November 2018

A Plant Pathology View of Signaling: a computational study of Fusarium oxysporum Kinomes and Downy Mildew Resistance in Sweet Basil

Gregory Deiulio

A Plant Pathology View of Signaling: a computational study of *Fusarium oxysporum* Kinomes and Downy Mildew Resistance in Sweet Basil

A Dissertation Presented

By

GREGORY DEIULIO

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 2018

Plant Biology

A Plant Pathologists View on Signaling: a computational study of *Fusarium oxysporum* Kinomes and Downy Mildew Resistance in Sweet Basil

A Dissertation Presented

By

GREGORY DEIULIO

Approved as to style and content by:

_____

Li-Jun Ma, Chair

_____

Jesse Mager, Member

_____

Rob Wick, Member

_____

Elizabeth Vierling, Member

_____

Li-Jun Ma, Graduate Program Director

Plant Biology

DEDICATION

To my family. To my friends.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Li-Jun Ma for her continuous support, guidance, and encouragement and for providing me with the freedom and environment to grow and learn as a person and as a scientist.

I would like to thank my other committee members Rob Wick, Elizabeth Vierling, and Jesse Mager for sticking with me and providing an outside critical view of my work.

I would like to thank Li Guo for helping to guide me through the field of computational biology and plant pathology.

I would like to thank the plant biology graduate program, the biochemistry and molecular biology department, and the biology department for funding in the form of TA ships and travel awards, and guidance navigating the university system.

Thank you to all the graduate students who I shared this journey with.

ABSTRACT

A PLANT PATHOLOGISTS VIEW ON SIGNALING: A COMPUTATIONAL STUDY OF *FUSARIUM OXYSPORUM* KINOMES AND DOWNY MILDEW RESISTANCE IN SWEET BASIL

SEPTEMBER 2018

GREGORY DEIULIO, B.S., SUNY ALBANY

PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Li-Jun Ma

This dissertation is composed of two projects that focus on pathogen and plant signaling within the framework of plant pathology. The first project targets protein kinases within the species complex *Fusarium oxysporum* based on genomic information and tracks their presence/absence and copy number variation across evolutionary time. We have predicted the kinomes of 19 Ascomycete fungi using the kinase annotating software Kinannote. Among Fusaria, kinases related to the perception of the environment, such as Histidine kinases, are proliferated. Similarly, I observed the expansion of Target of Rapamycin (TOR) kinase that regulates cell growth and development in responding to environmental cues. The second project compares metatranscriptomics of a resistant (MRI) versus a susceptible (SB22) *Ocimum basilicum*) cultivars, both infected with the pathogen *Peronospora belbahrii*. Using *de novo* approach, I predicted MRI unique genes, including R-genes and malectin-domain containing receptor-like kinases and genes involved in the biosynthesis of plant secondary metabolites. Based on the observation of the unique upregulation of isochorismate synthase in MRI, I hypothesize that salicylic acid is involved in the resistance response of MRI.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE MOLECULAR INTERACTION BETWEEN PLANTS AND THEIR PATHOGENS

A. Overview

Microscopic pathogens cause the majority of the world's crop and human diseases. They are borne on the wind, in the water, and on the backs of unwitting carriers as they are shuttled across the landscape in search of a suitable environment. Most of these disease causing agents are beyond our ability to see and because of this the vast majority of the events which take place between the pathogen and its host are completely unobservable to us. It is infinitely fascinating that things which occur at scales far exceeding our every day notice can have such a staggering impact on our lives.

By modifying the environment in some way an organism improves or maintains its chances of survival. As a general rule, all organisms perform the following cycle: 1) A signal from the environment is detected, 2) that signal is transmitted, 3) a change in transcription occurs, 4) a change in the abundance of effective molecules occurs, and 5) the environment is altered. In these cases the environment can refer to conditions both within as well as without the cell. Both projects described below study this cycle within the framework of plant disease and organisms which cause plant disease. One study examines the tools a cell has which function in signal detection, signal transmission, and effective output (1, 2, and 4), and the second project examines changes in translation and seeks to identify proteins which act in signal detection and

effective output (1, 3 and 4). Both projects are joined by three things; the questions both pertain to the above described cycle, the questions asked are answered by a unified methodology, and both questions aim to understand plant disease and disease causing agents.

Microscopy and molecular biology, which includes nucleic acid sequencing, are the primary tools with which we probe this tiny world. Microscopy enhances our own vision, allowing us to literally "look" with our own eyes. In the past, the vast majority of plant pathology, and biology as a whole, was conducted with microscopy. However microscopy is limited by its resolution and the events which can be the most useful to a more full understanding of the infection process take place in a plane which is currently beyond its capacity. We have therefore developed different methods of extracting information from an organism. Here I am concerned with DNA and RNA sequencing with a focus on two simple questions: What is the sequence of the nucleic acid is there? How much is there of each sequence?

The genome of a cell is its software. Like much of human developed software, not all of the code is run at once. The exact phenotype of any given organism is a reflection of one of many possible states it is capable of forming. Examining the code which provides all possible states and which parts of the code are running at any given time provides us with information about what an organism *can* do, and *when*.

Sequencing allows us to probe two of the most critical systems within an organism. DNA, the software, is the primary information storage device for a cell. Without a stable DNA molecule which codes for functional units and informs the next generation, evolution would have nothing to mold. The sequence of DNA could provide, if we were sophisticated enough to understand it, everything we could possible want to know about an organism. Every tool available to a cell, every action that could be taken, every response possible, every form and structure which it would construct could be predicted from DNA.

We can also probe RNA which represents which parts of the software are running at any given time. Predominantly used to collect messenger RNA (mRNA) information, RNA sequencing answers a set of overlapping and different questions from DNA analysis. While mRNA does not contain *all* possible genes, and likely never will given any single sequenced natural cellular state, it allows us to know which of the set of genes are being used *now* (more or less, mentioned below). RNA sequencing is the technology of understanding, or at the very least approximating, what the cell is actually doing in real time. We can track shifts in gene expression over time, the appearance of transcripts which were previously absent, the disappearance of transcripts which the cell no longer needs, and the relative speed at which these things happen. As the interaction of a pathogen and its host is molecular in nature, we need a molecular tool. DNA and RNA sequencing technology, and those technologies like it including proteomics and metabolomics, are tools which convert biological reality

into words. Numbers and letters are the output of sequencing technologies, allowing us to use computers to do work which may otherwise have taken lifetimes.

This thesis is composed of two projects. One uses DNA sequencing to target a family of tools, protein kinases, and ask how the presence and number of these kinases changes across evolutionary time and what their numbers can tell us about the subtle differences among a closely related group of pathogens, 12 sequenced members of the *Fusarium oxysporum* species complex. The second project is far more practical in its goal, but has less overarching knowledge to provide. Using mRNA sequencing of infected plants we predict genes within Sweet Basil which may allow it to resist infection by the oomycete pathogen *Peronospora belbahrii*. In both cases we are comparing sequencing information between individuals, or individuals across time and asking.

To respond properly to constantly shifting environmental conditions both the pathogen and the plant have developed signaling pathways which transmit instructions to the nucleus. These instructions modify gene transcription which in turn affects the protein content of the cell. Many of the signaling pathways of the cell flow through kinases of various types which modify their targets through phosphorylation, changing the activity of the target. Plants have developed entire families of receptors which act through signaling cascades including the mitogen activated protein kinase pathways which respond with various defense programs (i.e. reinforcing molecules, antifungals, and hormone signaling pathways). Pathogens respond to environmental

stimuli and activate different morphological shifts (i.e. germination, sporulation, and shifts between necrotrophy and biotrophy when applicable) and induce the expression of infection specific genes including cell wall degrading enzymes, toxic secondary metabolites, and effectors.

Comparative sequencing of both DNA and RNA can be used to answer large and small questions. In chapter 1, I review what we know about the molecular events which take place when a plant and its pathogen engage in their struggle, usually ending in the death of one of the pair. In chapter 2 I discuss my work with kinases in the *F. oxysporum* species complex and address what we have learned through kinome prediction. In chapter 3 I cover the prediction of basil downy mildew resistance genes and explore what the changing transcriptome tells us about the way in which Sweet Basil responds to infection. Finally in the conclusion I summarize key findings and discuss how the last 7 years have shaped my understanding of biology and what I have learned.

B. Preface

There are an estimated 250,000 species of higher plants and 1.5 million species of fungi. Most fungi are saprotrophic, living off of dead or decaying plant matter. In general the cohabitation between plants and fungi is beneficial to both parties, the classic examples being mycorrhizal and endophytic fungi (Grayer and Kokubun, 2001). Over evolutionary time, and arguably in some cases partly due to human

activity, some fungi have shifted the balance within this interaction in their favor, becoming parasites.

To coordinate such an intimate relationship between two organisms diverged by over a billion years, involves a combination of strategy and brute force. Both parties contain sets of genes which are dedicated almost exclusively to this interaction, subverting the immune system and degradation in the case of pathogenic fungi and oomycetes (Bozkurt et al., 2012), and generating diverse defense strategies as employed by plants (El Hadrami et al., 2012). Much about the dynamic nature of the interaction between a pathogen and its host can be observed by sequencing mRNA. Although mRNA transcript levels do not always correlate well with the amount of detectable resulting protein, as is the case for yeast (Gygi et al., 1999), the analysis of total protein content remains less high-throughput than mRNA analysis and so mRNA sequencing has been dominant.

We can probe not only shifts in the production of effective components, but ask which exist within an organism. The presence and absence of tools tells us the absolute effective capacity of an organism, while mRNA tells us how that capacity changes over time and in response to the environment. In the introductory chapter below I aim to provide background information in order to facilitate an understanding of the methods used. In addition I provide some of the thought processes behind some of the steps.

C. Fungi and signaling

1. The genus *Fusarium* and the *Fusarium oxysporum* species complex

The ascomycete fungi are one of the main branches of terrestrial fungi and are sister to the mushroom forming basidiomycetes. Ascomycetes are defined by structures formed by the colony within which sexual spores develop called the ascus. The ascomycete fungi are among the world's best understood, encompassing the genera *Saccharomyces*, *Aspergillus*, *Neurospora*, *Penicillium*, and *Fusarium* among hundreds of others. They represent some of the most useful fungi to humans due to their industrial uses, but also cause a large portion of the world's plant disease (Dean et al., 2012). A better understanding of the mechanisms which drive their biology and the strategies which they use to attack their plant hosts can only improve our ability to use and combat ascomycete fungi.

The genus *Fusarium* contains a number of well-studied fungal pathogens (Kistler, 1997) including *Fusarium graminearum* (Fg), a pathogen of wheat, *Fusarium verticillioides* (Fv), a pathogen of wheat and corn, and the *Fusarium solani* and *Fusarium oxysporum* (Fo) species complexes which each infect many plant hosts. The genus *Fusarium* is defined by a set of spores shared by all species (Leslie and Summerell, 2006). Fusaria produce both macro- and micro-conidia, with macroconidia containing roughly 5 individual cells organized into a banana shape and microconidia existing as a single cell roughly circular in shape.

The exact lifestyles of Fusaria differ, with *Fg* and *Fv* infecting aboveground reproductive structures and *F. oxysporum* generally infecting the vascular tissue via entry through the roots belowground. Unlike *Fv* and *Fv*, *Fo* has no known sexual cycle and reproduces only asexually. A key aspect of the fungi within the species complex *Fusarium oxysporum* (FOSC) is the existence of "lineage specific" (LS) chromosomes that define host-specific pathogenicity (Ma et al., 2010; Ma et al., 2013). Defined by comparing *Fg*, *Fv*, and *Fo*, the lineage specific portions of the genome of Fo4287 were seen to be absent from *Fg* and *Fv* even though the three species share a majority of the genome sequence. Each strain contains a set of LS DNA which although generally is in the form of distinct chromosomes can also be attached to the ends of conserved chromosomes.

The LS genome is transposon rich and is enriched for genes with annotated disease related functions. The LS genome was defined by regions unique to Foxy, but not all strains share the same material. Since LS chromosomes determins host-specificity within the FOSC, strains pathogenic to the same host may share LS material, which is the case for Fo4287 and FoMN25. Both Fo4287 and FoMN25 are tomato pathogens, and their genomes share at least chromosome 14 and some of chromosome 3 and 6 (personal observation). Fascinatingly, Ma et al showed that the transfer of chromosome 14 from Fo4287, a tomato pathogen, into a non-pathogenic strain was able to confer the ability to infect tomato. I understand the LS genome to be a "unique

SET of material", where although individual pieces of the set may overlap the set as a whole appears to be a unique combination.

This unique combination of material is the cornerstone of the strength of the FOSC as a model system to study disease. The large majority of the genome is shared among strains, but each has a LS component which we know can drive host specificity. In addition, excluding some notable exceptions like Fo32931 (an opportunistic human pathogen) (Nucci and Anaissie, 2007) and Fo47 (an endophyte)(Aimé et al., 2013), most sequenced FOSC isolates share a common lifestyle. A resting spore resides within the soil. The spore germinates and infects the roots of the host plant, eventually making its way into the vascular tissue. The presence of the pathogen within the xylem leads to a thickening of the vascular material which inhibits water transfer. As water transfer is inhibited the plant begins to wilt, causing the distinctive "Fusarium wilt disease" symptoms. The fungus kills the plant and produces resting spores called chlamydospores and micro- and macroconidia, and the cycle begins again. The presence of the resting spores is a significant hindrance to disease management as they can survive for extended periods within the soil (Couteaudier and Alabouvette, 1990). In general Fusaria also tend to be particularly fungicide resistant, a complex problem requiring the continuous development of new drugs to prevent both plant and human disease.

2. Kinases and cellular signaling

In order to respond to the environment around it an organism must be able to sense molecules in its surroundings, both within and without the cell, and then transmit a signal which reprograms the cell (Fig. 1.1). This reprogramming allows the cell to modify itself or the environment around it to best suit its needs at that particular moment.

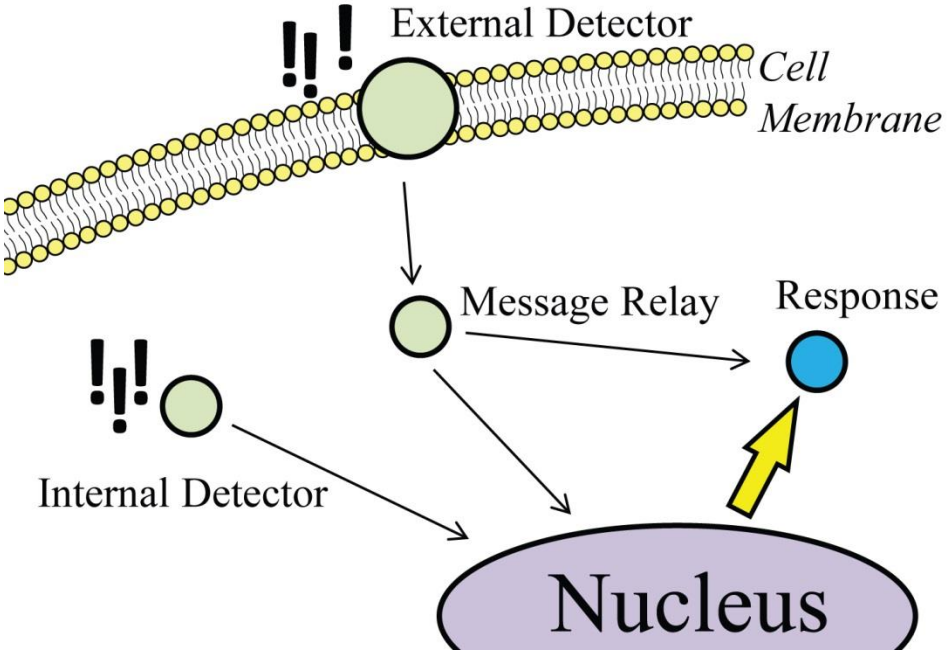

Figure 1.1: Cellular signaling. Cellular detectors either within the membrane or within the cytosol transmit signals down through the cell to modify the functions of the cell. In this way cells adapt to their surroundings.

Most kinases fill the role of signal transducers acting within the interior of the network. Some kinases, like the histidine kinases (Defosse et al., 2015), also act at the periphery of the network where they directly sense a target molecule and initiate a signaling cascade through, in many cases, other kinases. In other cases kinases act at

the terminal phase of the signaling network where they phosphorylate a protein which itself is not a kinase and whose function acts to directly reprogram the cell (e.g. a transcription factor). Kinases are not the only signaling molecules a cell uses for these functions, but they are the predominant mechanism.

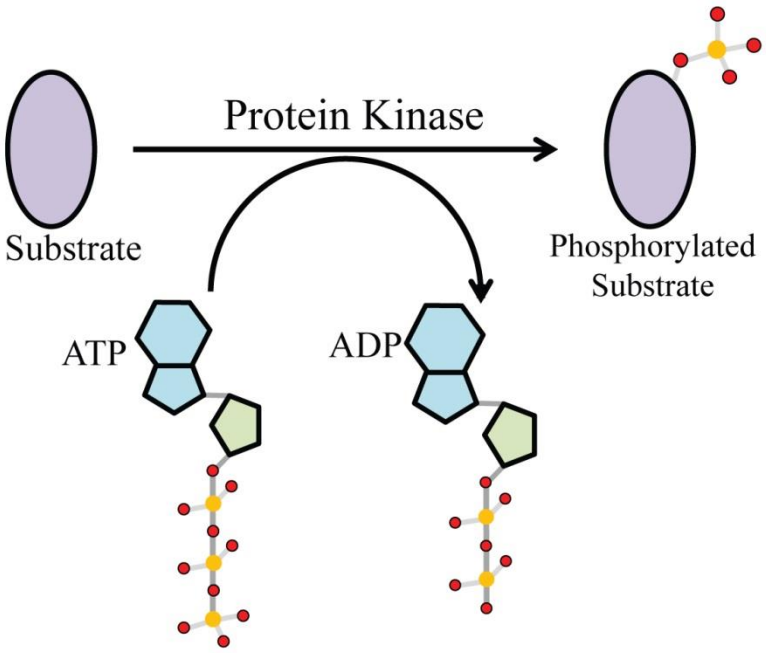

Figure 1.2: Basic kinase function. The action of a protein kinase removes a phosphate group from ATP and attaches it to a substrate protein.

As a whole kinases are diverse in form and function but are separated into individual families which generally share a similar sequence to their kinase domain (Hanks and Hunter, 1995). They function specifically by binding to a target protein and transferring a phosphate group onto that target protein which is received from ATP (Fig 1.2). The phosphorylation of the target protein changes the activity of the target. Although grouped into families based on their kinase domain, the remainder of the protein tends to differ significantly even among members of the same family.

3. Pathogen Signaling Pathways

All pathogen offensive programs require proper signaling to initiate. Among the most central signaling pathways are the map kinase (MAPK) cascades. In *Saccharomyces cerevisiae*, MAPK cascades signal biological shifts in mating, filamentous growth, cell wall integrity, osmolarity stress, and spore assembly (Seger and Krebs, 1995). In plant pathogenic fungi, MAPK cascades are known to regulate virulence related phenotypes including appressorium formation, invasive growth, and secondary metabolism (Di Pietro et al., 2001; Rispail and Di Pietro, 2009). Many of these pathways transmit signals from sensors on the membrane to the nucleus where they cause changes in transcription.

Upstream receptors signal through MAPK pathways, which regulate morphological stages and growth patterns. In *M. oryzae*, the MAPK Mor-Kss1 is required for proper appressorium formation and invasive growth, but is dispensable for normal growth and sexual and asexual development in culture (Xu and Hamer, 1996). The same kinase was shown in *F. oxysporum* to regulate invasive growth, hydrophobicity, and the generation of penetration hyphae, but also plays no role in vegetative growth (Di Pietro et al., 2001).

The two upstream kinases to Kss1, Ste7 (MAP2K) and Ste11 (MAP3K) are critical to this pathway and the loss of both also abolishes appressorium formation and virulence (Zhao et al., 2005). The same MAPK in *Colletotrichum gloeosporioides* is also required for appressorium formation and pathogenicity, but was also shown to be required for melanin synthesis (He et al., 2017). Fungal signaling is complex and

crossover between pathways appears common. In *S. cerevisiae*, the scaffold protein Ste5 helps to shuttle signaling from the virulence related Ste7 and Ste11, into a separate pathway acting through the Fus3 MAPK, instead of the Kss1 MAPK, which instead activates mating programs downstream of pheromone signaling (J Flatauer et al., 2005). Downstream of Kss1 signaling in *Fusarium oxysporum* f.sp. *lycopersici*, the Kss1 activated transcription factor Ste12 in is required for invasive growth but not for adhesion to tomato roots, vegetative hyphal fusion, or the secretion of pectinolytic enzymes (Rispail and Di Pietro, 2009).

Signaling also controls the production of fungal secondary metabolites that promote infection and survival. In *Aspergillus nidulans*, FadA through GTP signaling acts through protein kinase A (PKA), which in turn represses the production of sterigmatocystin and gliotoxin and promotes vegetative growth (Shwab and Keller, 2008). LaeA in *Aspergillus*, downstream of PKA, has been shown through microarray analysis to control the expression of roughly 100 secondary metabolite genes including those which produce pigments, alkaloids, and gliotoxin (Perrin et al., 2007). In the rice pathogen *Fusarium fujikuroi*, the TOR kinase likely acts upstream of the AreA transcription factor controlling biosynthesis of the virulence related secondary metabolites gibberellin and bikaverin (Teichert et al., 2006). In *Fusarium graminearum* the TOR kinase was shown to aid in controlling cell wall integrity, virulence, and secondary metabolism (Fangwei et al., 2014).

D. Plant disease

1. Sweet basil cultivars as a model for disease gene identification

In order to combat plant disease the breeding of plants for resistance has been the biggest push within the field of biology. However classical breeding can be time consuming as many generations of crosses are regularly needed to select for only those traits which lead to resistance while selecting away commercially unwanted characteristics. Although currently "genetically modified" crops are largely prohibited, the time will come when it may be our only course of action. Until then we can only aid in the development of new resistant crops by reducing as much as possible the time needed to go from concept to product. This is the primary focus of the second chapter below.

Peronospora belbahrii causes the basil downy mildew disease which affects the leaves of sweet basil (*Ocimum basilicum*) plants (Wyenandt et al., 2015). A spore of the pathogen lands on the surface of a leaf, germinates, and gains access to the mesophyll of the leaf (Fig 1.3). Once inside it begins to grow through the tissue and forms feeding structures within individual cells, called haustoria. The haustoria extract necessary molecules from the cells which the oomycete pathogen requires for growth and at the same time the pathogen uses the haustoria to deliver immune-suppressing effectors. After a matter of days given the proper environmental conditions, generally warm and humid, the pathogen reemerges from the stomata of the leaf producing sporangiophores. The sporangiophores extend roughly perpendicular from the surface of the leaf where they bare sporangia. The sporangia then detach from the sporangiophore and restart the cycle of infection.

Basil downy mildew is a particularly problematic disease for sweet basil as in addition to recurring roughly every year, the disease affects the leaves of the plant. For a crop like sweet basil the leaves are the source of its economic value as they houses the vast array of desirable secondary metabolites and are the consumed portion of the plant. During the course of the disease the leaves begin to yellow and sporulation also produces a dark gray discoloration on the under-side, and sometimes upper-side, of the leaf. These factors act to destroy the economically valuable portion of the crop.

Breeding for disease resistance to downy mildew in sweet basil is difficult as generally individual cultivars are not sexually compatible. In addition the broad phenotypic differences between cultivars, primarily in their secondary metabolite capacity, make selecting away unwanted traits complex.

In the study below I describe our efforts to identify resistance genes against basil downy mildew. To do this we have used a bioinformatics approach to predict genes unique to a sweet basil cultivar which shows strong resistance to the disease. The cultivar, Mirihani (MRI), was initially purchased from Horizon Seed Co., Williams, OR and selfed numerous times to generate a stably inbred line. To identify unique genes we compare MRI to a sexually compatible and therefore likely closely related sweet basil cultivar, SB22.

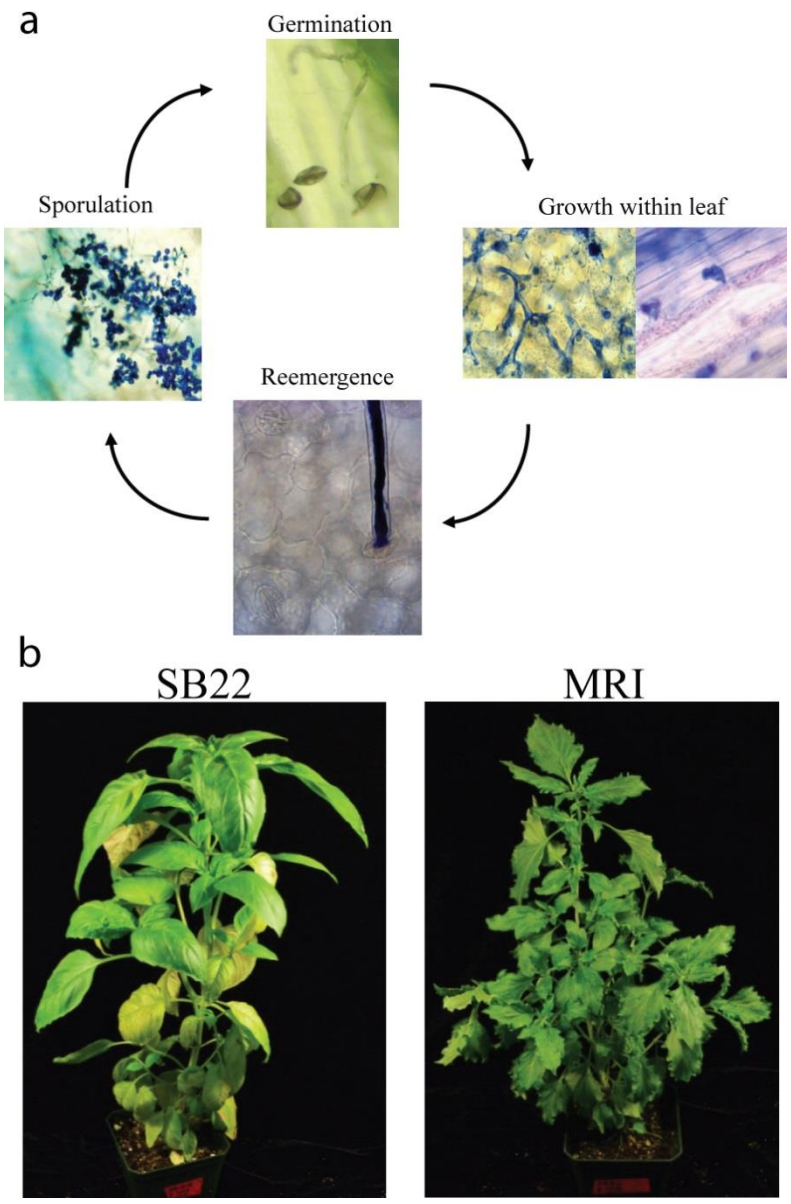

Figure 1.3: Basil downy mildew and basil. (a) Images taken of P. belbahrii during various growth phases in the SB22 cultivar. Blue staining is Trypan Blue. (b) The SB22 and MRI cultivars.

Although MRI exhibits strong disease resistance it is significantly different phenotypically from SB22, complicating breeding for resistance. The leaves of SB22 are broad, smooth edged, and downward cupped while the leaves of MRI are roughly flat with serrated leaf margins. The density of leaves in physical space is greater in

MRI than SB22, a problem for preventing basil downy mildew in susceptible plants. In addition, and most importantly, the secondary metabolites produced by MRI are in smell and taste much more alike to black licorice, very distinct from the classical basil taste consumers have come to expect. We know that MRI produces predominantly methylchavicol, while SB22 produces predominantly eugenol and 1,8-cineol (Robert Pyne, unpublished data).

Two studies by our collaborators identified at least 3 loci which are believed to contribute to resistance against basil downy mildew. One study examined the resistance phenotypes of various MRI and SB22 crosses in field studies in New Jersey in 2013 and 2014 (Pyne et al., 2015) where they predicted a dominant allele controlling resistance. In a second study examining QTL in this population they detected one dominant and two additive loci likely controlling resistance (Pyne et al., 2017).

