 colonization of a lateral root xylem of *Arabidopsis thaliana* at 5 days post-infection. Fungal structures (shown in green) were stained with WGA-Alexa Fluor 488 and detected with excitation at 488 nm and emission at 500 to 540 nm. Scale bar, 20 µm.
Figure 1.2 *Fusarium oxysporum* genome structure.

This diagram demonstrates the genome compartmentation of *F. oxysporum*. Core chromosomes contain genes that are highly conserved and fulfill all essential housekeeping functions. Accessory chromosomes contain genes that are lineage- or strain-specific and often perform functions related to adaptation to a unique host (e.g. host specific virulence). Active cross-talks are going on between core and accessory chromosomes by molecular regulations (MRs).
Figure 1.3 Characteristics that differentiate accessory chromosomes from core chromosomes in *Fusarium oxysporum* (This figure was published in https://doi.org/10.1094/PHYTO-03-20-0069-IA).

This diagram compares a small fraction of a single core chromosome with a complete accessory chromosome, demonstrating that core chromosomes are much richer in genetic content than accessory chromosomes. (A) Functional division of coding sequences. While core genes, which are highly conserved and fulfill all essential housekeeping functions, are vertically transmitted in diverse taxa, accessory genes are lineage- or strain-specific and often perform functions related to adaptation to a unique host. Gene family expansions were observed when similar functions were predicted for some accessory and core genes. For gene families (green boxes) that are represented in both core and accessory chromosomes, it is postulated that the core genes tend to be the predominant functional genes in terms of regulating downstream targets such as effector genes, but this is not always the case. (B) Disproportionate distribution of transposable elements (TEs). Accessory chromosomes (ACs) are enriched in TEs. In the *F. oxysporum* genome, core chromosomes have fewer repetitive sequences relative to genome size, and they are often ancient, mostly decayed class I RNA TEs, whereas ACs are densely populated with active class II DNA TEs. Effector genes are usually associated with TEs such as miniature inverted-repeat transposable elements. (C) Structural divergence of epigenetic signals. Core and accessory chromosomes have different histone markers. Fungal genes associated with the same type of histone markers may be coordinately expressed and could be under similar evolutionary pressures.
CHAPTER 2

CONSERVATION AND EXPANSION OF TRANSCRIPTIONAL FACTOR REPERTOIRE IN THE *Fusarium oxysporum* SPECIES COMPLEX

2.1 Introduction

The fungal species complex of *Fusarium oxysporum* (FOSC) has been used as a model to study cross-kingdom fungal pathogenesis. Members within FOSC can cause devastating fusarium wilt diseases among economically important crops (Michielse and Rep 2009; Dean et al. 2012; Ma et al. 2013; Ma 2014; Ploetz 2015; Halpern et al. 2018; Edel-Hermann and Lecomte 2019; Pegg et al. 2019; Yang et al. 2020; Viljoen et al. 2020; Rahman et al. 2021) and is listed among the top five most important plant pathogens (Dean et al. 2012). With strong host specificity, plant pathogenic *F. oxysporum* strains are further grouped as *formae speciales* (Armstrong and Armstrong 1981). For instance, tomato pathogens are named *F. oxysporum* f.sp. *lycopersici*, cotton pathogens *F. oxysporum* f.sp. *vasinfectum* (Halpern et al. 2018), and banana pathogen *F. oxysporum* f.sp. *cubense* (Viljoen et al. 2020). Recently, members within FOSC have also been reported to be responsible for fusariosis, the top emerging opportunistic mycosis (Ma et al. 2013; Yang et al. 2020), and fusarium keratitis, one of the major causes of cornea infections in the developing world and the

leading cause of blindness among fungal keratitis patients (Kredics et al. 2015; Hassan et al. 2016).

Comparative genomics studies on this cross-kingdom pathogen revealed that the FOSC genomes, both human and plant pathogens, are compartmentalized into two components: the core chromosomes (CCs) and accessory chromosomes (ACs). While CCs are conserved and vertically inherited to execute essential housekeeping functions, horizontally transmitted ACs are lineage- or strain-specific and related to fungal adaptation and pathogenicity (Rep et al. 2004; Ma et al. 2013; Yang et al. 2020; Yu et al. 2023).

To coexist and function within the same genome, ACs and CCs coordinate their gene expression. One intriguing cross-regulation example, reported in the reference genome of \textit{F. oxysporum} f.sp. \textit{lycopersici} Fol4287, includes transcription factors Sge1 (SIX Gene Expression 1), Ffts, and virulent factors SIX (Secreted in Xylem) proteins. Sge1 is a highly conserved, CC-encoding TF. By name definition, Sge1 regulates the expression of SIX proteins (Michielse et al. 2009; van der Does et al. 2016). The Fol4287 genome encodes an AC-encoding Ftf1 protein and one CC-encoding Ftf2 (Ftf1 CC homolog) (van der Does et al. 2016). Constitutive expression of either \textit{Ftf1} or \textit{Ftf2} induced the expression of effector genes (van der Does et al. 2016). Furthermore, It was documented that DNA binding sites of Sge1 and Ftf1 are enriched among the cis-regulatory elements of \textit{in planta} transcriptionally up-regulated genes (van der Does et al. 2016). Another example of CCs and ACs cross-talking is the alkaline pH-responsive transcription factor PacC/Rim1p reported in \textit{F. oxysporum} clinical
strains (Zhang et al. 2020). In addition to the full-length PacC ortholog (PacC\_O), located on a CC, the clinical isolate NRRL32931 genome encodes three truncated PacC homologs, named PacC\_a, PacC\_b, and PacC\_c in ACs (Zhang et al. 2020).

To thoroughly understand the coordination of the crosstalk between genome compartments and their contribution to the cross-kingdom fungal pathogenesis, this study compared the repertoire of TFs (i.e., TFome) among 15 *F. oxysporum* and 15 other ascomycete fungal genomes. Remarkably, we discovered a strong positive correlation \( y = 0.07264x - 190.9, r^2 = 0.9361 \) between the number of genes \( x \) and TFome size \( y \) of an organism. Primarily due to the acquisition of ACs, we observed increased TFome sizes among FOSC genomes. All TFs were organized into 48 families based on the InterPro classification of proteins. Fourteen families, involved in transcription/translation regulations and cell cycle controls, were highly conserved. Thirty families, accounting for \( \frac{3}{4} \) of all families, were expanded in various degrees among FOSC genomes. Unique TF expansions driven by ACs include members of Zn2-C6 fungal-type (Zn2-C6) and Zinc Finger C2H2 (Znf\_C2H2) families. This comparative study highlighted conserved regulatory mechanisms. The signature of conservation established the foundation to study various impacts of additional AC TFs on existing regulatory pathways. In combination with existing expression data, this study provides insights to the fine-tuning of environmental adaptation performed by this group of diverse organisms in order to engage in cross-kingdom interactions with different hosts.
2.2 Materials and Methods

2.2.1 Generation of fungal TFomes

The annotation pipeline is briefly summarized in Figure S2.1A-B. The fungal proteomes of 30 strains were downloaded from the JGI MycoCosm portal (Grigoriev et al. 2014). Protein annotation was performed using InterProScan/5.38-76.0 (https://www.ebi.ac.uk/interpro/search/sequence/) (Jones et al. 2014). Annotations of proteins that putatively serve as TFs were filtered out using a table containing InterPro terms related to transcriptional regulatory functions summarized by literature (Park et al. 2008; Shelest 2017), with further addition by manual curation (Table S2.1). Orthologous analysis was done with OrthoFinder 2.5.4 (https://github.com/davidemms/OrthoFinder) (Emms and Kelly 2019) to probe orthologs of functionally validated TFs (Table S2.3-4 and Table 2.3) in Fusarium.

2.2.2 RNA-seq analysis

The RNA-seq datasets were previously described (Guo et al. 2021; Redkar et al. 2022) and deposited by those authors to the NCBI Short Read Archive with accession number GSE87352 and to the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10597, respectively. For data reprocessing, reads were mapped to reference genomes of Arabidopsis [annotation version Araport11 (Cheng et al. 2017)], Fo5176 (Fokkens et al. 2021), Fo47 (Wang et al. 2020) and Fol4287 (Ma et al.
2010) using HISAT2 version 2.0.5 (Kim et al. 2019). Mapped reads were used to quantify the transcriptome by StringTie version 1.3.4 (Pertea et al. 2015), at which step TPM (transcript per million) normalization was applied. Normalized read counts were first averaged per condition, transformed by log2 (normalized read count + 1) and Z-scaled. This was then visualized in pheatmap (version 1.0.12).

2.2.3 Genome partition

The genome partition results for chromosome-level assemblies were retrieved from previous reports for Fol4287 (Ma et al. 2010), FoII5 (Zhang 2019), Fo5176 (Fokkens et al. 2021), and Fo47 (Wang et al. 2020). Fo47 has a clear genome partition with 11 core chromosomes and one accessory chromosome, therefore serving as the reference for the genome partition of other *F. oxysporum* genomes. mummer/3.22 was applied to align scaffolds of genome assemblies against 11 core chromosomes of the reference genome Fo47 using default parameters. The scaffolds aligned to the core chromosomes of Fo47 with a coverage larger than 5% were annotated as core scaffolds. The rest of the scaffolds are partitioned as accessory scaffolds. Genes residing on core and accessory scaffolds were annotated as core and accessory genes, respectively.

2.2.4 Phylogenetics analysis

Protein sequences were aligned via MAFFT/7.313 (Katoh and Standley 2013). The iqtree/1.6.3 (Nguyen et al. 2015; Minh et al. 2020) was run on the
sequence alignment to generate the phylogeny (by maximum likelihood method and bootstrapped using 1000 replicates) (Hoang et al. 2018). Visualization was done via the Interactive Tree of Life (Letunic and Bork 2021) to produce the phylogram. OrthoFinder 2.5.4 (Emms and Kelly 2019) was used for orthogroup determination. To build a species phylogram, randomly selected 500 conserved proteins (single-copy orthologs) were aligned first. The alignment was then concatenated, and phylogeny was determined and visualized using the above methods.

2.2.5 Expansion index calculation

To understand genome regulation among FOSC, we developed two expansion index scores. The first one uses two yeast lineages as the baseline \((E_{ly})\):

\[
E_{ly} = \frac{\text{Average number of TFs in FOSC} + 1}{\text{Average number of TFs in yeasts} + 1}
\]

By second index score \((E_{lf})\), we directly compared \(F. oxysporum\) with its \(Fusarium\) relatives to calculate the expansion index as follows:

\[
E_{lf} = \frac{\text{Average number of TFs in FOSC} + 1}{\text{Average number of TFs in FOSC sister species} + 1}
\]

2.3 Results

2.3.1 FOSC TFome expansion resulted from the acquisition of ACs

We compared 30 ascomycete fungal genomes (Figure 2.1 and Table 2.1), including 15 strains within the FOSC, nine sister species close to \(F. oxysporum\), two yeast genomes (\(Saccharomyces cerevisiae\) and \(Schizosaccharomyces\))
pombe), four other filamentous fungal species (Neurospora crassa, Aspergillus nidulans, Aspergillus acrisatulus, and Magnaporthe oryzae). To maintain consistency, the protein sequences for all these genomes were retrieved from the MycoCosm portal (Grigoriev et al. 2014).

For a comprehensive TFome annotation, we used InterProScan (IPR) terms associated with fungal transcriptional regulation (Park et al. 2008; Shelest 2017) and curated a mapping with updated IPR classification (interproscan version: 5.38-76.0) (Blum et al. 2021). In addition, we searched the IPR classification of protein families and obtained all other terms related to the transcriptional regulation activity. This resulted in 234 TF-related IPR terms (Table S2.1). Since most terms were initially defined in the mammalian systems, fungal genomes included in this study were associated with 71 out of the 234 TF-related IPR terms (Table S2.1, Materials and Methods, and Figure S2.1A-B for annotation pipeline). After removing 13 terms for redundancy (two terms describing the identical domain) and 10 terms for minimal presentation (< 4 among the 30 genomes), this comparative TFome study focused on 48 IPR terms in 27,967 TFs (Table S2.1-S2.2). Notably, ¼ of these terms were not reported to be affiliated with fungal transcriptional regulation by either Park et al. 2008 or Shelest 2017 (Table S2.1).

Comparing the total number of protein-coding genes (x) and the total number of TFs (y) within the same genome, we observed a strong positive correlation ($y = 0.07264x - 190.9$, $r^2 = 0.9361$) (Figure 2.2A). FOSC TFomes were larger than other genomes included in this study, with an average of 1144 TFs
per genome (Figure 2.2A, Table 2.1). After partitioning each FOSC genome into core and accessory regions (see Materials and Methods for details), we observed a positive correlation between the number of TFs encoded in the accessory chromosomal region of each strain (defined as accessory TFs hereafter) with the size of accessory genomes (Mb) \( y = 17.239x + 3.553 \) (Figure S2.2). This suggests that accessory chromosomes contribute directly to the expanded TFome.

Based on this index value, we classified TF families into three major groups (Table 2.2, Table S2.1). Group 1 contained 14 TF families with an expansion score of 1, indicating high conservation. Group 2 included four families with an index score below 1, reflecting some level of gene family contraction. Group 3 contained 30 families with an expansion index greater than 1, indicating gene expansion.

To understand genome regulation among FOSC, we developed an expansion index score using two yeast lineages as the baseline \( (EI_y) \):

\[
EI_y = \frac{\text{Average number of TFs in FOSC} + 1}{\text{Average number of TFs in yeasts} + 1}
\]

Based on this index value, we classified TF families into three major groups (Table 2.2, Table S2.1). Group 1 contained 14 TF families with an expansion score of 1, indicating high conservation. Group 2 included four families with an index score below 1, reflecting some level of gene family contraction. Group 3 contained 30 families with an expansion index greater than 1, indicating gene expansion.
2.3.2 Conserved TF families that are primarily associated with general/global transcription factors

Fourteen TF families, accounting for 30% of our annotated TF families, each had a single ortholog in all genomes included in this study (Figure 2.2B; Table 2.2; Table S2.1), suggesting their functional conservation across the Ascomycota. These 30% conserved TF families accounted for less than 2% of the total TFomes. Annotation based on S. cerevisiae and other model organisms suggested their involvement in transcription/translation regulation and cell cycle controls.

2.3.2.1 Transcription/Translation regulation

Nine TF families were annotated to be related to transcription and translational regulation, including TATA box-binding protein (TBP), TBP-associated factors (TAFs), RNA polymerase II elongation regulator Vps25, and CCAAT-Binding Factors (CBFs) related to ribosomal biogenesis.

One of the most conserved TF families, Transcription initiation TBP binds directly to the TATA box to define the transcription start and initiate transcription facilitated by all three RNA polymerases. In fact, the function of TBP is so conserved that the yeast homolog can complement TBP mutations in humans (Roberts and Winston 1996; Yamaguchi et al. 2001). Seven conserved TF families are classified as transcription positive/negative regulators, and transcription elongation. TAF12 and TAFII28 are parts of the transcription factor TFIIID complex. Interacting with TBP, TAFs form the TFIIID complex positively
participate in the assembly of the transcription preinitiation complex (Green 2000). Similarly, TFIIH works synergistically with TFII D to promote the transcription (Fribourg et al. 2000). In contrast, Negative cofactor 2 (Ncb2) inhibits the preinitiation complex assembly (Goppelt et al. 1996). Other factors include the CNOT1, a global regulator involved in transcription initiation and RNA degradation (Chalabi Hagkarim and Grand 2020), and Vps72/YL1 that contributes to transcriptional regulation through chromatin remodeling as reported in the yeast (Latrick et al. 2016; Liang et al. 2016). Vps25 is a subunit of the ESCRT-II complex, which binds to RNA polymerase II elongation factor to exert transcriptional control in mammalian systems (Kamura et al. 2001). One TF family is suggested to be involved in translational regulation. CCAAT box is a common cis-acting element found in the promoter and enhancer regions of genes in the eukaryotes (Vuorio et al. 1990; Becker et al. 1991). CBFs are necessary for the 60S ribosomal subunit biogenesis and therefore involved in the translational control (Edskes et al. 1998; Milkereit et al. 2001; Fromont-Racine et al. 2003). This family, including Noc3, Noc4, and Mak21 in S. cerevisiae, had three members in each genome, and a clear single-copy orthologous relationship can be observed for each member (Figure S2.3A).

2.3.2.2 Cell cycle control

Five TF families are related to cell cycle control, including cell cycle progression, DNA repair, and machinery/cell integrity maintenance.
One conserved TF family, **APSES-type HTH**, was reported to be involved in cell-cycle control and crucial to the development (Xin et al. 2020). Every genome included in this study encoded four copies of APSES-type HTH gene (Figure S2.3B) that formed four clades of single-copy orthologs in all genomes except yeasts. Genes in Clade 1 included StuA homologs. As a target of the cyclic AMP (cAMP)-dependent protein kinase A (PKA) signal transduction pathway, StuA was reported to be involved in dimorphic switch (Gimeno and Fink 1994; Pan and Heitman 2000), fungal spore development and the production of secondary metabolites (Lysøe et al. 2011). Genes in Clade 2 and Clade 3 included *S. cerevisiae* Swi4 and Swi6, which were reported to form a protein complex regulating cell cycle progression from G1 to S phase (Koch et al. 1993), as well as meiosis (Son et al. 2016). Genes in Clade 4 included homologs of *S. pombe* Bqt4 anchoring telomeres to the nuclear envelope (Chikashige et al. 2009).

The conserved TF family, **DTT**, represented by the *S. cerevisiae* homolog Itc1, is recognized as a subunit of ATP-dependent lsw2p-ltc1p chromatin remodeling complex and required for repression of early meiotic gene expression during the mitotic growth (Sugiyama and Nikawa 2001).

The other conserved TF family, **RFX**, was reported to be involved in DNA repair. Each strain encoded two orthologous copies, except *F. venenatum* encoding two copies within the RFX1 clade (Figure S2.3C). Being a major transcriptional repressor of DNA-damage-regulated genes in *S. cerevisiae*, Rfx1 functions in DNA damage repair and replication checkpoint pathways (Lubelsky et al. 2005). In *F. graminearum*, Rfx1 was reported to be essential in maintaining
the genome integrity (Min et al. 2014). The other copy, Rsc9 in *S. cerevisiae*,
was reported to be a member of the chromatin structure-remodeling complex
RSC involved in transcription regulation and nucleosome positioning (Cairns et

The conserved TF family **NFYA** was reported to bind to the CCAAT box.
All strains maintained a single copy of this family. Its yeast homolog Hap2 was
reported to induce the expression of mitochondrial electron transport genes
(Olesen et al. 1991) and its *F. verticillioides* homolog NFYA Hap2 was reported
to be essential for fungal growth and the virulence on maize stalks (Ridenour and
Bluhm 2014).

The **MADS MEF2-like** TF family, including *S. cerevisiae* Rlm1, was
reported to be a component of the protein kinase C-mediated MAP kinase
pathway involved in maintaining cell integrity (Jung et al. 2002). Having a paralog
from the whole genome duplication in *S. cerevisiae*, Rlm1 was detected as a
single copy gene in all filamentous fungi included in this study. Its member in *F.
verticillioides*, Mef2, was reported to play a vital role in the sexual development
(Ortiz and Shim 2013).

**2.3.3 Gene family contractions in FOSC partially caused by whole genome
duplication in yeast**

We detected expansion score less than 1 for four TF families,
MATalpha_HMGbox, NOT4, MADS_SRF-like, and HSF (Heat Shock Factor),
reflecting some level of gene family contraction among members of FOSC compared to the two yeast genomes (Figure S2.4).

TF family **MATalpha_HMGbox**, including *S. cerevisiae* mating type protein alpha 1, was reported to be a transcription activator that activates mating-type alpha-specific genes (Martin et al. 2010). Reflecting the potential heterothallic mating strategy, all *F. oxysporum* Mat1-1 type strains contained this TF, but not Mat1-2 strains, even though sexual reproduction has not been observed in FOSC (Arie et al. 2000).

TF family **NOT4** was reported to be a component of the multifunctional CCR4-NOT complex, a global transcriptional repressor of the RNA polymerase II transcription (Albert et al. 2002). Most genomes included in this study encoded one copy of this TF family, but some filamentous fungal genomes, including *A. nidulans*, *F. redolens*, *F. oxysporum* strains NRRL26365, MRL8666, and PHW726 lost it. The functional implication of this loss remains to be discovered.