2. Secondary Metabolites

Downstream of these high level signaling responses the plant must begin to produce molecules to protect itself, including reinforcing polymers like those associated with lignin, and small molecules through secondary metabolism which are antagonistic against the pathogen.

Upon detection of an invading pathogen the plant begins to deposit a number of modified cell wall materials that act to prevent disease progress. In addition, changes

in hydrophobicity and wax crystal structure of the cuticle are important, as differences reduced the overall germination and infection progress of rust and anthracnose pathogens on *Medicago truncatula* (Uppalapati et al., 2012). Once through the cuticle, the plant produces polymers that slow the progress of the invader. Callose is found in most sites of infection. The deposition of callose, the structure of which is called a papillum, is variable and depends on the environmental conditions and the infecting organisms. Its function seems to be that of a conduit matrix, supporting the delivery of antimicrobial compounds (Luna et al., 2011). Other hypotheses propose that callose acts as a physical barrier or that its primary function is to protect the plant from toxic metabolites within the deposition and from the pathogen (An et al., 2014). Additional defense compounds are produced in the deposition, called a papillae, including phenolic polymers. Some of these phenolics are thought to be structural and act through crosslinking within the deposition. Others may be have direct antimicrobial properties (von Röpenack et al., 1998).

The synthesis of a large number of important plant secondary metabolites lies downstream of phenylalanine ammonium lyase (PAL), effectively the beginning of the phenylpropanoid synthesis pathway. Various branches of this pathway are responsible for creating molecules important for plant defense including lignin and an entire family of phytoalexins. In many plants the expression of the phenylpropanoid pathway or sub-pathways are induced during some pathogen interactions including tomato, eucalyptus, cotton, lettuce, grape, barley, *Arabidopsis*, tanoak, and rice (Boddu et al., 2006; De Cremer et al., 2013; Fung et al., 2008; Hayden et al., 2014;

Kawahara et al., 2012; Meyer et al., 2016; Tan et al., 2015; Xu et al., 2011; Zhu et al., 2013). Fungi have undoubtedly developed methods to suppress this pathway as it is solely responsible for producing much of the plant's defense related compounds. In the case of cotton, infection of a resistant cultivar with *V. dahliae* induced the expression of PAL more rapidly and to a greater extent than in a susceptible cultivar (Xu et al., 2011). However this is not always the case, as after powdery mildew infection a susceptible grape cultivar showed upregulation of some phenylpropanoid pathways the same as its resistant grape partner (Fung et al., 2008).

The plant must redirect resources from pathways that can be sacrificed for the short term to other pathways that directly combat infection. An example that might support this hypothesis is the regularly observed downregulation of genes involved in light reactions and photosynthesis in general.

3. R Genes and Plant Resistance Genes

Plant resistance (R) genes have come to be a main focus of molecular plant pathology. Different in concept from genes that respond to the presence of molecules present in large groups of pathogenic organisms, e.g. flagellin and chitin, R genes seem to respond to an elicitor which is a protein or the action of a protein that indicate the presence of a specific pathogen. Many R genes fall within the family of nucleotide binding site leucine rich repeat (NB-LRR) proteins (Fig 1.4a), which are an ancient and widely dispersed gene group (Meyers et al., 2003). NB-LRRs primarily consist of an N-terminal domain of generally coiled-coil or Toll like

domain, an ATP binding domain, and a C-terminal LRR domain. Both the N- and C-terminal domains are thought to act in substrate and protein-protein interactions (DeYoung and Innes, 2006). Many R gene NB-LRR proteins act by monitoring defense signaling pathways which are targeted by pathogen effectors. The NB-LRR detects the effects of the effector protein and responds by activating the stronger pathogen triggered immunity.

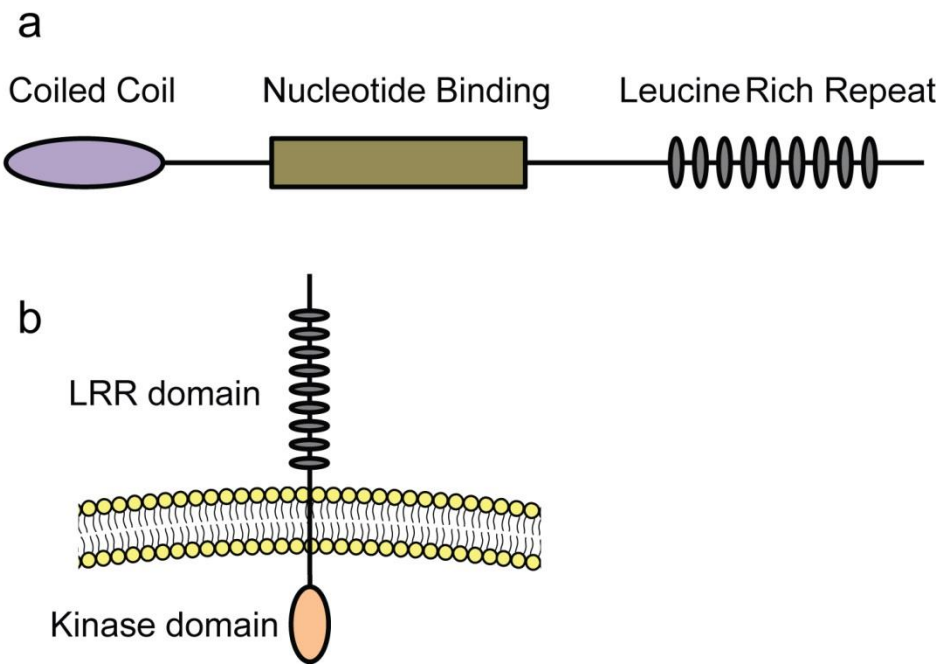

Figure 1.4: Typical defense signaling molecules. A large majority of defense signaling molecules are either (a) nucleotide binding leucine rich repeat proteins and (b) leucine rich repeat receptor like kinases. The general overall architecture is shown.

In some cases direct evidence for the interaction between an effector and an R gene exist. The *Arabidopsis* R gene Pto triggers a defense response when the plant is infected by strains of *Pseudomonas syringae* carrying the AvrPto gene. The interaction of Pto and AvrPto has been shown to occur even in the absence of FLS2 and both FLS2 and Pto compete for AvrPto binding, indicating that Pto detects the

effector directly (Xiang et al., 2008). In rice, plants containing the R gene Pi-ta are resistant to strains of *Magnaoprthe grisea* carrying the effector AVR-Pita. Direct binding of AVR-Pita to the LRR domain of Pi-ta was shown and is believed to be the triggering mechanism for plant defense (Jia et al., 2000). The exact function and target of AVR-pita during infection is still unknown.

The evolution of these proteins in plants and their pathogens has been referred to as an "evolutionary arms-race" (Anderson et al., 2010; Clay and Kover, 1996). When one individual develops a new R gene or effector, the counterpart organism is at the disadvantage until random mutation gifts it with an escape. This is likely the reason for the observation that both R genes and effector proteins exist within large, diverse families in their respective genomes. The Arabidopsis resistance to *Peronospora parasitica* (RPP) group of R genes illustrates the complexity and evolution of resistance to a pathogen through R genes. Characterization of different *Arabidopsis thaliana* accessions and pathogen strains clearly identified many segregating loci that contributed to resistance to different fungal strains for different cultivars (Holub et al., 1994).

The second class of defense genes discussed below are the receptor like kinases (RLKs) (Fig 1.4b). Unlike most NB-LRR proteins RLKs tend to be transmembrane where the kinase domain extends into the cell with, in the case of this study, a leucine rich repeat domain extending into the extracellular space.

4. Plant Hormones

Among the most regularly observed and important, but still fairly mysterious, functions of the plant defense network are the hormone signaling pathways. Hormone signaling directs shifts in gene expression in response to infection and prepares the plant for invasion in local, and in some cases, distal tissues. Different hormone pathways respond in different ways to particular types of pathogens based on the lifestyle of the pathogen.

The three major defense signaling plant hormones are salicylic acid (SA), jasmonic acid (JA), and ethylene (ET). SA regulates responses primarily to pathogens with biotrophic phases and is responsible for signaling for the initiation of systemic acquired resistance (SAR). SA defense responses are associated with the hypersensitive response. Generally opposite to the effects of the SA pathway, the JA and ET pathways are induced in response to necrotrophic pathogens and to insect attack.

Verticillium wilt of cotton, a necrotrophic root-infecting fungus, induces primarily a response in the ET pathway with some fluctuation in the GA, ABA, Auxin, and SA pathways (Xu et al., 2011). During infection with another necrotrophic root-infecting fungus, *F. oxysporum* 5176, the ET pathway in *Arabidopsis* responds by 1 dpi with the SA, JA, and ABA pathways lagging until as late as 6 dpi (Zhu et al., 2013). *Alternaria brassicicola* was seen to specifically not trigger a SA based response in *Arabidopsis* while *P. syringae* induced a response from the SA, JA, and ET pathways.

Interestingly, in the same study the aphid insect pest *M. persicae* did not induce any of the hormone pathways (De Vos et al., 2005). Plant hormone shifts can also differ between resistant and susceptible interactions for pathogens of the same species. Grape vine infected with powdery mildew undergoing an incompatible biotrophic interaction showed a SA-related transcript accumulation not seen in the compatible biotrophic interaction (Fung et al., 2008). Some interactions take significantly longer to develop, but react much the same way. At 9 dpi, around its shift from biotrophy to necrotrophy, infection of wheat by *Zymoseptoria tricici* induced ET and JA pathways (Rudd et al., 2015).

Like fungi, plants contain well conserved MAPK signaling cascades which orchestrate cellular function downstream of signal detection. In Arabidopsis MAPKs, MAP2Ks, and MAP3Ks have been shown to control lateral root formation, leaf senescence, flower architecture, stomatal development, embryo development, root cell elongation and division plane orientation, meristem maintenance, auxin transport, microtubule stabilization, among other processes (Xu and Zhang, 2015). Additionally, the MKK1 and MKK2 MAP2Ks function upstream of MEKK1 and MPK4, where they are implicated in controlling JA and SA regulated defense responses (Qiu et al., 2008).

In the study below we predict resistance genes by examining unique mRNA transcripts within MRI. We also examine the response by both plants to the disease in order to attempt to understand the broad mechanism for MRI resistance. In this, we

focus predominantly on disease signaling plant hormones. Although the hormones are known to crosstalk in order to fine tune the plants response they each have their own role to play in resistance. In particular interest to us is the hormone salicylic acid which is known to govern the response to biotrophic pathogens, like Peronospora belbahrii. SA initiates various pathways but is predominantly responsible for generating the hypersensitive cell death response. As biotrophic pathogens feed off of living cells, the death of the food source is the primary mode of combat these pathogens.

5. Pathogen Effectors

In response to the wide array of plant produced secondary metabolites and enzymes, pathogens have developed sets of genes that prevent plant defense signaling and damage to the fungal cell. The most well-known of these are the effectors. Many fungal secreted effectors interfere with plant signaling pathways controlling either the pathogen detection or the upregulation of plant defense products. These pathogen genes serve absolutely critical roles as their functions slow or completely prevent the plant's defense response. As a result they are generally observed to be some of the most highly expressed pathogen genes.

One of the hallmark features of effectors is their almost universal status as secreted proteins. As the plant and pathogen are separated by two walls and two membranes, secretion is required to deliver the effectors to where they function. Many contain a signal peptide and various conserved motifs, including an RXLR motif, which has

been used to predict effectors from oomycete genomes. Early evidence indicated that the RXLR-EER motifs of the *P. infestans* effector Avr3a were required not for secretion from haustoria, but for proper translocation into plant cells (Whisson et al., 2007). However, recent evidence indicates that this signal is cleaved from the protein prior to secretion making it unlikely that it has any function outside the cell (Wawra et al., 2017). Surprisingly, although some effectors have been characterized and many identified, it is still unclear how they translocate into plant cells when secreted into the extracellular space. Some effectors act outside individual cells, including the *Fusarium oxysporum* secreted in xylem genes which are known to act within the xylem vessels to promote disease (Thatcher et al., 2012).

The classic effector example is the interaction of the bacterial pathogen *Pseudomonas syringae* effector AvrPto and FLS2 receptor in *Arabidopsis thaliana* (Ronald et al., 1992). AvrPto, a kinase inhibitor, inhibits the function of the FLS2 receptor, blocking the downstream innate immunity response to bacterial flagellin. AvrPto has been shown to bind multiple receptor kinases including potato FLS2, potato elongation factor-Tu receptor, and the tomato LeFLS2. Likely in direct response to this, the Pto gene has evolved in *Arabidopsis,* which competes for binding with AvrPto and together with the Prf gene, triggers defense (Xiang et al., 2008).

Although many effectors seem to target plant proteins, some play a defensive role for the pathogen itself. The best characterized of these are the LysM domain effectors. Chitin, a molecule which can activate the plants defense, is a critical structural

component of the fungal cell wall. Plants have receptors that detect the presence of chitin, such as the Arabidopsis LYK5 receptor (Cao et al., 2014). Plant chitinases degrade the fungal cell wall, releasing chitin monomers that can trigger a defense response. To combat this threat, fungi have developed a set of effectors that bind to chitin monomers, effectively preventing them from binding to the plants chitin receptors or possibly preventing dimerization of those receptors. The *Magnaporthe oryzae* Slp1 LysM protein was shown to be dispensable for appressorium formation both is required for invasive growth (Zhang and Xu, 2014). In addition to possibly sequestering chitin monomers, *Cladosporium fulvum* Avr4 proteins bind to chitin and shield it from the effects of plant chitinases (van Esse et al., 2007).

The expression of effector genes appears to be tailored to the particular infection stage. During infection of wheat by *Z. tritici*, three candidate effector genes were similarly expressed by four closely related pathogen strains. A peptidase inhibitor showed decreased expression from 7 to 14 dpi (days post innoculation), and then a strong induction from 14 to 25 dpi. Two LysM domain containing proteins showed increasing expression from 7 to 14 dpi, with a strong reduction of expression from 14 to 25 dpi (Palma-Guerrero et al., 2017). Interestingly these shifts correlate perfectly with the shift from the biotrophic, requiring living plant cells for survival, to necrotrophic, living off of dead plant cell material, mode at 14 dpi. Genes required for masking the pathogen's presence are no longer needed after 14 dpi (LysM proteins), and genes required to resist the plant's activated defenses become activated after 14 dpi (peptidase inhibitor). We can see the effects of these effectors mirrored in the

plant's gene expression profiles. In the early phases of infection an overall downregulation of plant defense genes was observed with almost 60% of downregulated genes having plant defense related functions. Among these were four classical pathogen recognition (PR) genes and a lignin metabolism enzyme. At 9dpi, when the pathogen switches from biotrophy to necrotrophy, the plant only then beings to upregulate defense pathways including the jasmonic acid (JA) and ethylene (ET) pathways. At the same time many of the most highly expressed effectors in the pathogen shut down (Rudd et al., 2015).

6. Basil Downy Mildew Effectors

With data generated from an infected plant we not only have access to plant transcripts but also those produced by the pathogen within the plant. I have chosen to use this data to predict effector genes from P. belbahrii which it may use to ensure disease development. Effectors are a fairly mysterious group of proteins within oomycetes. Although they generally share a common base architecture, their functions are diverse and largely unknown beside their ability to reduce the plant defense response. Effectors come broadly in two types, apoplastic effectors and cytosolic effectors. Although both types are generally small, usually less than 100 AA, cytosolic effectors in many cases appear to require additional stabilization of the protein form as they are generally found to contain an even number of cysteines greater than 2. In contrast, cytosolic effectors instead have a characteristic motif within the first 30 AA which may play a role in gaining access to the host cytosol.

The canonical motif is the RXLR sequence followed sometimes 10-15 AA downstream by EER, RXLR-EER.

E. Methodology

1. Time point choice

Below we examine gene expression in infected plants in order to predict resistance genes based on expression. After landing on the surface of the leaf the spore germinates and penetrates likely within 24 hours. In my own experiments using MRI and SB22 I have observed obvious symptoms in SB22 but never in MRI (data not shown). Using microscopy and staining I have clearly observed heavy colonization of SB22 leaves but have never observed a single hyphae within MRI. Therefore resistance in MRI likely happens early during infection. In order to best capture the interaction and identify genes contributing to it we have sampled both 12 and 24 hours post inoculation (hpi). Samples taken at roughly a week after inoculation have shown heavy SB22 colonization. We have therefore also sampled timepoints at 2 and 3 days post inoculation to observe the response by the plant as colonization continues.

2. Mapping to the MRI combined assembly for gene clustering

In order to simplify MRI unique gene identification we map both the MRI and SB22 sequencing data to the MRI combined assembly. This step is only possible because of the close relationship between cultivars. We know of this close relationship because of their sexual compatibility and an experiment to identity nucleotide identity between orthologous genes described below. In mapping all data to a single assembly we do lose some information as SB22 unique genes should not map to an assembly

containing only MRI genes. However this method provides us with the ability to detect MRI unique genes visually, as genes unique to MRI should show no expression in SB22 samples.

3. The decision to search for MRI unique genes

In attempting to find resistance genes among MRI unique genes we exclude two possibilities. It is possible that the gene of interest is contained within both cultivars but that a mutation between the two creates the resistance phenotype. It is possible to screen for such mutations but the time required to complete the task would have been significantly longer. In addition any list of candidate genes would likely be much larger due to the divergence of the two cultivars, however little that divergence may be. In addition, without any prior knowledge of genes which do contribute to resistance or susceptibility the search space for the gene could be vast. The second possibility is that the resistance gene is in fact a susceptibility gene unique to SB22. In this case SB22 unique genes are of interest. However we do examine SB22 unique genes although they are not the focus of the manuscript.

CHAPTER II

EXPANDED SIGNALING NETWORKS RESPONDING TO ENVIRONMENTAL STRESSES HELP DEFINE THE *FUSARIUM OXYSPORUM* SPECIES COMPLEX

A. Introduction

A constantly evolving genome provides the genetic foundation for an organism to adapt to challenging environments. The fungal species complex of *Fusarium oxysporum* represents an exceptional model to study the relationship between genome evolution and organism adaptation. Phylogenetically related, members within the *Fusarium oxysporum* species complex (FOSC) include both plant and human pathogens and are collectively capable of causing plant wilt diseases in over one hundred plant species. Individual fungal isolates often exhibit a high degree of host specificity, reflecting rapid adaptation to particular host environments in a very short evolutionary time frame (less than 30 Mya) (Ma et al., 2013). *Forma specialis* has been used to describe strains that are adapted to a specific host. Comparative studies revealed that horizontally acquired lineage-specific (LS) chromosomes contribute to the host-specific pathogenicity (Ma et al., 2010; Schmidt et al., 2013; Vlaardingerbroek et al., 2016).

Protein kinases (which I simplify here as, kinases) are key regulators within cellular regulatory networks. They transduce extracellular and intracellular signals by modifying the activity of other proteins including transcription factors, enzymes, and other kinases via phosphorylation (Martin et al., 2015; Smith et al., 2010; Turrà et al., 2014). A

"protein kinome" encompasses all protein kinases within a genome (Goldberg et al., 2006; Hindle et al., 2014; Kosti et al., 2010; Li et al., 2015; Manning et al., 2002b; Miranda-Saavedra et al., 2007) and can be divided into several families, including the STE (homologs of the yeast Sterile kinases), CK1 (Casein kinase 1), CAMK (Ca^{2+}/calmodulin-dependent protein kinase), CMGC (cyclin-dependent kinases (CDKs), AGC (protein kinase A, G, and C families), HisK (Histidine kinase), Other, and Atypical families (Manning et al., 2002a). Some fungal kinomes have been functionally characterized (Bharucha et al., 2008; Breitkreutz et al., 2010; Coito et al., 2004; De Souza et al., 2013; Mok et al., 2010; Sharifpoor et al., 2012; Wang et al., 2011).

To study the variation and evolution of kinases with respect to FOSC host-specific adaptation, I compared kinomes of twelve *Fusarium oxysporum* isolates including ten plant pathogens, one human pathogenic strain, and one non-pathogenic biocontrol strain. In addition, we have included seven ascomycete fungal genomes available in public domains. Our study revealed a clear correlation between genome size and the size of the kinome of the organism. Due in part to the acquisition of LS chromosomes, the sizes of FOSC genomes are larger than other genomes included in this study, and so are their kinomes. Regardless of kinome size, I observed a highly conserved kinome core of 99 kinases among all fungal genomes examined. In contrast to this remarkably stable core, variation among kinase families and sub-families was observed across levels of taxonomic classification. Most interestingly, I observed the expansion of the target of rapamycin (TOR) kinase and histidine kinases among FOSC genomes. Monitoring nutrient availability and integrating intracellular and extracellular signals, the TOR kinase

and its associated complex serves as a central regulator of cell cycle, growth, proliferation and survival (Loewith and Hall, 2011). Increasing copy number of certain kinases may enable new functions, or add new temporal variations to existing pathways. The repeated, but independent, expansion of certain kinase families among FOSC genomes may suggest a fine tuning of similar pathways in responding to different host defenses or abiotic environmental challenges. This work is published in mSPHERE under the title "Kinome Expansion in the *Fusarium oxysporum* Species Complex Driven by Accessory Chromosomes". Dr. Goldberg predicted the kinomes of the included fungi using Kinannote, Yong Zhang divided the genomes of the FOSC isolates into core and lineage specific, Dr. Guo sequenced mRNA from temperature stressed Fo4287, I completed the remainder of the work.

B. Results

1. Overall kinase conservation defines a core kinome among ascomycete fungi.

I compared 19 Ascomycete fungal genomes (Table 2.1), including 12 strains within the FOSC, two sister species close to *F. oxysporum (F. graminearum* and *F. verticillioides*) (Fig 2.1a), two yeast genomes (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), two model fungal species (*Neurospora crassa* and *Aspergillus nidulans*), and an additional plant pathogen, *Magnaporthe oryzae*. With the exception of the two yeasts, genomes were annotated at the Broad Institute (Cambridge, MA) using the same genomic annotation pipeline (Galagan et al., 2003). All kinomes were predicted by Dr. Goldberg using the kinome prediction pipeline (Goldberg et al., 2013a) with the same parameters. Kinases were classified by Kinannote into the STE, CK1, CAMK, CMGC, AGC, HisK,

Other, or Atypical families (Fig 2.1b, Supp. Table 2.1). Kinases that could not be classified within specified statistical parameters were categorized as unclassified.

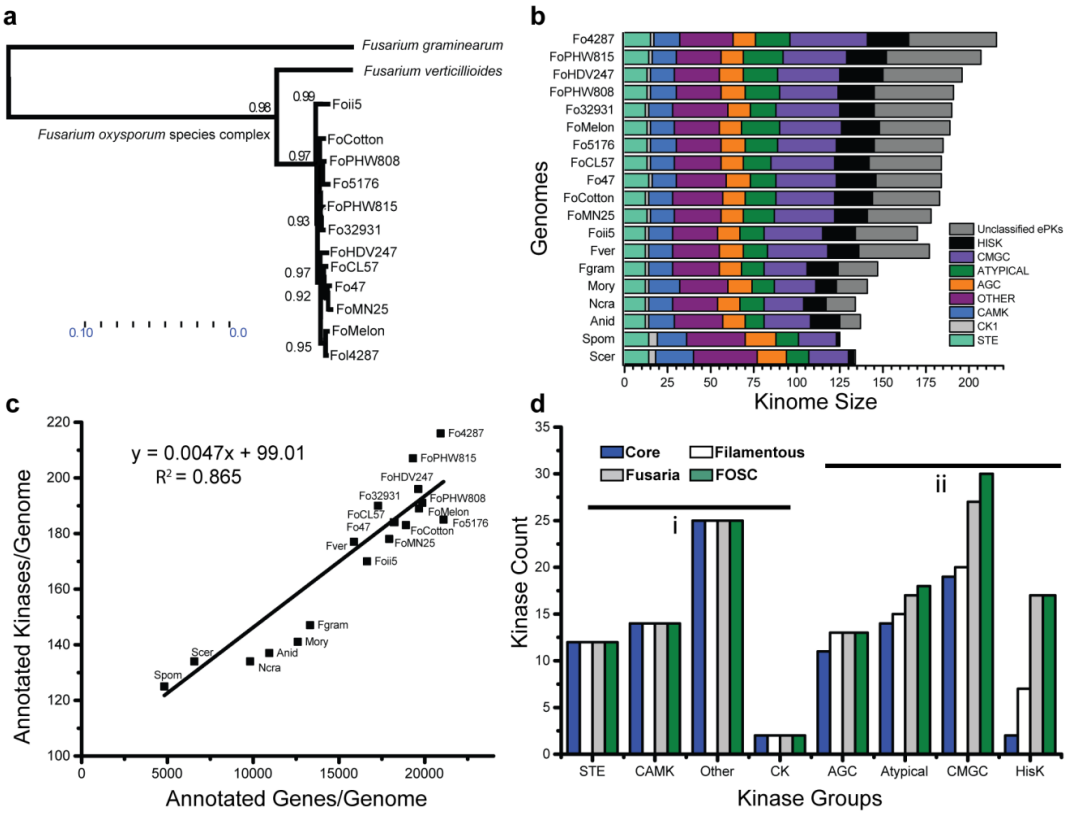

Figure 2.1. Kinomes across ascomycota. (a) A neighbor joining tree constructed from conserved genome genes showing the phylogenetic relationship among the Fusaria used in this study. (b) Kinases were broken up into major families and abundance per family plotted per species. Colors correspond to individual groups. (c) Total gene count plotted against kinome size in each genome. Origin coordinates are y=100, x=0. (d) By eliminating kinases missing from more than one species, I compiled "conserved kinomes" for Ascomycetes (all fungi here), Filamentous (all but *S. cerevisiae* and *S. pombe*), the genus *Fusaria*, and the FOSC. Some families remain relatively stable across species (group i; AGC, CAMK, CK, Other, STE), while others are expanded due to increased copy number or increased sub-family number (group ii; Atypical, CMGC, HisK). Total size of conserved kinomes: Ascomycete (99), Filamentous (108), Fusaria (126), FOSC (128).

Table 2.1 Fungal genomes analyzed in this study

Genome (Abbreviation[#])	Strain IDs*	Genome (MB)[**]	Total Genes	Kinome Size	Hosts	NCBI accession Number
F. oxysporum f. sp. *lycopersici* 4287 (Fo4287)	NRRL 34936	61.35	20,925	216	Tomato	GCA_000149955.2
F. oxysporum Fo47 (Fo47)	NRRL 54002	49.66	18,191	184	-	GCA_000271705.2
F. oxysporum Fo5176 (Fo5176)	NRRL 66176	54.94	21,087	185	Arabidopsis	GCA_000222805.1
F. oxysporum f. sp. *radicis-lycopersici (FoCL57)*	NRRL 26381	49.35	18,238	184	Tomato	GCA_000260155.3
F. oxysporum f. sp. *vasinfectum (FoCotton)*	NRRL 25433	52.91	18,905	183	Cotton	GCA_000260175.2
F. oxysporum HDV247 (FoHDV247)	NRRL 54007	55.18	19,623	196	Pea	GCA_000260075.2
F. oxysporum f. sp. *cubense* (Foii5)	NRRL 54006	46.55	16,634	170	Banana	GCA_000260195.2
F. oxysporum f. sp. *melonis* (FoMelon)	NRRL 26406	54.03	19,661	189	Melon	GCA_000260495.2
F. oxysporum f. sp. *lycopersici* MN25 (FoMN25)	NRRL 54003	48.63	17,931	178	Tomato	GCA_000259975.2
F. oxysporum f. sp. *conglutinans* (FoPHW808)	NRRL 54008	55.57	19,854	191	Brassica	GCA_000260215.2
F. oxysporum f. sp. *raphani* (FoPHW815)	NRRL 54005	53.49	19,306	207	Brassica	GCA_000260235.2
F. oxysporum 32931 (Fo32931)	NRRL 32931	47.90	17,280	190	Human	GCA_000271745.2
F. graminearum PH-1 (Fgram)	NRRL 31084	36.44	13,321	147	Wheat	GCA_000240135.3
F. verticillioides 7600 (Fvert)	NRRL 20956	41.77	15,869	177	Corn	GCA_000149555.1
Magnaporthe oryzae 70-15 (Mory)	FGSC 8958	41.02	12,593	141	Rice	GCA_000002495.2
Neurospora crassa OR74A (Ncra)	FGSC 987	41.03	9,820	134	-	GCA_000182925.2
Aspergillus nidulans FGSC A4 (Anid)	NRRL 194	30.06	10,937	147	-	GCA_000149205.2
Saccharomyces cerevisiae S288C (Scer)	ATCC 204508	12.07	6,572	134	-	GCA_000146045.2
Schizosaccharomyces pombe (Spom)	ATCC 24843	12.57	4,820	125	-	GCA_000002945.2

Overall, there is a positive correlation between the total number of proteins encoded in a genome and the total number of protein kinases within that genome ($y = 0.00473x + 99.01$ $R^2 = 0.86$) (Fig 2.1c). The dependency of these two variables (number of kinases versus number of proteins) also points to a potential minimal number of 99 kinases for

each ascomycete genome. Independently, I identified a core ascomycete kinome based on amino acid sequence conservation, taking a conservative approach by excluding kinase subfamilies that were missing from more than a single species. Here I use "conserved" to indicate kinases within a subfamily which occur in more than one species. This conserved core kinome has all major kinase families, including 12 kinases in STE, 2 CK1, 14 CAMK, 11 AGC, 19 CMGC, 2 HisK, 14 Atypical, and 25 Other families (Fig 2.1d), totaling 99 conserved kinase orthologs, in agreement with the prediction of the regression model. In addition to the ascomycete core kinome, I also compiled conserved kinomes for filamentous fungi (excluding both *S. cerevisiae* and *S. pombe*), the genus Fusaria, and the FOSC, each containing 108, 127 and 131 kinases respectively. Interestingly, the sizes of the STE, CK1, CAMK, and Other families remain effectively constant across phylogenetic divisions, while AGC, Atypical, CGMC and HisK families exhibit trends of continuous expansion, illustrated as categories i and ii in Figure 1d.

a. STE family

The STE kinases transduce diverse signals including osmolality, pheromone recognition and cell wall integrity, and alter cell growth patterns in response to extracellular changes (Widmann et al., 1999). The family includes MAP kinase kinases (MAPKK), MAP kinase kinase kinases (MAPKKK), and their upstream activators that function in MAPK signaling cascades. Evolutionarily conserved, a MAPKKK phosphorylates a MAPKK, and activated MAPKK phosphorylates a MAPK; activated MAPKs, belonging to the CMGC family described below, phosphorylate other proteins to further control gene expression and cellular function. The 12 conserved fungal STE kinases are divided into

nine subfamilies with a single kinase each, except the STE/PAKA and STE/YSK subfamilies, which have two and three kinases respectively (Supp. Table 2.2). Except *S. cerevisiae*, which lacks a single STE/YSK ortholog, all other genomes contain 12 orthologs.

b. CK and CAMK families

Like the STE kinases, the CK and CAMK kinases are highly conserved among the ascomycete fungal genomes. Each genome has 2 CK1 and 14 CAMK conserved subfamilies respectively. All filamentous fungal genomes have a single copy of each sub-family. The *S. pombe* genome lacks an ortholog of both the CAMK/CMK and the CAMK/Rad53 kinases, while the *S. cerevisiae* genome lacks an ortholog of the CAMK/CAMK1 kinase. An ancient kinase family of serine/threonine-selective enzymes, CK1 kinases are present in most eukaryotic organisms and are involved in important signal transduction pathways, including regulating DNA replication and circadian rhythm. In *S. cerevisiae*, the two CK1 kinases function in morphogenesis, proper septin assembly, endocytic trafficking, and glucose sensing (Robinson et al., 1992)). Found in almost all eukaryotic cells, CAMK kinases are activated by Ca^{2+} fluctuations within the cell and control tip growth, branching, spore production, cell cycle progression, and secretion, among other processes (Tamuli et al., 2011).

c. Other kinase family

The Other kinase family includes several unique eukaryotic protein kinases (ePKs) that cannot be placed into any of the major ePK groups based on sequence similarity. There

are 25 Other kinases, which are further divided into 24 subfamiles. Except for the Other/CAMKK subfamily, which contains two kinases, each subfamily of the Other kinase group has a single copy within each genome. The Other kinases account for a quarter of the ascomycete core kinome and many kinases within this group are involved in basal cellular functions including cell cycle control (Donaldson et al., 1998; Lundgren et al., 1991; Taylor and McKeon, 1997).

d. AGC family

The AGC family contains 11 subfamilies with a single orthologous copy each. All subfamilies are present in all genomes, except the AGC/YANK kinase is absent from the *S. cerevisiae* genome. AGC family kinases are cytoplasmic serine/threonine kinases regulated by secondary messengers such as cyclic AMP (PKA) or lipids (PKC) (Pearce et al., 2010). They are involved in signaling pathways that orchestrate growth and morphogenesis, as well as response to nutrient limitation and other environmental stresses (Sobko, 2006).

e. Atypical kinase family

There are 11 Atypical kinases, which are further divided into 10 subfamilies with a single copy in each subfamily, except the Atypical/ABC-1B subfamily which has two. Distinctively, kinases of the atypical family lack the canonical ePK domain, but have protein kinase activity (Manning et al., 2002b). This family also includes many functionally important kinases including the TOR kinase, a major regulatory hub within the cell controlling nutrient sensing, cell cycle progression, stress responses, protein

biosynthesis, and various mitochondrial functions (Baldin et al., 2015). Other subfamilies of this group also play significant roles in cell cycle progression (RIO), mRNA degradation (PAN), and the DNA damage response (ATM, ATR, TRRAP).

f. CMGC family

The CMGC family is an essential and large group of kinases found in all eukaryotes, accounting for roughly 20% of most kinomes. The group comprises diverse subfamilies that control cell cycle and transcription, as well as kinases involved in splicing and metabolic control (Malumbres, 2014), including cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAP kinases), glycogen synthase kinases (GSK), serine/arginine protein kinase (SRPK), and CDK-like kinases. Mitogen-activated protein kinases (MAPK) form the last step in the three step MAPK signaling cascades, which regulate functions from mating and invasive growth, to osmosensing and cell wall integrity (Hamel et al., 2012). Cyclin-dependent kinases (CDK) are widely known as controllers of the cell cycle and transcription (Malumbres, 2014). Kinannote reported 23 subfamilies of CMGC kinase, of which 18 were conserved and included in the conserved kinome. Each subfamily contains a single kinase, except for the CMGC/ERK1 subfamily that contributes two. The *S. cerevisiae* kinome lacks 3 out of the 18 conserved subfamilies (CMGC/CDK11, CMGC/DYRK2, CMGC/PRP4).

g. HisK family

Differing from other kinase groups discussed above, the HisK family is widely distributed throughout prokaryotes and eukaryotes outside the metazoans. This group of

kinases sense and transduce many intra- and extracellular signals (Catlett et al., 2003; Defosse et al., 2015; Li et al., 2010; Santos and Shiozaki, 2001). Distinctively, this family has the most significant expansion from yeast to filamentous fungi and is further expanded in the FOSC genomes (Figure 1d). All HisKs are classified into 11 families, or classes, based on the conservation of the H-box domain (Catlett et al., 2003). However, only two classes (class V and class X) can be considered orthologous among most ascomycete genomes, present in all genomes here except *S. cerevisiae* (Supp. Table 2.2). The class V HisK is orthologous to the *S. pombe* Mak1, while the single conserved class X kinase is orthologous to the *S. pombe* Mak2/Mak3 kinases. Mak2 and Mak3 are known peroxide sensors (Buck et al., 2001), while all Mak1/2/3 are predicted to have a central role controlling the stress responses network (Pancaldi et al., 2012).

2. LS chromosomes contribute to the individual expansion of FOSC kinomes.

The conserved kinome represents defining pathways shared among Ascomycetes across evolutionary time. Our data support the hypothesis that the unique additions to each kinome enable species-specific adaptation. Among species here *F. oxysporum* kinomes are the largest. At roughly twice the size of the core ascomycete kinome, the average *Fusarium* kinome contains 185 kinases in 112 subfamilies (Table 2.1, Supp. Table 2.2). In addition, I observed a positive correlation between kinases encoded the LS region of each strain, defined as LS kinases here after, with the total number of LS genes ($y = 0.0049x + 6.11$, $R^2=0.57$) (Fig 2.2a), suggesting that linage specific (LS) chromosomes

contribute directly to the expanded kinomes.

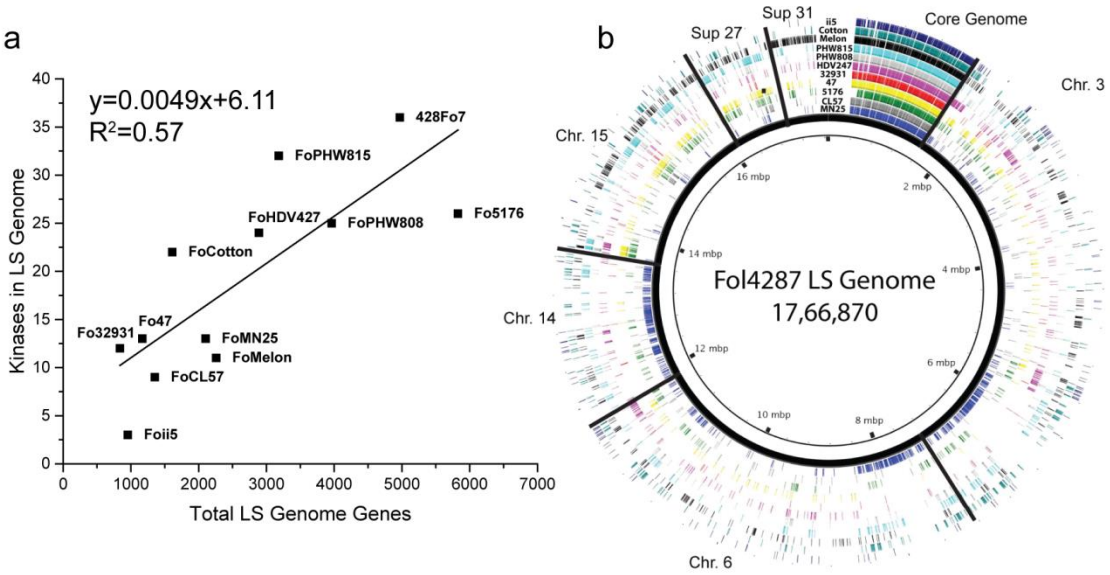

Figure 2.2: The LS genome contributes to kinome expansion. (a) The number of genes in the LS genome was plotted against the number of kinases found with the LS genome of each Fo strain. (b) The LS genome of each FOSC strain was compared to the LS genome of the reference strain. Colored lines indicate regions of conservation. Genomes were mapped to core genome contig 14 (roughly the first 1/5th of chromosome 1, a conserved chromosome) as a control. Large regions of color indicate potentially shared LS genome contigs. Names listed at the top of the ring correspond to the strain Fo strain for that particular colored ring (i.e. FoCotton is listed as Cotton). For brevity the Fo identifier was left off of the strain ID labels.

This correlation shares a similar slope as that of the whole genome, but effectively without a minimum kinase requirement.

In agreement with a comparative study among three *Fusarium* genomes (Ma, Does et al. 2010), little conservation was observed among the LS genomes (Fig 2.2b). The genome of each FOSC strain was partitioned into core and lineage specific (LS) using an "eliminating core" method (See Methods for detail). On average, the coverage of shared sequences among core genomes is over 70% (> 90% sequence similarity), as illustrated by supercontig 14 from the core. The average overlap for LS chromosomes of the

reference strain 4287 was 6.3%. The most significant conservation was observed in chromosome 14 between two tomato pathogens MN25 (race 3) and the reference genome Fo4287 (race 2) (Supp. Table 2.3). This closeness was the result of the transfer of a pathogenicity chromosome, chromosome 14, suggested by comparative studies (Ma et al., 2010; van Dam et al., 2016). Other conserved fragments also observed, however their functional significance remains elusive.

To evaluate the overall functional importance of LS kinases, Dr. Guo sequenced mRNA isolated from the reference strain Fo4287 under two experimental conditions, one at room temperature and the other shifted to 37 ^{o}C (Methods for details) (Supp. Table 2.5). Roughly half of the annotated Fo4287 genes (9,914) were expressed in either condition, including 85% of kinases within the conserved ascomycete kinome (85 out of the total 99 core kinases). In contrast, among the 44 LS kinases of Fo4287 (Supp. Table 2.4), the expression of only 8 (18%) was detected. Expressed LS kinases include 2 HisK (FOXG_14953, FOXG_15045), 2 Atypical/FunK1 (FOXG_12507, FOXG_14032), 2 Other/HAL (FOXG_06573, FOXG_07253), 1 STE/YSK (FOXG_14024), and 1 Unclassified kinase (FOXG_16175). For most cases, I saw higher levels of expression for the copies in the core compared to the LS copies.

I compared identified kinases to existing RNA-sequencing data generated by the lab during heat stress on solid media of Fo4287. Interestingly, heat stress at 37°C uniquely induced the expression of 4,394 genes, accounting for 44% of all expressed genes (Supp. Table 2.5). Similarly, about 44% of all expressed core kinases are induced under the heat

stress, consistent with the functional conservation of the core genome. In contrast, 6 out of 8 expressed LS kinases were induced at 37°C, including a HAL kinase (FOXG_06573), a Class IV HisK (FOXG_14953), one Atypical/FunK1 (FOXG_12507) and three Unclassified kinases. The overall expression pattern of the Fo4287 kinome supports the potential function of some LS kinases, especially in coping with different stresses. Additional functional studies under diverse stress conditions may capture the expression of other, potentially condition specific, LS kinases.

3. Expanded FOSC kinases enhance signaling transduction in cell cycle control and environmental sensing.

Expanded families belong to the category ii families (Fig 2.1d), enhancing functions related to cell cycle control and environmental sensing. For a soil-borne pathogen with strong host-specificity, like *F. oxysporum*, the adjustment of growth and cell cycle control in response to environmental cues are likely essential for survival.

a. The target of rapamycin (TOR) kinase

One of the most interesting expansions is the TOR kinase, which occurred in 7 out of 10 sequenced plant pathogenic FOSC strains. The TOR kinase is a top regulator of nutrient-sensing which dictates cellular responses according to the levels of nutrients and oxygen (Heitman et al., 1991). A member of the Atypical/FRAP kinases, the TOR kinase is highly conserved in nearly all eukaryotic organisms from fungi to humans with few exceptions (Shertz et al., 2010). Even though there are two TOR paralogs in the *S. cerevisiae* and the *S. pombe* genomes, almost all euascomycete fungal genomes have only

one copy (Teichert et al., 2006). Our study revealed a TOR kinase expansion in 7 out of 12 sequenced FOSC strains, with 5 strains containing 1 and two strains containing 2 additional copies in addition to the single orthologous TOR kinase (Fig 2.3a). The orthologous copies of the TOR kinase (Marked as Core clade in Fig 2.3a) form a monophyletic group with almost identical amino acid sequences (>99.9%). A total of nine TOR paralogs (marked as LS clade in Fig 2.3a) clustered together with an average 95.9% amino acid identity compared to the orthologous copies, and an average 97% among LS paralogs (Fig 2.4), arguing against a recent duplication within an individual genome as a mechanism for the expansion of this gene family. However, there does appear to have been a recent duplication of one pair of the paralogs in a pea pathogenic strain (FOVG_18014 and FOVG_19124).

Six out of the nine TOR paralogs encode full length proteins containing all functional domains (Fig 2.3b). One of the paralogous copies in the melon pathogen lacked the kinase domain and two paralogous copies (FOXG_15946 and FOTG_17386) had a truncation at the N-terminus (part of the HEAT repeats) of the protein. With the highly conserved domain structure, I was able to identify a shared mutation, M2345 to lysine, among all TOR paralogs near the very end of the catalytic loop of the kinase domain, while all other sequence motifs essential for TOR function were conserved (Fig 2.3b). Based on the protein structure, this catalytic loop forms the back of the ATP binding pocket, an invariant site for all conserved TOR kinases (Sauer et al., 2013).

López-Berges et al. (López-Berges et al., 2010) reported that the truncation of FOXG_15946 was the result of a transposon insertion, and very low level of expression of this truncated copy was detected in rich medium. Our RNAseq data on the Arabidopsis pathogen Fo5176 during the course of infection also support the expression of both the ortholog (FOXB_14471) and the expanded copy (FOXB_02991) of the TOR kinases and the expression of the orthologous copy is roughly three times higher than the paralogous copy (Ma, unpublished data).

As the TOR kinase is the direct target of the antibiotic rapamycin, I tested the sensitivity of 6 Fo strains, 4 with one TOR copy, and two with 2 TOR copies, to rapamycin. Rapamycin (50 ng/mL) reduced the rate of growth of all strains ($p < 0.005$) (Fig 2.3c, d). For the four strains containing the single copy of the TOR kinase, reduction in growth rate ranged from between 64% (MN25) to 82% (Fo47). The two strains carrying an additional TOR kinase, Fo5176 and Fo4287, showed increased resistance to the rapamycin treatment, resulting in 50% and 42% reduced growth, respectively. These results are consistent with an additional TOR kinase providing enhanced resistance to the drug rapamycin.

Functional studies suggested that the TOR signaling pathway may control pathogenic phenotypes, such as virulence in *F. oxysporum* infecting tomato (López-Berges et al., 2010) and toxin production in *F. fujikuori* (Teichert et al., 2006). Even though I don't have direct evidence on the specific function of the expansion of TOR kinase among FOSC, the high frequency (70% plant pathogenic strains), the preserved protein domain

organization, and the increase resistance to rapamycin all suggest its potential functional involvement in the adaptation to diverse environments.

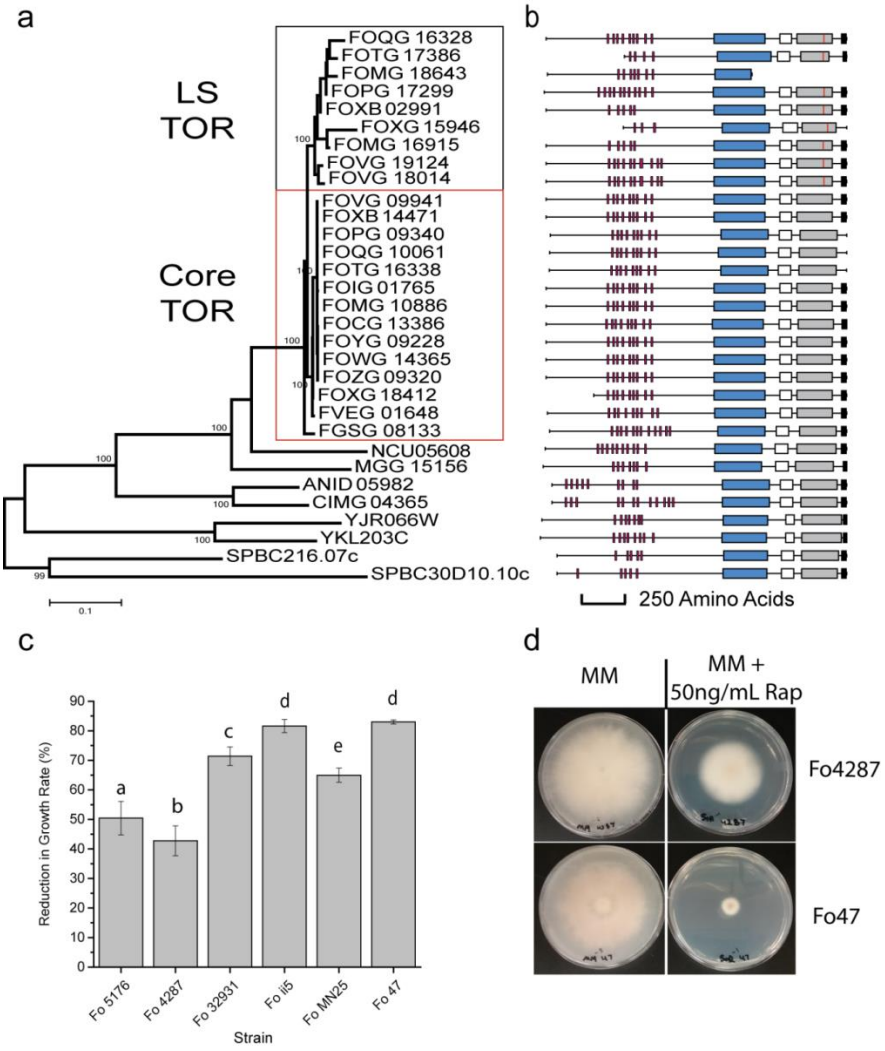

Figure 2.3: TOR kinase expansion through the LS genome. (a) A phylogenetic tree of TOR kinases constructed by NJ algorithm using TOR kinase domain protein sequence. The core and LS clades of TOR kinase in Fo are shown in boxes. (b) On the right of the tree is a protein domain image showing the relative structure of each TOR kinase. A red vertical line inside the kinase domain indicates the mutation M2345L. Domain shown are as follows: HEAT Repeats (Red), FAT Domain (blue), RFB Domain (white), Kinase Domain (gray), FATC Domain (black). (c) Measurement in reduction of radial growth rate of *F. oxysporum* 6 strains to 50 ng/mL Rapamycin on minimal media plates amended with antibiotic compared to no-antibiotic MM plate controls. Three biological replicates were done for each treatment. Letters (a-d) indicate groups which are statistically different from one another using a T-test comparison between all groups. (d) Pictures of fungal growth on minimal media with and without antibiotic at 7 days post inoculation. Plates for the most resistant (Fo4287) and least resistant (Fo47) strains are shown.

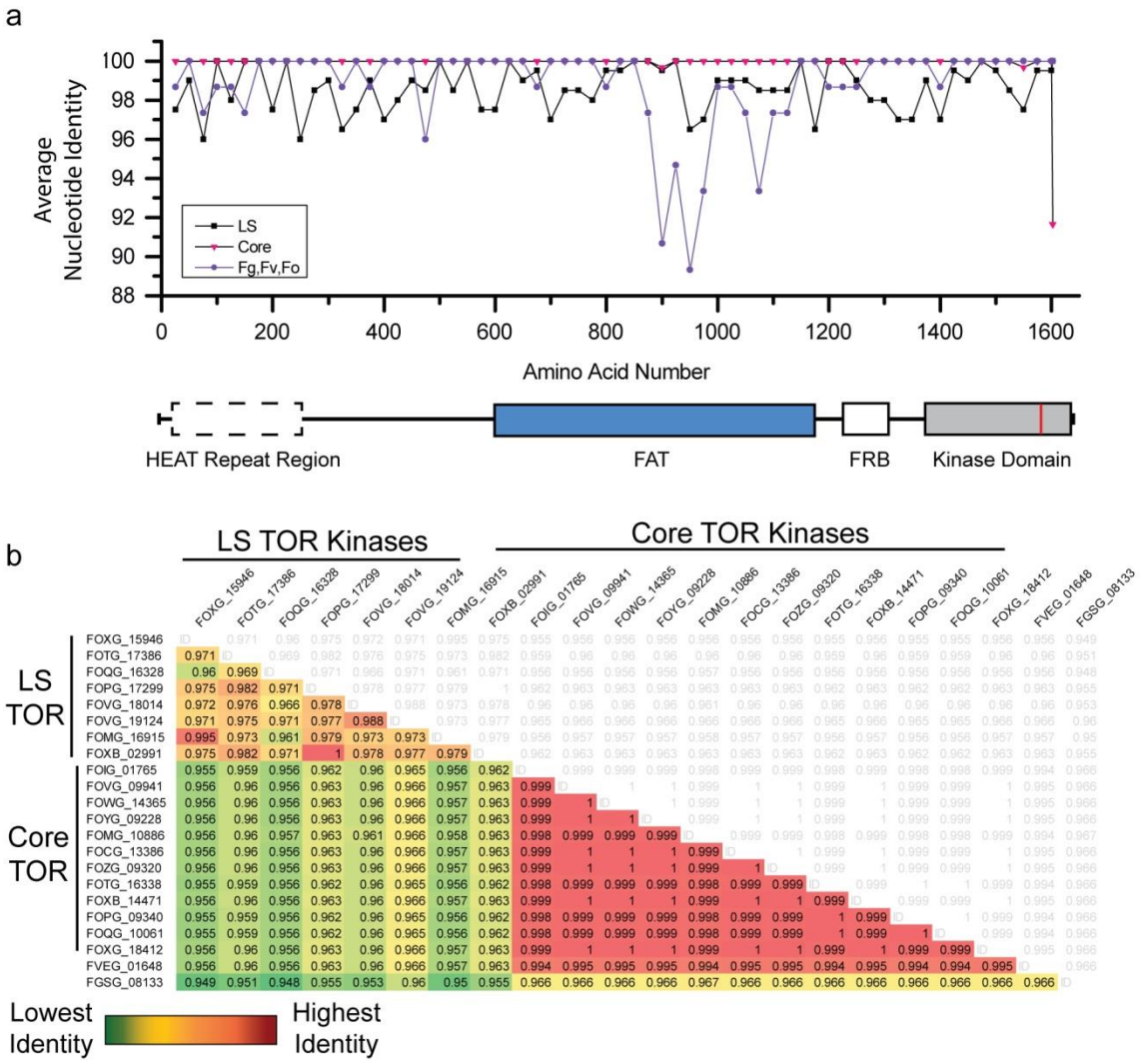

Figure 2.4: Nucleotide conservation among TOR kinases confirms single origin for LS TOR kinases. (a) The domain structure of the TOR kinase is shown with the M2345L marked as a red band. Above the graph indicates the percent amino acid conservation in 10 bp windows across a gap deleted protein sequences for Fg, Fv, and Fo4287 (purple), all FOSC core genome TOR kinases (blue), and FOSC LS genome TOR kinases (black). (b) The nucleotide identity between TOR kinases in Fusaria was compared. Gene IDs are organized into LS, Core, and Fg and Fv at the bottom. Colors are relative representations of the nucleotide identity.

TOR and its complex modulate cell growth patterns by partnering with other signaling components to control a number of regulatory subnetworks (Laplante and Sabatini, 2009). Interestingly, I observed the expansion of several kinase families functioning as TOR partners, including CDC2, CLK and the BCK1 MAPK.