The contractions of the other two TF families, **MADS SRF-like** and **HSF**, were primarily caused by the whole genome duplication in yeast. In contrast to the contraction at the global scale, both TF families were expanded among some FOSC strains when compared to other filamentous fungi (Figure S2.4).

Reported in *M. oryzae*, **MADS SRF-like** TF, essential for transcriptional regulation of growth-factor-inducible genes (Messenguy and Dubois 2003), is important for microconidium production and virulence in host plants (Ding et al. 2020). Due to the event of whole genome duplication, the *S. cerevisiae* genome contained 2 copies of this TF family, while all filamentous genomes encoded a
single copy. However, we detected an average of 2.73 copies among phytopathogenic FOSC strains. There are 6 copies in the genome of Fo5176, a pathogen of Brassicaceae plants including A. thaliana (Table S2.1).

TF family HSF was reported to activate the production of heat shock proteins that prevent or mitigate protein misfolding under abiotic/biotic stresses (Feder and Hofmann 1999). The S. cerevisiae genome contained five copies of HSF TF family, while all non-FOSC filamentous fungi had three copies. Members of FOSC exhibited some level of expansion to 4 or 5 copies (Fo47:4, Fol4287: 5, Il5: 4, HDV274: 4, and Fo5176: 4), with 1-2 copies encoded in ACs. Phylogenetically, all HSF TFs were clustered into three major clades, named as Skn7, Sfl1, and Hsf1 (Figure 2.3A-B). All AC-encoding HSFs were phylogenetically close to Hsf1 (Figure 2.3A). Based on the study in M. oryzae, the family Sfl1 is essential for vegetative growth, conidiation, sexual reproduction, and pathogenesis (Li et al. 2011). Based on a study in F. graminearum, the family Skn7 involved in regulating the oxidative stress response and is essential for pathogenicity (Jiang et al. 2015). Our expression data generated during the plant colonization (Guo et al. 2021) supported the involvement of all three core genes during plant colonization (Figure 2.3C). However, the Hsf1 accessory copies of these two strains were distinct, as the Fo47 AC-encoding Hsf1 was up-regulated and the Fo5176 AC-encoding Hsf1 was down-regulated, post inoculation (Figure 2.3C), suggesting their distinct regulatory function involved in these two distinct interactions.
2.3.4 Significant FOSC TFome expansion driven by a few exceedingly expanded TF families

2.3.4.1 Gain-of-Function among Filamentous Ascomycete Fungi

Three TF families, CP2 ($E_{ly} = 2.73$), HTH_AraC ($E_{ly} = 2$), and HTH_Psq ($E_{ly} = 3.53$), were absent in both yeast genomes, suggesting a gain of function among filamentous ascomycete fungi (Table S2.1). TF family CP2 was studied in animal and fungal kingdoms with a function related to differentiation and development (Paré et al. 2012). Both HTH_AraC and HTH_Psq are part of the helix-turn-helix (HTH) superfamily. HTH_AraC was first reported in bacteria as a positive regulator regulating the arabinose operon regulatory (Gallegos et al. 1993; Bustos and Schleif 1993; Schleif 2010). HTH_Psq, as part of the eukaryotic Pipsqueak protein family, was reported in vertebrates, insects, nematodes, and fungi to regulate process involved in cell death (Siegmund and Lehmann 2002). Most FOSC genomes encoded a single copy of HTH_AraC, while the count of HTH_Psq-containing proteins ranged from 0 to 9 in the FOSC and 0 to 3 in other Fusarium genomes. Since the HTH_Psq domain also exists in transposases (Siegmund and Lehmann 2002), and ACs in FOSC are transposon-rich, it remains to be studied whether proteins containing the Psq domain are bona fide TFs.
2.3.4.2 Seven exceedingly expanded TF families

Among the families containing minimally one yeast ortholog, seven TF families had expansion indexes greater than 2 (Table 2.2 and Figure 2.2B), including Zn2-C6 ($El_y = 15.09$), bZIP ($El_y = 5.80$), Znf_C2H2 ($El_y = 4.15$), Homeobox ($El_y = 2.28$), PAI2 ($El_y = 3.42$), NDT80 ($El_y = 3.47$), and bHLH ($El_y = 3.48$). Based on the number increment, the most significantly expanded TF families were Zn2-C6 (44 in yeasts versus 671 in FOSC) and Znf_C2H2 (40 in yeasts versus 167 in FOSC) (Figure 2.2C and Table S2.1). Because of their large expansion, these seven families accounted for more than 75% of the total TFome. All seven families exhibited a gradual expansion, following the pattern that yeasts < non-<i>Fusarium</i> filamentous fungi < non-FOSC <i>Fusarium</i> < FOSC (Table S2.1). Annotating large TF families could be challenging. Here we described some examples based on literature.

**Zn2-C6**, a fungal TF family (MacPherson et al. 2006), is detected as the most significant expanded TF family, reaching over 600 members among FOSC genomes and accounting for more than half of the total TFome. Able to form a homodimer, this group of TFs are able to bind to the specific palindromic DNA sequence through direct contact with the major groove of the double-stranded DNA molecules (MacPherson et al. 2006). The versatility of this group of TFs could be achieved by domain shuffling and by changing the nucleotide binding specificity. In addition to the well-documented Ftf1 (Ramos et al. 2007; Niño-Sánchez et al. 2016; van der Does et al. 2016; Zhao et al. 2020; Zuriégat et al. 2021), five additional TFs within this family were reported in *F. oxysporum,*
including Ctf1 (Rocha et al. 2008), Ctf2 (Rocha et al. 2008), Fow2 (Imazaki et al. 2007), XlnR (Calero-Nieto et al. 2007) and Ebr1 (Jonkers et al. 2014). Their functions were reported to be involved in the development, metabolism, stress response, and pathogenicity.

**Znf_C2H2** was reported to be the most common DNA-binding motif found in the eukaryotic transcription factors (Fedotova et al. 2017). Five reported *F. oxysprum* TFs are Czf1 (Yun et al. 2019), Con7-1 (Ruiz-Roldán et al. 2015), PacC (Caracuel et al. 2003; Zhang et al. 2020), ZafA (López-Berges 2020) and St12 (Rispail and Di Pietro 2009; Asunción García-Sánchez et al. 2010). Classified in the **Znf_C2H2** family, PacC was linked to the fungal virulence in both plant and human hosts (Caracuel et al. 2003; Zhang et al. 2020).

Other five families were **bZIP**, **Homeobox**, **PAI2**, **Ndt80** and **bHLH**. The **bZIP** domain contains a region for sequence-specific DNA binding followed by a leucine zipper region required for dimerization (Bader and Vogt 2006). Three *F. oxysporum* bZIP TFs were reported, including Atf1 (Li et al. 2013), Hapx (López-Berges et al. 2012), and MeaB (López-Berges et al. 2010), all of which are important for fungal pathogenicity. **Homeobox** is a DNA binding motif with a helix-turn-helix structure. In *S. pombe*, a homeobox-domain containing protein Phx1 was reported to be a transcriptional coactivator involved in yeast fission. In *M. oryzae*, a homeobox-domain containing protein Hox played roles in the conidiation and appressorium development (Kim et al. 2009). The TF family **PAI2** is involved in the negative regulation of protease synthesis and sporulation of the *Bacillus subtilis* (Honjo et al. 1990). The TF family **Ndt80** is essential for
completing meiosis in *S. cerevisiae* (Pierce et al. 2003; Tsuchiya et al. 2014) and *Ustilago maydis* (Doyle et al. 2016) by promoting the expression of sporulation genes for the fulfillment of meiotic chromosome segregation (Hepworth et al. 1998). The TF family bHLH forms a superfamily of transcriptional regulators found in almost all eukaryotes and involved in diverse developmental processes (Jones 2004). In *F. graminearum*, a bHLH-domain containing protein Gra2 was reported to regulate the biosynthesis of phytotoxin gramillin (Bahadoor et al. 2018), while a bHLH-domain containing protein SreA in *Penicillium digitatum* is required for anti-fungal resistance and full virulence in citrus fruits (Liu et al. 2015).

### 2.3.4.3 Other families

The other 20 TF families (expanded but with $El_y \leq 2$) accounted for 20% of the TFome, with an average 9.6 copies in each genome examined (Table S2.1).

Four TF families were functionally linked to chromatin remodeling, including Bromodomain ($El_y = 1.52$), CBFA_NFYB ($El_y = 1.35$), Znf_RING-CH ($El_y = 1.11$), and ARID ($El_y = 1.25$). The TF family **Bromodomain** contained Spt7. As a crucial part of the SAGA complex in yeast, Spt7 was reported to recognize acetylated lysines of histones and eventually lead to chromatin unwinding (Donczew et al. 2020). The **CBFA_NFYB** domain was found in proteins (*e.g.*, *S. cerevisiae* Dls1) that regulate RNA polymerase II transcription through controlling chromatin accessibility (*e.g.*, telomeric silencing) (lida and
Araki 2004). The TF family Znf_RING-CH also had a functional connection to chromatin modification (e.g., S. cerevisiae Rkr1) (Braun et al. 2007). The domain ARID, a 100 amino acid motif, was found in many eukaryotic TFs (Iwahara 2002), such as Swi1 in S. cerevisiae, playing an important role in chromatin remodeling. This domain is also required to transcribe a diverse set of genes, including some retrotransposons (Breeden and Nasmyth 1987; Hirschhorn et al. 1992).

TFs belonging to the Ste12 family are only found in the fungal kingdom. Except S. pombe, every genome encodes one copy. Binding to a DNA motif that mediates pheromone response, Ste12 TFs were reported to regulate fungal development and pathogenicity (Rispail and Di Pietro 2010) and are involved in mating and pseudohyphae formation (Gancedo 2001). In F. oxysporum, Ste12, downstream of the Fmk1-mediated MAPK cascade, is involved in the control of invasive growth and fungal virulence (Rispail and Di Pietro 2009).

Among others, Znf_NFX1 domain was found in NK-X1, a repressor of the human disease-associated gene HLA-DRA (Song et al. 1994). The HMG_box (high mobility group box) in S. cerevisiae is seen in three proteins: Spp41, which is involved in negative expression regulation of spliceosome components (Maddock et al. 1994); Nhp6a, which is required for the fidelity of some tRNA genes (Braglia et al. 2007); and Ixr1, a transcriptional repressor that regulates hypoxic genes (Vizoso-Vázquez et al. 2012). Fep1, an example of Znf_GATA, was reported to be a transcriptional repressor involved in the regulation of some iron transporter genes under high iron concentrations (Kim et al. 2016).
*cerevisiae* Mb1, belonging to **Cro/C1-type HTH**, is a transcriptional coactivator (Takemaru et al. 1997).

### 2.3.5 Orthologous survey of TF families that were manually curated

To further understand expanded TFs and their impacts on transcriptional regulation, we curated a list of 102 TFs reported in literature focusing on *F. oxysporum*, *F. graminearum*, and other phytopathogenic fungi (Table S2.3 and examples as described in the previous section). Compared to this list of curated TFs using Orthofinder, we define 80 orthologous groups among *Fusarium* genomes (Table S2.4). 62 out of the 80 orthogroups have been identified using the above IPR-annotated pipeline, including 17 in Zn2C6, 9 in Znf_C2H2, and 1 containing both Znf_C2H2 and Zn2-C6 domains (Table S2.4). This helps add to the functional annotation of these large TF families while also adding additional annotation to 18 TF families (Table S2.4) accounting for 32 genes per genome (3% of average *Fusarium* TFome size). These newly annotated TFs include homologs of those without domain annotation, *e.g.*, disordered proteins *F. oxysporum* Ren1 (Ohara et al. 2004), *M. oryzae* Som1 (Yan et al. 2011), and homologs of those with noncanonical TF domains such as **Ankyrin_rpt** and **WD40_repeat**.

We then directly compared *F. oxysporum* with its *Fusarium* relatives to calculate the expansion index as follows:

\[
EI_f = \frac{\text{Average number of TFs in FOSC} + 1}{\text{Average number of TFs in FOSC sister species} + 1}
\]
The $EI_f$ ranged from the highest score 3.54 (Fug, AreA_GATA) to the lowest score 0.5 (Fox1, Fork_head) (Table S2.4). Among these 80 orthogroups, 36 groups were conserved ($EI_f = 1$), with one gene per genome. Ten of these conserved groups were functionally validated in *F. oxysporum* (Table S2.4). Twenty four groups had scores less than 1, while 20 groups had score greater than 1 (Table 2.3, Table S2.4). Expanded groups included Fug1 (AreA_GATA, $EI_f = 3.54$), Cos1 (Znf_C2H2, $EI_f = 2.8$), Ftf1/Ftf2 (Zn2-C6, $EI_f = 2.7$), Ebr1/Ebr2 (Zn2-C6, $EI_f = 2.5$) and Ren1 (disordered, $EI_f = 2$). We also identified PacC ($EI_f = 1.57$) as the second most expanded group within the highly expanded Znf_C2H2 family. We will further discuss these six groups (highlighted in bold, Table 2.3).

Both Ftf1/Ftf2 and Ebr1/Ebr2, belonging to the Zn2-C6 family, contributes directly to the fungal virulence (Ramos et al. 2007; Michielse et al. 2009; van der Does et al. 2016). Deletion of AC-encoding Ftf1 reduced the pathogenicity of *F. oxysporum* f. sp. *phaseoli* (Ramos et al. 2007), highlighting the direct functional involvement of AC TF in virulence. In Fol, deleting either Ftf1 (AC encoding) or Ftf2 (CC encoding) reduced the fungal virulence (de Vega-Bartol et al. 2011; Niño-Sánchez et al. 2016). Constitutive expression of either Ftf1 or Ftf2 induced the expression of effector genes (van der Does et al. 2016). The core copy Ftf2 was conserved among all *Fusarium* species, and the AC copy Ftf1 was only found in *F. oxysporum* and *F. redolens* (Figure 2.4). Ebr1 had multiple homologs in *F. oxysporum* but a single copy in *F. graminearum* (Jonkers et al. 2014). The *F. oxysporum* genome had three AC-encoding paralogs: Ebr2, Ebr3, and Ebr4. Interestingly, these AC-encoding paralogs are regulated by core copy Ebr1.
(Jonkers et al. 2014). It is worth noting that the Ebr2 coding sequence driven by an Ebr1 promoter was able to rescue the Ebr1 knockout mutation, indicating some functional redundancy of this family.

Both Cos1 and PacC are part of the Znf_C2H2 family. In M. oryzae, Cos1 was reported to be involved in conidiophore development (Li et al. 2013) and functioned as a negative regulator reducing fungal pathogenicity (Zhou et al. 2009). PacC has been reported as an important pH-responsive TF in F. oxysporum (Caracuel et al. 2003; Zhang et al. 2020). This TF family was expanded in clinical strains, showing an average accessory copy number 3.7 of FOSC, whereas the non-clinical strains showed an average accessory copy number 0.5. All Fusarium relatives’ genomes examined only contained a single copy of core PacC. Our previous study using one F. oxysporum clinical isolate revealed that the expression of all PacC genes can be induced with a pH shift from 5.0 to 7.4 (the mammalian physiological pH) indicating a potential role in host adaptation (Zhang et al. 2020). Interestingly, the induction of AC-encoding PacC genes was CC-encoding PacC gene-dependent as the induction disappeared in the CC-encoding PacC knockout mutant, further supporting a cross-talking between core and accessory TFs (Yang 2020). Similar to EBR1, the expression of AC-encoding PacC genes was much lower than that of the CC-encoding PacC gene, and knockouts of one AC PacC gene affected a small subset of genes compared with the CC PacC knockout, which had a broader effect on cellular processes (Yang 2020).
**Fug1** has a role in pathogenicity (maize kernel colonization) and fumonisin biosynthesis in *F. verticillioides* (Ridenour and Bluhm 2017). The deletion of Fug1 increased sensitivity to the antimicrobial compound 2-benzoxazolinone and to hydrogen peroxide, suggesting its role in mitigating stresses associated with the host defense (Ridenour and Bluhm 2017). Neither CC nor AC-encoding copies of these two genes are experimentally examined in FOSC. **Ren1**, a disordered protein without IPR functional domain, was expanded with a Ei score of 2 among FOSC. However, the only reported study on its function is in *F. oxysporum f. sp. melonis*, regulating the development of the conidiation (Ohara et al. 2004).

2.3.6 Transcriptome analysis to probe the essential TFs during host colonization

To understand the functional importance of FOSC TFs, we took advantage of two recently reported transcriptomics datasets (Guo et al. 2021; Redkar et al. 2022), including pathogenic interactions (Fo5176 infecting Arabidopsis and Fol4287 infecting tomato) and endophytic interactions (Fo47 colonizing Arabidopsis) (Supplemental Dataset 2).

By examining patterns of expression (Table S2.5), we found that almost all genes within the conserved category (58 out of 60) (Group 1) were consistently expressed (TPM > 1 across all conditions), supporting their general roles in controlling life processes. Within the expanded category (Group 3), the proportion of genes being consistently expressed ranged from 41% to 59% for core TFs and only from 5% to 16% for AC-encoding TFs. With a less strict filter
(TPM > 1 at minimum 1 condition), we found that all genes within the conserved category were expressed. Within the expanded category, 93% of core TFs and between 49% to 67% of AC-encoding TFs were expressed. The significant increase of AC-encoding TFs with lower stringency further supported their conditional involvements in niche adaptation.

We further reviewed the expression patterns of reported TFs in Fol4287 (Table S2.6). Out of 27 TFs encoded on the core genome, 18 were up-regulated (defined by up-regulation under at least three out of four in planta conditions compared to the axenic growth) during plant colonization, which is consistent with their reported roles in pathogenicity. The AC-encoding Ftf1 was reported to play essential functions in fungal pathogenicity (Niño-Sánchez et al. 2016). Of the ten accessory Ffs, eight were upregulated during plant colonization.

Using a higher stringency filter, selecting TFs up-regulated under all in planta conditions, we were searching: 1) conserved core TFs that may be related to plant colonization and 2) expanded AC TFs that could be related to host-specific pathogenicity. In Fol4287, Fo5176, and Fo47 genomes, 95, 62, and 44 core TFs were upregulated during plant colonization. Among them, ten copies are highly conserved (Table S2.7), as they were single-copy orthologs across all 15 F. oxysporum strains, including Fow2 and Sfl1. Fow2, a Zn2C6 TF, is required for full virulence but not hyphal growth and conidiation in F. oxysporum f. sp. melonis (Imazaki et al. 2007). Sfl1 was reported to be essential for vegetative growth, conidiation, sexual reproduction, and pathogenesis in M. oryzae (Li et al. 2011).
Fol4287, Fo5176, and Fo47 contained 29, 34, and 9 upregulated accessory TFs, including *Ftf1* and *Ren1* (Figure 2.4 and Table S2.8). *Ftfs* were reported to play an essential role in pathogenicity in Fol4287, though their involvements in other interactions are not reported. The Fol4287 genome encoded 10 accessory *Ftfs* and 8 were upregulated during plant colonization. The Fo5176 genome encoded 6 accessory *Ftfs*, but only 1 copy was upregulated during plant colonization. Interestingly, 8 upregulated Fol4287 and 1 upregulated Fo5176 *Ftfs* were clustered together (Figure 2.4). The unique expansion with regulatory adaptation (i.e., fine-tuned expression regulation) seemed to be restricted to Fol4287 and not the other pathogenic strain, Fo5176. In contrast, the Fo5176 genome encoded 7 accessory *Ren1* TF and 2 were upregulated during plant colonization (Figure 2.4), while the Fol4287 genome had only 1 accessory *Ren1* not involved in host colonization. While functional validation is needed, the strain-specific expansion followed by fine-tuned expression regulation when infecting host species exists and likely contributes to the host-specific pathogenicity.

### 2.4 Discussion

For a soilborne pathogen with strong host specificity like FOSC, the adjustment of growth and cell cycle control in response to environmental cues is likely essential for survival. Expanded TF families likely contribute to the enhanced functions related to niche adaptation as these TF families play important roles in transmitting external and internal signals and regulating complex cellular signaling responses to the sensed stimuli. Therefore, it is not
surprising to see that genomes of FOSC have larger TFome sizes than other fungi included in the study. The expansion of TFs among FOSC resulted in a positive correlation between the total number of proteins and the size of the fungal TFome, observed in other instances as well (Shelest 2017).

A total of 14 TF families that control the global transcriptional event, such as TBP, are highly conserved within the ascomycete fungal lineages. Conserved regulatory mechanisms revealed through this study suggest that the plant colonization process could be a common process among FOSC strains regardless of their host-specific pathogenesis. This notion is also supported by recent studies that highlighted the ability of FOSC strains as root colonizers regardless whether they cause disease or function as endophytes (Redkar et al. 2022; Martínez-Soto et al. 2022).

In contrast to these stable TFs, 30 families are expanded in various degrees, and most significant expansions occurred in Zn2-C6 and Znf_C2H2 TF families among FOSC genomes. The number of Zn2-C6 TFs increases significantly (with the highest expansion score) and makes up most of the TFs (56.7%) found within the FOSC TFome. For example, Ftf1, a TF belonging to the Zn2-C6 and involved in the tomato pathogenicity, is most significantly expanded to 10 copies of accessory Ftfs in the tomato pathogen Fol4287 genome. Eight out of 10 were induced during plant colonization.

Unique expansion of some TFs, driven by ACs, may provide a clue to host-specific interactions. Acquiring additional TFs will modify existing regulatory pathways, and this will require fine-tuning of existing networks for this group of
organisms to successfully adapt to different hosts under diverse environments. A previous survey of kinome (the complete set of protein kinases encoded in an organism's genome) among FOSC and other Ascomycetes also revealed a positive correlation between the size of the kinome and the size of the genome (DeIulio et al. 2018), identical to what we reported here for TFomes. As kinases and TFs are key regulators that modulate all important signaling pathways and are essential for the proper functions of almost all molecular and cellular processes, strong correlations among kinome and TFome suggest an ordered recruitment and establishment of ACs among FOSC genomes.

This realization further emphasizes the importance of additional functional studies. Reverse genetics is a powerful tool in defining the functional importance of a TF. For example, TF Ren1, a disordered protein, was identified by genetic and molecular characterization (Ohara et al. 2004). This TF is most significantly expanded (seven copies of accessory Rens) in the Arabidopsis pathogen Fo5176 genome and is involved in plant colonization. High throughput approaches, such as chromatin immunoprecipitation sequencing (CHIP-Seq) and DNA affinity purification sequencing (DAP-seq), can be used to profile the cis-regulatory elements globally for a better understanding of transcriptional regulation in the fungal model F. oxysporum. Gene regulatory networks (Guo et al. 2016, 2020) can add more resolution to these complex regulatory processes. However, the ultimate understanding of the regulatory roles of each TF will come from careful molecular and biochemical characterization.
Our study offers a comprehensive look at the regulation from the evolutionary perspective while also providing an easily implemented computational pipeline to compare TFs and other functional groups in fungi. A better understanding of functions of TFs will not only inform *Fusarium* biology but can also be extrapolated to other filamentous fungal systems.

2.5 References


Gimeno CJ, Fink GR (1994) Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. MOL CELL BIOL 14:13


Olesen JT, Fikes JD, Guarente L (1991) The Schizosaccharomyces pombe homolog of Saccharomyces cerevisiae HAP2 reveals selective and stringent conservation of the small essential core protein domain. 11:9


Figure 2.1 Phylogeny of fungal genomes included in this study.

Both left and right phylograms were constructed by concatenated alignment of randomly selected 500 single-copy orthologous proteins, followed by the maximum likelihood method with 1000 bootstraps. Left shows a phylogram of FOSC (represented by the reference genome Fol4287) together with the other 15 ascomycetes. The right shows a phylogram of members within FOSC, rooted by *F. verticillioides* (not shown). Only bootstrap values not equal to 100 are shown.
Figure 2.2 TFome conservation and variation among ascomycete fungi: baseline description.

(A) There is a positive correlation between the number of protein-coding genes and TFome size of an organism. JGI fungal genome identifiers were used as labels. (B) Histogram illustrates the distribution of expansion indexes among different families. (C) Average number of TFs of two most drastically expanded families (Znf_C2H2 and Zn2-C6) within each genome set. Genome Set 1 (G1) includes two yeast genomes (S. cerevisiae and S. pombe). Genome Set 2 (G2) includes four filamentous fungal species (N. crassa, A. nidulans, A. acristatus, and M. oryzae). Genome Set 3 (G3) includes nine sister species close to F. oxysporum. Genome Set 4 (G4) includes 15 FOSC genomes.
Figure 2.3 Evolutional trajectory of heat shock factors (HSFs) suggesting genome expansion and adaptation.

(A) Phylograms of HSFs were constructed by maximum likelihood method with 1000 bootstraps. Branches of *Fusarium* HSFs were colored in yellow. Accessory HSFs of FOSC are shared in red. (B) Number of accessory HSFs in some FOSC genomes. (C) Expression of *HSF* genes during plant colonization (hpi indicates hours post inoculation), compared to axenic growth. Transcriptome data was previously described in Guo et al. 2021. See Materials and Methods for details of data reprocessing and visualization.
Figure 2.4 Unique expansion of some TFs, driven by ACs, may provide clues to host-specific adaptation.

RNA-seq data were previously described (Guo et al., 2021; Redkar et al., 2021). (A) Ftf1, the TF involved in the tomato pathogenicity is most significantly expanded (10 copies of accessory FTFs) in the tomato pathogen Fol4287 genome and the expression of eight out of 10 were induced during plant colonization. (B) Ren1 is most significantly expanded (seven copies of accessory RENs) in the Arabidopsis pathogen Fo5176 genome, and two of them were induced during plant colonization.
Table 2.1 Fungal genomes used in this study

<table>
<thead>
<tr>
<th>Fungal species or strains</th>
<th>MycoCosm identifier</th>
<th>Genome Size (MB)</th>
<th>No. of TFom genes</th>
<th>Host</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>Sacce1</td>
<td>12.07</td>
<td>6575</td>
<td></td>
<td>(Goffeau et al. 1996)</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>Schpo1</td>
<td>12.61</td>
<td>5134</td>
<td></td>
<td>(Wood et al. 2002)</td>
</tr>
<tr>
<td>Aspergillus nidulans</td>
<td>Aspnid1</td>
<td>30.48</td>
<td>10680</td>
<td></td>
<td>(Galagan et al. 2005)</td>
</tr>
<tr>
<td>Aspergillus acrisstatulus</td>
<td>Aspacri1</td>
<td>32.59</td>
<td>11221</td>
<td></td>
<td>(Vesth et al. 2018)</td>
</tr>
<tr>
<td>Neurospora crassa</td>
<td>Neucr2</td>
<td>41.04</td>
<td>9730</td>
<td></td>
<td>(Galagan et al. 2003)</td>
</tr>
<tr>
<td>Magnaporthe oryzae</td>
<td>Magor1</td>
<td>40.49</td>
<td>12673</td>
<td>Rice</td>
<td>(Dean et al. 2005)</td>
</tr>
<tr>
<td>Fusarium solani</td>
<td>Fusso1</td>
<td>52.93</td>
<td>17656</td>
<td>broad hosts</td>
<td>(Mesny et al. 2021)</td>
</tr>
<tr>
<td>F. pseudograminearum</td>
<td>Fusps1</td>
<td>36.33</td>
<td>12395</td>
<td>Wheat</td>
<td>(Cuomo et al. 2007)</td>
</tr>
<tr>
<td>F. graminearum</td>
<td>Fusgr1</td>
<td>36.45</td>
<td>13321</td>
<td>Wheat</td>
<td>(Mesny et al. 2021)</td>
</tr>
<tr>
<td>F. venenatum</td>
<td>Fusven1</td>
<td>37.45</td>
<td>12845</td>
<td></td>
<td>(Mesny et al. 2021)</td>
</tr>
<tr>
<td>F. tricinctum</td>
<td>Fustr1</td>
<td>43.69</td>
<td>14106</td>
<td>Broad hosts</td>
<td>(Ma et al. 2010)</td>
</tr>
<tr>
<td>F. verticillioides</td>
<td>Fusve2</td>
<td>41.78</td>
<td>15869</td>
<td>Corn</td>
<td>(Wiemann et al. 2013)</td>
</tr>
<tr>
<td>F. fujikuroi</td>
<td>Fusfu1</td>
<td>43.83</td>
<td>14813</td>
<td>Broad hosts</td>
<td>(Mesny et al. 2021)</td>
</tr>
<tr>
<td>F. redolens</td>
<td>Fusre1</td>
<td>52.56</td>
<td>17051</td>
<td>Broad hosts</td>
<td>(Mesny et al. 2021)</td>
</tr>
<tr>
<td>F. commune</td>
<td>Fusco1</td>
<td>48.37</td>
<td>15731</td>
<td></td>
<td>(Zhang 2019)</td>
</tr>
<tr>
<td>F. oxysporum t.sp. cubense (II5)</td>
<td>FoxII5</td>
<td>49.43</td>
<td>16048</td>
<td>Banana</td>
<td>(Delulio et al. 2018)</td>
</tr>
<tr>
<td>F. oxysporum f. sp. radicis-lycopersici (CL57)</td>
<td>Fusoxrad1</td>
<td>49.36</td>
<td>18238</td>
<td>Tomato</td>
<td>(Wang et al. 2020)</td>
</tr>
<tr>
<td>F. oxysporum Fo47 (Fo47)</td>
<td>FusoxFo47_2</td>
<td>50.36</td>
<td>16207</td>
<td></td>
<td>(Delulio et al. 2018)</td>
</tr>
<tr>
<td>F. oxysporum f. sp. lycopersici (MN25)</td>
<td>Fusoxlyc1</td>
<td>48.64</td>
<td>17931</td>
<td>Tomato</td>
<td>(Yang 2020)</td>
</tr>
<tr>
<td>F. oxysporum NRRL26365 (NRRL 26365)</td>
<td>Fox26365_1</td>
<td>48.46</td>
<td>16047</td>
<td>Human</td>
<td>(Ma et al. 2014)</td>
</tr>
<tr>
<td>F. oxysporum f. sp. melonis (FoMelon)</td>
<td>Fusoxmel1</td>
<td>54.03</td>
<td>19661</td>
<td>Melon</td>
<td>(Ma et al. 2014)</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. lycopersici</strong> (Fol4287)</td>
<td>Fusox2</td>
<td>61.36</td>
<td>20925</td>
<td>1292</td>
<td>Tomato</td>
</tr>
<tr>
<td><strong>F. oxysporum NRRL32931</strong> (NRRL32931)</td>
<td>Fusox32931</td>
<td>47.91</td>
<td>17280</td>
<td>1072</td>
<td>Human</td>
</tr>
<tr>
<td><strong>F. oxysporum MRL8996</strong> (MRL8996)</td>
<td>FoxMRL8996</td>
<td>50.07</td>
<td>16631</td>
<td>1057</td>
<td>Human</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. matthioliæ</strong> (PHW726)</td>
<td>FoxPHW726</td>
<td>57.22</td>
<td>17996</td>
<td>1157</td>
<td>Brassica</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. vasinfectum</strong> (FoCotton)</td>
<td>Fusoxvas1</td>
<td>52.91</td>
<td>19143</td>
<td>1189</td>
<td>Cotton</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. pisi</strong> (HDV247)</td>
<td>Fusoxpis1</td>
<td>55.19</td>
<td>19623</td>
<td>1229</td>
<td>Pea</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. raphani</strong> (PHW815)</td>
<td>Fusoxrap1</td>
<td>53.5</td>
<td>19306</td>
<td>1132</td>
<td>Brassica</td>
</tr>
<tr>
<td><strong>F. oxysporum f. sp. conglutinans</strong> (PHW808)</td>
<td>Fusoxcon1</td>
<td>53.58</td>
<td>19854</td>
<td>1142</td>
<td>Brassica</td>
</tr>
<tr>
<td><strong>F. oxysporum Fo5176 (Fo5176)</strong></td>
<td>FoxFo5176</td>
<td>67.98</td>
<td>19130</td>
<td>1236</td>
<td>Arabidopsis</td>
</tr>
</tbody>
</table>
**Table 2.2 Expansion index (\(E_{Iy}\)) of 48 TF families**

<table>
<thead>
<tr>
<th>IPR</th>
<th>Term</th>
<th>(E_{Iy})</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR000814</td>
<td>TBP</td>
<td>1</td>
</tr>
<tr>
<td>IPR003228</td>
<td>TFIID_TAF12</td>
<td>1</td>
</tr>
<tr>
<td>IPR004595</td>
<td>TFIIH_C1-like</td>
<td>1</td>
</tr>
<tr>
<td>IPR006809</td>
<td>TAFII28</td>
<td>1</td>
</tr>
<tr>
<td>IPR042225</td>
<td>Ncb2</td>
<td>1</td>
</tr>
<tr>
<td>IPR008570</td>
<td>Vps25</td>
<td>1</td>
</tr>
<tr>
<td>IPR008895</td>
<td>Vps72/YL1</td>
<td>1</td>
</tr>
<tr>
<td>IPR007196</td>
<td>CNOT1</td>
<td>1</td>
</tr>
<tr>
<td>IPR005612</td>
<td>CBF</td>
<td>1</td>
</tr>
<tr>
<td>IPR001289</td>
<td>NFYA</td>
<td>1</td>
</tr>
<tr>
<td>IPR018004</td>
<td>APSES-type HTH</td>
<td>1</td>
</tr>
<tr>
<td>IPR003150</td>
<td>RFX</td>
<td>1</td>
</tr>
<tr>
<td>IPR033896</td>
<td>MADS_MEF2-like</td>
<td>1</td>
</tr>
<tr>
<td>IPR018501</td>
<td>DDT</td>
<td>1</td>
</tr>
<tr>
<td>IPR006856</td>
<td>MATalpha_HMGbox</td>
<td>0.8</td>
</tr>
<tr>
<td>IPR039515</td>
<td>NOT4</td>
<td>0.9</td>
</tr>
<tr>
<td>IPR00232</td>
<td>HSF</td>
<td>0.98</td>
</tr>
<tr>
<td>IPR003163</td>
<td>Tscript_reg_HTH_APSES-type</td>
<td>1.04</td>
</tr>
<tr>
<td>IPR001766</td>
<td>Fork_head</td>
<td>1.05</td>
</tr>
<tr>
<td>IPR011016</td>
<td>Znf_RING-CH</td>
<td>1.11</td>
</tr>
<tr>
<td>IPR001965</td>
<td>Znf_PHD</td>
<td>1.11</td>
</tr>
<tr>
<td>IPR009071</td>
<td>HMG_box</td>
<td>1.12</td>
</tr>
<tr>
<td>IPR004181</td>
<td>Znf_MIZ</td>
<td>1.24</td>
</tr>
<tr>
<td>IPR001606</td>
<td>ARID</td>
<td>1.25</td>
</tr>
<tr>
<td>IPR000679</td>
<td>Znf_GATA</td>
<td>1.3</td>
</tr>
<tr>
<td>IPR001005</td>
<td>SANT/Myb</td>
<td>1.32</td>
</tr>
<tr>
<td>IPR000818</td>
<td>TEA/ATTTS</td>
<td>1.33</td>
</tr>
<tr>
<td>IPR003120</td>
<td>Ste12</td>
<td>1.33</td>
</tr>
<tr>
<td>IPR003958</td>
<td>CBFA_NFYB</td>
<td>1.35</td>
</tr>
<tr>
<td>IPR001083</td>
<td>Cu_fist</td>
<td>1.37</td>
</tr>
<tr>
<td>IPR000967</td>
<td>Znf_NFX1</td>
<td>1.4</td>
</tr>
<tr>
<td>IPR006565</td>
<td>Bromodomain</td>
<td>1.52</td>
</tr>
<tr>
<td>IPR001387</td>
<td>Cro/C1-type_HTH</td>
<td>1.6</td>
</tr>
<tr>
<td>IPR001841</td>
<td>Znf_RING</td>
<td>1.64</td>
</tr>
<tr>
<td>IPR000571</td>
<td>Znf_CCCH</td>
<td>1.74</td>
</tr>
<tr>
<td>IPR001878</td>
<td>Znf_CCHC</td>
<td>1.83</td>
</tr>
<tr>
<td>IPR010666</td>
<td>Znf_GRF</td>
<td>2</td>
</tr>
<tr>
<td>IPR018060</td>
<td>HTH_AraC *</td>
<td>2</td>
</tr>
<tr>
<td>IPR001356</td>
<td>Homeobox</td>
<td>2.28</td>
</tr>
<tr>
<td>IPR007604</td>
<td>CP2 *</td>
<td>2.73</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>IPR007396</td>
<td>PAI2</td>
<td>3.42</td>
</tr>
<tr>
<td>IPR024061</td>
<td>NDT80</td>
<td>3.47</td>
</tr>
<tr>
<td>IPR011598</td>
<td>bHLH</td>
<td>3.48</td>
</tr>
<tr>
<td>IPR007889</td>
<td>HTH_Psq *</td>
<td>3.53</td>
</tr>
<tr>
<td>IPR013087</td>
<td>Znf_C2H2</td>
<td>4.15</td>
</tr>
<tr>
<td>IPR004827</td>
<td>bZIP</td>
<td>5.8</td>
</tr>
<tr>
<td>IPR001138</td>
<td>Zn2-C6</td>
<td>15.09</td>
</tr>
</tbody>
</table>
Table 2.3 Ortholog copy number and expansion index (EI) of the characterized and expanded TFs in *F. oxysporum*

<table>
<thead>
<tr>
<th>TF</th>
<th>Reported species</th>
<th>References</th>
<th>Family</th>
<th>Overlap*</th>
<th>Average_Fo</th>
<th>Average_non-Fo</th>
<th>EI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ftf1/Ftf2</td>
<td><em>F. oxysporum</em></td>
<td>(Niño-Sánchez et al. 2016)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>4.80</td>
<td>1.11</td>
<td>2.75</td>
</tr>
<tr>
<td>Ebr1/Ebr2</td>
<td><em>F. oxysporum</em></td>
<td>(Jonkers et al. 2014)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>5.27</td>
<td>1.56</td>
<td>2.45</td>
</tr>
<tr>
<td>Znf1</td>
<td><em>M. oryzae</em></td>
<td>(Yue et al. 2016)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>6.47</td>
<td>2.78</td>
<td>1.98</td>
</tr>
<tr>
<td>Ctf2</td>
<td><em>F. oxysporum</em></td>
<td>(Bravo-Ruiz et al. 2013)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>2.93</td>
<td>1.33</td>
<td>1.69</td>
</tr>
<tr>
<td>Fow2</td>
<td><em>F. oxysporum</em></td>
<td>(Imazaki et al. 2007)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>2.07</td>
<td>1.00</td>
<td>1.53</td>
</tr>
<tr>
<td>Dep6</td>
<td><em>A. brassicicola</em></td>
<td>(Wight et al. 2009)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>0.93</td>
<td>0.67</td>
<td>1.16</td>
</tr>
<tr>
<td>Pf2</td>
<td><em>A. brassicicola</em></td>
<td>(Jones et al. 2019)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>1.20</td>
<td>1.00</td>
<td>1.10</td>
</tr>
<tr>
<td>Art1</td>
<td><em>F. verticillioides</em></td>
<td><em>(Oh et al. 2016)</em></td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>1.00</td>
<td>0.89</td>
<td>1.06</td>
</tr>
<tr>
<td>Cita1</td>
<td><em>C. lindemuthianum</em></td>
<td><em>(Dufresne et al. 2000)</em></td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>1.07</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>Fhs1</td>
<td><em>F. graminearum</em></td>
<td>(Son et al. 2016a)</td>
<td>Zn2-C6</td>
<td>Yes</td>
<td>1.07</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>Cos1</td>
<td><em>M. oryzae</em></td>
<td>(Li et al. 2013)</td>
<td>Znf_C2H2</td>
<td>Yes</td>
<td>1.80</td>
<td>0.00</td>
<td>2.80</td>
</tr>
<tr>
<td>PacC</td>
<td><em>F. oxysporum</em></td>
<td>(Caracuel et al. 2003)</td>
<td>Znf_C2H2</td>
<td>Yes</td>
<td>2.13</td>
<td>1.00</td>
<td>1.57</td>
</tr>
<tr>
<td>Fug1</td>
<td><em>F. verticillioides</em></td>
<td><em>(Ridenour and Bluhm 2017)</em></td>
<td>AreA_GATA</td>
<td>No</td>
<td>7.27</td>
<td>1.33</td>
<td>3.54</td>
</tr>
<tr>
<td>Ren1</td>
<td><em>F. oxysporum</em></td>
<td>(Ohara et al. 2004)</td>
<td>disordered</td>
<td>No</td>
<td>3.00</td>
<td>1.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Tri10</td>
<td><em>F. graminearum</em></td>
<td>(Jiang et al. 2016)</td>
<td>Fun_TF</td>
<td>No</td>
<td>1.13</td>
<td>0.33</td>
<td>1.60</td>
</tr>
<tr>
<td>Ltf1</td>
<td><em>B. cinerea</em></td>
<td>(Schumacher et al. 2014)</td>
<td>Znf_GATA</td>
<td>Yes</td>
<td>4.00</td>
<td>2.44</td>
<td>1.45</td>
</tr>
<tr>
<td>Ndt80</td>
<td><em>U. maydis</em></td>
<td>(Doyle et al. 2016)</td>
<td>NDT80</td>
<td>Yes</td>
<td>1.73</td>
<td>1.11</td>
<td>1.29</td>
</tr>
<tr>
<td>Hap3p</td>
<td><em>F. verticillioides</em></td>
<td><em>(Ridenour and Bluhm 2014)</em></td>
<td>CBFA_NFYB</td>
<td>Yes</td>
<td>1.33</td>
<td>1.00</td>
<td>1.17</td>
</tr>
<tr>
<td>Sod1</td>
<td><em>F. oxysporum</em></td>
<td>(Wang et al. 2021)</td>
<td>SOD_Cu_Zn</td>
<td>No</td>
<td>1.47</td>
<td>1.22</td>
<td>1.11</td>
</tr>
<tr>
<td>Prf1</td>
<td><em>F. oxysporum</em></td>
<td>(Mendoza-Mendoza et al. 2009)</td>
<td>HMG_box</td>
<td>Yes</td>
<td>1.07</td>
<td>1.00</td>
<td>1.03</td>
</tr>
</tbody>
</table>
CHAPTER 3
METATRANSCRIPTOMIC COMPARISON OF ENDOPHYTIC AND PATHOGENIC FUSARIUM–ARABIDOPSIS INTERACTIONS REVEALS PLANT TRANSCRIPTIONAL PLASTICITY