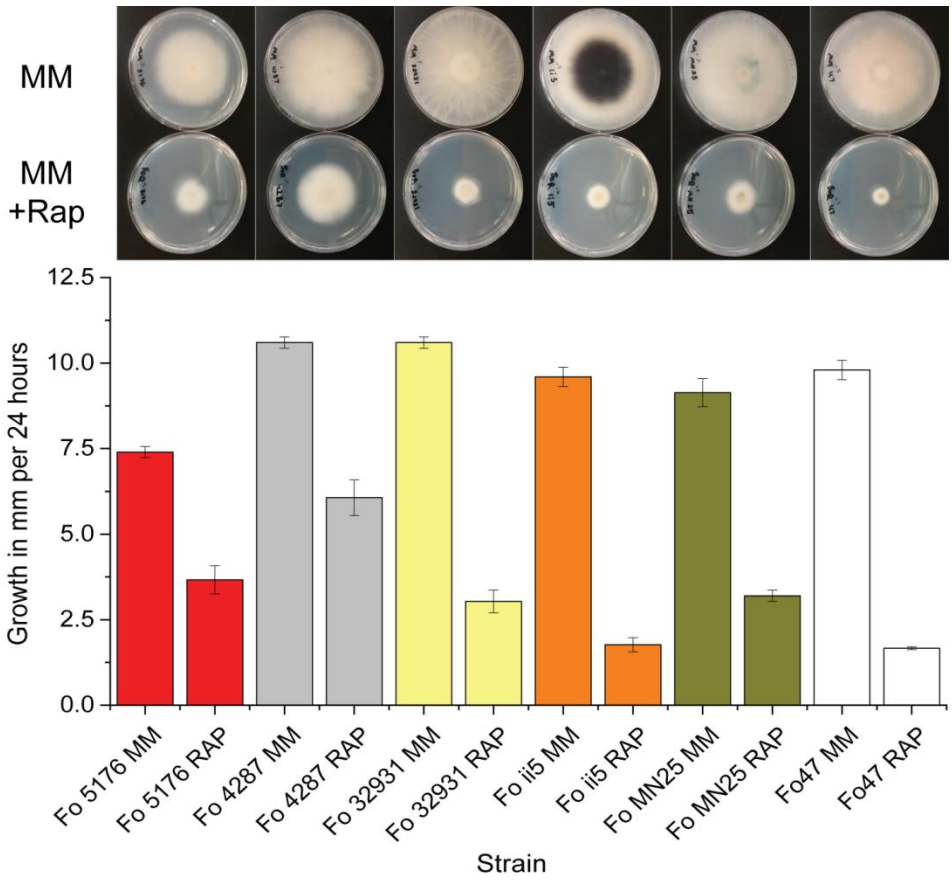

Figure 2.5: Differential inhibition of FOSC isolate growth by rapamycin. (top) Images for one representative pair of plates were taken for each strain at 7 dpi. Strains were cultured on minimal media (top plate) or minimal media with 50 ng/mL Rapamycin (bottom plate). (below) The growth rate in millimeters per 24 hours was calculated for each strain on both media. All MM and MM+Rap pair differences are statistically significant (t-test, P.value < 0.005). All plates were done in triplicate.

b. BCK1

The *Fusarium* BCK1 kinase family is expanded among 8 out of the 10 plant pathogenic isolates. All BCK1 kinases within FOSC fall into two clades, including an orthologous clade conserved in all *Fusarium* genomes, and a LS group only found within nine FOSC isolates (Fig 2.6a). BCK1 is a member of the STE group kinase and is the only expanded family belonging highly conserved category I families (Fig 2.1d). Part of the MAPK

signaling cascades under the control of the Rho1 GTPase and PKC1, BCK1 interacts with the TOR complex through direct binding with one of the subunits, LST8 (Breitkreutz et al., 2010; Sharifpoor et al., 2012) and regulates transcription during the G1 to S phase transition as reported in *S. cerevisiae* (Madden et al., 1997).

c. CMGC families

Several kinase families of the group CMGC (Fig 2.1b), including CDC2, CDC2-like kinase (CLK), and SRPKL kinases, are expanded in FOSC genomes. Similar to TOR and BCK1 kinases, CLK gene expansion was only observed among FOSC genomes. An LS CLK was present in 7 out of 10 sequenced FOSC plant pathogenic strains. The core and the LS CLKs are phylogenetically distinct with strong bootstrap support (Fig 2.6b). However, the expansion of CDC2 (Fig 2.7a) and SRPKL kinases (Supp. Table 2.2) already occurred within the genus *Fusarium* before the split of *Fusarium* species. Collectively, SRPKL kinases constitute 4% to 7% of the total kinome among FOSC genomes and expansion happened multiple times. All other fungal genomes examined here contain a single CDC2 kinase, while *F. graminearum*, *F. verticillioides* and all FOSC genomes have two. Expansion continued further within the FOSC. The tomato pathogen Fo4287 contains 10 additional LS CDC2 kinases and a pea pathogen FoHDV247 contains one extra copy.

Not surprisingly, the two CDC2 genes detected in all *Fusarium* species are located in the core of the genome and recent duplication events resulted in the 10 LS copies in the Fo4287 (Ma et al., 2010).

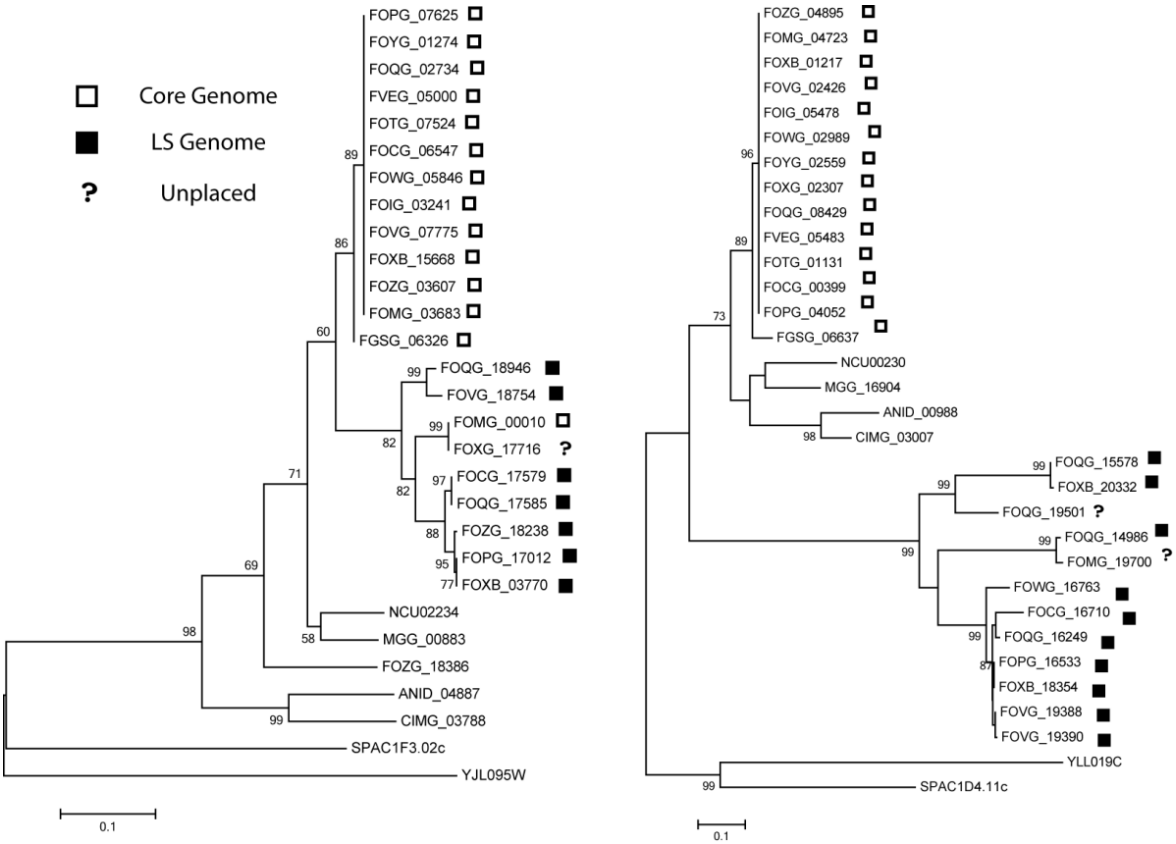

Figure 2.6: BCK1 and CLK kinase sub-families. Phylogenetic trees (Maximum Likelihood) using nucleotide sequences from BCK1 (a) and CLK (b) kinase sub-families from all species show the clear separation of core and LS genome orthologs. Scale bar is substitution rate. Black boxes, LS genome genes. White boxes, core genome genes. Question marks, unknown contigs.

These expanded CMGC kinase families all have functions related to cell cycle control directly or indirectly linked through the TOR signaling pathways. The CDC2 kinase is a major regulator of the cell cycle. In *S. pombe*, the CDC2 kinase (SPBC11B10.09) directly regulates the G1-S to G2-M transitions and DNA damage repair (Caspari and Hilditch, 2015; Den Haese et al., 1995). CLK kinases are involved in various cellular functions, including regulating the cell cycle in *S. pombe* (Yu et al., 2013), controlling ribosome and tRNA synthesis in response to nutrient limitation and other cellular stresses in *S. cerevisiae* (Lee et al., 2012), and regulating cell wall biogenesis, vegetative growth, and

sexual and asexual development in *Aspergillus nidulans* (Choi et al., 2014; Kang et al., 2013). Like the CLK kinases, the SRPK kinases have been implicated in the control of SR protein mediated splicing in a TOR dependent manner in *S. cerevisiae* (Siebel et al., 1999).

Functional importance of these expanded kinases in *Fusarium* genomes has been indicated by a few, however solid, functional studies. Deletion of the orthologous CDC2 kinases (FGSG_08468) in *F. graminearum* resulted in profound pleiotropic effects including reduced virulence and decreased ascospore production. Deletion of the *Fusarium* specific copy (FGSG_03132) resulted in a milder, but still significant phenotype (Wang et al., 2011). In *F. graminearum*, the CLK kinase was downregulated during conidial germination (Wang et al., 2011). Removal of one SRPKL kinase in *F. graminearum* (FGSG_02488 a SRPKL2) reduced the production of the toxic secondary metabolite deoxynivalenol production by half (Wang et al., 2011). The expression of the same gene was suppressed significantly during sexual development. Among the 9 core SRPKL kinases, 5 are expressed and four (two SRPKL1 kinases: FOXG_08977 and FOXG_10022; one SRPKL1 kinase: FOXG_21922; and one SRPKL3 kinase: FOXG_19803), were upregulated at 37°C. Two CDC2 kinases within the core were expressed under both conditions, while the expression of the 10 LS copies was not detected. Of the 51 Unclassified kinases in Fo4287, 12 core kinases and one LS kinase (FOXG_16175) had detectable expression in the RNAseq data generated from the reference strain.

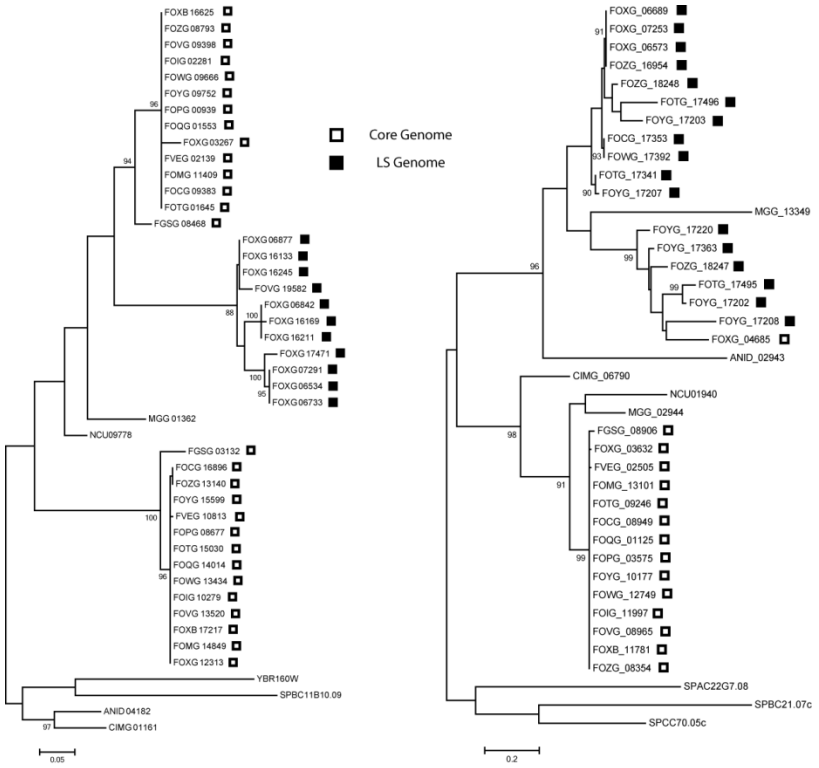

Figure 2.7: CDC2 and HAL kinase sub-families. Phylogenetic trees (Maximum Likelihood) using nucleotide sequences from CDC2 (a) and HAL (b) kinase sub-families from all species show the clear separation of core and LS genome orthologs. Scale bar is substitution rate. Black boxes, LS genome genes. White boxes, core genome genes. Question marks, unknown contigs.

d. Enhanced environmental sensing accomplished through HisKs

One of the most significant kinase expansions occurs in the HisK group (Fig 2.3) known to play important roles in sensing and transducing many intra- and extracellular signals (Catlett et al., 2003; Defosse et al., 2015; Li et al., 2010; Santos and Shiozaki, 2001). The two yeast genomes, *S. cerevisiae* and *S. pombe* encode one and three HisKs respectively, while the HisK group consistently expands across the filamentous fungi (Fig 2.1d). All filamentous ascomycete fungal genomes included in this study have more than 10 HisKs and FOSC isolates have by far the most HisKs, with the tomato wilt pathogen Fol4287 and the tomato root rot pathogen FoHDV247 both predicted to have 23 HisKs. Fungal

HisK have been divided into 11 classes (Catlett et al., 2003). Excluding *F. graminearum,* which lacks the class IV HisK (TcsA kinase), all *Fusarium* genomes contain 10 distinct classes of HisK, only lacking the class VII that was reported in the *C. heterostrophus* and *Botrytis cinerea* genomes (Catlett et al., 2003). Interestingly, the class II HisK is uniquely present in all *Fusarium* genomes and few other plant pathogenic Ascomycetes including *Cochliobolus heterostrophus* and *Bipolaris maydis* (Catlett et al., 2003). The most significantly expanded FOSC HisKs are in class I and class IV (Figure 5), for instance class I HisKs continue to expand from 5 in *F. verticillioides* and 6 in *F. graminearum* to 7 or more among FOSC genomes.

According to protein domain organization, class I and class IV contain a GAF and a PAS N-terminal domain, respectively. PAS domains detect signals including light and oxygen, usually through binding to an associated cofactor (Taylor and Zhulin, 1999). GAF domains form small binding pockets with potentials to bind to small signal molecules, such as cyclic GMP, cyclic AMP, or both. The motif is ubiquitously present in hundreds of signaling and sensory proteins from all kingdoms of life (Ho et al., 2000). Between class I HisKs, AA conservation of the GAF domain is low, likely indicating non-redundant function. Unlike the class I HisK, class IV copies have very similar N terminal domain sequences. Other than class I HisKs, class II, VIII, X and XI Hisks also contain GAF-related domains, while class IV, V, IX and XI also contain PAS-related domains.

First reported in prokaryotes, bacterial HisKs involve two-components where the first component senses environmental signals and autophosphorylates a conserved histidine

residue, transducing the signal to the second component through phosphate transfer (Thomason and Kay, 2000). Among almost all eukaryotic systems, these two components are fused to form hybrid HisKs, containing the conserved histidine residue, ATP-binding and Response Regulator (RR) Receiver domains function in three-step "phosphorelay" reactions. Upon activation, the HisK phosphorylates itself twice, ending with phosphorylation of the response receiver domain at the C-terminus. This phosphate is then transferred via a histidine phosphotransfer (Hpt) protein to a response regulator, which carries out downstream signaling functions. Each *Fusarium* isolate was found to have two RR proteins, orthologous to the yeast SKN7 and SSK1 RR, which are both involved in the osmotic stress response downstream from the SLN1 HisK and YPD1 Hpt proteins (Krems et al., 1996; Posas and Saito, 1998). The single yeast Hpt protein YPD1 has a single counterpart in *F. graminearum*, *F. verticillioides*, and each FOSC strain except for strains FoPHW808, FoPHW815, FoMelon, and FoCL57, which each have two. In these four cases, strains contain a copy that is located within the core genome and one within the LS genome. Core genome copies share a 100% AA identity across the 143 AA protein, while LS genome copies share roughly 91% AA identity across 154 AA.

e. HAL kinases

Regulators of the cell's primary potassium pumps (Forment et al., 2002; Mulet et al., 1999), the HAL kinases, are expanded in 6 out of the 12 sequenced FOSC genomes. In addition to the orthologous copies located in core genome regions, the second phylogenetically supported group contains all LS HAL kinases. Different from other

FOSC specific expansions, this subfamily is also expanded in *M. oryzae* and *A. nidulans* (Figu 2.7b).

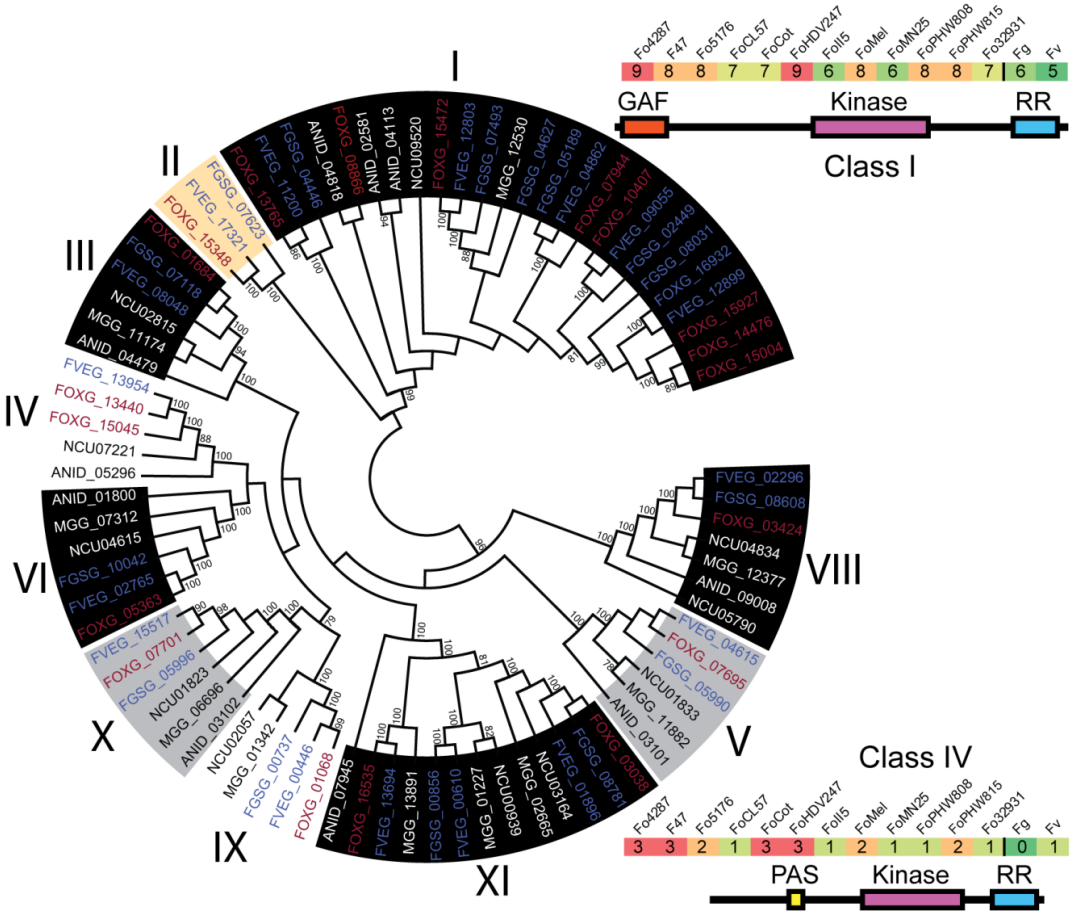

Figure 2.8: Histidine kinases are expanded in the Fusaria. A phylogenetic tree was constructed for all two-component signaling histidine kinase nucleotide sequences and labeled according to their class (family)(Neighbor Joining). Some kinases were excluded as they lacked a large portion of the aligned conserved sequence among HisK. Gray highlight, part of the conserved ascomycete kinome. Black highlight, part of the filamentous kinome. White highlight, part of the Fusaria kinome. Orange highlight, only found in Fusaria in this study. Fo gene IDs are colored red, Fg and Fv gene IDs are colored blue. All bootstrap values above 80% are shown. The domain structure for the class I and class IV HisK are shown along with a heatmap of the copy number for each class among the Fusaria.

Noticeably, the HAL kinases are most significantly expanded to 6 copies in the strain Fo32931, a *F. oxysporum* strain isolated from an immunocompromised patient**.** Among the five HAL kinases in the reference genome Fo4287, all except one (FOXG_04685)

were expressed (no change of expression with the shift of temperature), but the single core genome HAL kinase was expressed at least 30 fold higher than the LS HAL kinases (Supp. Table 5).

f. Unclassified protein kinases

In addition, I observed a large number of kinases in the Unclassified family among all *Fusarium* genomes. On average, each *Fusarium* genome contains 41 unclassified kinases, accounting for roughly half of the kinome beyond the core ascomycete kinome and constituting 30 to 50 percent of the LS kinome among FOSC genomes. No Unclassified kinases were present among all ascomycete fungal genomes examined, but four are conserved within the genus *Fusarium* (not shown) and more conservation was observed among FOSC genomes (Fig 2.9). Among these unclassified kinases, eight were included in a *F. graminearum* knockout study and six mutants exhibit measurable phenotypes (Wang et al., 2011), including decreased production of the mycotoxin deoxynivalenol in five mutants (FGSG_13509, FGSG_02153, FGSG_00792, FGSG_06420, FGSG_10591), and significant upregulation during sexual development for another (FGSG_12132).

g. Atypical/FunK1

Among the expanded families one of the least understood is the Atypical/Funk1, only found in complex Basidiomycota and Pezizomycota and possibly involved in the switch to multicellularity (Stajich et al., 2010). Among Fusaria, copy number ranged from a single copy to 9, including one that is highly conserved among all *Fusarium* genomes. The expression of this conserved Funk1 gene in *F. graminearum* (FGSG_03499) is

strongly upregulated during sexual development (Wang et al., 2011). Since no sexual reproduction was reported in *F. oxysporum*, the conservation of this gene within the *Fusarium* genus suggests functions other than sexual reproduction.

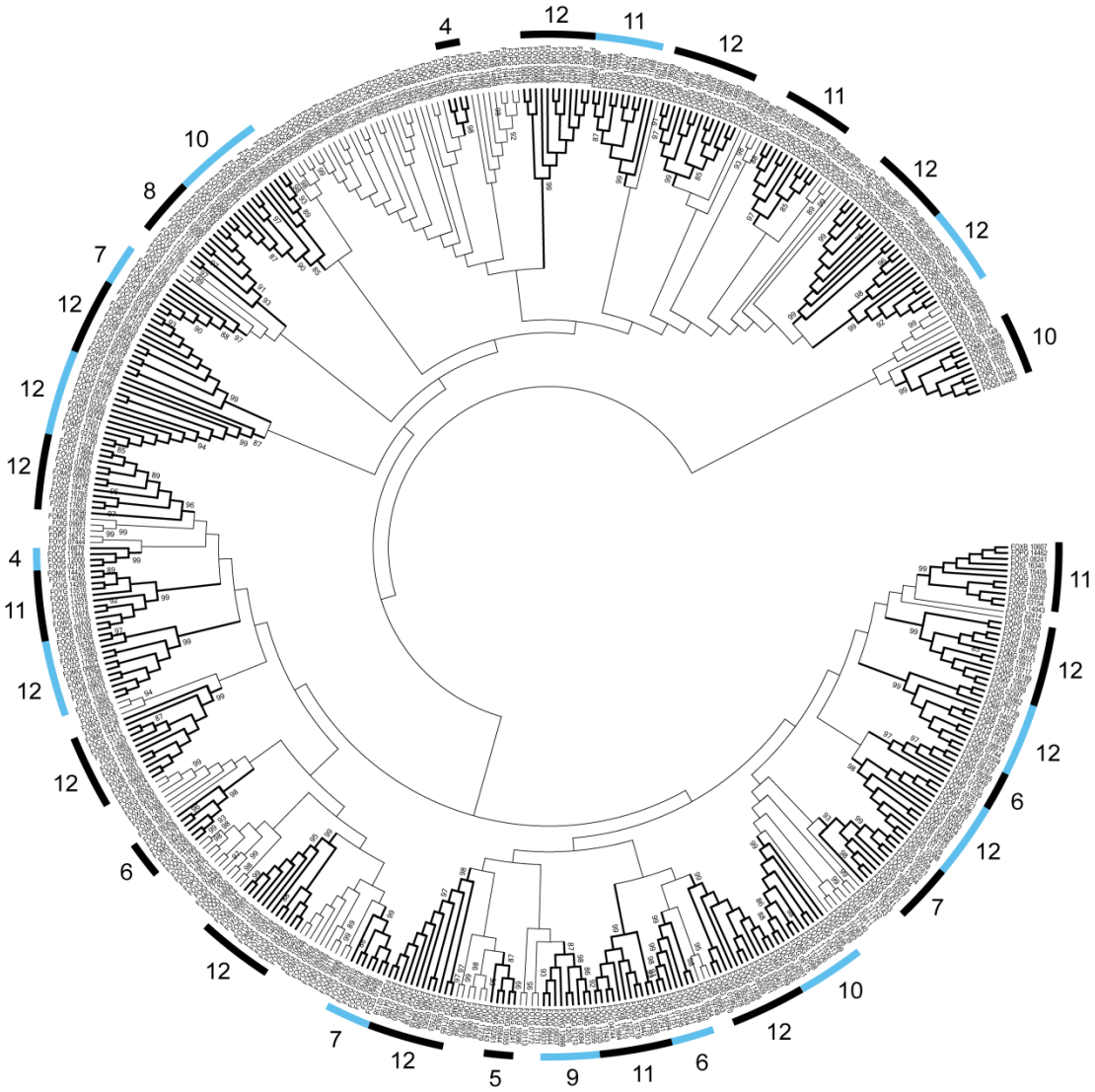

Figure 2.9: Shared unclassified kinases among the FOSC. Phylogenetic trees (Maximum Likelihood) using nucleotide sequences from all FOSC unclassified kinases. Bars around the outside of the tree denote clades supported with high bootstrap values. When 2 clades are adjacent one is colored blue to aid visually. The number next to each bar indicates the number of FOSC strains containing a gene from that clade.

C. Discussion

Kinases play key roles in transmitting external and internal signals and regulating complex cellular signaling responses. Genomes within the genus *Fusarium* have large kinomes compared to most fungi. As it is crucial for a pathogen to adapt to stresses encountered both outside and inside its host it is not surprising to see expansion of kinases among FOSC, as the species complex thrives in diverse hosts. The positive correlation between the total number of proteins and the size of the fungal kinome reported here was also reported in pathogenic *Microsporidia* species and several model organisms (Li et al., 2015). Distinctively, this study defined a core kinome of the ascomycete fungi consisting of about 100 kinases, in agreement with a prediction based on gene to kinase count. This core kinome outlines the most fundamental kinase signaling pathways supporting the largest phylum of fungi.

Within the ascomycete fungal lineages some kinase families appear more recalcitrant to change. The MAP Kinase cascades, Cell Kinase I, and Calcium/Calmodulin regulated kinases changed very little within the phylum over millions of years of evolutionary time. However, I observed the expansion of Cyclin-dependent Kinases and other close relatives, Histidine kinases, and Atypical kinases, suggesting their potential roles in species-specific adaptation along the various evolutionary trajectories of ascomycete fungi.

The number of Unclassified kinases increased significantly within the genus *Fusarium* and made up a majority of the kinases found within FOSC kinomes beyond

those conserved among the filamentous fungi. In contrast, most non-Fusaria had relatively few Unclassified kinases. Additional research is necessary to understand how these Unclassified kinases contribute to *Fusarium*-specific evolution and adaptation. However, it is clear that many Unclassified kinases are important, judging by the results from a reverse genetic screen (Wang et al., 2011), observed expression in our RNAseq data (Supp. Table 5), and their conservation within the genus and across the FOSC.

Both the histidine kinase groups and the SPRKL family kinases were largely absent from the yeasts, moderately expanded in the non-*Fusarium* filamentous fungi, and appeared in large numbers in the FOSC. Little is known about the function of SRPKL kinases in fungal biology. HisKs are extensively used by bacteria and archaea to sense and respond to a variety of biotic and abiotic stimuli (Mascher et al., 2006). In addition to regulating stress responses as reported in yeasts, the filamentous class III HisK were reported to regulate fungal morphogenesis and virulence in various human, plant, and insect pathogenic fungi (for detailed review see: (Defosse et al., 2015)). Overall, the function of histidine kinases is understudied in filamentous fungi. As HisKs are absent in mammals and some are essential for virulence in fungal pathogens; they represent interesting fungal targets for the discovery of new antifungal drugs as illustrated by Shor et al. (Shor and Chauhan, 2015). Although limited functional studies exist for HisK or SRPKL kinases, their continuous expansion suggests functional importance of these understudied protein kinases. A better understanding of their functions would not only inform *Fusarium* biology, but could be extrapolated to other filamentous fungi and complex basidiomycetes.

Although LS kinases differ from their core counterparts, catalytic domains are generally conserved. For BCK1 kinases, the catalytic kinase domain of all nine LS BCK1 kinases is highly conserved (~91% between groups), but all of them are roughly 800 AA shorter than their core genome counterparts, missing a portion of the N-terminus. Similarly, sequences of the additional copies of CLK differ significantly from that of the orthologous copy, however the invariant residues within the catalytic kinase domain were conserved among all. In the case of CDC2 kinases, two of the three phosphorylation sites, Y15 and T14, are mutated in all LS CDC2 kinases. Since phosphorylation of these residues inhibits CDC2 kinase function (Morgan, 1995), mutation at these two residues may lead to CDC2 kinase activity without tight control.