3.1 Introduction

Over millions of years of coevolution, plants and microbes have established intimate relationships, forming beneficial, neutral, or antagonistic partnerships. Plant pathogens threaten agricultural production and global food security (Dean et al. 2012; Strange and Scott 2005), but beneficial microbes such as rhizobia, mycorrhizae, and endophytes limit plant pests and promote plant growth through nutrient mineralization and availability (Rashid et al. 2016; White et al. 2019). How plants recognize and react differently to friends versus foes is an intriguing topic of research.

Our understanding of plant immunity has been revolutionized by the recent increase in the breadth of genomic data available. The classical, binary view of plant immunity consists of pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) (Cui et al. 2015; Jones and Dangl 2006). Plant PTI relies on plasma membrane (PM)-localized pattern recognition receptors (PRRs),

---

which are often receptor-like proteins or kinases (RLPs or RLKs) that sense conserved microbe-associated molecular patterns (MAMPs) or damage-associated molecular patterns and induce downstream defense reactions (Bigeard et al. 2015; Dodds and Rathjen 2010; Jones and Dangl 2006; Zhou et al. 2020). Plant ETI employs an intracellular nucleotide-binding site and leucine-rich repeat domain receptors (NLRs) that recognize specific microbial effectors as a result of ongoing host-pathogen coevolution (Cesari 2018; Cui et al. 2015; Jones and Dangl 2006; Monteiro and Nishimura 2018; W. Wang et al. 2020). The blurry boundary between MAMPs and effectors prompted a new spatial immunity model based on extracellular or intracellular locations of pattern recognition (van der Burgh and Joosten 2019; W. Wang et al. 2020).

This study investigated transcriptome reprogramming of the same plant host and explored how PTI and ETI or extracellular and intracellular immunity are involved in both beneficial and antagonistic interactions. We established the Fusarium oxysporum–Arabidopsis thaliana model system, which includes an endophyte, F. oxysporum Fo47, and a pathogen, F. oxysporum Fo5176. Arabidopsis plants infected by these two F. oxysporum strains display distinctive phenotypes, with Fo5176 causing typical vascular wilt diseases and Fo47 colonizing plants endophytically without any disease symptoms. Their distinct effects on plants, combined with their minimal genetic diversity, should facilitate the identification of meaningful genotype–phenotype correlations.

In addition to being a good model system, F. oxysporum is of great agricultural importance, as it is listed among the top 10 most-researched fungal
pathogens for food production (Dean et al. 2012). Collectively, this group of filamentous fungi causes devastating vascular wilt diseases in over 100 crop species, leading to annual yield losses of billions of dollars (Ma et al. 2013). One notorious example is the recent Panama disease outbreak in banana caused by \textit{F. oxysporum} f. sp. \textit{cubense} tropical race 4 (Viljoen et al. 2020). Information accumulated over the past 10 years has provided a clear picture of compartmentalization of the \textit{F. oxysporum} genome: A core genome component that is conserved and vertically transmitted performs essential housekeeping functions, and an accessory genome that is believed to have been initially acquired horizontally mediates unique host-fungal interactions (Armitage et al. 2018; Delulio et al. 2018; Galazka and Freitag 2014; Hane et al. 2011; Ma et al. 2010, 2013; van Dam et al. 2016; Vlaardingerbroek et al. 2016a and b; Williams et al. 2016; Yang et al. 2020; Zhang et al. 2020).

Using an unbiased approach and taking advantage of two recently released high-quality genome assemblies of Fo47 and Fo5176 (B. Wang et al. 2020; Fokkens et al. 2020), we employed metatranscriptomics to dissect how Arabidopsis plants react to two \textit{F. oxysporum} isolates with distinct lifestyles during the early course of infection. We demonstrated that endophytic infection suppresses host immunity but activates plant nutrient assimilation. By contrast, pathogenic infection activated defense response but suppressed plant developmental functions. Genome comparison of the two isolates revealed unique accessory chromosomes that harbor genes enriched for fungal virulence and detoxification in Fo5176 and cell signaling and nutrient sensing in Fo47. Our
study showed that, while for both plants and *F. oxysporum*, most genes displayed a similar expression pattern during infections, a small number of genes displayed transcription plasticity between endophytic and pathogenic infections, perhaps leading to the different interaction outcomes.

**3.2 Results**

**3.2.1 A pathosystem that reveals both endophytic and pathogenic interactions**

To dissect beneficial versus pathogenic fungal-plant interactions, we inoculated Arabidopsis plants with two *F. oxysporum* strains, the beneficial (endophytic) strain Fo47 and the pathogenic strain Fo5176. The pathogenic fungus Fo5176, initially isolated in Australia (Chen et al. 2015; Thatcher et al. 2012), causes vascular wilt in several Brassicaceae plants, including *A. thaliana* (Ma et al. 2010; Thatcher et al. 2009). The endophytic strain Fo47 was originally isolated from disease-suppressing soils (Alabouvette 1999) and has been used as a biocontrol agent to prevent disease from soil-borne pathogens by inducing the production of plant secondary metabolites and priming host resistance (Aimé et al. 2013; Benhamou and Garand 2001; Benhamou et al. 2002; Olivain et al. 2006; Veloso and Diaz 2012).

We adopted a robust and reproducible root-dipping protocol to inoculate 14-day-old Columbia-0 (Col-0) plants with a suspension of fungal spores (Thatcher et al. 2012). Plants inoculated with Fo5176 developed typical yellowing
and wilting symptoms, visible at 6 days postinoculation (DPI) (Figure 3.1A). Almost all Fo5176-infected plants died within 3 weeks of inoculation (Figure 3.1B). By contrast, plants inoculated with Fo47 not only stayed healthy (Figure 3.1A and B) but also showed increased aboveground biomass (Wilcoxon rank-sum test, P < 0.001) when compared to mock-inoculated plants (Figure 3.1C), suggesting that Fo47 may have a growth-promoting effect.

By comparison with a sister species, *F. verticillioides*, accessory chromosomes were identified in both the Fo47 (B. Wang et al. 2020) and Fo5176 (Fokkens et al. 2020) genomes in addition to the 11 core chromosomes (Figure 3.2), vertically inherited from the common ancestor shared between these two sister species 10 to 11 million years ago (Ma et al. 2013). The Fo47 genome had one accessory chromosome (chromosome 7, with a length of 4.25 Mb), while the Fo5176 genome had four (chromosomes 2, 14, 16, and 18) (Figure 3.2). The combined length of accessory chromosome regions in Fo5176 was 21.63 Mb, including large segments (size >1 Mb) of chromosomes 4, 10, 11, and 13 that shared no syntenic block with the *F. verticillioides* genome. Fo47 and Fo5176 accessory chromosomes were enriched in repetitive sequences (Figure S3.1), a common property observed from all accessory chromosomes (Yang et al. 2020). Fo47 accessory genes were significantly enriched for cell signaling and nutrient sensing functions, whereas Fo5176 genes were enriched for functions relating to virulence and detoxification (Tables S3.1 and S3.2). As these two genomes share an almost identical core sequence, we hypothesized that distinct
accessory chromosomes in each genome may play important roles in the distinct phenotypic outcomes (disease versus growth promotion).

3.2.2 Reprogramming of the plant transcriptome in response to a fungal pathogen or endophyte

To examine the transcriptional regulation underlying the distinct endophytic and pathogenic interactions of the two strains (Figure 3.1), we sequenced the fungal and host plant transcriptomes from Arabidopsis plants inoculated with Fo47 or Fo5176. Infected plants were sampled at 12, 24, 48, and 96 h postinoculation (HPI), in parallel with plants mock-inoculated with water at 12 HPI as a control. We harvested root tissues for transcriptome deep sequencing (RNA-seq). Dual RNA-seq data were analyzed using an in-house pipeline to calculate the transcript levels of plant and fungal genes.

About half of all annotated Arabidopsis genes (16,544 of a total of 32,833) were differentially regulated in at least one of 18 comparisons between different timepoints for the same interaction type (12, 24, 48, and 96 HPI; 12 comparisons), between different interaction types at the same timepoint (beneficial versus pathogenic; four comparisons), or between endophytic or pathogenic interactions and the mock control at 12 HPI (two comparisons). These differentially expressed genes (DEGs) revealed several interesting patterns (Figure 3.3A).

First, we observed a strong correlation between patterns of gene expression for both treatments at the same timepoints, despite clearly distinctive
endophytic and pathogenic phenotypes (Figure 3.1). The Pearson’s correlation coefficients (PCCs) for the four comparisons between plants infected with either the beneficial or the pathogenic fungal strain at each timepoint were very high, with values of 0.95, 0.94, 0.97, and 0.96 at 12, 24, 48, and 96 HPI, respectively (Figure 3.3A, labeled in red), suggesting that a small subset of genes contribute to the observed phenotypic differences. Global clustering analysis using the 16,544 Arabidopsis DEGs yielded 24 co-expression gene clusters (Figure 3.3B; Figure S3.2; Table S3.3). A total of 10,014 genes within 12 clusters had similar expression patterns at all timepoints (Table S3.3), accounting for 60.5% of all DEGs. Since 16,289 genes were either not expressed or not changed, we concluded that 6,544 (about 20% of all) genes held answers to the transcriptional reprogramming between these two treatments.

We also observed significant transcriptional reprogramming within each interaction over time. For samples inoculated with Fo47, PCC scores decreased from 0.94 (between 12 and 24 HPI) to 0.84 (between 12 and 96 HPI) as infection progressed.

Similarly, PCC values dropped from 0.93 (between 12 and 24 HPI) to 0.81 (between 12 and 96 HPI) for Fo5176-inoculated plants. We then compared each fungal interaction pairwise at each timepoint to identify reciprocal DEGs (Figure S3.3A), yielding 1,009, 642, 59, and 403 genes that were preferentially expressed in Fo47-infected plants and 868, 1,172, 604, and 425 plant genes in Fo5176-infected plants at 12, 24, 48, and 96 HPI, respectively (Figure S3.3B). Notably, plant genes that were preferentially expressed during the endophytic
interaction were enriched in gene ontology (GO) terms such as cell cycle, cell growth, development, response to stimuli, and cellular transport. Moreover, the genes associated with each enriched GO term showed a temporal wave as the infection course progressed, with genes involved in cell cycle highly enriched at the early stages of infection but with a diminishing contribution that was consecutively replaced by genes related to development at around 24 HPI, response to stimuli at 48 HPI, and transport at 96 HPI (Figure S3.3C). Conversely, genes preferentially induced in response to the pathogenic fungus were consistently enriched in GO terms mainly related to defense responses, with no obvious underlying temporal pattern (Figure S3.3C).

Second, when compared to the mock-inoculated samples, plants inoculated with Fo47 or Fo5176 both displayed drastic transcriptional reprogramming at the earliest timepoint of this study (12 HPI), as these comparisons had the lowest PCCs of 0.85 for Fo47 and 0.83 for Fo5176. As time from initial inoculation progressed, however, the transcriptomes of all plants became much more similar, with PCCs rising to 0.98 for Fo47 and 0.96 for Fo5176 (Figure 3.3A, labeled in green). This observation indicated that the outcome of the plant-host interaction might be decided as early as 12 HPI. To begin to dissect the critical transcriptional reprogramming taking place at 12 HPI in both endophytic and pathogenic interactions, we conducted a careful analysis to identify genes that are not only differentially expressed between the two treatments but also differentially expressed relative to mock-inoculated samples. This analysis resulted in the identification of genes that were specifically
upregulated or downregulated in fungus-infected samples. These four plant gene sets consisted of 140 upregulated and 422 downregulated genes specifically in response to Fo47 infection and 286 upregulated and 767 downregulated genes in response to Fo5176 infection.

Functional analysis of these genes using GO enrichment and network analyses (Figure 3.4; Tables S3.4 through S3.7) confirmed previous observations of fungal-plant interactions but also revealed unexpected findings. As expected, we observed significant suppression of genes related to plant growth by the pathogenic strain Fo5176 (Figure 3.4A), including genes associated with the cell cycle, cell-wall organization, plant-type cell-wall biosynthesis, and microtubule-based processes. Genes upregulated early in response to Fo5176 infection were highly enriched in toxin and indole metabolism as well as small molecule biosynthesis (Figure 3.4B), possibly reflecting the initial upheaval brought upon by the infection. For the endophytic interaction, we noticed a significant suppression of immunity-related functions, including plant defense and immunity and jasmonic acid response (Figure 3.4C). This data therefore also suggested that the endophytic strain Fo47 attenuates plant defenses. Among the genes induced by the endophyte, we were pleased to see that several define a module related to nitrate metabolism and anion transport (Figure 3.4D), which would be consistent with the promotion of plant growth by Fo47 (Figure 3.1D).

A nitrate-CPK (Ca$^{2+}$-sensor protein kinase)-NLP (Nin-like protein) signaling pathway was previously reported (Liu et al. 2017) that activated the expression of 394 genes and repressed another 79 genes in response to
exogenous nitrate treatment. We examined whether our clusters of DEGs (Figure 3.3B) showed an overrepresentation of genes differentially regulated by this nitrate signaling pathway. Cluster 21 included the most downregulated genes from this pathway, with eight genes (P value = 2.56e-04, two-sided Fisher’s exact test) that were downregulated in both interactions, with stronger suppression by the endophyte (Figure S3.2; Table S3.8). Of the 394 upregulated genes in the nitrate pathway, 329 were differentially expressed in our dataset, with 251 assigned to clusters. Of those, over half were significantly enriched in five clusters: C5 (20 genes, P value = 2.12e-05), C6 (14 genes, P value = 1.07e-05), C8 (40 genes, P value = 1.46e-14), C16 (16 genes, P value = 7.37e-04), and C23 (38 genes, P value = 6.62e-21). These 251 genes, representing a majority of the genes upregulated in the nitrate signaling pathway, were induced by both the endophyte and the pathogen but exhibited stronger responses in the context of endophytic inoculations (Figure S3.2).

Notably, cluster C23 included NLP1, encoding a transcription factor involved in the nitrate-CPK-NLP signaling pathway (Liu et al. 2017), as well as NITRATE TRANSPORTER2.1 and 2.2 (NRT2.1, NRT2.2), NITRATE REDUCTASE1 (NIA1), and NITRITE REDUCTASE1 (NIR1), all major components of the pathway that were upregulated when compared to the mock-inoculated sample (Figure S3.4). Of 24 previously reported transcription factors that control transcriptional regulation of nitrogen-associated metabolism and growth (Gaudinier et al. 2018), 16 were assigned to our clusters (Table S3.9), including WUSCHEL RELATED HOMEOBOX14 (WOX14) and LOB DOMAIN-
CONTAINING PROTEIN4 (LBD4) in cluster C23. Collectively, this analysis suggests that nitrogen signaling is involved in the *F. oxysporum*–Arabidopsis interaction and the endophyte may enhance the nitrogen signal and hence change the course of the plant response.

### 3.2.3 Perturbation of plant immunity

To better understand how the endophyte and the pathogen perturb plant immunity via shared and distinct responses, we carefully investigated the 24 co-expression clusters, based on their global patterns of expression. Four clusters, C7, C15, C16, and C21, showed enrichment (P < 0.05) for GO terms related to immunity and defense responses; the same clusters also lacked GO terms related to development (Figure 3.3C). Compared to plant PTI and ETI networks (consisting of 1,856 PTI-related and 1,843 ETI-related genes) previously constructed using a machine learning algorithm (X. Dong et al. 2015), three clusters (C15, C16, and C21) were enriched for both PTI and ETI genes, whereas cluster C7 was primarily enriched in PTI response genes (two-sided Fisher’s exact test, P < 0.05) (Table S10 and 11). This suggests a transcriptional plasticity of plant immunity in responding to the endophytic and pathogenic *F. oxysporum*.

#### 3.2.3.1 Conserved immune response toward an endophyte and a pathogen

Cluster C15 comprised 1,290 genes and was the largest immunity-related cluster, with nearly identical plant transcriptome responses following Fo47 and
Fo5176 inoculation. Indeed, genes from cluster C15 were initially strongly upregulated at 12 HPI in both interactions and gradually returned to an expression level comparable to that of mock-inoculated plants as infection progressed (Figure 3.5A). C15 was most significantly enriched in PTI genes (P value = 2.66e-72), reflecting the general plant perception of fungal signals derived from both pathogenic and symbiotic organisms (e.g., MAMPs). Cluster C15, indeed, included many immunity-related genes involved in fungal perception, signal transduction, and transcriptional regulation, including ERECTA, an RLK that regulates stomatal patterning and immunity (Sopeña-Torres et al. 2018), RECOGNITION OF PERONOSPORA PARASITICA5 (RPP5), which encodes a putative NLR protein that confers resistance to Peronospora parasitica (Noel et al. 1999; Parker et al. 1997), RESPONSIVE TO DEHYDRATION 21A (RD21A), which encodes a cysteine proteinase with peptide ligase and protease activity that is involved in immune responses against the necrotrophic fungal pathogen Botrytis cinerea (Lampl et al. 2013), and NUCLEAR FACTOR Y, SUBUNIT B3 (NF-YB3), which encodes a transcription factor activated by endoplasmic reticulum stress responsible for the regulation of stress responses (Liu and Howell 2010).

Cluster C15 was also highly enriched in genes with functions related to the chloroplast/plastid (P value = 9.2e-89) (Figure 3.5B; Table S3.12). An organelle essential for plant photosynthesis, chloroplasts have recently come to the forefront as key players in plant immune responses (Gohre et al. 2012; Serrano et al. 2016), possibly functioning as a signaling hub that links the initial
recognition of diverse pathogens at the PM and signal transduction to the nucleus to orchestrate transcriptional reprogramming in response to infection (Chan et al. 2016; de Souza et al. 2017; de Torres Zabala et al. 2015; Liu 2016; Medina-Puche et al. 2020; Wang et al. 2016). It is unlikely that chloroplast-related genes from cluster C15 represent artifacts caused by the manipulation of roots in the light during harvesting, as these genes were expressed at low levels in mock-inoculated plants, although they were subjected to the same inoculation and harvesting procedure. Our observations are also consistent with a previous report in which strain Fo5176 was shown to induce the expression of Arabidopsis genes normally involved in photosynthesis in root tissues at 1 DPI (Lyons et al. 2015). Pathogens may thus interfere with host chloroplast/plastid functions to manipulate host immunity in their favor. We know very little about the possible role played by chloroplasts during endophytic colonization.

3.2.3.2 Stronger induction of plant PTI responses by the pathogen

We hypothesized that a subset of plant immune responses against the pathogen and endophyte would differ, given their distinctive phenotypes, even though most clusters showed the same pattern during endophytic and pathogenic responses. Indeed, three immunity-related clusters, C7, C16, and C21, exhibited distinct patterns between the pathogen and the endophyte (Figure 3.5A). Cluster C7 (422 genes), which was primarily enriched in genes associated with PTI, exhibited a stronger induction by Fo5176 infection than by Fo47, despite being induced by both strains (Figure 3.5A). Several GO terms were
shared between clusters C7 and C15, such as chloroplast/plastid-related functions (Figure 3.5B), possibly reflecting fine-tuning of the initial recognition of conserved fungal signals.

Genes uniquely induced by the pathogen included *PEROXIDASE37* (*PRX37*), encoding a putative apoplastic peroxidase that generates H$_2$O$_2$ primarily in the vascular bundles for host defense (Pedreira et al. 2011), and *PENETRATION2* (*PEN2*), encoding an atypical tyrosinase required for broad-spectrum resistance to filamentous plant pathogens (Fuchs et al. 2016). Also included in cluster 7 were genes with dual functional roles in immunity against different pathogens. For instance, *PATATIN-LIKE PROTEIN2* (*PLP2*) promotes cell death and facilitates *Botrytis cinerea* and *Pseudomonas syringae* infection in Arabidopsis (La Camera et al. 2005), whereas it confers host resistance to cucumber mosaic virus (Camera et al. 2009). *KUNITZ TRypsIN INHIBITOR1* (*KTI1*), a trypsin inhibitor referenced as an antagonist involved in the negative regulation of programmed cell death that mediates susceptibility in *Erwinia carotovora* but has an opposite function in *Pseudomonas syringae pv tomato* DC3000 (Li et al. 2008).

### 3.2.3.3 Suppressed plant immunity in the presence of the endophyte

In contrast to clusters C7 and C15, both clusters C16 (615 genes) and C21 (766 genes) exhibited stronger suppression of expression by the endophyte (Figure 3.5A). We also observed a unique and specific suppression of plant immunity by the endophyte from the GO term enrichment and network analyses
at 12 HPI described above (Figure 3.4C). While clusters C7 and C15 showed minimal overlap of enriched GO terms, clusters C16 and C21 shared many terms, including signal perception and transduction, protein-protein interaction, and PM localization (Figure 3.5B). Of 70 genes identified as contributing to danger-sensing and signaling systems (Zhou and Zhang 2020), we detected 12 genes in cluster C16 and another 12 genes in cluster C21 (Table 3.1; Table S3.13). For instance, PRRs and downstream components in cluster C16 include *EFR, BAK1, LYK5*, and *CERK1* and, in cluster C21, *PEPR1, PEPR2, FERONIA*, and *RBOHD*. NLRs and downstream signaling components in cluster C16 include *ADR1-L1/ADR1-L2* and *RPM1, PAD4*, and *RPS4* in cluster C21. In summary, a strong suppression of diverse immunity-related genes is unique to the endophytic interaction, suggesting that modulation of plant immunity may contribute to the different outcomes.

**3.2.4 PTI- and ETI-sensing genes induced by the endophyte and the pathogen**

Overall, we observed strong host immune responses, when challenged with either the endophyte or the pathogen, involving complex signal perception and signal transduction cascades. Distinct responses included the suppression of plant growth and the induction of plant defenses by the pathogenic strain Fo5176 and the attenuation of host immunity with the concomitant induction of nitrogen metabolism by the endophytic strain Fo47 (Figure 3.4). To further dissect the plant immunity pathways involved in these two interactions, we conducted a
systematic analysis of the expression profiles of genes encoding RLPs or RLKs and NLR proteins (Table 3.1; Tables S3.14 and S3.15).

### 3.2.4.1 RLP and RLK genes

The Arabidopsis genome encodes 533 RLPs and RLKs, as determined by the MAPMAN Mercator annotation (Schwacke et al. 2019). Of those, 311 were assigned to our clusters of DEGs (Table S3.14). In addition to the two immunity clusters, C16 (37 genes, P value = 4.76e-11) and C21 (30 genes, P value = 2.11e-05), whose expression is repressed by the endophyte, these RLP and RLK genes were also enriched in clusters C11 (14 genes, P value = 6.31e-04) and C12 (11 genes, P value = 6.31e-04). Their expression appeared to be repressed by both the endophyte and the pathogen to varying degrees. Characterized defense-related RLK genes include *SUPPRESSOR OF BIR1 (SOBIR1)* and *EFR* in cluster C16 and *RESISTANCE TO FUSARIUM OXYSPORUM1 (RFO1)* and *PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASE1 (PERK1)* in cluster C21. *RFO1* encodes a protein that confers a broad-spectrum resistance to *Fusarium* spp. (Diener and Ausubel 2005). Cluster C16 also included genes encoding LysM receptor-like kinases CERK1 and LYK5 (Cao et al. 2014), which are essential for the perception and transduction of the chitin oligosaccharide elicitor. Although not significantly enriched, 26 RLP/RLK genes grouped in cluster C15, including *FLG22-INDUCED RECEPTOR-LIKE KINASE1 (FRK1)* and *HERCULES RECEPTOR KINASE1 (HERK1)*, which are activated in response to both pathogenic and endophytic *F. oxysporum* strains.
3.2.4.2 NLR genes

Many plant NLRs are commonly identified as resistance proteins that act as surveillance molecules recognizing pathogen effectors that target the host machinery. In accordance with the "guard" model, NLRs then trigger an ETI response (Cao et al. 2014). The Arabidopsis genome encodes 160 NLR proteins (Baggs et al. 2020), of which we identified 119 genes among our Dataset of DEGs (84 were assigned to clusters). Among these NLR genes, about 40% clustered across four immunity clusters, cluster C21 (18 genes), C16 (14 genes), C15 (11 genes), and C7 (four genes), and were uniquely enriched in clusters 21 (P value = 2.58e−07) and 16 (P value = 6.82e−06), both of which include genes specifically repressed in the endophyte (Table S3.15). The 11 NLR genes in cluster C15 included five known resistance genes (AT1G63880, AT1G61190, RPP4, RPP5, and RPP8) against oomycete and fungal pathogens (Goritschnig et al. 2012; McDowell et al. 2005; Staal et al. 2006; van der Biezen et al. 2002).

Notably, most of the NLR genes that were enriched in endophyte-suppressed clusters C16 and C21 are not functionally characterized. Nevertheless, characterized NLRs represented by genes in cluster C16 included two apoplast/chloroplast-localized ADP-binding immune receptors (ADR1-L1 and ADR1-L2) (Dong et al. 2016). Also belonging to cluster C16 were the two effector-induced resistance genes LAZARUS5 (LAZ5) and HOPZ-ACTIVATED RESISTANCE1 (ZAR1) (Barbacci et al. 2020; Baudin et al. 2017), conferring resistance to a Pseudomonas syringae strain expressing the AvrRPS4 and Hop effectors, respectively. NLR genes in cluster C21 included disease resistance
proteins RPS3, RPS4, and RPS6, which provide specific resistance against *P. syringae* pv. tomato carrying the avirulence genes *AvrRPS3*, *AvrRPS4*, and *AvrRPS6*, respectively (Bisgrove et al. 1994; Kim et al. 2009; Narusaka et al. 2009). Repression of the expression of these NLR genes by the endophyte again supports the idea that the danger-sensing and signaling systems underlying the responses to Fo47 and Fo5176 are distinct.

### 3.2.5 Accessory chromosomes in two strains harbor genes induced during infection and with distinct biological functions

The distinct plant responses at both the phenotypic and transcriptome levels, resulting from inoculation with the two *F. oxysporum* isolates, is no doubt related to genomic differences between the two strains. Even though both strains belong to the same species complex and share an approximately 46-Mb core genome, each strain also carries distinct accessory chromosomes (Figure 3.2). The Fo47 accessory chromosome 7 harbored 1,299 predicted genes (7.2% of total predicted genes) (Table S3.16), 757 of which were expressed and 160 were strongly induced (false discovery rate [FDR] < 0.05) during one or more timepoints of the infection.

To explore the function of these accessory genes encoded in the endophytic strain Fo47, we next analyzed the functional domains they encoded. After excluding genes that encoded proteins with transposase-like domains or unknown domains, we highlighted five enriched PFAM domains, *i.e.*, the regulator of G-protein signaling domain (PF00615), nitric oxide (NO)-binding
membrane sensor involved in signal transduction (PF03707), basic leucine zipper (bZIP) transcription factor (PF00170), chromodomain (PF00385), and bromodomain (PF00439) (Figure 3.6A; Figure S3.5A). The enrichment of functional domains involved in cell signaling and the apparent lack of enrichment for domains related to virulence suggested that the Fo47 accessory chromosome contains genes with functions that are well-suited to a non-pathogenic lifestyle. Transcriptome analysis showed that nine Fo47 genes encoding proteins with the bacterial signaling protein domain (PF03707) are most highly induced at 24 and 48 HPI. Domain PF03707 plays a role in sensing oxygen, carbon monoxide, and NO (Galperin et al. 2001). The Fo47 genome has the highest number of genes encoding a PF03707 domain, with nine genes, compared to other filamentous fungi, such as Aspergillus nidulans (1), Neurospora crassa (1), Pyricularia grisea (1), F. graminearum (2), F. verticillioides (2), and F. solani (3) (Cuomo et al. 2007; Dean et al. 2005; Galagan et al. 2005; Ma et al. 2010), as well as other F. oxysporum species complex members (three to about seven, average of four) (Delulio et al. 2018). Notably, six of the nine genes reside on accessory chromosome 7, making it a major contributor to the expansion of this gene family within this strain (Figure S3.6).

By contrast, accessory chromosomes and regions from the pathogenic strain Fo5176 contributed 4,136 predicted genes (23% of total predicted genes) (Table S3.17), of which 3,502 were expressed and 1,140 were strongly induced during one or more timepoints of the infection. Genes located in accessory regions in Fo5176 encoded proteins that were enriched for 42 PFAM domain...
terms. We noticed six PFAM domains that were highly enriched at different stages of the infection course and whose encoding genes were highly expressed, namely, cysteine-rich secretory protein family (PF00188), calpain family cysteine protease (PF00648), peptidase M16 (PF16187), poly (ADP-ribose) polymerase regulatory domain (PF02877), cyclin C-terminal domain (PF02984), and WGR domain (PF05406) (Figure 3.6B; Figure S3.5B). Most of these domains are likely associated with microbial pathogenesis or detoxification; their associated genes were induced during infection (Figure 3.6B). In particular, members of the cysteine-rich secretory protein (CAP) superfamily (PF00188) have a wide range of biological activities, including fungal virulence, cellular defense, and immune evasion (Schneiter and Di Pietro 2013). For example, the *F. oxysporum* CAP family protein Fpr1 is a PR-1-like protein that is important for the virulence of strain Fol4287 (Prados-Rosales et al. 2012). The Fo5176 genome encodes 15 CAP family members, significantly more than Fo47 and other comparable fungal species (average five members) (Figure S3.7). Phylogenetic analysis of CAP proteins showed that four CAP members formed a core group shared by Fo5176 and Fo47. However, a separate clade of six CAP family proteins was expanded in Fo5176 and encoded by Fo5176 accessory chromosomes (Figure S3.7). These results highlight the distinctive functions of and roles played by accessory chromosomes in the nonpathogenic strain Fo47 and the pathogenic strain Fo5176. These differences might provide the mechanistic basis that allows Fo47 to specialize in host-sensing and benefit its host as an endophyte, while the pathogenic Fo5176 specializes in host invasion and killing.
3.3 Discussion

We performed a comparative study of infection by an endophytic (Fo47) and a pathogenic (Fo5176) strain of *F. oxysporum* in the context of the *F. oxysporum*–Arabidopsis system that revealed the transcriptional plasticity of plant defense responses. The pathosystem we developed combines the extensive knowledge of plant immunity in Arabidopsis and one of the most damaging fungal pathogens for agriculture, *F. oxysporum*. Strain-specific interactions with a common host are likely dictated by the accessory chromosomes from each *F. oxysporum* genome, which allows a comparative study that minimizes genetic differences between strains to address the underlying mechanism that results in distinct phenotypes (growth promotion or disease or even death). Up to 50% of crop losses in the United States can be attributed to soil-borne pathogens (Raaijmakers et al. 2009), and our results provide a foundation for the development of technologies to enhance plant health, sustain a healthy ecosystem, and feed a continuously growing human population.

We employed comparative metatranscriptomics over the course of early infection to systematically capture temporal transcriptional changes in both the host and the interacting microbes. Our results may be summarized along four main axes, as illustrated in Figure 3.7.

First, host transcriptional responses were strikingly similar at all timepoints, regardless of the obvious phenotypic differences seen after infection of Arabidopsis plants by Fo47 and Fo5176. Of all clusters of DEGs, cluster C15
exhibited a strong and early induction of genes, followed by a return to an expression level comparable to that of control samples. This cluster captured 26 Arabidopsis RLP and RLK genes as well as 269 PTI and 159 ETI response genes, suggesting that both strains initially elicit a similar MAMP response, which would not be surprising, as they belong to the same species.

Second, our data revealed rapid transcriptional reprogramming at the beginning of the interactions. While most plant genes exhibited a common expression pattern during the two infections, a small subset of plant genes displayed divergent gene expression profiles. By far the most striking difference was observed at 12 HPI, when the GO biological processes for genes uniquely induced by Fo47 or Fo5176 reflected almost opposite responses. The endophytic strain Fo47 stimulated nitrogen metabolism and suppressed host immunity, whereas the pathogenic strain Fo5176 stimulated host immune responses and toxin metabolism but repressed functions related to plant growth and development. We propose that this distinct expression profile, reflected in the early divergence of the host transcriptome, is the result of plasticity of the host transcriptome when facing an endophyte or a pathogen.

We hypothesize that the perception by the host of distinct fungal signals occurs shortly after inoculation and is followed by the rapid activation of downstream signaling cascades. Our results also stress the importance and necessity of sampling early during the establishment of a fungal-host interaction to better capture the full extent of the underlying temporal dynamics.
Third, Fo47 inoculation resulted in suppression of genes related to plant defense and induced genes related to plant growth, in agreement with the tradeoff between growth and defense. It has been reported that plants can channel nitrogen resources towards production of defense-related compounds when confronted with pathogens (Ullmann-Zeunert et al. 2013). For instance, allele polymorphism at the single locus ACCELERATED CELL DEATH6 (ACD6) can dictate a distinct difference between growth and defense among different Arabidopsis ecotypes (Todesco et al. 2010). Further characterizing the F. oxysporum–Arabidopsis pathosystem should illuminate the mechanism or mechanisms by which nutrients are allocated in relation to plant defense.

Finally, we observed an agreement between plant infection phenotypes and distinctive gene functions associated with fungal accessory chromosomes. While upregulated fungal accessory genes were primarily enriched in proteins with roles in cell signaling and nutrient transport in the endophyte Fo47, they were enriched for virulence and detoxification in the pathogen Fo5176, likely contributing to the contrasting phenotypes of plants infected by these two F. oxysporum strains.

In conclusion, time-resolved comparative metatranscriptomics can be used to characterize transcription regulation when the model plant Arabidopsis is challenged with an endophyte and a pathogen of the same fungal species. We showed both the conservation and plasticity of the plant and fungal transcriptomes and how they may relate to the distinctive genomic features associated with each fungal genome. The Arabidopsis and F. oxysporum
pathosystem developed here is likely to become an ideal system to characterize plant recognition and response mechanisms against soil-borne root fungi. We believe this system will be pivotal in enriching our understanding of the molecular mechanisms necessary to enhance vascular wilt resistance not only in Arabidopsis but also in crops that are under threat by *F. oxysporum* pathogens.

3.4 Materials and Methods

3.4.1 Plant and fungal growth

*F. oxysporum* Fo5176 and Fo47 were routinely cultured on potato dextrose agar (BD) at 28°C under a 12-h light and 12-h dark photoperiod. Fungal spores were collected from 5-day-old cultures in potato dextrose broth (BD) by passing the liquid culture through a double layer of sterile cheesecloth, followed by centrifugation of the flow-through at 3,000 × g for 15 min at room temperature. Fungal spores were mixed with an appropriate volume of sterile deionized water to prepare the spore suspension (concentration 1 × 10⁶ spores/ml) for infection assays.

3.4.2 Plant infection assay

Seeds of *A. thaliana* accession Col-0 were obtained from the Arabidopsis Biological Resource Center (Ohio State University) and were surface-sterilized in 1 ml of 70% (vol/vol) ethanol three times, 5 min each, followed by one wash with 50% (vol/vol) bleach for 5 min. After removing the bleach solution, seeds were
rinsed with 1 ml of sterile distilled and deionized water and were stratified for 3 to 4 days in darkness at 4°C. Seeds were planted into pots filled with an autoclaved mixture of fine-grain play sand/MetroMix 360/vermiculite in a 1:2:1 ratio, were watered with distilled deionized water, and were covered with a clear plastic lid to retain a high humidity for 3 days in the growth chamber at 24°C, with 14 h of light and 10 h of dark and a light intensity (T8 fluorescent and incandescent bulbs) ranging from 89 to 94 μmol·m⁻²·s⁻¹. After 3 days, the plastic lid was removed and seedlings were allowed to grow for 11 additional days, prior to inoculation with *F. oxysporum* microconidia. Plants were 14 days old at the time of inoculation and had at least four fully expanded true leaves. For *F. oxysporum* infection, the roots of 14-day-old Arabidopsis plants were dipped for 30 s in a 1 × 10⁶ fungal spores/ml suspension of Fo5176 or Fo47 or in sterile distilled H₂O for the mock control. Inoculated plants were planted in autoclaved potting mix and moved to a growth chamber set to 28°C with the same photoperiod as above.

### 3.4.3 RNA preparation, sequencing, and data analysis

Roots from infected plants at 12, 24, 48, and 96 HPI were harvested from five plants per treatment and timepoint for total RNA isolation. For control samples, roots from the same number of control plants were collected at 12 HPI. Fungal cultures from Fo5176 and Fo47 fungal mycelia were harvested after 5 days from liquid cultures for RNA extraction. Three biological replicates were produced for each treatment. Total RNA was extracted using the ZR Soil/Fecal RNA microprep kit (Zymo Research), following manufacturer protocol, and the
RNA quantity and quality were assessed using a NanoDrop 2000 and Agilent 2100 Bioanalyzer. Illumina TruSeq stranded mRNA libraries were prepared and sequenced on an Illumina HiSeq2000 platform at the Broad Institute. One replicate each for infected plant samples inoculated with Fo47 at 12 HPI and Fo5176 at 24 HPI failed, as did one replicate for Fo47 and Fo5176 mycelia samples; these four conditions are therefore only represented by two replicates and were used for downstream processing and analysis.