Most interestingly, I observed the repeated expansion of TOR kinases among the FOSC genomes. Most fungi have a single TOR gene, however two TOR paralogs were observed in both yeast genomes *S. cerevisiae* and *S. pombe* (Heitman et al., 1991). Duplication of TOR kinases was also reported in *Batrachochytrium dendrobatidis*, an amphibian pathogenic chytrid (Shertz et al., 2010). Seated at the center of many signal transduction pathways, TOR integrates the input from upstream pathways, sensing cellular nutrient, oxygen, and energy levels and dictates cellular responses. The convergent evolution toward TOR kinase duplication in the fungal kingdom might reflect selection for more finely tuned environmental response pathways. Interestingly I found that two FOSC strains containing an additional TOR copy were more resistant to rapamycin that those containing only one copy. Although our data for Fo4287 indicated

that the TOR paralog was unexpressed (Supp. Table 2.5), it may be expressed under other, unobserved, conditions.

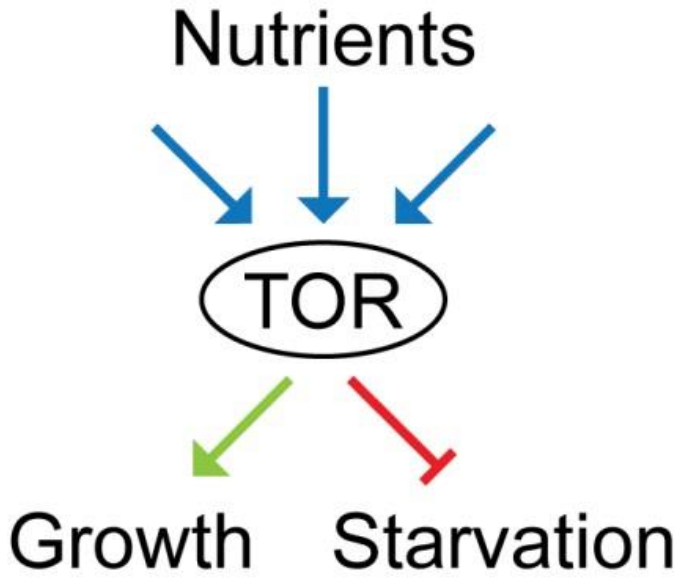

Figure 2.10 TOR signaling and growth. Nutrient availability indirectly regulates the activity of TOR. Under high nutrients the pathway suppresses starvation responses and promotes growth and development.

The expansion of families like the TOR kinases has been, in many cases, facilitated in part by the LS genome, which contributes to the unique and specific expansion of subfamilies, such as HisK, CLK, Atypical/FunK1, HAL, CMGC/SRPKL and Atypical/FRAP. Many of these kinases have either a known, or proposed, function in responding to environmental signals or cell cycle control and in many cases function downstream of TOR. Most expanded subfamilies can be linked to environmental signaling or cell cycle control; many under control of the TOR nutrient sensing complexes. Through the lenses of kinases, these pathogens seem to be enhancing their ability to sense their environment and tighten their regulation of the cell cycle, mediated primarily through the TOR signaling pathways (Fig 4.1).

D. Methods

1. Generation of fungal kinomes

The two yeast genomes *Saccharomyces cerevisiae* S288C (Sc), *Schizosaccharomyces pombe* strain 972h- (Spom) were downloaded from NCBI. All other fungal genomes were downloaded from The Broad Institute of MIT and Harvard and used for kinome analysis. Using the established Kinannote pipeline (Goldberg et al., 2013b) Dr. Goldberg generated the kinomes of the above fungal genomes. Briefly, Kinannote uses hidden Markov models generated from the manually aligned complete kinome of the slime mold *Dictyostelium discoideum* to search the given genome for kinases. It then identifies both well conserved eukaryotic protein kinases as well as unusual protein kinases. Finally Kinannote uses a BLAST search of amino acid sequences to classify the kinases based on family.

Kinannote assigned all kinases in all genomes into 135 different classes (Sup. Table 1). To create the conserved kinomes I removed any kinase subfamilies for which, under a given phylogenetic grouping, more than a single species was missing the family entirely. The groupings consisted of all species (Ascomycetes), all species except *S. cerevisiae* and *S. pombe* (Filamentous Fungi), only the genus *Fusarium* (Fusarium), and only members of the FOSC (FOSC). The number of conserved subfamily members was set as the lowest number among all species, excluding one.

2. Gene alignment and BLAST

Fungal kinase protein and nucleotide sequences were downloaded from either The Broad Institute or from the National Center for Biotechnology Information (NCBI). Sequences

were aligned using either Muscle or Clustal through MEGA6 (Tamura et al., 2013). Alignments were inspected manually and adjusted based on known conserved sequences. Sequence phylogeny was constructed using MEGA6 (Tamura et al., 2013) and maximum likelihood and bootstrapped using 100 replicates.

Identification of conserved LS regions was done by comparing each FOSC genome to the full LS gnome of Fo4287 including one core supercontig as a control for mapping (supercontig 14). BLAST results were filtered for regions matching greater than 90% nucleotide identify and alignment lengths greater than 3 kb. BLAST results were mapped to the Fo4287 LS reference sequence using BRIGS (Alikhan et al., 2011). Total overlap of each strains LS region with Fo4287 was calculated by summing all regions returned by BLAST.

3. Generation of RNA samples and data analysis

Fusarium oxysporum f. sp. *lycopersici* 4287 gene expression during temperature stress was assayed using RNA-sequencing. *F. oxysporum* spores ($1x10^9$ spores) were cultured for 14 hours at 28°C (in the case of 28° growth experiment) or for 10 hours at 28°C and switched to 37°C for 4 hours (in the case of 37° growth experiment) in 200 ml of minimal medium supplemented with 25 mM GluNa and buffered with HEPES (20 mM final concentration) to pH 7.4 at 170 rpm. Three replicates of each condition were done for each species, with 36 RNA samples in total. Fungal tissue was collected with filter paper and RNA was extracted using a standard TRIzol RNA Isolation Reagent extraction (Life Technologies, Carlsbad CA). A RNA library was constructed using Illumina TruSeq

Stranded mRNA Library Prep Kit (Illumina, CA) following the manufacturer's protocol and sequenced using the Illumina HiSeq platform (Illumina, CA).

Resulting data files were trimmed using Trimmomatic to remove poor quality reads (Bolger et al., 2014) resulting in a total 73.8 million read-pairs with an average 12.3 million read-pairs per dataset. Reads were then aligned into BAM files using Rsubread (Liao et al., 2013). Gene expression and differentially expressed gene calls were then made using limma and edgeR (Ritchie et al., 2015; Robinson et al., 2010a). Genes were called as expressed if both replicates had an RPKM value greater than 1.0. Genes were called as differentially expressed genes (DEGs) if the adjusted Pearson correlation value was less than 0.05.

4. Rapamycin resistance screen

Cultures of Fo4287, Fo5276, Fo47, Foii5, and Fo32931 were grown in Potato Dextrose Broth (Becton, Dickinson and Company, Sparks, MD), the strain MN25 was grown on Potato Dextrose Agar (Becton, Dickinson and Company, Sparks, MD), for 5 days and spotted into the center of either minimal media plates (2006) or minimal media plated with a final concentration of 50 ng/mL rapamycin. Plates were stored at 28 degrees for 48 hours and then transferred to room temperature. After transfer to room temperature the diameter of each colony was measured at 24 hour intervals. Growth rate was determined as the average increase in size, in millimeters, per 24 hours. Reduction in growth rate was calculated as (100 – ((growth rate on rap. plate)/(growth rate on control plate)*100)). In order to generate a range of error, all 9 comparisons of growth rate between the 3 control

plates and 3 rapamycin plates were used to calculate reduction in growth. All species/condition plates were done in triplicate.

5. Data Availability

The *F. oxysporum* temperature RNAseq data from this study have been deposited in the NCBI GEO repository under the accession number GSE113332.

CHAPTER III

COMPARATIVE TRANSCRIPTOMICS OF *PERONOSPORA BELBAHRII* INFECTED *OCIMUM BASILICUM* CULTIVARS REVEALS SALICYLIC ACID SIGNALING RELATED DEFECTS AND CAIDIDATE RESISTANCE GENES

A. Introduction

Basil (genus *Ocimum*) is a major herb crop of the *Lamiaceae* family, with a center of diversity in tropical and subtropical regions. Within the genus, basil plants are highly diverse in physical forms and essential oil content (Vieira and Simon, 2006). Sweet Basil (*Ocimum basilicum),* is the most commonly grown commercial basil and produces essential oils used in applications from medicine and health care products to soft drinks. In 2014, revenue generated in the US from sweet basil and other culinary herbs was estimated to be 76.8 million dollars (USDA, 2014).

Basil suffers from infection with several plant pathogens that pose control challenges and reduce yield. The most important over the last 10 years has been Basil Downy Mildew (BDM), *Peronospora belbahrii*, which is an obligate parasitic oomycete pathogen responsible for BDM disease. The pathogen enters basil through open stomata, or possibly by direct penetration of the upper cuticle, and colonizes leaf tissue intercellularly. Given proper conditions the pathogen can reproduce by producing sporangiophores bearing sporangia through stomata, creating a gray to dark-gray discoloration with yellowing typically between the veins of a leaf. Eventually the leaf

turns brown and falls from the plant. All these symptoms render the plant unfit for commercial sale.

First reported in Uganda in 1930, BDM began to attract attention in 2001 when disease instances were increasingly reported across the world with reports now from the Americas, Asia, and Europe (Choi et al., 2016; Garibaldi et al., 2005; Garibaldi et al., 2004; Kanetis et al., 2014; Khateri et al., 2007; Martinez de La Parte et al., 2010; McLeod et al., 2006; Nagy and Horvath, 2011; Roberts et al., 2009; Ronco et al., 2009; Šafránková and Holková, 2014). In 2014, BDM was reported in 36 US states (Wyenandt et al., 2015), threatening productivity in every growing region.

Although some resistance measures have shown promise, including fungicides (Homa et al., 2014; Raid, 2015), nocturnal fanning (Cohen and Ben-Naim, 2016), nocturnal illumination (Cohen et al., 2013), and daytime heating (Cohen and Rubin, 2015), these measures can be labor intensive and costly. The Simon group at Rutgers University had previously identified the *O. basilicum* cultivar Mrihani (MRI) (Horizon Seed Co., Williams, OR) displayed significant resistance to BDM compared to a typical commercial sweet basil (SB22)(Pyne et al., 2015). However, several undesirable traits prevent its direct use as commercial culinary product. Physically the plant leaf is smaller, has upturned margins, and is serrated compared to the large downward cupped smooth leaves of commercial sweet basil. Most importantly, the taste and smell of MRI (effectively a chemotype) differs considerably.

Data obtained by Pyne et al. (Pyne et al., 2017; Pyne et al., 2015) suggest a dominant effect of the resistance in MRI. Furthermore, QTL analysis of an F_2 mapping population from a cross between MRI and SB22 (Pyne et al., 2017) identified a single locus *(dm11.1)* that explains 20%-28% of the variance in resistance observed among the F_2 population. In addition, this study identified two minor loci *(dm9.1 and dm14.1)* that explained 5-16% and 4-18% of the F_2 population's phenotypic variation. The underlying causative sequences are unknown.

While traditional breeding may prove useful given that resistance genes do exist in a compatible cultivar, it is expensive, as well as time and labor intensive. To speed classical genetic breeding this study was designed to identify genes in MRI that may confer resistance, and therefore serve as molecular markers for selection. We have sequenced meta-transcriptome data generated by the Simon group of the resistant cultivar MRI and the susceptible commercial cultivar SB22 challenged with the pathogen *P. belbahrii* at 12, 24, 48, and 72 hours post inoculation (hpi) representing roughly germination (12 h), penetration (24 h), and intracellular growth (48 & 72 h) stages.

B. Results

1. Transcriptome Sequence

An average of 24.8 million (MRI – resistant cultivar) and 27.6 million (SB22 – commercial cultivar) 71 base-pair high quality read-pairs, with an average base quality score greater than 30 (Fig 3.1a), were generated per replicate, per infection time-point condition. Additionally 12.8 million (MRI), 14.3 million (SB22), and 9.9 million

(Sporangia) high quality paired-reads per replicate were obtained from the controls (Table 3.1).

Due to the lack of a sweet basil reference genome, I used 240.2 million paired-reads were used to de-novo generate the MRI combined assembly and contained 341,633 unique transcripts corresponding to 133,441 genes as called by Trinity (Grabherr et al., 2011). The SB22 combined assembly was generated using 263.9 million paired-reads and contained 118,296 genes and a total of 322,696 unique transcripts. The MRI and SB22 plant-only control assemblies contained 66,930 and 52,844 genes respectively; the sporangia control had less with 10,144 genes. As expected there are more genes assembled among infected samples, representing both host and pathogen mRNA. The larger transcriptome of the infected plants likely reflects genes contributed by the pathogen and genes expressed during host-pathogen interaction phases, as well as assembly errors introduced as transcriptome assembly complexity has increased.

Table 3.1: Summary of assembled transcripts by timepoint

Dataset (organism)	Read pairs (millions)	Mapping Percentage	Average base quality	Number of Genes	Average gene length	Average coverage
Sporangia (*P. belbahrii*)	29.7	84.1	37.28	10144	1603	230.4
MRI Water (*O. basilicum* MRI)	42.7	62.0	37.20	66930	893	66.4
MRI 12 hpi (*O. basilicum* MRI, *P. belbahrii*)	75.9	63.3	37.21	81873	897	98.1
MRI 24 hpi (*O. basilicum* MRI, *P. belbahrii*)	75.3	63.3	37.22	89302	864	92.6
MRI 48 hpi (*O. basilicum* MRI, *P. belbahrii*)	77.0	63.1	37.23	85832	860	98.7
MRI 72 hpi (*O. basilicum* MRI, *P. belbahrii*)	69.8	63.4	37.22	78061	909	93.5
SB22 Water (*O. basilicum* SB22)	38.3	62.9	37.18	52844	977	69.9
SB22 12 hpi (*O. basilicum* SB22, *P. belbahrii*)	85.9	63.9	37.18	74918	934	117.6
SB22 24 hpi (*O. basilicum* SB22, *P. belbahrii*)	83.2	63.3	37.19	72699	949	114.5
SB22 48 hpi (*O. basilicum* SB22, *P. belbahrii*)	75.9	65.2	37.15	84478	870	100.9
SB22 72 hpi (*O. basilicum* SB22, *P. belbahrii*)	86.6	66.6	37.20	80675	903	117.7
MRI Combined Assembly	370.4	64.8	37.22	133,441	765	352.6
SB22 Combined Assembly	399.6	66.0	37.19	118,296	692	477.7

Even though we sequenced fewer reads for both water and sporangia control samples the sporangia control has the highest sequence coverage (>90x) and longest average gene length (1,603 bp), likely due to the smaller genome/transcriptome size of the pathogen. The average assembled gene length of *O. basilicum* transcripts from a previously published transcriptome was 1,363 bp (Rastogi et al., 2014), smaller than our average assembly size, and the average gene size of the oomycete *Phytophthora infestans* was 1,523 bp (Haas et al., 2009), slightly smaller than this assembly.

To assess the genetic similarity between MRI and SB22, I did a BLAST search between MRI and SB22 water control assemblies and identified 20,943 best bi-directional hits, likely representing orthologs between these two host plants. These orthologs are highly similar with an average pair-wise sequence identity of 98.66% (Figure 3.1b).

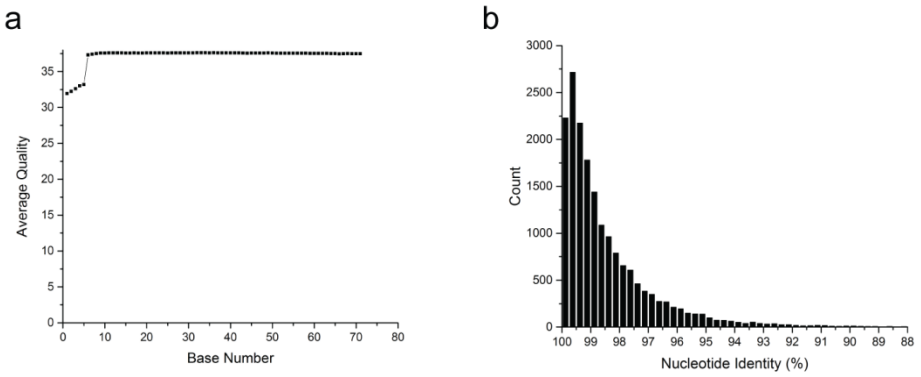

Figure 3.1: Average base quality and cultivar identity. (a) Average base quality across 75 bp reads as reported by FastQC is graphed for all reads among all libraries. (b) Bi-directional blast hit among 20,943 genes shared between the MRI and SB22 water controls. The nucleotide identity for each top BLAST hit is graphed here with a bin size of 0.25%.

Using RPKM in a given assembly, the MRI transcripts can be divided into 36,414 plant genes (present in the MRI water control), 9,988 pathogen genes (present in the sporangia control), and 29,502 infection unique genes (absent in both plant and pathogen controls)(Table 3.2). The SB22 transcripts include 31,702 plant genes, 9,426 pathogen genes, and 26,486 infection unique genes.

Table 3.2: Expressed genes across timepoints

Dataset	Number of Genes FPKM > 1			
	Total Genes	Plant Control	Sporangia Control	Infection Unique
MRI Water	36,528	36,414*	0	0
MRI 12 hours	39,349	29,247	472	9,534
MRI 24 hours	40,171	29,074	389	10,606
MRI 48 hours	40,973	28,852	179	11,840
MRI 72 hours	39,279	28,365	825	9,987
Sporangia	10,102	0	9,988*	0
Total unique	**76,018**	**36,414**	**9,988**[+]	**29,502**
SB22 Water	31,794	31,702*	0	0
SB22 12 hours	35,561	25,770	274	9,431
SB22 24 hours	36,732	25,584	699	10,364
SB22 48 hours	40,216	25,746	1,788	12,598
SB22 72 hours	40,838	25,414	5,375	9,962
Sporangia	9,518	0	9,426*	0
Total unique	**67,706**	**31,702**	**9,426**[+]	**26,486**

*- When the data was split into 3 sets 114 MRI points and 92 SB22 points were removed due to ambuiguity
+- The numbers displayed in in the chart above are total genes in each category as split by FPKM based separation after mapping to each master assembly. If the gene had a non-zero FPKM in the sporangia samples it was included in the "sporangia" sample. The difference in gene count is a reflection of mapping to two seperate assemblies.

Consistent with SB22 susceptibility, there was a significant increase in expressed pathogen genes in the susceptible host, with an almost 20-fold increase from 12 to 72 hpi compared to only 2-fold in MRI. Using an imperfect measure of total pathogen mRNA abundance (the number of reads mapping to roughly 4,553 pathogen genes), I detected

roughly 43-fold more mapped pathogen reads from the SB22 72hpi samples than the MRI 72hpi samples (Fig 3.2).

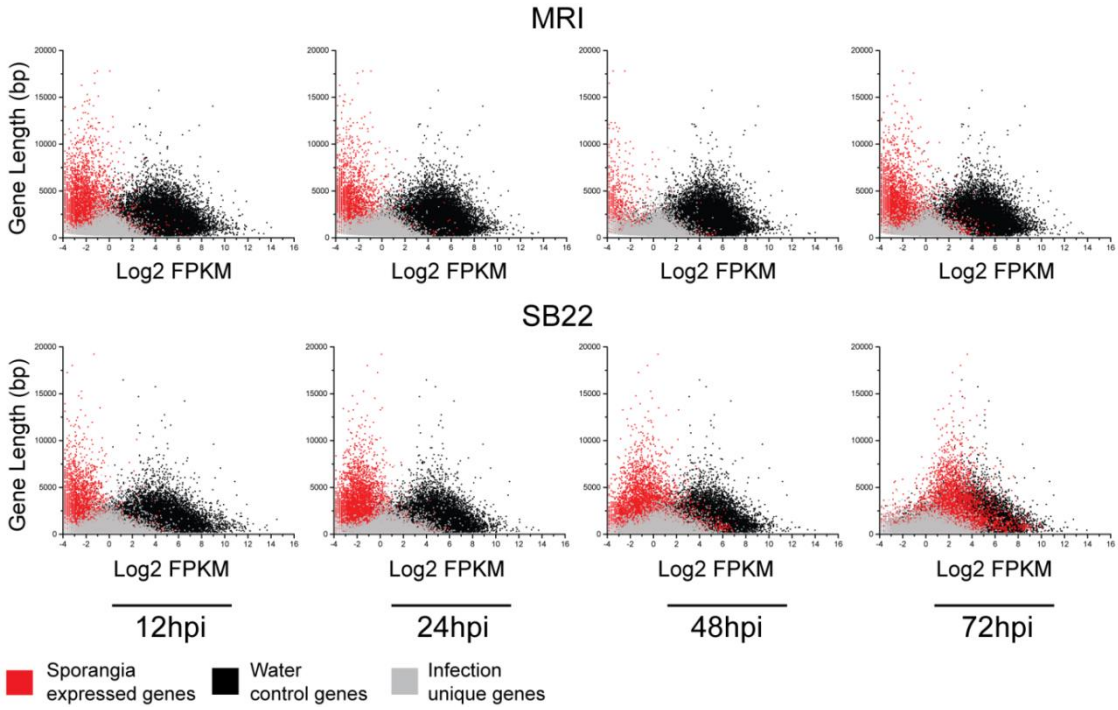

Figure 3.2: Pathogen mRNA builds within the susceptible cultivar. All expressed MRI (75,904) and SB22 (67,613) genes were broken as described into Sporangia expressed, Water control, and Infection unique genes. For each gene the log2 FPKM was graphed opposite the longest transcript sequences belonging to that gene. Expression and length were plotted per timepoint, with timepoints listed below each column. Red dots represent sporangia genes, black plant genes, and gray infection unique genes.

2. Gene Clustering

Based on overall ortholog identity, I was confident that SB22 reads from orthologous genes would be aligned to the MRI assembly. Using the assembled MRI transcripts as a reference, I mapped both the SB22 and MRI sequences using RSEMs default parameters (Li and Dewey, 2011). Overall, 63.2% and 58.7% of reads from MRI and SB22 were mapped to the MRI reference, respectively. This method resulted in the loss of 5.9% of

SB22 reads that mapped to the SB22 assembly but not to MRI, suggesting SB22-unique genes or orthologs of high sequence divergence. Therefore, this approach may introduce false-positive identification of MRI unique sequences and careful downstream examination is necessary.

Using MRI and SB22 data mapped to the MRI assembly I initially coarsely clustered all genes, resulting in 12 clusters with an average 2,810 genes per cluster (Figure 3.3). Cluster size ranged from 2 to 11,389 genes. Clusters D-1 (11,389 genes), D-2 (1214 genes), and D-3 (2 genes) represented transcripts that lacked consistency among biological replicates and were not included in the broad level analysis (Figure 3.3 D).

a. Oomycete and biological contaminate transcripts

Three clusters showed non-zero expression in the pathogen control (Figure 3.3 panel A). Both A-1 (6,368 genes) and A-2 (213 genes) represent the genes maximally expressed in the sporangia control with expression increasing in SB22. Cluster A-1 in MRI shows a strong decrease in the expression of pathogen genes from 12-72 hours, while cluster A-2 displays only a moderate downturn. GO enrichment of these clusters compared to all genes indicated that while cluster A-2 is enriched for many processes including metabolism, cluster A-1 is enriched specifically for membrane components.

Cluster A-3, containing 450 genes, appeared to show a profile consistent with pathogen genes, which are expressed at a low level in the sporangia control, but are highly expressed when the pathogen is in contact with the plant. Cluster A-3 is enriched for

many functions, among the most enriched being translation and nitrogen related metabolism and biosynthesis. In MRI, cluster A-3 gene expression steadily rose from 24 to 72 hours, inconsistent with our understanding of the resistant phenotype of the cultivar. Therefore the genes in this cluster were examined in more detail later.

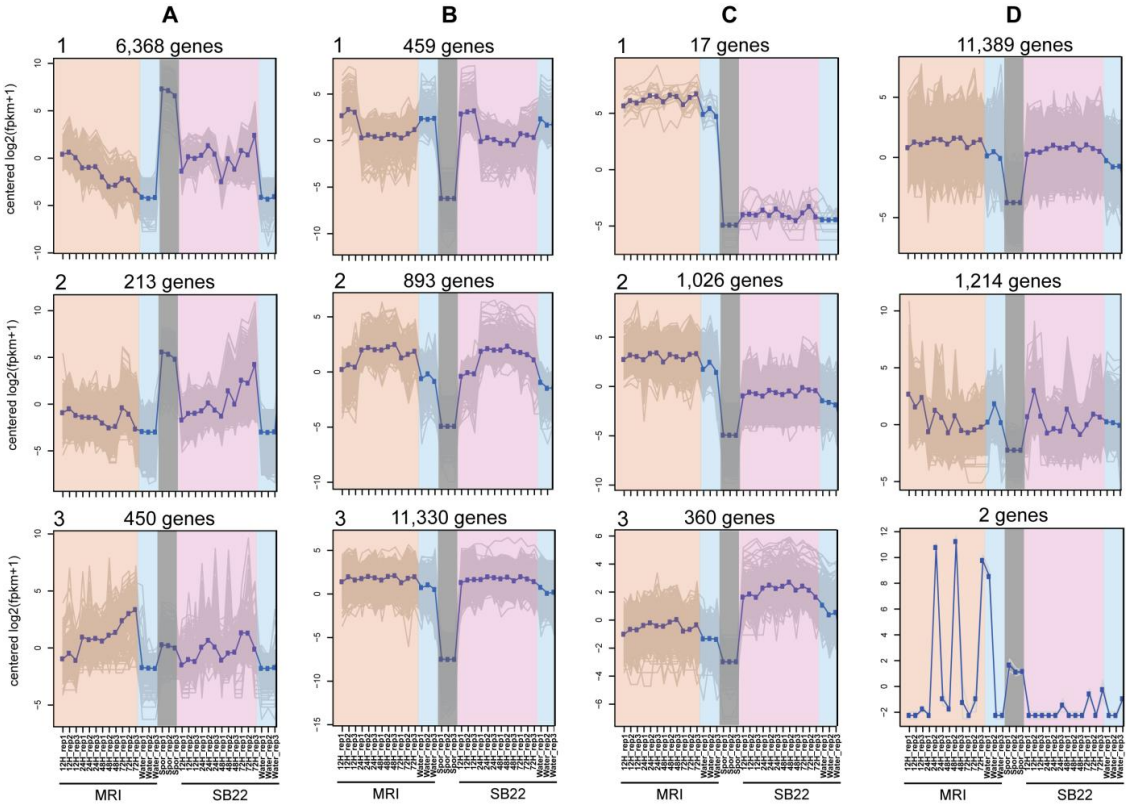

Figure 3.3. Patterns of gene expression. In total 12 clusters were generated using a p=40 value. The cluster number and the number of genes in that cluster are displayed above each cluster image. Expression value, on the y-axis, is on a log2 scale. Dataset and replicate are listed on the bottom X-axis. Data from MRI datasets are highlighted in orange, SB22 datasets colored pink, both water controls are labeled blue (MRI left, SB22 right), and the pathogen control is colored gray.

b. Plant transcripts

Six clusters showed plant gene expression profiles (Fig 3.3 column B, C). Genes in clusters B-1, B-2, and B-3 had comparable expression profiles between MRI and SB22. Both B-1 and B-2 clusters show a pattern consistent with a 12 hour shift in photoperiod

but which respond in opposite directions. Cluster B-1 was enriched for metabolism, oxidation-reduction, and photosynthesis functions. No GO functions were statistically significantly enriched in cluster B-2. However, of those categories returned by Blast2Go metal ion functions were predominant. The largest cluster, B-3, represents roughly stably expressed plant genes. Cluster B-3 is enriched for many categories including various metabolic processes, protein modification, and protein localization, among others.