Paired-end RNA-seq reads were first assessed for quality by FastQC 0.10.1 (Andrews 2010). RNA-seq data were analyzed using the HISAT, StringTie, and DESeq2 pipelines (Love et al. 2014; Pertea et al. 2016). Briefly, reads were mapped to reference genomes of Arabidopsis (annotation version Araport11 [Cheng et al. 2017]), Fo5176 (Fokkens et al. 2020), and Fo47 (B. Wang et al. 2020) using HISAT2 version 2.0.5 (Kim et al. 2015). Mapped reads were used to quantify the transcriptome by stringTie version 1.3.4 (Pertea et al. 2015). Read count normalization and differential gene expression analysis were conducted using DESeq2 version 1.27.32 with a maximum FDR of 0.05 (Love et al. 2014). Corrplot version 0.84 was used to visualize the correlation in gene expression profiles between different conditions. Read counts of DEGs were first averaged per condition and were then normalized by log transformation as $\log_2(\text{normalized read count} + 1)$ and, then, correlations were calculated. Clustering analysis on per-condition averaged, log-transformed, and Z-scaled read counts was performed using the K-means clustering algorithm ‘Lloyd’ (R function K-means) and then visualized in ggplot2 version 3.3.0.
3.4.4 Functional analysis and visualization

GO enrichment analysis (plant GO slim) of Arabidopsis gene clusters and reciprocal DEG analysis were conducted with the singular enrichment analysis (SEA) tool of agriGO v2 (Du et al. 2010; Tian et al. 2017), using the Arabidopsis TAIR10 annotation. We applied a hypergeometric test, combined with the Hochberg (FDR) multi-test adjustment method to discover enriched GO terms at a significance level of 0.01 with a minimum of three mapping entries. Comparisons of different enrichment results were performed using cross-comparison of SEA (SEACOMPARE). We generated PFAM annotations for *F. oxysporum* 5176 and 47 by InterproScan, following a standard annotation pipeline (Jones et al. 2014). PFAM enrichment in proteins encoded by fungal genes was performed in TBtools v.1.0692 (Chen et al. 2020) using Fisher’s exact test with FDR < 0.05. The expression heatmaps of PFAM domain were visualized by TBtools. We performed a custom analysis in Metascape (Zhou et al. 2019), with the options minimum overlap of 3, P value cutoff of 0.01, and minimum enrichment of 1.5, for the discovery of GO term enrichment and network visualization of Arabidopsis DEGs at 12 HPI. The top five terms (with the smallest P values) were selected, and the terms that shared the gene entries (forming edges) were visualized. The visualization was further polished in Cytoscape version 3.8.0 (Shannon 2003).
3.4.5 Synteny and phylogenetic tree construction

Synteny was detected by the Basic Local Alignment Search Tool for nucleotides (BLASTN), with parameters above 50-kb coverage and 98.5% sequence identity, and were visualized as a Circos plot (Krzywinski et al. 2009). MEGA v. 7.0 (Kumar et al. 2016) was used to generate maximum likelihood phylogeny trees for proteins with the PF03707 and PF00188 domain with the JTT (Jones, Taylor, and Thorton) amino acid substitution model. Statistical support for phylogenetic grouping was assessed by 1000 bootstrap re-samplings.

3.5 References


Figure 3.1 Compatible versus incompatible Arabidopsis interaction with an endophytic (Fo47) versus a pathogenic (Fo5176) *Fusarium oxysporum* strain.

*F. oxysporum* Fo5176 causes chlorosis on Arabidopsis Col-0 plants, while Fo47-infected plants and plants mock-inoculated with water do not exhibit any symptoms. Photographs were taken at 7 days postinoculation (DPI) and representative plants are shown. B, Survival analysis assay illustrating the survival rates of Arabidopsis plants mock-inoculated with water or infected with Fo47 or Fo5176 at six timepoints, from 4 to 28 DPI. Ninety plants were assayed per treatment. C, Summary of shoot dry biomass of Arabidopsis plants mock-inoculated with water or infected with Fo47 or Fo5176 at 6 DPI. Statistical significance was determined by Kruskal–Wallis and Wilcoxon rank-sum tests. Asterisk indicates statistical significance at P < 0.001. Thirty-six plants were assayed per treatment.
Figure 3.2 Comparative genomics reveals unique sets of accessory chromosomes in *Fusarium oxysporum* Fo47 and Fo5176.

Track a shows distribution of karyotypes of assembled chromosomes; track b, GC density; track c, density of transposable elements (TEs) calculated in 10-kb windows; track d, gene density calculated in 100-kb windows. Track e shows syntenic blocks. Relationships are shown through linking syntenic block genes (gene number >10) in each genome pair. Core chromosomes can be identified through synteny between *F. verticillioides* and each *F. oxysporum* strain, whereas accessory chromosomes and regions show no or reduced synteny. Chromosomes 2, 14, and 18, and large segments (size >1 Mb) of chromosomes 4, 10, 11, 13, and 16 in Fo5176, and chromosome 7 in Fo47 show no synteny with the *F. verticillioides* genome and are thus identified as accessory regions, characterized by their high TE density and low gene density.
Figure 3.3 Expression profiling of Arabidopsis roots inoculated with an endophytic versus a pathogenic *Fusarium oxysporum* strain.

A. Extent of correlation between Arabidopsis differentially expressed genes (DEGs) in Fo47- and Fo5176-infected plants across the different timepoints at 12, 24, 48, and 96 h postinoculation (HPI). Correlation coefficients (converted to percentages) are scaled to the sizes and colors of the circles. B. Gene ontology (GO) enrichment of 24 gene clusters from K-means clustering of Arabidopsis DEGs. The color scale of the heatmap represents the significance level of GO enrichment for biological processes related to stimuli response and developmental processes, expressed as $-\log_{10}$(false discovery rate [FDR]). Four clusters, C7, C15, C16, and C21, highly enriched for stimuli responses and deprived of developmental regulation, are highlighted in red and defined as immunity clusters.
Figure 3.4 A summary of transcriptomic changes occurring at 12 h postinoculation.

Gene ontology (GO) enrichment analysis and visualization were performed on four datasets representing up- and downregulation by Fo47 and Fo5176, respectively. Nodes represent the GO categories with enrichment, while edges exist when two GO categories share the same genes. The nodes labeled in the same color represent the GO terms that belong to a master term as labeled. The size of the nodes is scaled to the number of genes within each GO term in each figure section. (A) Pathogen suppression: Arabidopsis genes with expression when infected by Fo47 are smaller than when infected by Fo5176 and when infected by water (Fo47 < Fo5176 and Fo47 < mock). (B) Pathogen induction: Fo5176 > Fo47 and Fo5176 > mock. C, Endophyte suppression: Fo47 < Fo5176 and Fo47 < mock. D, Endophyte induction: Fo47 > Fo5176 and Fo47 > mock.
Figure 3.5 Expression and gene ontology (GO) enrichment of immunity gene clusters.

(A) Expression profile of immunity gene clusters. Color scale indicates the correlation of expression between genes and the cluster centroids. Genes that were removed from the clusters before functional analysis, due to the expression correlation with centroid lower than (or equal to) 0.8, are shown in gray. Enrichment of pathogen-associated molecular pattern-triggered immunity and effector-triggered immunity genes (X. Dong et al. 2015) is indicated (in all labeled cases, $P < 1E^{-7}$, by Fisher's exact test). Gene number within the four clusters is as follows: $C7 = 422$, $C16 = 615$, $C21 = 766$. (B) GO enrichment analysis of immunity gene clusters for biological processes, molecular functions, and cellular components. Color scale of the heatmap represents the false discovery rate. Stimuli responses, developmental processes, and redundant GO terms were removed.
Figure 3.6 Distinct biological functions for induced accessory chromosome (AC) genes in the endophyte Fo47 and the pathogen Fo5176.

Fold enrichment refers to the ratio of the proportion of genes on the ACs with a specific term over the proportion of genes in the whole genome with a particular term (adjusted P value < 0.05). (A) Significantly induced Fo47 accessory genes are represented in five enriched PFAM domains, including regulator of G-protein signaling domain (PF00615), nitric oxide–binding membrane sensor involved in signal transduction (PF03707), bZIP transcription factor (PF00170), chromo domain (PF00385), and bromodomain (PF00439)-containing proteins. (B) In Fo5176, six PFAM domains are significantly enriched and induced at different stages of the infection course, i.e., cysteine-rich secretory protein family (PF00188), Calpain family cysteine protease (PF00648), peptidase M16 (PF16187), poly(ADP-ribose) polymerase regulatory domain (PF02877), cyclin C-terminal domain (PF02984), and WGR domain (PF05406).1078/ Molecular Plant-Microbe Interactions
Figure 3.7 Model of transcriptomic plasticity in beneficial and antagonistic plant–fungal interactions.

Molecular response of Arabidopsis thaliana plants challenged with an endophyte Fo47 and a pathogen Fo5176, two Fusarium oxysporum strains sharing a core genome of about 46 Mb, in addition to their unique accessory chromosomes. Distinct responding genes depict the transcriptional plasticity, as the pathogenic interaction activates plant stress responses and suppresses plant growth and development-related functions, while the endophyte attenuates host immunity but activates plant nitrogen assimilation. The differences in reprogramming of the plant transcriptome are linked to accessory genes encoded by the two closely related fungal genomes. NO = nitric oxide, TF=transcription factors; MAMPs = microbe-associated molecular patterns, PRRs = pattern recognition receptors, NLR = nucleotide-binding site and leucine-rich repeat domain receptors, and JA = jasmonic acid.
Table 3.1 Genes that play important roles in danger sensing and signaling suppressed by the endophytic interaction

<table>
<thead>
<tr>
<th>PRRs and downstream components</th>
<th>RLCKs</th>
<th>MAP kinase cascades</th>
<th>NLRs and downstream signaling</th>
<th>SA biosynthesis and signaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster C16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFR</td>
<td>PBL1</td>
<td>MEKK1</td>
<td>ADR1-L1</td>
<td>CBP60g</td>
</tr>
<tr>
<td>BAK1</td>
<td>PBL2</td>
<td>MPK11</td>
<td>ADR1-L2</td>
<td>SARD1</td>
</tr>
<tr>
<td>LYK5</td>
<td></td>
<td></td>
<td>LAZ5</td>
<td>ZAR1</td>
</tr>
<tr>
<td>CERK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOBIR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster C21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEPR1</td>
<td></td>
<td>MPK3</td>
<td>RPM1</td>
<td>CAMTA3</td>
</tr>
<tr>
<td>PEPR2</td>
<td></td>
<td></td>
<td>PAD4</td>
<td>NPR1</td>
</tr>
<tr>
<td>FER</td>
<td></td>
<td></td>
<td>RPP1</td>
<td>NPR3</td>
</tr>
<tr>
<td>RBOHD</td>
<td></td>
<td></td>
<td>RPS4</td>
<td></td>
</tr>
<tr>
<td>PERK1</td>
<td></td>
<td></td>
<td>RPS3</td>
<td></td>
</tr>
<tr>
<td>RFO1</td>
<td></td>
<td></td>
<td>RPS6</td>
<td></td>
</tr>
</tbody>
</table>

PRRs = pattern recognition receptors; RCLKs = receptor-like cytoplasmic kinases; MAP = mitogen-activated protein; SA = salicylic acid.
CHAPTER 4
UNVEILING THE TRANSCRIPTIONAL DYNAMICS OF ARABIDOPSIS-
FUSARIA M OXYSPORUM INTERACTIONS THROUGH SINGLE-NUCLEUS
RNA SEQUENCING

4.1 Introduction

Plant health is crucial for maintaining a healthy ecosystem and ensuring food security for a rapidly growing human population. Unfortunately, some pathogens target the host vasculature, causing systemic infections and posing a serious threat to plants. *Fusarium* wilt is a notorious disease caused by the vascular colonization of the pathogen *Fusarium oxysporum*, which attacks nearly every crop except for cereals (Berendsen et al. 2012; Gordon 2017). Despite extensive efforts to control this disease, a highly aggressive clone known as tropical race 4 rapidly spreads and threatens to wipe out the world's industrial Cavendish banana (*Musa acuminata*) production (Ordonez et al. 2015; Viljoen et al. 2020). These have highlighted the devastating impact of *Fusarium* wilt on crops and the urgent need for effective prevention and management strategies.

Developing effective strategies to manage plant diseases such as *Fusarium* wilt requires a comprehensive understanding of the molecular basis of pathogenesis and plant defense mechanisms. Soil-borne pathogens like *F. *

*F. oxysporum* invade the plant through the root epidermis and propagate toward the root vasculature with the help of effector proteins (Rep et al. 2004; Tintor et al. 2020). These effectors are secreted into the apoplastic space, and some are translocated inside plant cells, modulating various host functions, including defense signaling and cell wall composition (Toruño et al. 2016). As the pathogen colonizes the plant's vascular system, it blocks water and nutrient transport, leading to devastating wilt disease (Gordon 2017). The *F. oxysporum* can infect over a hundred crops; however, individual isolates have a narrow host range and are categorized into *forma speciales* (van Dam et al. 2016; Edel-Hermann and Lecomte 2019). This host-specific pathogenicity is linked to accessory genomic regions that encode distinct combinations of virulence factors, which can be horizontally transferred between isolates, distinctive from core chromosomes which are vertically transmitted (Ma et al. 2010; van Dam et al. 2016; Yang et al. 2020; Yu et al. 2023).

To defend against microbic invaders, plants have evolved a sophisticated defense system that relies on pattern recognition receptors to detect pathogen-associated molecular patterns (PAMPs) (Jones and Dangl 2006; Ngou et al. 2021). Wall-associated kinases (WAKs), which act as cell wall integrity sensors, are among the receptors upregulated in response to *F. oxysporum* infection (Wang et al. 2022). Upon PAMP recognition, plants activate downstream responses, including the production and signaling of two major phytohormones, salicylic acid (SA) and jasmonic acid (JA) (Nishad et al. 2020). SA primarily combats biotrophs and induces long-lasting systemic acquired resistance, while
JA is effective against necrotrophs (Nishad et al. 2020). Notably, the strongest upregulation of JA signaling was observed at 6 days post-inoculation (DPI), coinciding with the shift of *F. oxysporum* from a biotrophic to a necrotrophic lifestyle (Wang et al. 2022).

In recent years, the Arabidopsis and *F. oxysporum* Fo5176 interaction has emerged as a model system for studying pathogenesis and plant defense signaling using the tools of molecular biology, genetics, and genomics (Wang et al. 2022). These studies often considered the plant or only the root part as an integral subject to characterize the biotic responses. However, as the pathogen interacts with the host plant, different root cell types will likely respond differently to the infection, as each cell type contains a unique molecular signature and exhibits high cell-type specificity in its response to environmental signals (Long 2011; Iyer-Pascuzzi et al. 2011). For example, the endodermis cell type can modify its cell wall to form a physical barrier against certain pathogens, while pathogens aim to weaken these barriers to establish infection (Salas-González et al. 2021). Furthermore, understanding how different cells coordinate their defense response is essential to better understanding the plant’s overall defense program (Cole et al. 2021).

Recent studies have shown that snRNA-seq can confidently identify plant roots’ most known cell types, including the epidermis, cortex, endodermis, and vasculature (Farmer et al. 2021). Established approaches, such as the 10X genomics pipeline, enable the measurement of genetic regulation in Arabidopsis roots at a single cell/nucleus resolution (Cuperus 2022). Such techniques could
help shed light on the specific responses of different cell types and the coordination between cells to the pathogen and provide a more detailed understanding of the molecular mechanisms underlying pathogenesis and plant defense signaling.

In this study, we performed single-nucleus RNA sequencing (snRNA-seq) to study the transcriptional responses and regulatory mechanisms involved in interactions between Arabidopsis and *F. oxysporum*. Using label transferring from the Arabidopsis root atlas (Shahan et al. 2022), we assigned nuclei identities to our de novo-derived nuclei clusters. Our integrated atlas captured the trajectory of plant development from the quiescent center (QC) to apical meristem tissue types, elongation zones, and mature cell types.

We identified a unique pathogen-induced cluster in our atlas. In pathogen-treated samples, the nuclei population in this cluster increased by 7-fold compared to mock-treated ones, reflecting a shift in cell destination upon fungal challenge. The upregulated genes among these "induced" nuclei are enriched for defense-related functions, suggesting that this population could potentially be used to develop new strategies for improving plant defense mechanisms. We also used an infection pseudo-time course to reveal a trajectory of this unique induced cluster derived from the mature cortex. Our fungal-plant interaction-based cell atlas also revealed regulatory networks coordinating development and defense. By searching for crucial receptor-ligand pairs (Xu et al. 2022), we identified negative regulators involved in the development and positively
regulated modules, including *Resistance to Fusarium 1 (RFO1)* (Diener and Ausubel 2005) and other partners, upon challenge by the *Fusarium* pathogen.

We also discovered clusters of fungal nuclei with distinct gene expression patterns during infection. We identified core chromosome-encoding infection marker genes showing consistent high expression across multiple time points and accessory chromosome-encoding infection marker genes induced explicitly in one cluster but at only 3 DPI. Gene ontology enrichment analysis revealed that genes positively correlated with infection pseudo-time were mainly related to cell wall degrading enzymes, suggesting their importance during infection. Overall, our snRNA-seq data provides valuable insights into the spatially resolved immune response of the host, and the infection-specific programs from the fungal partner.

4.2 Results

4.2.1 Capturing key transcriptomic changes in *Arabidopsis* response to *F. oxysporum* at bulk and single-nucleus resolutions

By colonizing the xylem, *F. oxysporum* is responsible for destructive vascular wilt diseases in crops. Within 3 DPI, *F. oxysporum* progresses from the epidermis to the endodermis (Gordon 2017; Martínez-Soto et al. 2022). Using *Arabidopsis* - Fo5176 pathosystem, our study focused on the early stages of infection (up to 3 DPI) before *F. oxysporum* fully advanced into the xylem. We aimed to investigate the critical genetic factors that determine the outcome of the interaction. To collect root tissues less disruptively and more efficiently, we used
a protocol that involved growing and inoculating Arabidopsis on an agar plate medium (Kesten et al. 2019; Huerta et al. 2020). We first confirmed that the infection progression in the agar plate system (Figure 4.1A) was similar to that under soil conditions (Martínez-Soto et al. 2022). Fo5176 germinated, attached to the root epidermis 12 hours post-inoculation (HPI) and penetrated the epidermis at 1 DPI. By 3 DPI, Fo5176 had colonized the root cortex and started progressing into the vasculature. We did not observe a significant difference in primary root growth length when comparing mock and Fo5176-treated samples at 1 and 3 DPI (Figure S4.1A), highlighting a comparable system with minimal confounding due to distinct developmental stages of the roots among treatments.

We collected Fo5176 and mock treated samples from 12 HPI, 1 DPI and 3 DPI, applied bulk and snRNA sequencing to examine immune signaling activation in host-localized tissues in the Arabidopsis-Fo5176 pathosystem. We first investigated the critical transcriptional responses during the biological interactions between Arabidopsis and F. oxysporum. We identified 5,448 differentially expressed genes (DEGs) in our bulk RNA-seq dataset between mock and Fo5176-treated samples, with at least one time point showing differential expression (Table S4.1; Figure S4.1B). Wang and colleagues previously compiled a list of key Arabidopsis genes involved in the Arabidopsis-Fo5176 pathosystem (Wang et al. 2022). We found that 68 genes reviewed by Wang and colleagues were significantly enriched in the DEGs we identified (35 out of 68 genes were DEGs, p-value = 1.8E-7, two-sided Fisher’s exact test) (Table S4.2). We also identified 160 consistently upregulated genes across all
three time points (Table S4.3). These 160 genes include a couple of JAZs (JAZ1, JAZ5, JAZ6, JAZ8, and JAZ10) that mediate the transcriptional responses to JA (Pauwels and Goossens 2011); ERF1 and ERF2 are involved in the ethylene signaling (Huang et al. 2016); WRKY53 is involved in senescence (Miao et al. 2004); PRX33 is associated with ROS production (Miao et al. 2004), and MYB15 plays a role in lignin biosynthesis (Kim et al. 2020).

We performed Gene Ontology (GO) enrichment network analysis on the 160 genes that were consistently upregulated and found that the top two enriched networks (with the smallest adjusted p-values) are related to toxin metabolism and indole-containing compound metabolism (Figure 4.1B; Table S4.4), which is consistent with our previous findings (Guo et al. 2021). The enriched modules are all associated with plant immune responses, including defense, JA response, ROS response, senescence, and metabolism. Our bulk RNA-seq dataset successfully identified the essential genes involved in Arabidopsis-Fusarium interactions.