Genes in clusters C-1, C-2, and C-3 display differing regulation between MRI and SB22 (Figu 3.3 C). In Cluster C-1, MRI expression is high but almost completely absent from the SB22 transcriptome. Within that cluster, 8 genes had no usable annotation, 2 were annotated as late blight resistance homologs (gi:848916018, gi:848932751), and 7 other proteins had various annotations. In cluster C-2, MRI expression was higher than SB22 and was enriched for genes related to defense response, response to stress, response to stimulus, and DNA integration. In cluster C-3, the expression of MRI genes is instead lower than SB22. No GO categories were enriched in cluster C-3.

3. Identifying MRI unique genes

Coarse clustering confirmed the feasibility of using clustering to identify genes of interest based on expression profiles. To narrow the possible search space, I repeated the clustering process with higher stringency (p=20, see methods for details) and generated 188 clusters. Cluster size varied from a single gene (multiple clusters) to 5800 genes (cluster 7), with an average 179 genes per cluster. Most clusters displayed a pattern

identical or related to the patterns described above.

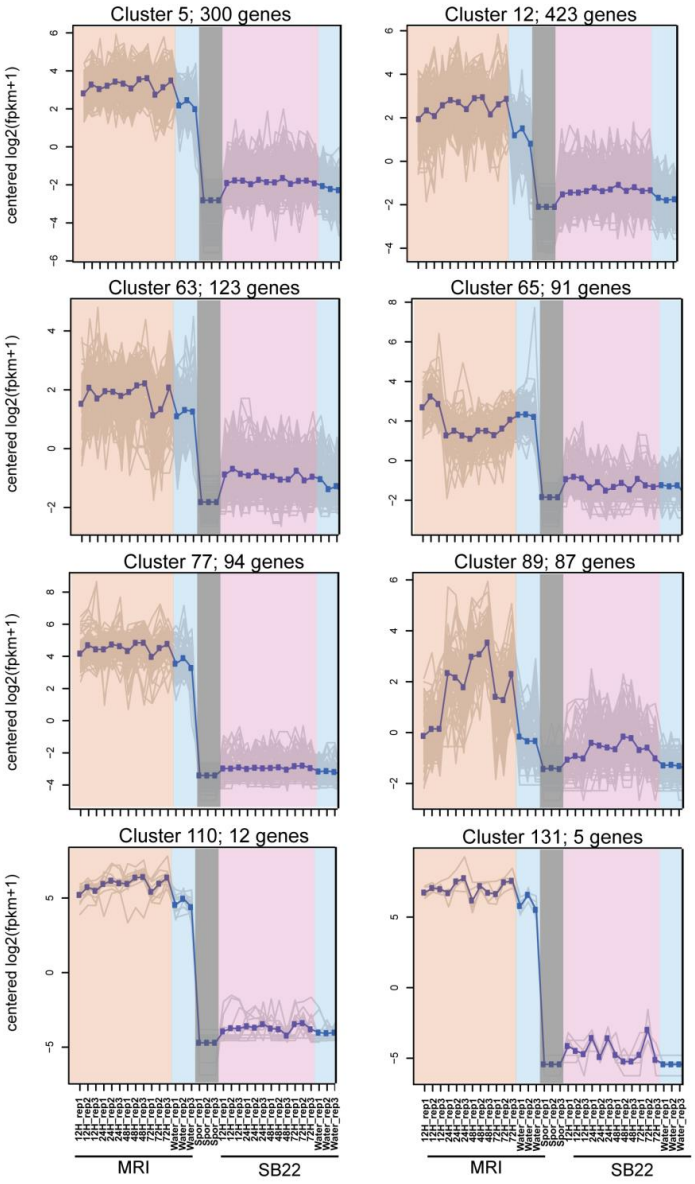

Figure 3.4: Candidate MRI unique gene clusters. Clusters were generated from MRI and SB22 data mapped to the MRI reference. Clusters were chosen which appeared expressed in MRI and to some extent not in SB22.

We were most interested in clusters showing expression patterns consistent with an MRI gene being uniquely expressed during infection, with upregulated genes being of particular interest. Eight clusters were chosen as they appeared expressed in MRI,

minimally expressed in SB22, and showed no expression in the sporangia control (Figure 3.4). I further refined the 1,135 genes in the 8 clusters by filtering out genes with mapped reads from SB22. I identified 562 candidate resistance genes which are uniquely expressed in MRI, including 159 genes in cluster 5, 225 in cluster 12, 29 in cluster 63, 47 in cluster 65, 66 cluster 77, 29 in cluster 89, 6 in cluster 110, and 1 in cluster 131. Due to the possibility of diverged orthologs in SB22 not mapping to the MRI assembly I removed transcripts with bi-directional BLAST hits between MRI and SB22, leaving 376 MRI unique candidate genes.

The B2G database offered annotation for 163 of 376 candidate genes. These genes could be grouped into five general categories; 1) homologs of the nucleotide-binding site leucine-rich repeat (NLR) disease resistance protein family (18 genes), 2) receptor-like kinases (RLK)(12 genes), 3) secondary metabolic enzymes (22 genes), 4) transposable elements (TEs) or TE-like sequences (11 genes), and 5) others, including hypothetical proteins, predicted proteins, transcription factors, and transporters, among others (100 genes)(Supp. Table 1). Uniquely expressed in MRI and related to pathogen detection and cellular signaling, NB-LRR and RLK genes are of particular interest and detailed analysis was conducted focusing on these candidate disease resistant genes.

a. Analysis of MRI unique NB-LRR family candidate resistance genes

The majority of plant resistance genes (R genes) belong to the NLRs, a large protein family (McHale et al., 2006). The NB (nucleotide binding) domain of the protein utilizes ATP for protein function while the N- and C-terminal domains aid in protein and

substrate interactions. Eighteen expressed MRI unique genes were annotated as NLR homologs. The most significant hits are resistance genes against oomycete pathogens, including three Arabidopsis homologs of Arabidopsis genes At5g43470 and AT3G46530, which encode RPP8 (Resistance to *Peronospora parasitica* 8) (McDowell et al., 1998; Meyers et al., 2003) and RPP13 (Bittner-Eddy et al., 2000), respectively. In addition, annotation identified eleven homologs of the potato R1 gene that confers resistance against *Phytophthora* pathogens (Botella et al., 1998). Both R1 and RPP8/13 proteins belong to the coiled coil type NLRs (CC-NLR).

To examine the RPP8/RPP13/R1 gene families identified as unique in MRI, both MRI and SB22 were screened for canonical NLR genes and sequences were compared between cultivars. I conducted a BLAST search using RPP8 and RPP13 along with the R1A gene against the 6 frame translated versions of both the MRI and SB22 transcriptomes.

The search of the MRI transcriptome returned 108 sequences in total with 16 hits unique to RPP8, 9 hits unique to RPP13, and 20 hits unique to R1A (63 overlap). SB22 returned 82 BLAST hits with 10 unique to RPP8, 3 unique to RPP13, and 17 unique to R1A (52 overlap). This tentatively places the number of CC-NLR disease resistance genes in each genome to 108 in MRI and 82 in SB22.

As described in relation to the assembly data, some truncation is expected among plant transcripts. Ignoring truncated sequences lacking two of the three domains (CC, NB-

ARC, LRR), I have identified 36 MRI and 28 SB22 RPP best confidence homologs and 11 MRI and 6 SB22 R1A top homologs.

Among the 47 top homologs of NLR sequences in MRI, BLAST between cultivar transcriptomes confirmed the pair-wise relationship between 39 sequences. Between MRI and SB22 sequence identity ranged between 93% and 100% with an average of 98.1%, in agreement with our estimation of overall nucleotide identity between cultivars. Three MRI RPP and five R1A homologs had no ortholog in SB22; two SB22 RPP and one R1A homologs had no ortholog in MRI.

All but 7 of 108 MRI and all but 2 of 82 SB22 NLR homologs were expressed in at least one sample (FPKM > 1.0). NLR expression on average increased over time peaking at 72 hpi (Figure 3.5a) in MRI. Average expression peaked earlier at 48 hpi in SB22. NLRs had an overall positive fold-change compared to the water control with 84 genes (77%) in MRI and 68 genes (82%) in SB22, greater than the genome at large (39% MRI, 46% SB22 with positive fold change). Of the 108 NLR homologues in MRI, 20 were called as DEGs (p.val < 0.05, FDR < 0.01) by edgeR. Thirty-two SB22 paralogs of MRI NLR homologs were called as DEGs. Consistent with raw expression data, DEG calls were predominantly upregulation (17 in MRI, 21 in SB22). This underscores the overall increase in expression during infection of NLR genes. In agreement with BLAST between transcriptomes, mri_179261c2, mri_168974c2, and mri_160460c0 had consistent detectable expression in MRI (average 30.3) and effectively no expression in SB22 (average 0.02). Among NLR sequences predicted to be unique to MRI, two R1A

homologs were excluded as candidates as they had significantly low expression across all MRI samples (on average 0.9 FPKM).

b. Top NLR Candidate Resistance Genes

We have identified three RPP and three R1A homologs uniquely expressed in MRI (Figure 3.5b). Among these, four appear to be upregulated during infection, however only mri_160460 was called as differentially upregulated. Although all 6 six candidates were required to contain at least 2 of 3 NLR domains, mri_160460c0 was the only predicted MRI unique expressed NLR candidate to contain all three domains.

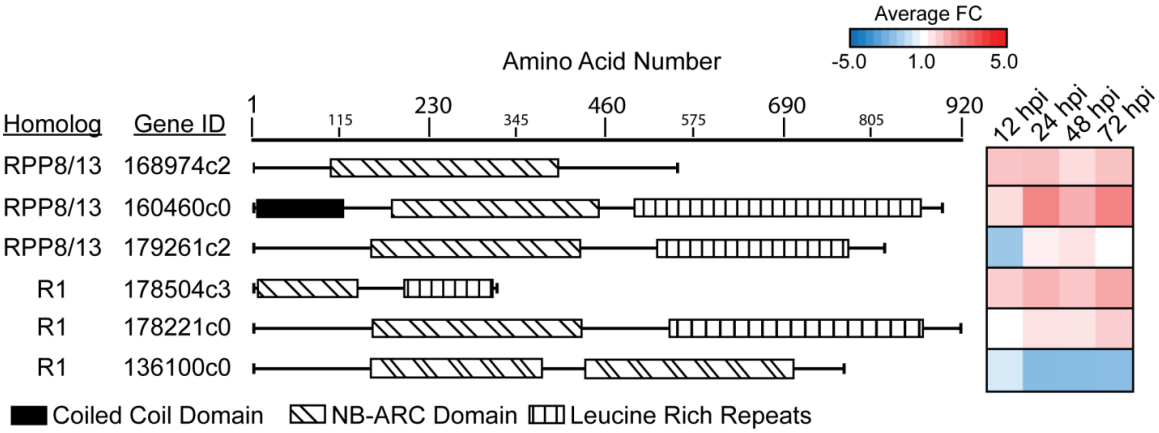

Figure 3.5: NLR resistance candidates. The domain structure of the top 6 candidate predicted resistance genes are shown. The closest homolog is listed on the left followed by the component ID in the MRI assembly. Box fill patterns indicating protein domains are labeled at the bottom.

We also identified a single PRR homolog that is expressed uniquely in SB22 (sb22_153191c3); it contains an NB-ARC domain without the RX-CC or LRR domains and is stably expressed across all timepoints.

c. Malectin-Like Receptor-Like Kinases

Twelve predicted candidate resistant genes belong to receptor-like kinases (RLK), another large gene family implicated in plant defense (Chern et al., 2016; Mendy et al., 2017). The RLK family has over 600 members in the Arabidopsis genome, and over one third are LRR RLKs (Shiu and Bleecker, 2001). Detailed inspection of the 12 MRI RLK hits revealed many truncated fragments. Three contained only the LRR domain, two contained only protein kinase domains, and two did not encode an identifiable domain. Additionally, two were annotated as containing "stress antifungal" domains, which contain a RLK-like domain and are present in proteins with known antifungal and salt tolerance related phenotypes (Sawano et al., 2007; Zhang et al., 2009), but little is known about their function. A search for protein domains using the NCBI conserved domain finder identified two sequences as transposable elements misannotated by our B2G pipeline. This left only a single protein, mri_comp170662, which encodes a full length malectin-like RLK, a member of which was shown to be involved in plant susceptibility to another downy mildew pathogen *Hyaloperonospora arabidopsidis* (Hok et al., 2011).

To further examine the unique nature of mri_170662, but also to avoid analyzing a large and diverse protein family, I searched the MRI and SB22 translated transcriptomes using mri_170662 as the query sequence. The MRI assembly returned more with 12 hits. The SB22 assembly returned 10 hits. Protein domain prediction indicated that all mri_170662 homologs except 1 contained at least 2 of the 3 domains. Three sequences lacked a large portion of the LRR domain, 3 proteins lacked the malectin domain, and 10 sequences lacked much or all of the kinase domain (Fig 3.6c).

Predicted MRI malectin-like RLKs were expressed across all timepoints. Expression in both cultivars ranged from roughly 1 up to over 500 FPKM. Four MRI and nine SB22 (roughly 60%) RLKs were differentially expressed up- and downregulated roughly equally in both cultivars and predominantly between 12 and 24hpi. Replicate variability for mri_comp170662 excluded it from a DEG call, but there appears to be a trend toward up upregulation at 12 and 24 hours compared to the water control (Figure 3.6b).

Phylogenic analysis using the malectin domains followed by a second targeted BLAST between transcriptomes identified 10 clear SB22/MRI orthologous pairs (Figure 3.6a), however three pair members lacked a malectin domain and were excluded from the tree. I confirmed that mri_comp170662 was unique to the MRI transcriptome but additionally identified, by nucleotide BLAST, mri_comp181331 as unique to MRI. It is 98% identical to SB22_comp162722 and shows SB22 expression in the combined mapped data. However, comparison between mri_181606 and sb22_162722 reveals them to be identical protein sequences with 99% nucleotide identity, identifying mri_181331 as a likely recent duplicate of mri_181606 and unique to MRI. These data support the unique nature of the receptor-like kinase candidate resistance gene mri_170622 and a paralog mri_181331.

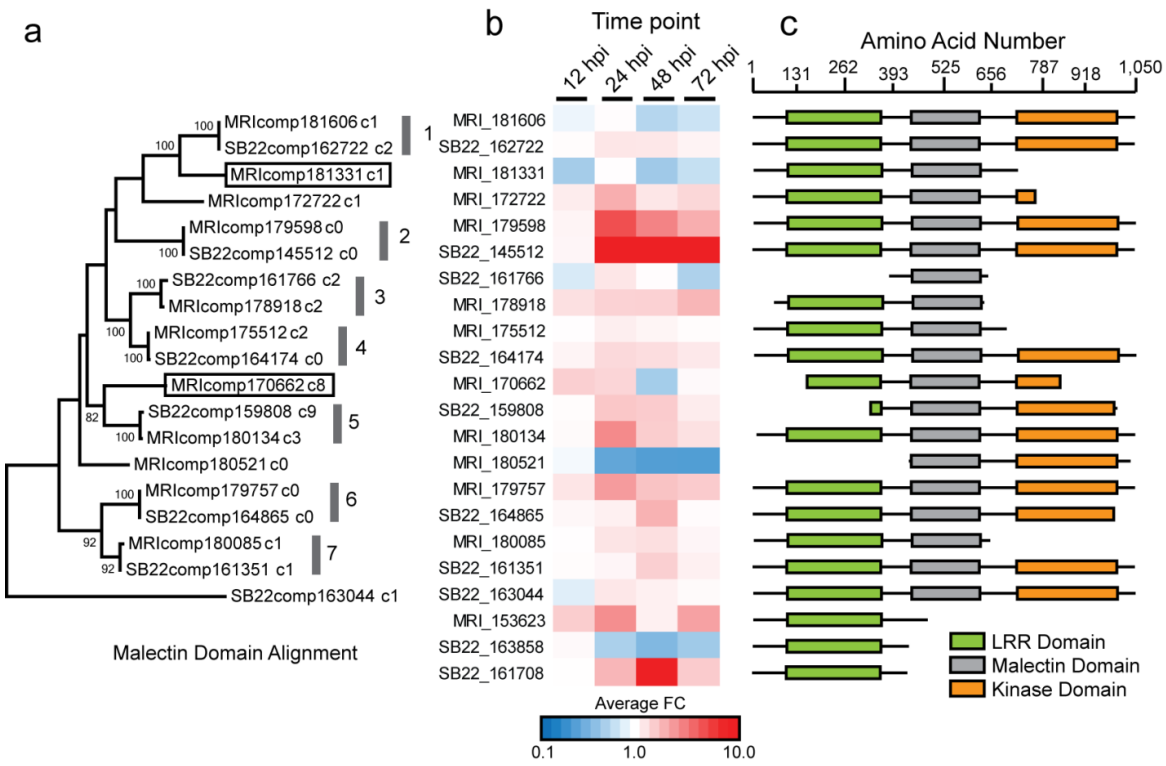

Figure 2.6: Malectin-like RLK proteins in MRI and SB22. (a) Alignment of 19 of 22 RLK proteins by their conserved malectin domain. Gray bars with numbers represent orthologous pairs, a box indicates a sequence absent from the SB22 cultivar. (b) Fold change compared to water of MRI and SB22 RLKs across infected plant samples. (c) Protein domain structure of 22 RLK hits generated from translated nucleotide sequences. Protein IDs are those of the adjacent fold change row.

d. Secondary Metabolite Genes

I examined the twenty-five secondary metabolite related genes predicted as MRI unique genes (Supp. Table 1). I found enzymes related to secondary metabolites that specifically differentiate the MRI and SB22 chemotypes including chavicol/eugenol O-methyltransferase and cinnamate p-coumarate carboxyl methyltransferase, as well as those involved in anthocyanin biosynthesis, among many other pathways. Overall the majority of identified secondary metabolite enzymes were not differentially expressed with only 7 of 25 called as DEG.

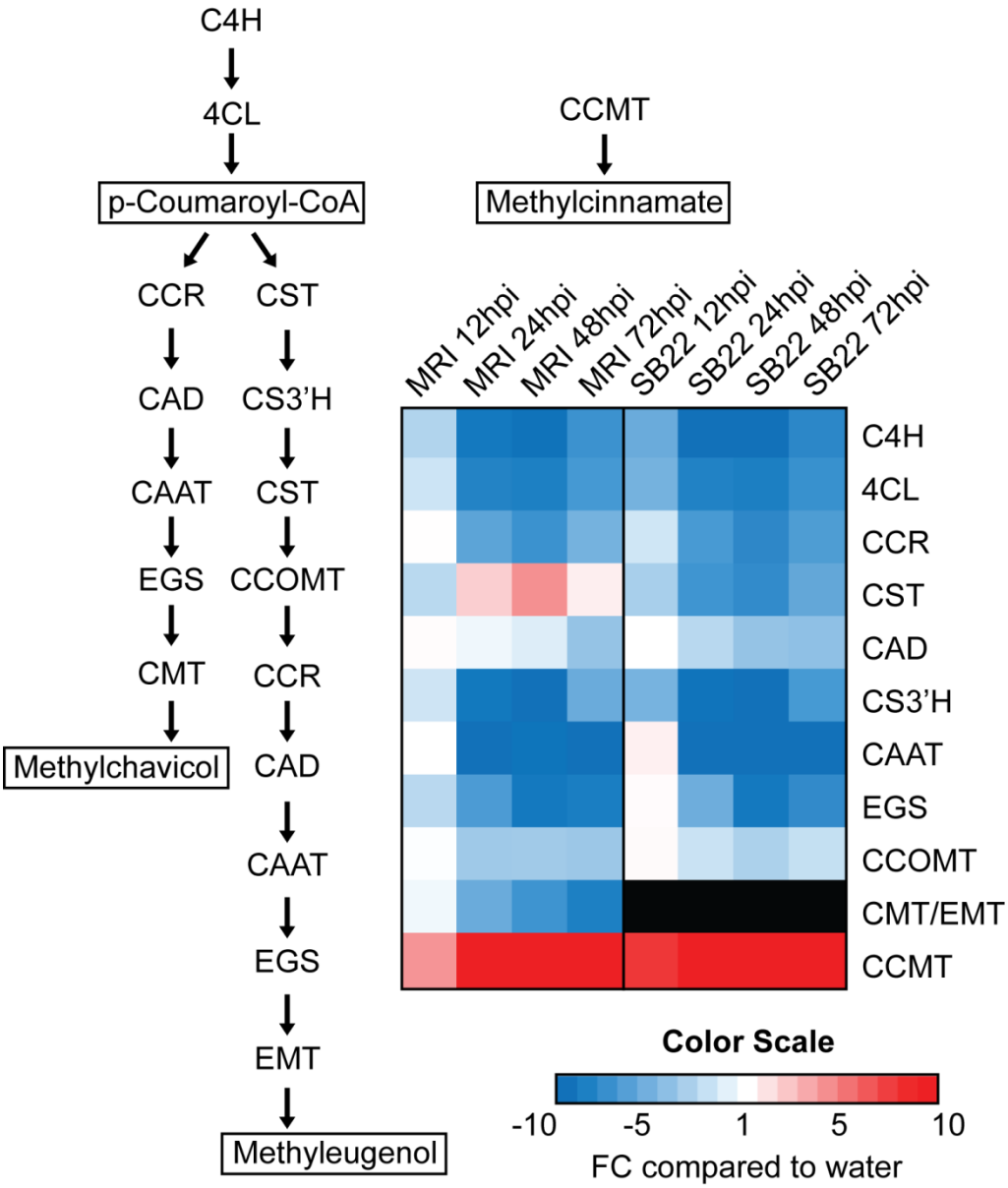

Figure 3.7: Methylchavicol/methyleugenol synthesis pathway. Genes involved in the synthesis of methylchavicol and methyleugenol and methylcinnamate were identified with BLAST. (a) Pathway leading to the synthesis of methylchavicol and methyleugenol. Enzymes are listed as 3 or 4 letter abbreviations. Synthesized molecules are listed in boxes. (b) Average fold change (FC) compared to water is plotted for both cultivars and all timepoints. Fold change; blue is downregulated, white no change, red upregulated compared to water. Black colored boxes indicate no gene. Enzyme abreviations: C4H (cinnamate4-hydroxylase), 4CL (4-coumarate:CoA ligase), CCR (cinnamoyl-CoA reductase), CST (p-coumaroyl shikimate transferase), CAD (cinnamyl alcohol dehydrogenase), EGS (eugenol/chavicol synthase), CCOMT (caffeoyl-CoA O-methyltransferase), CAAT (coniferyl alcohol acetyl transferase), CMT (chavicol O-methyltransferase), EMT (eugenol O-methyltransferase).

Chemotype is one of the major phenotypic traits that prevents MRI from immediate commercial use. SB22 accumulates a significant amount of eugenol, only sparsely converted into methyleugenol, while MRI predominantly accumulates methylchavicol (Rob Pyne, unpublished data). I were able to identify genes responsible for the synthesis of methylchavicol and methyleugenol and identified 11 enzymes from the pathways that lead to their biosynthesis (Kapteyn et al., 2007)(Fig 3.7). Compared to uninfected samples, infection resulted in the overall downregulation of the pathway. Only a single gene, responsible from shifting precursor from the methyleugenol to the methylchavicol pathway, was differentially regulated between cultivars.

4. SB22 Unique Genes

While a gene unique to, or uniquely expressed in, MRI is a likely candidate for the significant resistance phenotype observed, two other hypotheses exist. The gene could be present and expressed in both MRI and SB22, but contain a mutation that affects its function, thereby causing resistance or susceptibility, or it is possible that SB22 uniquely contains or expresses a gene that confers susceptibility. I address SB22 unique genes here.

As previously described for MRI, I conducted a clustering, expression, and annotation screen for sequences uniquely present in the SB22 transcriptome. SB22 clustering (p=30) resulted in 38 clusters. Four clusters containing a total of 1,287 sequences displayed a pattern consistent with SB22 unique genes including cluster 11 (30 genes), cluster 12 (354 genes), cluster 17 (677 genes), and cluster 19 (226 genes). After filtering individual

genes by counts mapping from MRI datasets, 230 sequences remained, which represent SB22 uniquely expressed sequences. Bi-directional BLAST data indicated that 70 genes contained a sequence homolog in MRI. Of the 160 remaining sequences B2G was able to provide an annotation for 62 sequences. Among these, 29 were called as either of transposable element origin, predicted unnamed protein, or hypothetical protein.

The remaining 33 sequences contained 5 annotated as late blight resistance genes, 3 receptor-like kinases, a calcium sensor, a homolog of the fungal and bacterial disease resistance regulating histidine kinase 5, a powdery mildew susceptibility gene MLO1-like, and others. As I conducted a more in-depth BLAST search for some MRI unique genes, I did the same for the 33 genes described here as having informative annotations. A BLASTn search was done against the SB22 transcriptome revealing only one gene to have an apparent paralog. This confirms the unique nature of 32 annotated SB22 sequences.

5. Hormone signaling pathways

In order to attempt to understand how susceptibility manifests transcriptionally in SB22 I targeted pathways known to play regulatory roles in plant defense for analysis (Fig 3.8). Both NLR and RLK genes interact with hormone signaling pathways to ensure host defense (McHale et al., 2006) including the RLKs CRK13 and FLS2 (Acharya et al., 2007; Gomez-Gomez and Boller, 2000) and the Arabidopsis NLR RPP8, which is induced by SA and aids in defense responses to downy mildew (Mohr et al., 2010).

To dissect the differences in the response to *P. belbahrii* by the two cultivars I examined the expression of genes involved in five major plant hormone defense signaling pathways. *A. thaliana* genes involved in ethylene, jasmonic acid (JA), abscisic acid (ABA), indole-3-acetic acid (Auxin), gibberellic acid (GA), and salicylic acid (SA) pathways were compared to 6 frame translated protein sequences for MRI combined assembly transcripts (1e-20, sequence similarity >40%). In most cases homologs recovered from MRI and SB22 were compared back to the NCBI database using BLASTp to confirm their identity.

No significant differences were observed between the MRI and SB22 pattern of expression in any Ethylene, Jasmonic acid, Abscisic acid, or Gibberellic acid pathway genes (Fig 3.8). There was a difference in expression profile for two of five auxin genes; YUC1 and TAA1 were both upregulated at early timepoints in MRI, but they were not statistically significantly differentially expressed.

a. The Salicylic Acid Pathway

In contrast to the pathways described above, some key genes involved in SA signaling pathways show significantly different expression patterns between MRI and SB22 (Fig 3.9). Here, I examined homologs of EDS1 and EDS5 (Enhanced Disease Susceptibility), PAD4 (Phytoalexin Deficient), ALD1 (agd2-like defense response protein), SID2 (Salicylic Acid Induction Deficient), NPR1 (Non-expressor of PR1), WRKY70 transcription factor, PR1 (Pathogenesis Related), LURP1 (LATE UPREGULATED IN RESPONSE TO HYALOPERONOSPORA PARASITICA 1),

PAL (Phenylalanine Ammonia-Lyase), BAK1 (Brassinosteroid insensitive 1-Associated Kinase), NDR1 (non-race-specific disease resistance 1), SAMT (Salicylate carboxymethyltransferase), SABP2 (Salicylic acid-binding protein 2), and SARD1 (SAR Deficient 1)(Figure 8).

I broke genes in to three categories, those related to the biosynthesis of SA, those related to predominantly upstream SA signaling, and those downstream of SA.

SID2 synthesizes isochorismate from chorismate and is one step from SA synthesis (Wildermuth et al., 2001). PAL synthesizes cinnamate from L-phenylalanine (Olsen et al., 2008) and is the first step in phenylpropanoid pathway which can lead to SA synthesis. The transcription factor SARD1 likely binds to the promoter of SID2, activating its expression (Zhang et al., 2010). EDS5 is a transporter required to export SA from the chloroplast (Serrano et al., 2013). SABP2 and SAMT act together to promote the SAR signal as SAMT converts SA to methyl-SA which is translocated to distal tissues where SABP2 acts to convert methyl-SA back into SA (Park et al., 2007; Tieman et al., 2010).