We then proceeded with our snRNA-seq data. After filtering out contaminants for each sample based on high organelle RNA contents, high spliced RNA contents, high debris score, low unique molecular identifier (UMI) count (< 500), and clusters with no marker genes (see Materials and Methods), we integrated eight samples into one integrated Seurat objective (using Seurat v4 integration method) containing 23,334 nuclei for mock-treated samples and 20,248 nuclei for Fo5176-treated samples. Our de novo clustering analysis generated 19 distinctive clusters, and the transcriptome lineages of nuclei were
captured by Uniform Manifold Approximation and Projection (UMAP) (Figure S4.2 and Figure 4.1C). We annotated each cluster based on 1) the most representative cell types resulting from the label transferring to a root atlas (Shahan et al. 2022) and 2) the developmental lineages of different clusters from meristem to mature. This analysis confidently labeled clusters such as phloem, with 89% of nuclei being transferred to phloem on the atlas (see Figure S4.2 and Table S4.5). However, some mature cells had low confidence in label transfer if the most representative cell types were used as the cluster label. Therefore, we relied more on the developmental lineages inferred from the UMAP and, in some cases, labeled clusters as mixed populations. For example, the Atrichoblast-M* denotes the Mature atrichoblast mixed with other cell types (asterisk* denotes the mixed cluster).

The snRNA-seq (pseudo-bulked transcriptomes from all nuclei for each sample) and bulkRNA-seq datasets were highly correlated, with a Spearman correlation of 0.83-0.94, as shown in Figure S4.3, supporting that snRNA-seq data represent the expression patterns in intact root tissue. We aggregated the expression of each gene across all nuclei within the same cluster for each sample and performed differential gene expression tests, treated each sample as one replicate (four Fo5176-treated versus four mock-treated), and identified 255 cluster-based differentially expressed events (150 genes as some genes were differentially expressed in multiple clusters) (Table S4.6). Interestingly, none of these 150 genes were discussed by Wang and colleagues (Wang et al. 2022), indicating the novelty of our snRNA-seq dataset and its potential to reveal
spatially restricted biotic responses. One notable finding was the upregulation of
CYP71A13 in one of the mature cortex populations (Cortex-M*) in the Fo5176-
treated samples. CYP71A13 is a cytochrome P450 that, together with
CYP71A12, produces dihydrocamalexic acid (DHCA), a precursor to the
defense-related compound camalexin (Müller et al. 2015). Camalexin
accumulates in the intercellular space and contributes to the resistance of
Arabidopsis to Pseudomonas syringae by inhibiting bacterial growth (Koprivova
et al. 2019). The same biochemical defense pathway may be conservatively
functional when Arabidopsis defend Fusarium.

To analyze the nuclei-based transcriptome at a more stringent level, we
further filtered out nuclei with unique molecular identifier (UMI) count < 2000 to
ensure that the baseline level of transcripts was captured in all the nuclei. This
resulted in 18,030 high-quality nuclei (10,157 Mock and 7,873 Treated samples),
as shown in Figure 4.1C. Since the treatment with Fo5176 may affect cell
identity, we examined if there were any changes in the number of nuclei
identified in each cluster (Figure 4.1D; Table S4.7). Our analysis revealed that
the Fo5176-treated samples have a significantly higher (adjusted p-value < 0.05,
Chi-squared test with Bonferroni correction) proportion (3.23%) of nuclei in
Cortex-M*, compared to mock-treated Cortex-M*, which only accounts for 0.46%
of the total nuclei. As we observed a 7-fold change (Figure 4.1D), we further
labeled Cortex-M* as "INDUCED", separated from another mature cortex cluster
(Cortex-M) (Figure 4.1D). In addition, our analysis revealed a significant
decrease in the number of nuclei associated with mature procambium and xylem
(Procambium-M and Xylem-M), together with a significant increase in the number of nuclei within Meristematic stele (Stele-Mer) upon Fo5176 treatment (adjusted p-value < 0.05, Chi-squared test with Bonferroni correction; Figure 4.1D). We also observed that the endodermis exhibited a similar trend, with mature nuclei numbers decreasing in two sub-populations and meristematic and elongating nuclei numbers increasing, though only statistical significance was found in the meristematic and Elongating endodermis (Endodermis-E.Mer) (adjusted p-value < 0.05, Chi-squared test with Bonferroni correction; Figure 4.1D).

Among genes involved in plant immunity, we observed interesting expression patterns. For example, JAZ9/TIFY7, which encodes a protein interacting with COI1 and mediating the JA signaling (Mosblech et al. 2011), had a significant expression induction (p-value < 0.01, Wilcoxon Rank Sum test) in the Cortex-M, Endodermis-M3 (sub-population 3 of mature endodermis), Pericycle-M.E, and Procambium-M (Figure 4.1E and Table S4.8). To validate these expression patterns, we used JAZ9 expression reporter lines (pJAZ9::NLS-3xmVENUS) (Zhou et al. 2020) to test whether Arabidopsis roots activated this gene in a localized manner in response to F. oxysporum as shown in our snRNA-seq data (Figure 4.1F). In the mature zone of the primary root, while we could not observe a clear signal for JAZ9 when the reporter line was mock treated with H2O at either 1 DPI or 3 DPI, we observed localized expression activation of JAZ9 in mature cortex cells (not in any other layers) when roots were challenged with Fo5176 at 1 DPI (JAZ9 positive cortex cells/total cortex cells observed = 0.07). At 3 DPI, Fo5176 not only induced cortex expression of JAZ9 (JAZ9
positive cortex cells/total cortex cells observed = 0.17) but also induced JAZ9 expression in inner layers (JAZ9 positive inner layer cells/total cortex cells observed = 0.66). This result was consistent with our snRNA-seq data showing JAZ9 induction in the endodermis, pericycle, and procambium within 3 DPI. Our findings confirmed that localized tissue activation of immune signaling occurs in the F. oxysporum–Arabidopsis pathosystem, and our further comprehensive analysis of the cellular program at the single-nucleus level promised to provide insights into spatially resolved immune responses.

4.2.2 Infection pseudo-time inferred by examining the trajectory of nuclei from mature cortex clusters

We conducted GO enrichment analysis on each cluster’s top 200 markers (based on adjusted p-values, from small to large, see Table 4.9) to better understand the underlying biology and investigate immunity signatures of each cluster (Figure 4.2A). Based on our findings, we classified the 19 clusters into three categories: 1) specific immunity cluster (INDUCED covering a sub-population of mature cortex), 2) general immunity clusters (markers are significantly enriched in more than two terms out of six that are related to stress responses) covering mature cells and developmental endodermis and pericycle (Atrichoblast-M*, Cortex-M, Endodermis-M3, Endodermis-M1, Endodermis-E.Mer, Pericycle-M.E, Procambium-M), and 3) developmental clusters (all the rest) including developmental cells and part of the mature endodermis (Endodermis-M2). The INDUCED cluster showed enrichment in effector-triggered
immunity related terms, including cell death (biological process), kinase activity (molecular function), and plasma membrane localization (cellular component) (Cui et al. 2015), and is depleted from all the terms related to development. The general immunity clusters were enriched in some stress response terms but also contained terms related to growth, such as primary metabolism, tropism, and cellular organization. The developmental clusters showed minimal stress response enrichment but included terms related to development. Our analysis suggested that mature cells are the frontier of interaction.

Notably, the INDUCED cluster was over-represented in the Fo5176 treated samples, which led us to investigate its underlying biology. One of the top markers of the INDUCED cluster is CCOAMT, which encodes S-adenosyl-L-methionine: transcaffeoyl coenzyme, a 3-O-methyltransferase that is involved in the lignin biosynthetic pathway (Xie et al. 2019). The expression of CCOAMT was significantly higher ($p < 0.01$, Wilcoxon Rank Sum test) in Fo5176-treated samples compared to mock-treated samples in the INDUCED cluster (Figure 4.2B). The second marker of INDUCED is CYP82C2, which is involved in the biosynthesis of 4-hydroxy indole-3-carbonyl nitrile (4-OH-ICN), a cyanogenic phytoalexin in Arabidopsis (Rajniak et al. 2015). This gene was only “representatively” expressed in the Fo5176 infected population of the INDUCED cluster (Figure 4.2B). Consistently, we searched for this gene in the reported Arabidopsis root atlas (Denyer et al. 2019; Ryu et al. 2019; Shahan et al. 2022) but could not identify it as an expressed gene, highlighting it is a defense-specific expressed gene.
We wanted to determine how the INDUCED cluster developed from the mature cortex. We combined nuclei from INDUCED with Cortex-M to investigate this and conducted the trajectory inference and pseudo-time analysis using Monocle 3 (Trapnell et al. 2014; Cao et al. 2019). Our analysis supported a clear lineage from Cortex-M to INDUCED and revealed a trajectory from the uninfected mature cortex to the infected and responsive mature cortex (Figure 4.2C and Figure S4.4). We defined the Monocle3 inferred pseudo-time as the “infection pseudo-time”, starting from Cortex-M and ending at INDUCED (Figure 4.2C). Mapping the marker genes of the INDUCED cluster on the trajectory from Cortex-M to INDUCED, we observed that CCOAMT expression increased as the infection pseudo time progressed (Figure 4.2C, Spearman Correlation Coefficient = 0.55, p = 4.57E-29). Instead of a steady increase, CYP82C2 was activated at the infection pseudo-time point when cells switched from Cortex-M to INDUCED (Figure 4.2C, Spearman Correlation Coefficient = 0.57, p = 3.04E-31). Our data supported that a clear shift in cell destination upon fungal challenge happens.

4.2.3 The regulatory networks coordinating development and defense

To have a better global view of our 18030 transcriptomes and infer biology from the variance, we computed the biotic response signature score (BRS), representing the extent to which each nucleus contains a defense program. A score in each nucleus summarized the expression of 160 signature genes identified through bulk RNA-seq (see 4.2.1). We observed that the mature cells have a higher BRS than the meristematic cells, irrespective of the treatment
We also found that the biotic response is global when comparing the average BRS of two treatments within each cell type (adjusted p-value < 0.05, Wilcoxon Rank Sum test and Bonferroni correction) in all pair-wise comparisons except Xylem-M) (Table S4.10).

We aimed to investigate the regulatory networks that underlie BRS variability. It is widely recognized that plants detect biotic threats through receptors that activate downstream cellular responses, which are ultimately transmitted to the transcriptional machinery in the nucleus, resulting in the expression regulation of defense-related genes. Therefore, we focused our network analysis on ligand-receptor signaling and transcriptional regulations.

We first investigated the ligand-receptor interactions (Xu et al. 2022) associated with BRS variability in the snRNA-seq data. To do this, we developed a ligand-receptor importance score (LR score) by taking the absolute value of the product of the Spearman correlation coefficients between the expression of ligand and receptor, and the Spearman correlation coefficients between the total expression of ligand and receptor, and BRS (See Materials and Methods). We then constructed a ligand-receptor network for the top 100 pairs based on LR scores (Figure 4.3B and Table S4.11). Notably, we found three interconnected modules. A module positively correlated with BRS variation in the network constitutive of a group of WAKs, including RFO1 (also known as WAKL22), which are potential cell wall integrity sensors (Kohorn and Kohorn 2012). RFO1 is the first WAK family member to be identified as a receptor for signaling from *F. oxysporum* and is responsible for the resistance to *F. oxysporum* f. sp.
Introducing the Col RFO1 allele into susceptible accessions is sufficient to confer resistance (Diener and Ausubel 2005). However, the role of RFO1 in resistance to *F. oxysporum* still needs to be completely understood, and the mechanism by which it functions as a cell wall integrity sensor remains unknown. Our ligand-receptor network analysis revealed that the RFO1 module includes other WAKLs and Glutathione S-Transferases of Phi Group (GSTFs) (Figure 4.3B), providing candidates for further experimental validation to understand the mechanism of RFO1-mediated resistance further. We identified another module that positively correlated with BRS variation related to root hair development. We observed a significant induction of RHS19 in Trichoblast-M.E. cells (p-value < 0.01, Wilcoxon Rank Sum test) (Figure 4.3C). Based on our findings, we hypothesized that Fo5176 treatment leads to promotion of root hair growth, and our experiments confirmed this by observing a significant longer root hair in mature trichoblast at 3 DPI (p-value < 0.05, Wilcoxon Rank Sum test; Figure 4.3D).

We also found negative regulators (the expression of the Ligand and Receptor are negatively correlated with BRS) involved in the development, which can be explained by the positive correlation between BRS (lower to higher) and the developmental trajectory (meristem to mature) of our atlas (Figure 4.2A and C). For example, AGP31 and AGP9 of this module, are two genes that encode for arabinogalactan proteins (AGPs), a family of highly glycosylated cell surface proteoglycans. AGPs play various roles in plant development, including cell proliferation, differentiation, and cell-to-cell communication (Pereira et al. 2015;
To probe the transcriptional network that correlated with BRS variability, we conducted Spearman correlation on all genes across 18030 transcriptomes to determine their relationship with the BRS. We found 193 genes significantly correlated with BRS (Spearman correlation p < 0.05) and with a correlation coefficient greater than 0.3 (Table S4.12). We combined this list of 193 genes with the 160 signature genes identified through differential gene expression analysis of the bulk RNA-seq, resulting in 325 genes (28 genes overlapped). We focused on the transcriptional regulation of these genes and extracted a gene regulatory network from the publicly available PlantRegMap database by these candidates (Tian et al. 2019). This network consisted of 8 regulators and 65 target genes with 107 regulations (Figure 4.3E and Table S4.12). To validate the network using our snRNA-seq dataset, we used Genie3 (Huynh-Thu et al. 2010), a decision-tree-based machine learning method, to independently construct a gene regulatory network (Genie3-GRN) for these 73 genes, resulting in 864 regulations (Table S4.13). The Genie3-GRN, de novo learned from the snRNA-seq dataset, validated all 107 regulations extracted from PlantRegMap dataset. Interestingly, 13 out of the 65 target genes are associated with JA pathways (Figure 4.3E), highlighting the importance of JA signaling in biotic stress responses. These genes include those involved in JA signaling, response to JA, and JA biosynthesis (Table S4.14). We also identified MYB15, an essential regulator for lignin biosynthesis during effector-triggered immunity (Kim et al. 2020), as a regulator of GSTF3 (a WAK interactor), linking our transcriptional
network and ligand-receptor network together (Figure 4.3B and E).

4.2.4 Fungal nuclei displayed distinct stages during infection

To investigate the feasibility of identifying fungal nuclei and associated programs, we mapped the sequencing reads from Fo5176-treated samples to a concatenated Arabidopsis and F. oxysporum genome and extracted a count of reads confidently mapped to the fungal genome only. Based on UMI distribution, we used different minimum UMI to filter samples (>250 for 12hpi and 1dpi replicate 2, and >100 for 1dpi replicated 1 and 3dpi, Figure S4.5A). We clustered the transcriptomes in each sample to identify a signature that may mark different infection stages (i.e., some fungal cells are more aggressive towards infection than others that contain less infection program) (Figure 4.4A). Although we clustered the 12hpi sample into two clusters, we could not identify infection markers for either cluster (Figure S4.5B). However, we found clusters with interesting infection marker genes for the 1dpi and 3dpi samples (Figure 4.4A). These samples each contained a cluster with marker genes putatively involved in the infection, defined as “infection clusters”. Genes shared within the infection cluster among three samples include gene-13930/Glycoside hydrolase and gene-1090/Ser-Thr-rich GPI anchored protein (Table S4.15; Figure 4.4B; Figure S4.5C). When we combined all the samples and conducted clustering, these two genes also marked one cluster (which we defined as an “Infection” cluster in the combined sample). The cluster bridge different samples, representing a common infection stage among different samples/time points (Figure 4.4A). Interestingly, in the 3 DPI sample only, we identified glycoside hydrolase (gene-11818), SIX9b
(gene-11133), Six4 (gene-14552), and homologs of Foa4 (Foa4-like, gene-18741) within the infection cluster, supporting the idea that the clusters we traced are related to pathogenesis (Table S4.15; Figure 4.4C; Figure S4.5C).

Focusing on the Infection cluster (Figure 4.4A), we conducted trajectory and pseudotime analysis by Monocle 3. A clear trajectory was inferred (Figure 4.4D). Among the top genes showing the highest Spearman correlation coefficients of their gene expression with the infection pseudotime within the Infection cluster, we identified Six4 and two glycoside hydrolases (gene-11818 and gene-14043) (Figure 4.4D). To further understanding the Infection cluster systematically, we identified 80 and 182 genes showing significant positive and negative correlations with the infection pseudotime, respectively (Table S4.16). The positive regulators are significantly enriched in polygalacturonase activity. In contrast, negative regulators are significantly enriched in housekeeping functions such as translation (Figure 4.4E). Our fungal nuclei analysis overall confirmed the feasibility of the identification of fungal nuclei and associated infection programs.

4.3 Discussion

Single-cell/nuclei RNA sequencing has been used to address the longstanding question of possible heterogeneity among cell types in response to abiotic stress. By applying heat stress to whole seedlings of plants, scRNA-seq has revealed subtle but significant differences in certain gene expressions among cell types (Jean-Baptiste et al. 2019), indicating its promise for studying plant stress response with unprecedented resolution. Similarly, scRNA-seq is
promising to unravel the events associated with transcription programming in Arabidopsis root biotic stress response to *F. oxysporum*, particularly in understanding the cell-type-specific regulation of Arabidopsis roots in response to *F. oxysporum*.

To probe the transcriptional dynamics of Arabidopsis in response to *F. oxysporum*, we generated and mined our snRNA-seq datasets. Our analysis revealed a unique pathogen-induced cluster enriched for defense-related functions that could aid in developing new strategies to improve plant defense mechanisms. Additionally, we identified regulatory networks that coordinate development and defense, including the *RFO1* gene and its partners. *RFO1*, a receptor-like kinase expressed in Col plants’ root vasculature, confers resistance to *Fom* (Diener and Ausubel 2005). We also discovered that fungal nuclei displayed distinct programs during infection and identified core and accessory chromosome encoding genes that mark the aggressiveness. Our findings demonstrate the potential of snRNA-seq to dissect spatially resolved immunity responses and fungal partners’ various programs during infection.

Effector proteins, such as Secreted In Xylem (SIX) effectors, are key in determining the pathogenicity (Rep et al. 2004). These effector proteins were originally identified in the xylem sap of infected tomato plants and are required for full virulence in some *F. oxysporum* (Houterman et al. 2007). However, how SIX effectors determine the pathogenicity has yet to be fully understood. Recent research has shown that some SIX effectors may suppress pattern-triggered immunity (PTI)-associated oxidative burst in *A. thaliana* (Tintor et al. 2020). Cell
type-specific expression of effectors can be achieved by fusing fungal genes with plant cell type-specific promoters and transformation (Brady et al. 2007). After expressing effectors in specific tissues, we could probe the molecular program in a tissue-specific manner by single cell/nuclei RNA-seq in the simplified model.

Aside from provoking vascular wilt disease on the main host, *F. oxysporum* isolates can also colonize the roots of alternative host plants, where they may grow asymptotically as endophytes or even act as biocontrol agents to protect the plant against pathogenic *F. oxysporum* forms or other root pathogens (Alabouvette 1999). It is commonly accepted that pathogenic strains of *F. oxysporum* can successfully penetrate the vasculature and propagate there, while non-pathogenic ones do not have the same effectiveness (Redkar et al. 2022; Martínez-Soto et al. 2022). Further study can focus on a comparative system, including pathogens and endophytes, to compare single-cell/nucleus-resolved molecular programs across different types of interactions.

Furthermore, Guillotin et al. (2023) recently illustrated how single-cell techniques could efficiently generate a pan-transcriptome of closely related plant species. Researchers employed a cell-by-cell comparative analysis to extract conserved modules from a comprehensive transcriptomics analysis. It provided new avenues for investigating the connection between genetic modules and cellular traits, such as those vital for crop functions among different crops (Guillotin et al. 2023). We believe a similar approach can be used to study plants–*F. oxysporum* interactions in multiple pathosystems, particularly focusing on different hosts.
High-resolution cellular profiling studies have demonstrated the ability to link cell-level transcriptional regulation with important agricultural traits and specialized cells often mediate these traits (Kajala et al. 2021). We not only identified that Arabidopsis generated a new specialized cell identity from existing cell types when challenged by *Fusarium*, but we have further probed ligand-receptor networks and mapped TF-target networks to explain the balance between growth and defense across the plant cell atlas in the context of *Fusarium*-Arabidopsis interactions. These networks linked known knowledge to a more mechanistic point of view, and thus provided new insights.