Significantly, but consistent between cultivars, the expression of PAL drops roughly 20 to 100-fold in both cultivars compared to the water control at 48 hpi. The expression of PAL is significant, being the 12^{th} most highly expressed gene in the MRI water control. SID2 and its transcription factor SARD1 were upregulated in MRI. Between water control and 24hpi, roughly consistent with the timing of the loss of PAL expression, the

relative normalized expression of SID2 in MRI rose 23-fold. In SB22, SID2 expression rose only 2.5-fold by 24 hpi and remained effectively stable onward. Although upregulated in MRI, SARD1 expression decreased 25% in SB22. Both EDS5 and SAMT were differentially expressed and downregulated in both cultivars during infection. SABP2 was downregulated at late timepoints in MRI but expression remained stable in SB22. All six SA biosynthesis related genes were called as DEGs.

Upstream of SA biosynthesis PAD4, EDS1 (Feys et al., 2001), and ALD1 (Cecchini et al., 2015) are required for the initiation of SA dependent responses. Overexpression of BAK1 leads to SA accumulation and the activation of some cell death related genes (Kim et al., 2017). NDR1 mutants lack SA accumulation in response to ROS and had impaired systemic acquired resistance responses after infection (Shapiro and Zhang, 2001). EDS2, BAK1, and ALD1 showed similar profiles between plants with overall stable or downregulated expression during infection. NDR1 showed the opposite profile, being overall upregulated. PAD4 was downregulated in MRI and to a lesser extent in SB22. Only ALD1 and BAK1 were called as DEGs.

Downstream of SA biosynthesis NPR1 and WRKY70 activate the expression of genes involved in the defense response, including the defense response marker PR1 and LURP1 (Kinkema et al., 2000; Knoth and Eulgem, 2008; Li et al., 2004). Although the functions of both PR1 and LURP1 are unknown, Arabidopsis mutants lacking LURP1 fail to fully induce defense responses mediated by RPP4 and RPP5 and LUPR1 expression is known to respond to infection by *H. parasitica* (Knoth and Eulgem, 2008). PR1, WRKY70, and

LURP1 were called as DEG by edgeR and also showed differing patterns of expression between MRI and SB22. WRKY70 was 3-fold upregulated in MRI at 12hpi, peaking again at 72hpi. In SB22 WRKY70 expression peaked only at 72hpi. Likewise, LURP1 expression was upregulated in MRI at 12, 24, and 72hpi while in SB22 expression peaked only at 72hpi. A classic defense response marker, PR1 was 2-fold upregulated in MRI at 24hpi. The expression of NPR1 remained effectively unchanged in both cultivars.

b. Genes downstream of SA signaling

Resistance to biotrophic pathogens generally results in programed cell death and ROS production. RBOH proteins (Respiratory Burst Oxidase Homolog) are responsible for controlling cell death, with members varying in their effects (Pogány et al., 2009; Torres et al., 2002). Glutathione S-transferases are known stress response genes acting in both biotic and abiotic stresses (Nianiou-Obeidat et al., 2017). Knockouts of accelerated cell death 11 (ACD11) auto-activate plant defense responses downstream of SID2 (Brodersen et al., 2005).

Both GST2 and ROBHD-1 showed a similar profile; upregulated in SB22 at 24 and 48hpi, even if only marginally, but down or unchanged in MRI. ACD11 expression was similar except for at 12hpi where the gene was downregulated in MRI and up in SB22. The profiles of GST1, ROBHD-2, ROBHE, and NUDT7 were similar between cultivars.

Together these data pinpoint a dysregulation of genes directly involved in the biosynthesis of Salicylic acid and some of the genes which respond to SA immediately downstream. However the downstream genes mechanistically closer to defense responses are less clear with few perturbations identified in defense-causative genes.

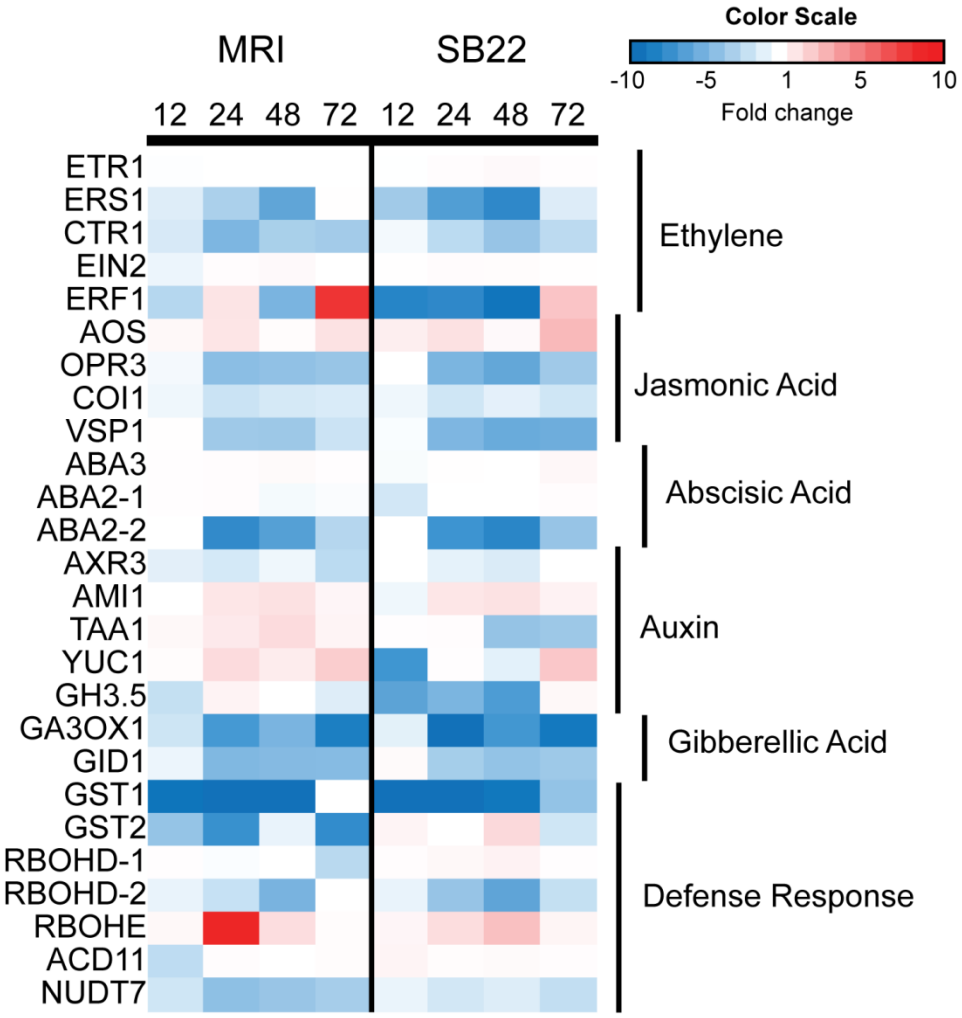

Figure 3.8: Plant hormone pathway gene expression. Genes identified by BLAST to known Arabidopsis hormone pathways. Genes are listed and grouped by their umbrella pathway. Fold change compared to water control is listed by timepoint. Fold change; blue is downregulated, white no change, red upregulated compared to water.

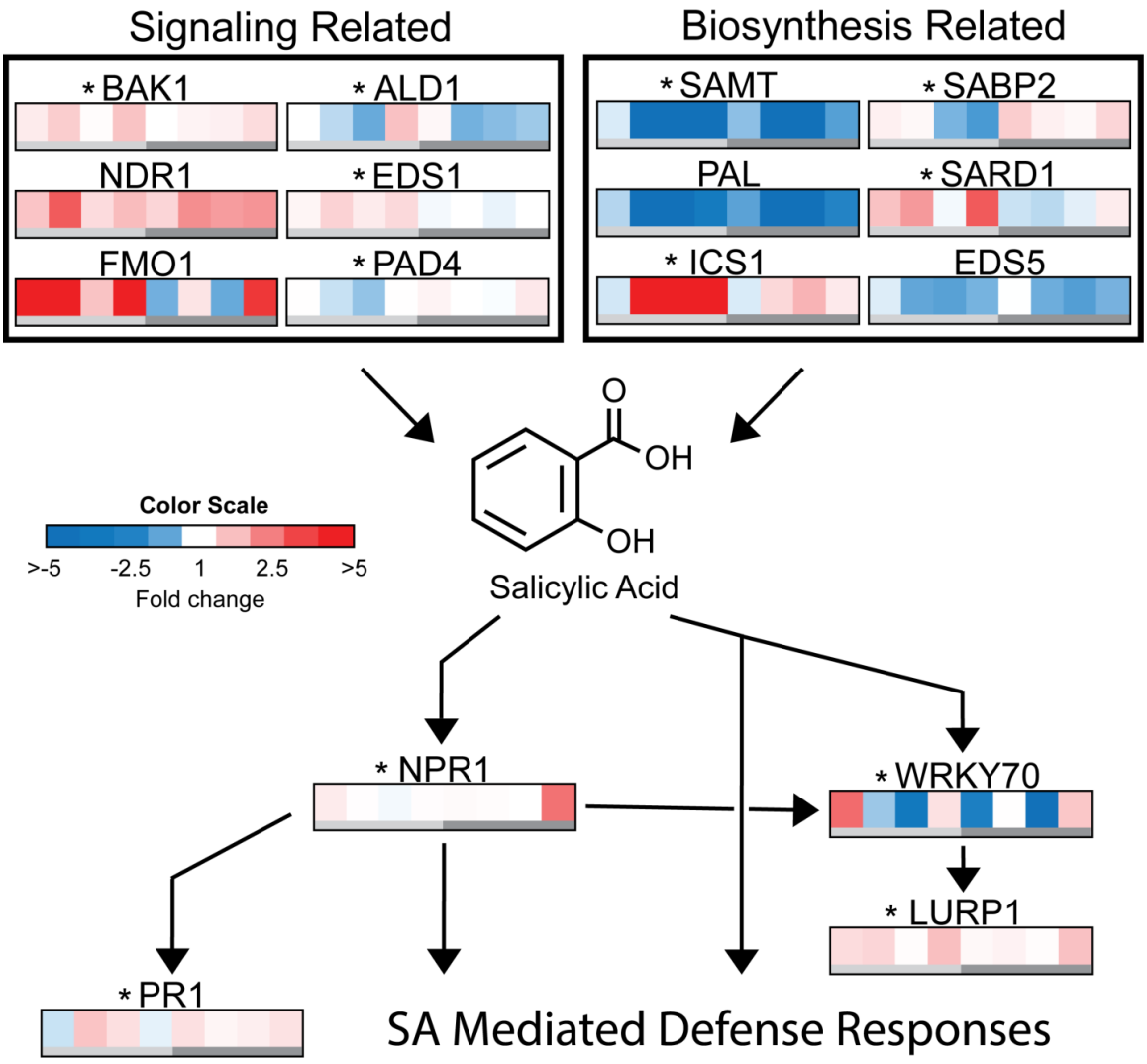

Figure 3.9: The salicylic acid signaling network. Genes known to be involved salicylic acid signaling or its biosynthesis. Fold-change as TMM FPKM average of 3 replicates compared to water control is plotted. Below each fold-change color bar; light gray represents MRI timepoints (left 4 boxes), dark gray represents SB22 timepoints (right 4 boxes). Timepoint blocks are ordered left to right 12 hpi, 24 hpi, 48 hpi, 72 hpi. Fold-change; blue is downregulated, white no change, red upregulated compared to water.

6. Pathogen Data Analysis and Leaf Microbiome

After filtering genes for expression in the water control and translated AA size I was left with 322,966 sequences that could represent secreted proteins. I put these proteins though a very stringent filtering process (see methods) that removed all but 2,770 sequences,

which were ideal candidates for secreted proteins as they 1) contained a signal peptide, 2) did not appear targeted to mitochondria, and 3) did not contain a transmembrane domain. Of the 2,770 candidate secreted proteins, 266 were called as cysteine rich and 119 contained a single RXLR motif in the correct position (Morgan and Kamoun, 2007)(Figure 3.10).

Sequences with an RXLR motif closer than 20 AA or farther than 100 amino acids from the N-terminus were removed. Additionally sequences which contained more than one RXLR domain were removed (13 sequences). Of the 101 remaining sequences only 11 were chosen as primary candidate cytoplasmic effectors (CCE) due to the presence of an RXLR motif just upstream of an EER motif. No sequence contained an EDR motif following an RXLR domain. The distance between the RXLR motif and EER motif ranged from 6 to 14 AA, and the distance of the RXLR motif to the N-terminus ranged from 34 to 51 AA with an average of 44 AA.

Expression data from the combined mapping indicates candidate secreted effectors increase within SB22, but predominantly not in MRI (Figure 3.11). In agreement with the resistant phenotype of the MRI cultivar, the expression of the 11 CCEs increased in SB22 from 12 to 72 hours, peaking at 72 hours and at a higher expression value than was seen in the sporangia control. This indicates that these genes are likely induced during infection specifically. Alternatively, this could be a reflection of the abundance of the pathogen within sequenced tissues (as is shown above for mapped pathogen reads in SB22). The average expression of a CCE in SB22 increased from roughly 1 FPKM at 12

hours to 80 at 72 hours; however large variations in replicate expression exist. In MRI the expression remained close to zero, with the exception of two genes whose expression jumped significantly at 72 hours in two of three replicates (an issue discussed before). Expression of these 11 genes was always absent or extremely low in the plant control and moderate in the sporangia control.

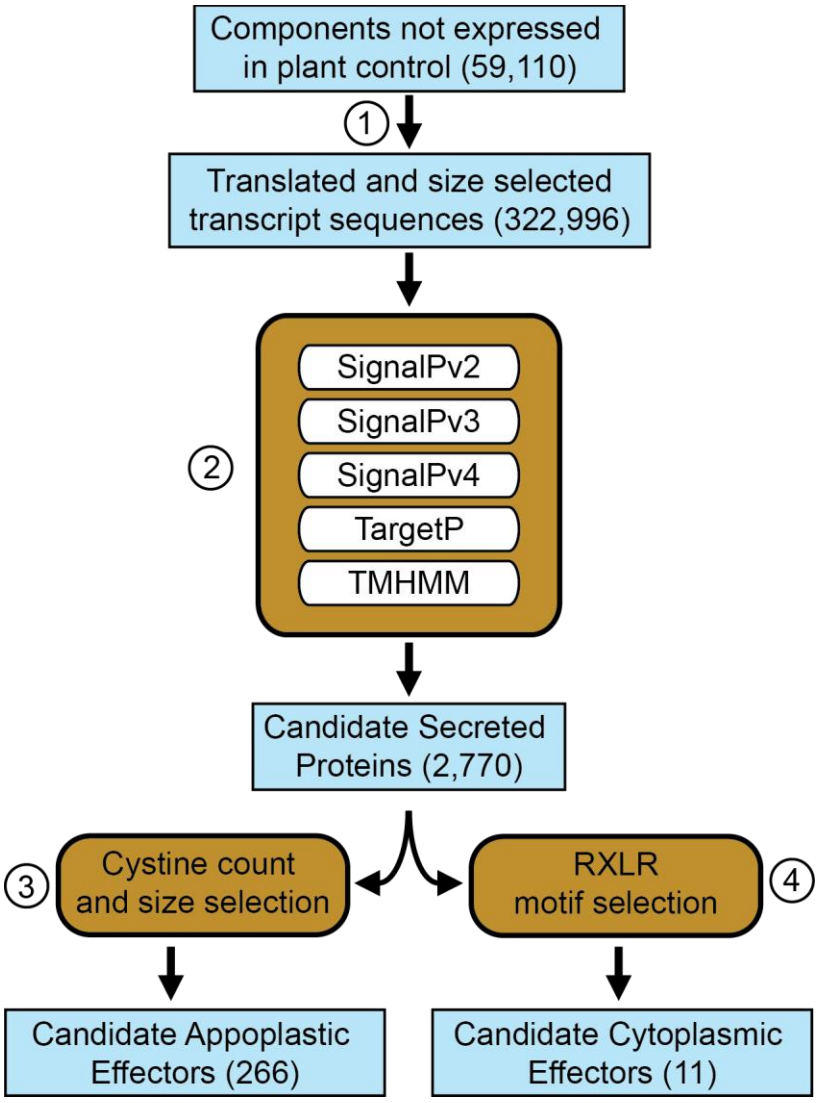

Figure 3.10: Candidate effector prediction pipeline. Genes not expressed in the plant control were (1) filtered for translated AA length, (2) passing filters for signal peptide, transmembrane domain, and mitochondrial signal peptide prediction, (3) cysteine count, and (4) present effector motifs. Numbers indicate the number of gene sequences at each step.

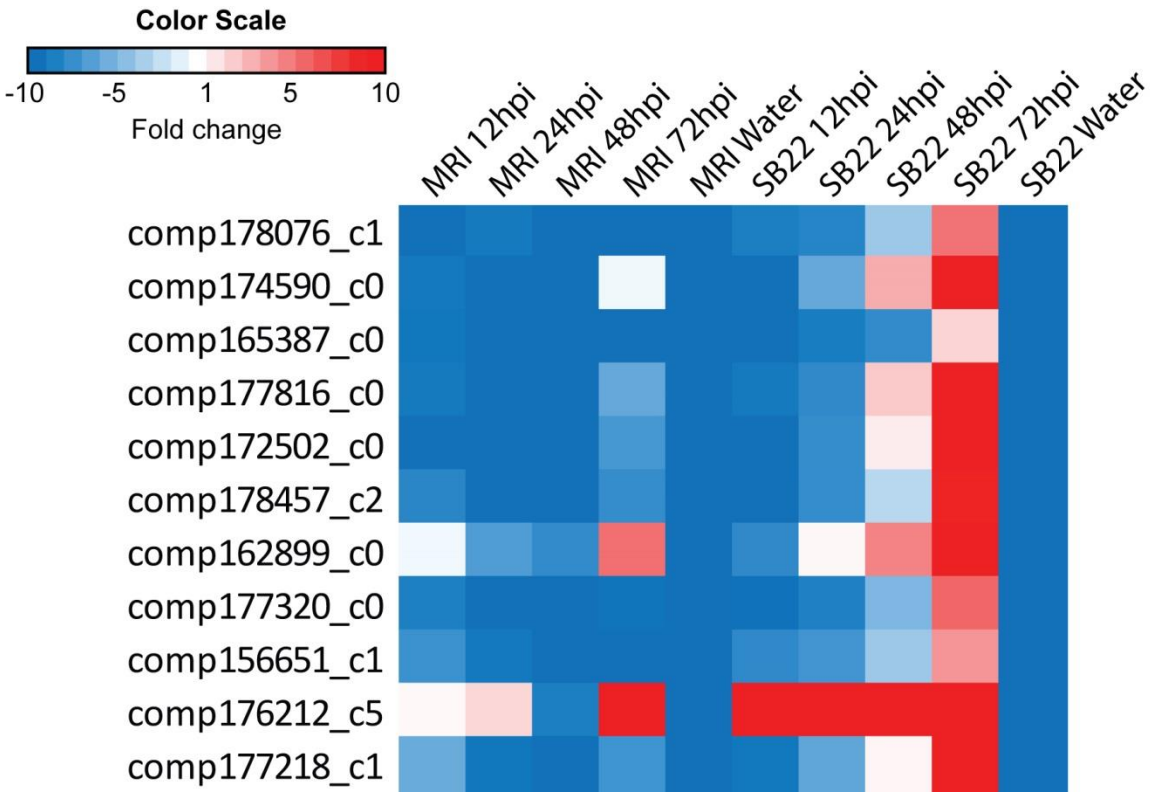

Figure 3.11, Eleven candidate RXLR effectors become abundant in SB22. Color indicates the relative fold change compared to expression in the sporangia control data. Blue indicating less expressed than the sporangia control. Gene ID is listed on the left of the graph and timepoints on the top.

The remaining RXLR domain genes were screened using protein BLAST for homology to known proteins. Three genes had a top 10 BLAST hit which was annotated as an RXLR-like or -putative protein, comp178155_c1, comp172394_c0, and comp174590_c0, all from cluster 1.

Fabro et al. 2011 used computational prediction and in vivo testing to show that a number of candidate effectors in the genome of *Hyaloperonospora arabidopsidis* were capable of increasing disease symptoms when plants engineered to express these candidates were

challenged with the *H. arabidopsidis* strain Noco2. A BLAST search for homologues of 7 *H. arabidopsidis* effectors, which were shown to improve disease symptoms and reduce callose deposition and ROS accumulation after infection, resulted in no significant BLAST hits.

Of the 2,770 candidate secreted genes, no sequence contained the conserved LXLFLAK or HVLVXXP domains typical of some effector proteins. Additionally a BLAST search against the combined MRI database using the *P. infestans* CRN1 and CRN2 nucleotide sequences returned no significant hits. As *P. belbahrii* is obligately biotrophic, this supports the hypothesis by Stam et al. 2013 that CRN effectors may be associated with the hemi-biotrophic and necrotrophic lifestyles.

7. Taxa Contamination

Examining genes in cluster A-3 uncovered sequence contamination from organisms likely living on the plants harvested for this study (Figure 3.12). In total 189 of 987 sequences from this cluster return BLAST hits to plant sequences (predominantly Sesame and Monkey Flower) and are likely sweet basil genes. Genes within this cluster that BLAST to plant sequences have stable expression across all infected samples with zero expression in the sporangia control and low to no expression in the plant control. Most sequences had a top hit that was a small nuclear ribonucleoprotein, ADP-ATP carrier protein, tRNA ligase, or uncharacterized protein. Sequences matching plant genes constituted 20.1% of hits, none matching with 100% nucleotide identity, as would be expected for RNA extracted from basil.

Among 987 sequences that returned nucleotide BLAST hits with an e value less than 1e-5, 60% of sequences returned a top hit to a fungal species. Though over 19 genera were returned, the most dominant species were *Tilletiaria* (308 sequences) and *Pseudozyma* (91 sequences), both basidiomycete yeasts. Two sequences matched with 100% nucleotide identity to *Pseudozyma* sequences (Alignment 280 nuc. or greater).

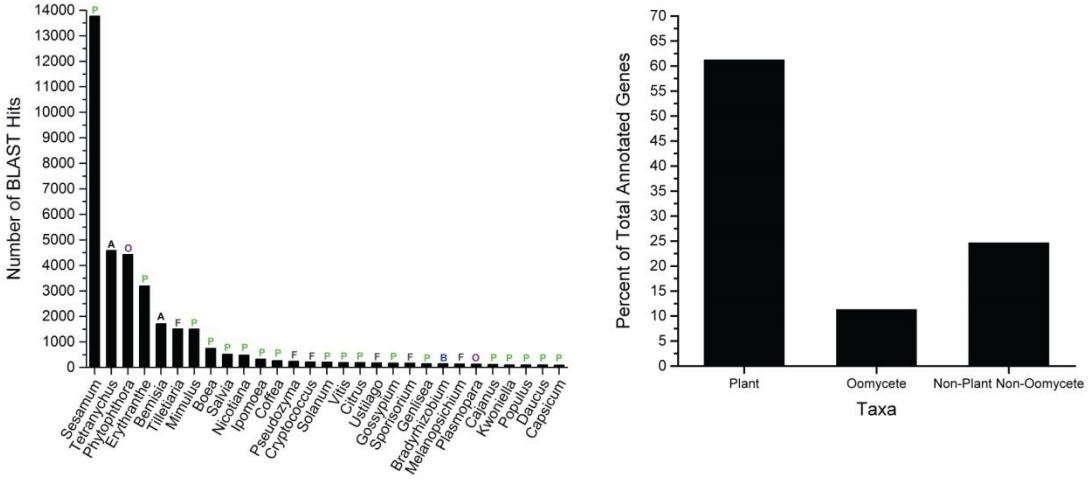

Figure 3.12. Biological contamination as uncovered by BLAST. Sequences were annotated with BLAST2Go. The genus of the top hit per sequence was used to assign taxonomic classification. (a) Number of genes as reported by trinity among annotatable sequences which belong to each genus listed. Letters above indicate broad classification. P, plant. I, Insect. A, arthropod. F, fungus. O, oomycete. B, bacteria. (B) Of gene sequences reported by trinity the percent of all sequences which belong to each of the listed Phyla.

81 sequences matched Phytophthora sequences, all with an identity between 68% and 86%. I predict these sequences belong to *P. belbahrii,* as no other oomycete sequences were recovered and the divergence implies sequences belonging to another species. Among 987, only 39 sequences matched to a bacterial sequence, likely due to the different function for polyadenylation in bacteria. Three genera matched with 100%

identity, *Clostridium*, *Propionibacterium*, and *Staphylococcus*. One replicate of MRI 72 hpi, one replicate of SB22 12 hpi and 48 hpi, and two replicates of SB22 24 hpi and 72 hpi returned non-zero FPKM values for silver leaf whitefly, a common greenhouse pest.

C. Discussion

Here I have used sequenced mRNA of basil cultivars infected with *Peronospora belbahrii* with the goal of identifying genes conferring resistance. We generated 33 sequence datasets. Although Trinity's estimate of gene count in the plant samples is greater than might be expected based on the average plant genome size I identified roughly 21,000 likely orthologous genes between cultivars with overall high sequence conservation.

Gene expression clustering identified patterns consistent with MRI unique infection-expressed genes. Annotated MRI unique candidates were found to be enriched for NLR, RLK, and secondary metabolic proteins. Analysis of the NLR and malectin-like RLK genes in both MRI and SB22 identified orthologous pairs and cultivar unique genes. Members of the NLR and RLK families are known to act alongside or upstream of disease signaling pathways and serve as good candidate resistance genes. Secondary metabolite genes found here in some cases correlate well with known chemotypic characteristics of MRI and SB22.

In order to understand the functional consequence of SB22 susceptibility I examined the hormone signaling pathways known to govern defense responses for significant

differences between cultivars. Of the three pathways examined only the SA pathway contained significant differences, particularly related to the biosynthesis of SA and markers generally associated with SA defense responses.

This study has taken mRNA sequenced over an infection time course and provided reasonable evidence linking susceptibility to known mechanisms that control defense responses and attempted to predict the genes behind the resistant phenotype. Transcriptomics without the need for a reference genome is a powerful tool for comparative analyses given that methods for data annotation and pattern identification are available. Maximizing phenotypic difference while minimizing genotypic difference, as was done here between two fertile but morphologically distinct members of the same species, is critical for analyses targeting phenotypic causative mutations. The strong resistance phenotype of MRI compared to SB22 likely led to the ability to detect gene expression differences between the cultivars. After validation, the sequences provided here can serve as molecular markers for the selection of disease resistant sweet basil breeding.

D. Methods

1. Sample preparation

Inbred *Ocimum basilicum* (sweet basil) genotypes SB22 (*P. belbahrii* susceptible) and MRI (*P. belbahrii* resistant) plants were grown from seed. Previously infected sweet basil leaves with fresh *P. belbahrii* sporulation were harvested and agitated for 5 minutes in diH2O. The inoculum mixture was filtered with 40 μm nylon mesh. A 1 mL subsample

from the filtered inoculum was pipetted into an eppendorf tube and frozen at -80C to serve as pathogen control. The remaining inoculum was centrifuged at 1,000g for 1 min and diH_20 decanted. The resulting sporangia pellet was resuspended in diH_2O and inoculum adjusted to 1 x 10^5 sporangia/mL. Plants were inoculated at the 6-leaf (3 true leaf set) growth stage with approximately ~1 mL/leaf and plants were incubated at 100% relative humidity. MRI and SB22 were sprayed with sterile diH_2O in triplicate to serve as the mock inoculated control.

Four disks per true leaf were sampled from both genotypes at 12, 24 and 72 hpi and immediately flash frozen in liquid nitrogen. The water control leaves were harvested at 12 hpi only. Total RNA was extracted from freshly ground tissue using the Spectrum™ Plant Total RNA Kit (Sigma Aldrich). RNA samples were used to generate sequence libraries using a library prep kit from the New England Biolabs (NEB #E7530). Pair-end sequence reads of 75 bp were generated at the TUFTs genomic center at the Tufts University School of Medicine using Hi-Seq Illumina platform.

Higher coverage was specifically designed for post inoculated samples that contain both the pathogen and the host due to the increased complexity of these samples, allowing the future study of pathogen expression.