4.4 Materials and Methods

4.4.1 Sample collection and library preparation

The *F. oxysporum*-Arabidopsis interaction system has previously been studied (Wang et al. 2022), and it has been observed that the pathogenic Fo5176 infects xylem elements in the root vascular stele (Martínez-Soto et al. 2022). To gain insights into the spatiotemporal disease progression, we modified a previous protocol (Kesten et al. 2019) and inoculated the roots of 7-day-old Arabidopsis seedlings with Fo5176 on a half-strength MS agar medium (pH = 5.7) supplemented with 5 mM MES monohydrate (CAT: 145224-94-8). We used filter paper strips soaked in *F. oxysporum* spore solution (concentration: $10^7$/ml) for positive treatment and water as a mock treatment. The strips were placed on the roots for 15 seconds before being removed. We collected samples containing 500 roots each for single-nucleus RNA-seq at 12 HPI, 1 DPI, and 3 DPI for mock and
Fo5176-treated roots, resulting in six samples. We added replicates for 1 DPI, aiming for eight samples. Additionally, we performed bulk RNA-seq for each condition with three replicates (200 roots were collected for bulk RNA extraction) with the same time point set-up. To assess the Fo5176 colonization within roots, we stained and examined the roots using confocal microscopy at 0.5, 1, and 3 DPI, as previously described (Martínez-Soto et al. 2022).

4.4.2 snRNA-seq data processing

To prepare the snRNA-seq libraries, we used a microfluidic system on the 10x Genomics platform at the Joint Genome Institute (Dr. Maxim Koriabine did library preparation). Nuclei were extracted from flash-frozen inoculated root samples and encapsulated in individual oil droplets with a barcoded bead and cell lysis buffer during 10x Genomics platform processing. The lysed nuclei were then used to synthesize a cDNA library with pooled barcoded transcripts. Sequencing of the cDNA generated transcriptomic data for thousands of nuclei in each sample. Data processing was conducted by JGI's in-house pipeline (Dr. Sharon Greenblum, Joint Genome Institute). Meta-data and mapping statistics are summarized in Table S4.17.

After filtering out contaminants for each sample, based on high organelle RNA contents, high spliced RNA contents, high debris score, low unique molecular identifier (UMI) (UMI < 500) (Table S4.18), and clusters with no marker genes, we integrated eight samples into one integrated Seurat objective based on the method described (Hao et al. 2021). Our de novo clustering analysis generated 19 distinctive clusters, while UMAP captured the transcriptome lineages of nuclei. A
previously constructed root atlas was used as a reference (Shahan et al. 2022), and label transfer was conducted to tentatively label the identity of individual nucleus. We then annotated each cluster by the most representative cell types based on the label transfer result and the lineages of different clusters. The final filtering based on UMI>2000 yielded 18030 high-quality nuclei (10157 Mock and 7873 Treated samples). SCT-transformed (v2 function) (Hafemeister and Satija 2019) values were used for all the above steps.

### 4.4.3 Differential gene expression analysis

DEseq2 (Love et al. 2014) detected differentially expressed genes (DEGs) for bulk and pseudo-bulk data. For bulk RNA-seq, DEGs were defined as fold change > 2 and adjusted p-value < 0.05. For pseudo-bulk data, DEGs were defined as adjusted p-value < 0.05. Markers of each cluster were called by Wilcox Rank Sum test using Seurat v4, with parameters: min.pct = 0.25, logfc.threshold = log(1.5) and adjusted p-value < 0.05. Nuclei-based differential gene expression analysis in each cluster was conducted by Wilcox Rank Sum test using Seurat v4 with parameters: min.pct = 0 and logfc.threshold = 0. DEGs were further defined as: p-value < 0.01. RNA assay was used for the snRNA-seq analysis above and the further expression visualization of genes discussed in the result.

### 4.4.4 Functional enrichment analysis

Gene Ontology (GO), KEGG, and Wiki pathway enrichment network analysis on 160 genes being consistently upregulated from bulk RNA-seq data set
were conducted by Metascape (Zhou et al. 2019) with default parameters: Min Overlap = 3; P value Cutoff = 0.01; Min Enrichment = 1.5 and Gene Prioritization by Evidence Counting (selected). In the network graph visualization, each node represents one GO term with similar GO terms labeled in the same color, and representative terms are shown. The size of the nodes is proportional to the genes associated with the term. When two terms shared the same genes, an edge was drawn.

Gene Ontology enrichment analysis on each cluster’s top 200 markers genes was done using agriGOv2 (Du et al. 2010; Tian et al. 2017). GO enrichment analysis (plant GO slim) of Arabidopsis marker genes and reciprocal DEG analysis was conducted with the singular enrichment analysis (SEA) tool using the ArabidopsisTAIR10 annotation (Swarbreck et al. 2008). We applied a hypergeometric test, combined with the Hochberg (FDR) multi-test adjustment method to discover enriched GO terms at a significance level of 0.01 with a minimum of three mapping entries. Different enrichment results were compared using cross-comparison of SEA (SEACOMPARE).

For GO enrichment of fungal genes positively correlated with infection pseudo time and negatively correlated with infection pseudo time, ClusterProfiler 4.0 (Swarbreck et al. 2008) was used with default parameters and an adjusted p-value cutoff of 0.05.

**4.4.5 Score functions and network construction**

The BRS is a calculation based on the expression values of 160 signature genes. The formula is as follows:
BRS = \frac{\log_2(\text{SUM(160 gene expression value) + 1})}{\text{the max value}}

The sum of the expression values was calculated and then log2-transformed. The value of 1 was added to the sum before taking the logarithm to avoid errors caused by zero expression values. The maximum value of the log-transformed sum was then calculated across all samples. Finally, the log-transformed sum was divided by the maximum value to obtain the normalized BRS for each nucleus (range from 0 to 1). The BRS quantitatively measured each nucleus’s overall activation of biotic stress response pathways. Higher scores indicate a stronger response to biotic stress, while lower scores indicate a weaker response.

The receptor-ligand analysis (LR) importance score was calculated using the following formulas:

\[ LR\ value = \log_2(\text{expression of Ligand} + 1) + \log_2(\text{expression of Receptor} + 1) \]

\[ Correlation_1 = \text{absolute(Spearman Correlation Coefficient } (LR\ value, \text{BRS})) \]

\[ Correlation_2 = \text{Spearman Correlation Coefficient} \]
\[ (\log_2(\text{expression of Ligand} + 1), \log_2(\text{expression of Receptor} + 1)) \]

\[ LR\ score = Correlation_1 \times Correlation_2 \]

We first calculated the LR value to summarize the expression of both receptors and ligands. The correlation between the LR value and the BRS was defined as Correlation 1 (absolute value of correlation coefficient), and the correlation between the expression levels of the ligand and receptor genes was also calculated and defined as Correlation 2. The LR importance score was then obtained by multiplying the two correlations. The LR importance score was used
to identify potential ligand-receptor interactions that may play a role in the observed biotic responses. All networks were visualized by Cytoscape 3.9.1 (Shannon 2003)

4.4.6 Experimental validation

Expression of JAZ9 (line: pJAZ9::NLS-3xVENUS, gifted from Dr. Niko Geldner, Université de Lausanne) was observed by confocal microscope (Olympus FV1000) with a consistent confocal parameter setting: VENUS signal was detected with excitation at 473 nm and emission at 500 to 540 nm (HV 680, Gain 1, Offset 6%, Laser 20%), while propidium iodide was detected with excitation at 559 nm and emission at 580 to 660 nm (HV 580, Gain 1, Offset 6%, Laser 2%). The plant's roots were mounted in 10 µg/mL propidium iodide solution before microscopy.
Figure 4.1 Capturing key transcriptomic changes in Arabidopsis response to *F. oxysporum* at bulk and single-nucleus resolutions

(A) Temporal colonization progression of *F. oxysporum* Fo5176 on Arabidopsis roots. Fungal colonization was tracked using confocal microscopy after WGA-Alexa Fluor staining of fungal cell walls and propidium iodide staining of plant cell walls. Fo5176 germinated and attached to the root epidermis at 12 hours post-inoculation (HPI), penetrated to the epidermis at 1 day post-inoculation (DPI), colonized the root cortex by 3 DPI, and progressed into the vasculature. Scale bar = 50 µm. (B) Network analysis of gene ontology (GO), KEGG, and Wiki pathways enriched in 160 consistently upregulated genes from bulk RNA-seq data. Each node represents a GO term, with similar terms labeled in the same color. Node size is proportional to the number of associated genes. Edges are drawn between terms that share genes. Analysis was done using Metascape (Zhou et al. 2019) with default parameters. (C) Uniform Manifold Approximation and Projection (UMAP) of transcriptome lineages of the nuclei. M denotes mature; E denotes elongation; Mer denotes meristem; an asterisk (*) denotes clusters with multiple identities. (D) UMAP of mock- versus Fo5176-inoculated samples and nuclei identity change statistics. The induction index indicates whether pathogen treatment altered nuclei identity by changing the relative ratio of nuclei identified in each cluster. Fo5176-treated samples (3.23%) contained significantly more cells (about 7-fold change) in one cluster (labeled as INDUCED) than mock-treated samples (0.46%), and the identity of this cluster is a mature cortex and mixed. an asterisk (*) above the columns denotes a significant difference (adjusted p-value < 0.05, Chi-squared test) comparing the relative ratio of nuclei in the same cluster of mock versus Fo5176 treated samples. The details of the statistics associated with this figure are in Table S4.7. (E) *JAZ9* expression under mock or Fo5176-treated conditions in different clusters. * above the violins denotes significant upregulation (p-value < 0.01) during fungal infection. (F) Fo5176 induced a localized *JAZ9* response in the primary root mature zone. Nuclear-localized signals of the *JAZ9* transcriptional reporter (shown in green) were co-visualized with propidium iodide (PI, shown in red).
Figure 4.2 Infection pseudo-time inferred by examining the trajectory of nuclei from mature cortex clusters

(A) Gene Ontology enrichment analysis on each cluster’s top 200 marker genes.

(B) Expression of the top two marker genes of the INDUCED cluster. (C) Infection pseudo-time transition between mature cortex and fungal-induced cluster. Monocle3 (Trapnell et al., 2014) was used to infer the trajectory and pseudo-time of the INDUCED cluster combined with Cortex-M. A clear linear trajectory from the nuclei of Cortex-M to the nuclei of INDUCED was identified. Expression and pseudo-time of the top two marker genes of the INDUCED cluster highly correlate with the infection pseudo-time progression.
Figure 4.3 The regulatory networks coordinating development and defense
(A) Biotic response score (BRS) summarizes the changes in 160 signature genes across transcriptomes of all nuclei. (B) Three interconnected modules of receptor-ligand networks for top 100 scored pairs. (C) Expression of RHS19. Fo5176 significantly induced both genes in mature and elongating trichoblasts. (D) Fo5176 significantly induced the root hair length at 3 days post-inoculation (DPI). Root hair (3 mm to 5 mm from the root tips) length was measured with sample size: Mock (46) and Fo5176 (82). Wilcoxon test with p-value = 6.508E-08 and Student t-test with p-value = 1.817E-09. (D) Regulatory network that correlates the variability of BRS. The network includes genes involved in JA signaling (JAZ1, JAZ5, JAZ6, AT5G19230, NPR3, ERF1, CB5-C, and WRKY53), response to JA (JAZ1, JAZ5, JAZ6, AT3G23570, AT5G53990, AT1G13520, and AT1G23710), and JA biosynthesis (AT4G08850), which are labeled in green.
Figure 4.4 Fungal nuclei displayed distinct stages during infection

(A) UMAP showing clusters of fungal nuclei in combined and individual samples. UMI count distribution was also plotted for the combined sample. (B) Core chromosome encoding marker genes are expressed conservatively induced at multiple times. (C) Expression of accessory chromosome encoding marker genes upregulated at 3 DPI. (D) Infection pseudo time of the Infection cluster. The expression patterns of the top three pseudo-time marker genes (based on the smallest p-value gained from the Spearman correlation test) were plotted along the trajectory. (E) GO enrichment of genes significantly positively and negatively correlated with infection pseudo time. There are 80 genes in the positive correlation category and 182 genes in the negative correlation category. Adjusted p-value cutoff: 0.05
4.5 References


APPENDIX I - Genome Sequence of *Fusarium oxysporum* f. sp. *matthiolae*, a Brassicaceae Pathogen

Strains of the filamentous fungus *Fusarium oxysporum*, a notorious plant pathogen, can infect hundreds of cultivated species and, at the same time, have distinct host specificity (Kistler 1997; Ma et al. 2013; Michielse and Rep 2009). This host specificity is used to classify *F. oxysporum* into formae speciales, and a forma specialis typically represents one to a few monophyletic clonal lineages that cause disease in a narrow range of taxonomically related plants (Kistler 1997). At the genomic level, host specificity corresponds to the presence of lineage-specific chromosomes (Ma et al. 2010). However, little is known about molecular mechanisms involved in these host-specific plant–fungus interactions.

One of the three formae speciales that are pathogenic to the crucifer *Arabidopsis thaliana* (Diener and Ausubel 2005; Provart et al. 2016), the genome sequence of *F. oxysporum* f. sp. *matthiolae* will enable the genetic analysis of fungal pathogenesis and host immunity using the model plant *A. thaliana*.

Phylogenic analyses indicate that *F. oxysporum* f. sp. *matthiolae* isolates form a single clonal lineage (Bosland and Williams 1987; Kistler and Benny 1989; Kistler et al. 1987, 1991; O'Donnell et al. 2009), although two races of *F. oxysporum* f. sp. *matthiolae* are distinguished by the differential susceptibility of varieties of *Matthiola incana* (Bosland and Williams 1988). Natural variation of

---

immunity is observed among wild accessions or ecotypes of A. thaliana toward Fusarium wilt (Diener and Ausubel 2005). Quantitative trait loci mapping in offspring of crosses between resistant and susceptible ecotypes has identified three RESISTANCE TO F. OXYSPORUM (RFO) genes—one receptor-like protein gene (RFO2) and two receptor-like kinase (RLK) genes (RFO1 and RFO3)—from different RLK gene subfamilies (Cole and Diener 2013; Diener 2013; Diener and Ausubel 2005; Shen and Diener 2013). Because receptor-mediated immunity is reported to be the major determinant of disease resistance to F. oxysporum f. sp. matthiolae (Cole and Diener 2013), investigation of the interaction of F. oxysporum f. sp. matthiolae and A. thaliana should lead to a fundamental understanding of receptor-mediated plant immunity, especially against fungal pathogens. The genome sequence described here comes from DNA purified from F. oxysporum f. sp. matthiolae race 2, isolated from wilted garden stock (M. incana), a cultivated plant in the crucifer or mustard (Brassicaceae) family, prized for its colorful flowers (Baker 1948; Tatsuzawa et al. 2012). This strain was previously deposited in the American Type Culture Collection (ATCC 16603) by G. M. Armstrong and subsequently designated by P. H. Williams as PHW726 (Kistler et al. 1987).

The pipeline for genome assembly was adapted from Ayhan et al. (2018). Genomic DNA was purified from the mycelium of PHW726, then sequenced by Illumina MiSeq and PacBio RS II platforms with ×119 and ×21 coverage, respectively. We used MiSeq paired-end sequencing with 150 cycles. The maximum size of the PacBio RS II reads was 59 kb while the mean size was 8.5
kb. Trimmomatic, version 0.32 (Bolger et al. 2014), was used to remove adaptors and trim ends of Illumina reads (parameters: ILLUMINAACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). FastQC, version 0.11.5, was used to check the quality of all reads. SPAdes, version 3.9.1 (Antipov et al. 2016), was used to combine PacBio subreads and trimmed Illumina reads into an initial hybrid assembly with default parameters. BWA, version 0.7.12 (Li and Durbin 2009), was used to map the Illumina reads to the assembly. Further cleaning, fixing, and sorting of mapping reads was done with Picard, version 2.0.1, and Samtools, version 1.3 (Li et al. 2009). A structural variant caller, GRIDSS, version 1.4.1 (Cameron et al. 2017), was used to identify links between scaffolds in the initial assembly. A custom script (scaffolding.m) was used for scaffolding. Minimap2, version 2.17 (Li 2018), was used to map PacBio subreads to new scaffolds, and links were manually inspected and, if necessary, fixed. Further polishing was performed by remapping Illumina reads to the assembly, during which FreeBayes, v0.9.10-3-g47a713e (Garrison and Marth 2012), was used to identify base variants between reads and the assembly (specially 70% support of minimal 10 alternate counts, with a minimal base mapping-quality greater than q30). Identified variants were used to correct the assembly by a custom script (FASTAeditWithVCF.m). RepeatMasker 4.0.5 (Tarailo-Graovac and Chen 2009) was used to screen the repeats. Mummer 3.22 (Kurtz et al. 2004) was used to align the assembly with the reference genome assembly for the tomato pathogen *F. oxysporum* f. sp. *lycopersici* Fol4287 (Ma et al. 2010).
The final assembly was 57.3 Mb in total length and comprised 583 scaffolds with an N\textsubscript{50} value of 0.77 Mb (Table APPENDIX I). The largest scaffold size was 3.6 Mb. The GC content was 47.4%. The size of total interspersed repeats was 3.1 Mb, which accounted for 5.4% of the assembly. A comparison with Fol4287 assembly (Ayhan et al. 2018) suggested a larger assembly size and higher interspersed repeat content of PHW726. The size of sequence mapped to the core chromosomes of Fol4287 which, including 66 scaffolds (defined as core scaffolds), was 43.8 Mb. The assembly also included a scaffold of 52,365 bp that captured the entire mitochondrial DNA. This Whole Genome Shotgun project has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession WJXY00000000. The version described in this article is version WJXY01000000. This assembly for the genome of PHW726 should facilitate future molecular genetics and genomic studies. Candidate \textit{F. oxysporum} f. sp. \textit{matthioliæ} genes that promote pathogenesis or elicit immune response in \textit{A. thaliana} and \textit{M. incana} can now be predicted, subcloned, and genetically characterized.

Table APPENDIX I. Summary of the \textit{Fusarium oxysporum} f. sp. \textit{matthioliæ} strain PHW726 genome assembly and a comparison with Fol4287 (Ayhan et al. 2018)

<table>
<thead>
<tr>
<th>Variables</th>
<th>PHW726</th>
<th>Fol4287</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly size (bp)</td>
<td>57,270,650</td>
<td>53,912,367</td>
</tr>
<tr>
<td>Core sequence size (bp)</td>
<td>43,818,233</td>
<td>42,239,438</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>583</td>
<td>499</td>
</tr>
<tr>
<td>Number of core scaffolds</td>
<td>66</td>
<td>55</td>
</tr>
<tr>
<td>Size of largest scaffold (bp)</td>
<td>3,557,637</td>
<td>5,733,288</td>
</tr>
<tr>
<td>Interspersed repeat content (bp)</td>
<td>5.35%</td>
<td>4.21%</td>
</tr>
<tr>
<td>N\textsubscript{50} (bp)</td>
<td>774,050</td>
<td>1,338,693</td>
</tr>
<tr>
<td>N\textsubscript{90} (bp)</td>
<td>47,752</td>
<td>49,310</td>
</tr>
<tr>
<td>L\textsubscript{50} (bp)</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>GC content</td>
<td>47.44%</td>
<td>47.68%</td>
</tr>
</tbody>
</table>
References:


REFERENCES


Fokkens L, Guo L, Dora S, et al (2020) A Chromosome-Scale Genome Assembly for the *Fusarium oxysporum* Strain Fo5176 To Establish a Model Arabidopsis-Fungal Pathosystem. G3 (Bethesda) 10:3549–3555. https://doi.org/10.1534/g3.120.401375


Gimeno CJ, Fink GR (1994) Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. MOL CELL BIOL 14:13


Olesen JT, Fikes JD, Guarente L (1991) The Schizosaccharomyces pombe homolog of Saccharomyces cerevisiae HAP2 reveals selective and stringent conservation of the small essential core protein domain. 11:9


207


Stover RH 1962. Fusarial wilt (Panama disease) of bananas and other Musa species. Commonwealth Mycological Institute, Kew, Surrey, UK.