2. Generating transcript assemblies and FPKM expression

FASTQC was used to assess average read quality. Paired end fastq files were provided to Trinity version 2.4.0 and assembled using default parameters (Grabherr et al., 2011).

Datasets were assembled including a single set using all MRI datasets and the sporangia control (MRI Combined Assembly), and all SB22 datasets including the sporangia control (SB22 Combined Assembly). Separately assembled control data provided organism specific databases of genes and transcripts.

The resulting output files served as the references for expression quantification. RSEM version 1.2.29 and bowtie version 1.0.0 were used to calculate FPKM (Fragments per Kilobase exon per Million mapped reads) values for assembled contigs while tracking replicate information. All RNAseq datasets from both MRI and SB22 were mapped to the infected MRI reference. This was done to standardize the reference, allow for the usage of already generated gene annotations, and produce an output which could be used to cluster both plant isolates together. In all cases the standard settings were used for assembly and transcript quantification. Additionally edgeR was used to calculate differential expression among genes. Expression data from both MRI and SB22 data mapped to the MRI combined assembly using Trinity and edgeR was used to assess differential expression between all timepoints within a single cultivar. Trinity DEG output data was filtered for genes with a p-value less than 0.05 and the FDR was less than 0.01.

3. Bi-Directional BLAST hits

In order to explore the overall sequence conservation between MRI and SB22 I performed a bidirectional BLAST using all sequences within the MRI and SB22 water control assemblies. Briefly, all MRI sequences were compared to the SB22

transcriptome, and all SB22 sequences were compared to the MRI transcriptome. The hit with the highest bit score for each gene was chosen. Results were compared and pairs of top scoring genes were considered bi-directional blast hits (i.e. if MRI gene X BLASTs to SB22 gene Z, and SB22 gene Z BLASTs to MRI gene X).

4. Sequence translation and annotation

We generated a database of sequence annotations for MRI genes. All MRI genes with an expression greater than 1 FPKM in at least one timepoint was chosen and the longest transcript associated with that gene was compared to the NCBI non-redundant database using cloud BLAST through BLAST2Go. Annotations were saved as a searchable database in text format. In some cases genes were filtered by taxonomic hit to verify their species of origin.

In order to facilitate easier searches for gene families of interest I translated the longest nucleotide sequence associated with each gene into six-frame translated protein sequences using EMBOSS, searching for only those translated sequences between START and STOP codons and longer than 30 AA. In many cases in order to verify protein domain structure, the nucleotide or protein sequence was analyzed using either the NCBI conserved domain finder (Marchler-Bauer et al., 2015), PFAM (Finn et al., 2016), or InterProScan (Jones et al., 2014). In many cases sequences were aligned using MEGA 6 for visualization (Tamura et al., 2013).

5. Expression clustering toward candidate resistance gene identification

Genes were clustered using Trinity version 2.2.0 based on read counts following the steps outlined in the Trinity manual (Robinson et al., 2010b). Expression data generated by mapping all datasets to the MRI infected reference were used for clustering. A matrix of all gene expression at all timepoints and replicates was used to define clusters with the edgeR function associated with Trinity, using p=50 and p=20 (effectively a measure of how large the clusters are created. Higher numbers form larger, broader, clusters). The resulting clusters, available in pdf format, were visually examined for clusters which displayed the target expression profile.

6. BLAST search for NB-LRR, RLK, hormone pathway, and secondary metabolite enzyme genes

Toward the analysis of MRI unique gene families and hormone signaling genes I screened MRI and SB22 using protein sequences. Arabidopsis, or in some cases sweet basil, protein sequences retrieved from NCBI were compared to MRI and SB22 translated nucleotide sequences using BLASTp. Generally the hit with the highest bit score was chosen as the top hit for each sequence. In cases of short alignment length or low sequence identity, the recovered MRI or SB22 hit was compared to the green plant database on NCBI. BLAST version 2.2.22 was used in all cases to compare protein sequences using the Massachusetts green-energy high performance computing center (Altschul et al., 1990).

E. Acknowledgements

We would like to acknowledge Kelly Allen for providing a conceptual biological framework around which to understand the sequencing data and for helping with bench experiments that have not been published. We would like to thank Kelly Allen and Kathryn Vescio for help reviewing the manuscript prior to submission.

F. Author Contribution

Jim Simon, Li-Jun Ma, Robert Pyne, and Li Guo planned the infection time-course experiments. Robert Pyne performed the infection time-course experiments, extracted and sequenced genetic material. Greg DeIulio performed all data analysis.

CHAPTER IV

CONCLUSIONS

A. A Study in kinomes

To call the signaling networks of cells complex understates their nature, but they are not infinitely complex. Each component plays a part in the greater whole while on or off, in one state or another. The complexity of these networks still prevents us from condensing the entirety of the network into a predictable set of equations and rules. At our current pace however, it is only a matter of time before this becomes reality.

The study of the kinomes of a subset of the *Fusarium oxysporum* species complex has told us that among hundreds of kinases that form the signaling backbone of the cell, kinases related to signaling environmental status are proliferated. All organisms living within the greater environment are at its mercy. An enhanced ability to detect and incorporate signals from outside, as well as from within the cell are undeniably an advantage for any living individual.

Kinases at the periphery of these networks, histidine kinases as highlighted in Chapter 2, are among the first line in providing an organism with information about its surroundings. Each in their own way provides additional signals with which the cellular machinery "decides" on the best course of action. Downstream of these, the TOR kinases exist as a hub of nutrient signaling through which much of the cells signaling of the environment flows. Signals from kinases like histidine kinases may flow through hubs like TOR and then be integrated into transcriptional responses. Additional TOR kinases could provide

the cell with novel signaling hubs or provide more flexibility to already existing ones. Further downstream of these critical choke points, other kinases like the CDC2 and SRPKL kinases function to enact changes in the output of the cell as directed from above and below, with broadened numbers possibly adding novel or more refined control.

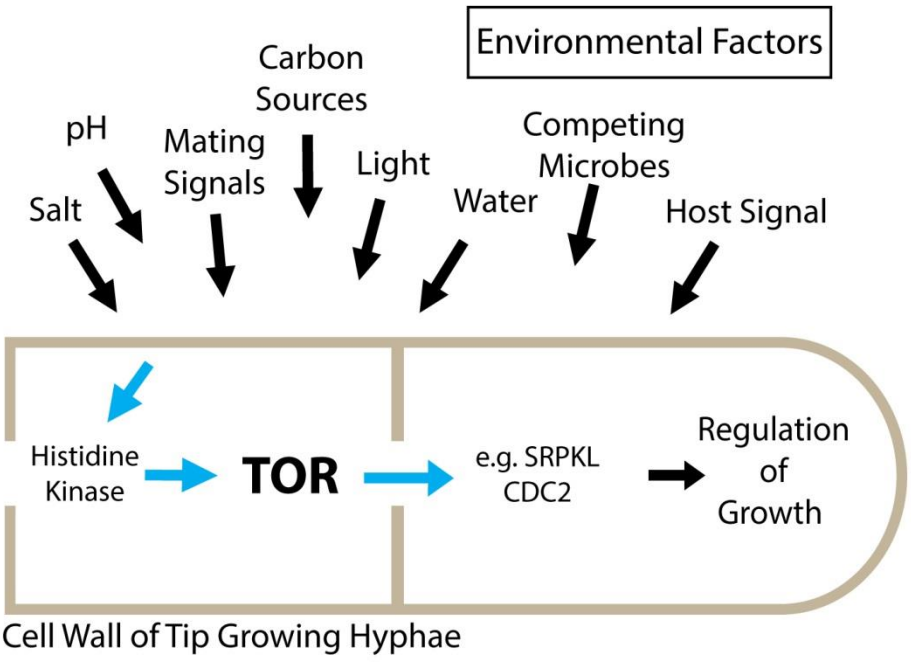

Figure 4.1 Signaling pathways and the regulation of growth

Large variation in kinase number among a wide array of families is fairly rare within the Ascomycota included in this study, with the large majority of pathways preserved perfectly across evolutionary time. Although the expansion among histidine kinases is not unique to Fusaria, it continues an observed trend extending from the yeast fungi to the FOSC (Catlett et al., 2003).

I hypothesize that the expanded pathways seen here are linked to cellular growth patterns. Hyphal fungi can move, although this movement is constricted to directional growth. Organisms like fungi grow in order to obtain the resources. For an organism that grows to move, especially one that grows by adding cells to the tip of an existing thread, control of the cell cycle is a critical. The CDC2 kinases are known to be primary cell cycle regulators. Furthermore, in mammals, some members of the SRPK family of kinases aid in cell cycle control in neurons through the regulation of mRNA splicing. Finally, the TOR kinase was a large focus of Chapter 2 and sits at a junction that controls the initiation of autophagy. This process begins large scale protein degradation and is intimately involved in the growth and development of fungal hyphae (Veneault-Fourrey and Talbot, 2007). It is tempting to hypothesize that the expansion of some kinase families from yeast to hyphal fungi, especially those which appear related to environmental awareness, are related in some way to directional growth and the ability to determine directionality.

Additionally, autophagy contributes to long term survival through the redistribution of cellular resources. With ubiquitous soil organisms like *Fusaria*, which are known to survive for extended periods in a state of stasis, perhaps these expanded pathways could contribute to their persistence.

The sequenced members of the FOSC are predominantly pathogens, an inevitable outcome of preferentially sequencing only those organisms that pose a problem to human welfare. While I saw that variation among some kinases exists even between closely

related FOSC isolates, I have not captured the full breadth of the species complex as it exists in nature. Studies like this can only be aided by additional sequencing of strains that show no pathogenicity or those that are pathogenic to diverse hosts. Of the over 100 hosts of *F. oxysporum* I have examined barely less than one tenth of that breadth. Likewise, although my study was broad in the sense that species from across Ascomycota were chosen, the inclusion of more taxa from various empty spaces in the phylogeny could only have strengthened this study. While the focus was the FOSC, to truly understand the significance of the differences among strains it is important to know how these same differences manifest across large evolutionary timescales.

B. Basil downy mildew

Basil downy mildew is a problem here and now. I hope that the candidate resistance genes outlined in the section above can help basil breeding or could perhaps be incorporated into genetically modified basil cultivars in the future when restrictions on GM-crops might be reduced.

The study of the transcriptomes of MRI and SB22 highlights interesting differences between two genetically similar cultivars. There are clear morphological and chemotypic differences between them, in leaf shape and color, leaf density, the plant volatiles, and taste. Additionally, we observe the single trait which is of most interest here, disease resistance. Fascinatingly, by examining the mRNA of these plants we clearly see the chemotypic and disease traits reflected in genes unique to each cultivar.

Although the cultivars share the vast majority of expressed genes, MRI was found to express more unique genes. Both cultivars express a unique set of NB-LRR type genes, which are annotated as being homologous to disease resistance genes, underscoring cultivar-unique disease resistance phenotypes. Additionally, MRI is annotated as having a large set of unique secondary metabolic genes, while SB22 contains effectively no unique members. This is interesting, as it tells us that MRI and SB22 not only have different capacities for secondary metabolism but also that MRI has an expanded capacity in relation to SB22. This could perhaps be a result of more heavy breeding in SB22 resulting in the removal of a larger pool of the genome in order to select for the specific traits desired.

I was also able to directly detect a difference in a critical disease signaling hormone pathway. Numerous studies across a wide array of plant species have shown SA to be an important part of the defense response to biotrophic pathogens. As a major driver of cell death pathways, SA is the perfect response to a microbe that requires living host cells to survive.

While the signal seen in the SA pathway appeared obvious and suggests possible mechanisms for resistance, transcriptional differences in genes which more directly cause resistance were less obvious (i.e. ROS production, Cell Death related proteins, etc.). It is possible that if single infected cells were sequenced I might have observed the expected strong contrast in expression profiles. Almost undoubtedly by sequencing a population of cells, and thereby an average of expression across hundreds of thousands of cells, some

signals are lost. However, likewise, we likely capture many signals what would be lost by sequencing only a single cell. We learn what we can with the technology we have.

We did however effectively ignore the possibility that the resistance causative sequence was a mutation in a gene shared between the cultivars. Without previous knowledge outlining the set of genes that cause disease resistance in sweet basil, a search for SNPs which cause resistance may never be successful. While the chosen approach does not cover all possibilities it was still able to provide an encouraging list of resistance candidates and explored the interaction between basil and its downy mildew pathogen.

C. Views of comparative sequencing

1. Lessons from the Sweet Basil project

Both projects have come at similar questions from different directions, attempting to link biology to nucleic acid sequence. In the case of the kinome project we had host variation and hoped to use DNA to identify the underlying causative sequences. While studying basil, we had disease resistance variation and hoped to use RNA to identify the sequences underlying the phenotype.

As discussed in the Preface, the fantastic strength of sequencing technology is its ability to turn something almost incomprehensible into something approachable; data, letters and numbers. This simple text-based data can be computed and quickly analyzed in ways

which would otherwise take entire human lifetimes. Genomes can be billions of bases long.

The basil project was enlightening as I realized most of the way through that it was actually irrelevant that I have any prior information on the organism itself. I did not *need* to understand the biology. The data itself could tell me everything I wanted to know, if I knew how to look and had the time.

I didn't need information on the phenotype, chemotype, species, or even kingdom within which it resided. Given two transcriptomes covering sufficient depth of coverage and breadth of conditions to elicit the expression of a majority of genes, the data itself can lead you. Based on BLAST to known species, the fact that the individuals were plants would have become clear first, followed by a location in the tree enough to identify close relatives. Due to the presence of some sweet basil sequences within NCBI it would have been possible to identify the plants as basil with strong confidence. High sequence identity between homologs would have highlighted their close association. It would have been possible to say that the plants, likely, differ in their resistance capacity to oomycete diseases given the overrepresentation of resistance to *Peronospora parasitica* (Arabidopsis downy mildew) genes among unique genes, and that the cultivars had different secondary metabolite outputs. Given the transcriptional response of the plants it would have been possible to predict that they may have been infected by a pathogen of some kind and, examining the SA response, that the pathogen was likely biotrophic.

However, everything seems clear in hindsight.

2. Handling of Genetic Information

DNA and RNA sequencing have the ability to provide answers to questions as broad as evolutionary relationships between kingdoms, to as simple as SNPs that cause observed phenotypic traits. The key to its success is the design of proper experiments and the experience and planning to know where and how to look.

Many RNA sequencing experiments are geared toward not only the determination of the sequences present, but also how their abundance changes over time. To do this, multiple timepoints or conditions are sequenced and compared. The exact timing of these sequencing experiments, especially in plants, but also in other organisms, can and will affect the data which are generated. The circadian cycle in plants is known to control a wide array of plant transcripts that fluctuate with the time of day. A 12 hour shift in photoperiod will result in changes in expression of many genes, which are otherwise likely not involved in the experiment being performed (assuming you are not studying circadian rhythm). This does not prevent the study from being done, but it does add additional information, which will need to be considered in analysis of the data. In some cases this is a necessity, as capturing time points below the 24 hour threshold is key.

My time spent in analysis of RNA sequencing data has shown me the need for adequate replicates, which I now believe to be at least 4. While 3 may be the magic number for statistical assessments of significance, firsthand experience shows that variability among

RNA sequencing replicates is unpredictable. In general replicates correlate well when tested by a simple Pearson correlation, but this is a grand average. Many transcriptomics studies at some point examine not only the average, but also individual genes. At this level variations in expression within a single gene can produce significant doubt in a result, with verification then needed. The goal in reality is not to convince the computer of the significance of your results, but yourself and your reader, to be confident going forward that you are correct.

Much of the strength in DNA and RNA is that they provide in many cases *all* the possibilities, and it is up to the scientist to narrow those possibilities. Once a target or set of targets is identified then a wide array of techniques can be used to probe the gene and those like it. With a sea of data to be sifted through it is very easy to get lost or to be tempted to inquire into every small interesting detail. It is critical to realize that with computational methods you will always have an amount of unknown. Being too stringent in removing possibilities could lead to false negative answers while being to lenient provides false positives. However, this is the field of large data analysis, its strengths and weaknesses should be remembered when definitive statements are made based on its results.

BIBLIOGRAPHY

Acharya, B.R., Raina, S., Maqbool, S.B., Jagadeeswaran, G., Mosher, S.L., Appel, H.M., Schultz, J.C., Klessig, D.F., and Raina, R. (2007). Overexpression of CRK13, an Arabidopsis cysteine-rich receptor-like kinase, results in enhanced resistance to Pseudomonas syringae. The Plant Journal *50*, 488-499.

Aimé, S., Alabouvette, C., Steinberg, C., and Olivain, C. (2013). The Endophytic Strain Fusarium oxysporum Fo47: A Good Candidate for Priming the Defense Responses in Tomato Roots. Molecular Plant-Microbe Interactions *26*, 918-926.

Alikhan, N.-F., Petty, N., Ben Zakour, N., and Beatson, S. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics *12*, 402.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

An, J., Shen, X., Ma, Q., Yang, C., Liu, S., and Chen, Y. (2014). Transcriptome profiling to discover putative genes associated with paraquat resistance in goosegrass (Eleusine indica L.). PLoS ONE *9*.

Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., and Singh, K.B. (2010). Plants versus pathogens: an evolutionary arms race. Functional Plant Biology *37*, 499-512.

Baldin, C., Valiante, V., Krüger, T., Schafferer, L., Haas, H., Kniemeyer, O., and Brakhage, A.A. (2015). Comparative proteomics of a tor inducible Aspergillus fumigatus mutant reveals involvement of the Tor kinase in iron regulation. Proteomics *15*, 2230-2243.

Bharucha, N., Ma, J., Dobry, C.J., Lawson, S.K., Yang, Z., and Kumar, A. (2008). Analysis of the yeast kinome reveals a network of regulated protein localization during filamentous growth. Mol Biol Cell *19*, 2708-2717.

Bittner-Eddy, P.D., Crute, I.R., Holub, E.B., and Beynon, J.L. (2000). RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. Plant J *21*, 177-188.

Boddu, J., Cho, S., Kruger, W.M., and Muehlbauer, G.J. (2006). Transcriptome analysis of the barley-Fusarium graminearum interaction. Molecular plant-microbe interactions *19*, 407-417.

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

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D. (1998). Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct Peronospora parasitica avirulence determinants. Plant Cell *10*, 1847-1860.

Bozkurt, T.O., Schornack, S., Banfield, M.J., and Kamoun, S. (2012). Oomycetes, effectors, and all that jazz. Current Opinion in Plant Biology *15*, 483-492.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., *et al.* (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043-1046.

Brodersen, P., Malinovsky, F.G., Hématy, K., Newman, M.-A., and Mundy, J. (2005). The Role of Salicylic Acid in the Induction of Cell Death in Arabidopsis acd11. Plant Physiology *138*, 1037-1045.

Buck, V., Quinn, J., Soto Pino, T., Martin, H., Saldanha, J., Makino, K., Morgan, B.A., and Millar, J.B. (2001). Peroxide sensors for the fission yeast stress-activated mitogen-activated protein kinase pathway. Mol Biol Cell *12*, 407-419.

Cao, Y., Liang, Y., Tanaka, K., Nguyen, C.T., Jedrzejczak, R.P., Joachimiak, A., and Stacey, G. (2014). The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1. eLife *3*, e03766.

Caspari, T., and Hilditch, V. (2015). Two distinct Cdc2 pools regulate cell cycle progression and the DNA damage response in the fission yeast S.pombe. PLoS ONE *10*, e0130748.

Catlett, N.L., Yoder, O.C., and Turgeon, B.G. (2003). Whole-genome analysis of two-component signal transduction genes in fungal pathogens. Eukaryotic Cell *2*, 1151-1161.

Cecchini, N.M., Jung, H.W., Engle, N.L., Tschaplinski, T.J., and Greenberg, J.T. (2015). ALD1 Regulates Basal Immune Components and Early Inducible Defense Responses in Arabidopsis. Mol Plant Microbe Interact *28*, 455-466.

Chern, M., Xu, Q., Bart, R.S., Bai, W., Ruan, D., Sze-To, W.H., Canlas, P.E., Jain, R., Chen, X., and Ronald, P.C. (2016). A Genetic Screen Identifies a Requirement for Cysteine-Rich–Receptor-Like Kinases in Rice NH1 (OsNPR1)-Mediated Immunity. PLOS Genetics *12*, e1006049.

Choi, Y., Choi, I., Lee, K., and Shin, H. (2016). First report of downy mildew caused by Peronospora belbahrii on sweet basil (Ocimum basilicum) in Korea. Plant Disease *100*, 2335-2335.

Choi, Y.K., Kang, E.-H., and Park, H.-M. (2014). Role of LAMMER kinase in cell wall biogenesis during vegetative growth of Aspergillus nidulans. Mycobiology *42*, 422-426.

Clay, K., and Kover, P.X. (1996). The red queen hypothesis and plant/pathogen interactions. Annual Review of Phytopathology *34*, 29-50.

Cohen, Y., and Ben-Naim, Y. (2016). Nocturnal Fanning Suppresses Downy Mildew Epidemics in Sweet Basil. PLoS ONE *11*, e0155330.

Cohen, Y., and Rubin, A.E. (2015). Daytime Solar Heating Controls Downy Mildew Peronospora belbahrii in Sweet Basil. PLoS ONE *10*, e0126103.

Cohen, Y., Vaknin, M., Ben-Naim, Y., and Rubin, A.E. (2013). Light Suppresses Sporulation and Epidemics of Peronospora belbahrii. PLoS ONE *8*, e81282.

Coito, C., Diamond, D.L., Neddermann, P., Korth, M.J., and Katze, M.G. (2004). High-throughput screening of the yeast kinome: identification of human serine/threonine protein kinases that phosphorylate the hepatitis C virus NS5A protein. Journal of Virology *78*, 3502-3513.

Couteaudier, Y., and Alabouvette, C. (1990). Survival and inoculum potential of conidia and chlamydospores of Fusarium oxysporum f.sp. lini in soil. Canadian journal of microbiology *36*, 551-556.

De Cremer, K., Mathys, J., Vos, C., Froenicke, L., Michelmore, R.W., Cammue, B., and De Coninck, B. (2013). RNAseq-based transcriptome analysis of Lactuca sativa infected by the fungal necrotroph Botrytis cinerea. Plant, cell & environment *36*, 1992-2007.

De Souza, C.P., Hashmi, S.B., Osmani, A.H., Andrews, P., Ringelberg, C.S., Dunlap, J.C., and Osmani, S.A. (2013). Functional analysis of the Aspergillus nidulans kinome. PLoS ONE *8*, e58008.

De Vos, M., Van Oosten, V.R., Van Poecke, R.M., Van Pelt, J.A., Pozo, M.J., Mueller, M.J., Buchala, A.J., Metraux, J.P., Van Loon, L.C., Dicke, M., *et al.* (2005). Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack. Mol Plant Microbe Interact *18*, 923-937.

Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., and Ellis, J. (2012). The Top 10 fungal pathogens in molecular plant pathology. Molecular plant pathology *13*, 414-430.

Defosse, T.A., Sharma, A., Mondal, A.K., Dugé de Bernonville, T., Latgé, J.-P., Calderone, R., Giglioli-Guivarc'h, N., Courdavault, V., Clastre, M., and Papon, N. (2015). Hybrid histidine kinases in pathogenic fungi. Molecular Microbiology *95*, 914-924.

Den Haese, G.J., Walworth, N., Carr, A.M., and Gould, K.L. (1995). The Wee1 protein kinase regulates T14 phosphorylation of fission yeast Cdc2. Mol Biol Cell *6*, 371-385.

DeYoung, B.J., and Innes, R.W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. Nature immunology *7*, 1243-1249.

Di Pietro, A., García-MacEira, F.I., Méglecz, E., and Roncero, M.I. (2001). A MAP kinase of the vascular wilt fungus Fusarium oxysporum is essential for root penetration and pathogenesis. Molecular microbiology *39*, 1140-1152.

Donaldson, A.D., Fangman, W.L., and Brewer, B.J. (1998). Cdc7 is required throughout the yeast S phase to activate replication origins. Genes Dev *12*, 491-501.

El Hadrami, A., El-Bebany, A.F., Yao, Z., Adam, L.R., El Hadrami, I., and Daayf, F. (2012). Plants versus fungi and oomycetes: pathogenesis, defense and counter-defense in the proteomics era. Int J Mol Sci *13*, 7237-7259.

Fangwei, Y., Qin, G., Yingzi, Y., Yanni, Y., Jin-Rong, X., Won-Bo, S., and Zhonghua, M. (2014). The TOR signaling pathway regulates vegetative development and virulence in Fusarium graminearum. New Phytologist *203*, 219-232.

Feys, B.J., Moisan, L.J., Newman, M.A., and Parker, J.E. (2001). Direct interaction between the Arabidopsis disease resistance signaling proteins, EDS1 and PAD4. The EMBO Journal *20*, 5400-5411.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., *et al.* (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research *44*, D279-D285.

Forment, J., Mulet, J.M., Vicente, O., and Serrano, R. (2002). The yeast SR protein kinase Sky1p modulates salt tolerance, membrane potential and the Trk1,2 potassium transporter. Biochimica et Biophysica Acta (BBA) - Biomembranes *1565*, 36-40.

Fung, R.W., Gonzalo, M., Fekete, C., Kovacs, L.G., He, Y., Marsh, E., McIntyre, L.M., Schachtman, D.P., and Qiu, W. (2008). Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. Plant physiology *146*, 236-249.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., *et al.* (2003). The genome sequence of the filamentous fungus Neurospora crassa. Nature *422*, 859-868.

Garibaldi, A., Minuto, A., and Gullino, M.L. (2005). First Report of Downy Mildew Caused by Peronospora sp. on Basil (Ocimum basilicum) in France. Plant Disease *89*, 683-683.

Garibaldi, A., Minuto, A., Minuto, G., and Gullino, M.L. (2004). First Report of Downy Mildew on Basil (Ocimum basilicum) in Italy. Plant Disease *88*, 312-312.

Goldberg, J.M., Griggs, A.D., Smith, J.L., Haas, B.J., Wortman, J.R., and Zeng, Q. (2013a). Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. Bioinformatics *29*, 2387-2394.

Goldberg, J.M., Griggs, A.D., Smith, J.L., Haas, B.J., Wortman, J.R., and Zeng, Q. (2013b). Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. Bioinformatics *29*, 2387-2394.

Goldberg, J.M., Manning, G., Liu, A., Fey, P., Pilcher, K.E., Xu, Y., and Smith, J.L. (2006). The Dictyostelium kinome—analysis of the protein kinases from a simple model organism. PLoS Genet *2*, e38.

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell *5*, 1003-1011.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology *29*, 644-652.

Grayer, R.J., and Kokubun, T. (2001). Plant–fungal interactions: the search for phytoalexins and other antifungal compounds from higher plants. Phytochemistry *56*, 253-263.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol *19*, 1720-1730.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T., *et al.* (2009). Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature *461*, 393-398.

Hamel, L.-P., Nicole, M.-C., Duplessis, S., and Ellis, B.E. (2012). Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. Plant Cell *24*, 1327-1351.

Hanks, S.K., and Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J *9*, 576-596.

Hayden, K.J., Garbelotto, M., Knaus, B.J., Cronn, R.C., Rai, H., and Wright, J.W. (2014). Dual RNA-seq of the plant pathogen Phytophthora ramorum and its tanoak host. Tree genetics & genomes *10*, 489-502.

He, P., Wang, Y., Wang, X., Zhang, X., and Tian, C. (2017). The Mitogen-Activated Protein Kinase CgMK1 Governs Appressorium Formation, Melanin Synthesis, and Plant Infection of Colletotrichum gloeosporioides. Frontiers in microbiology *8*, 2216.

Heitman, J., Movva, N.R., and Hall, M.N. (1991). Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. Science *253*, 905-909.

Hindle, M.M., Martin, S.F., Noordally, Z.B., van Ooijen, G., Barrios-Llerena, M.E., Simpson, T.I., Le Bihan, T., and Millar, A.J. (2014). The reduced kinome of Ostreococcus tauri: core eukaryotic signalling components in a tractable model species. BMC Genomics *15*, 1-21.

Ho, Y.-S.J., Burden, L.M., and Hurley, J.H. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. The EMBO Journal *19*, 5288-5299.

Hok, S., Danchin, E.G.J., Allasia, V., PanabiÈRes, F., Attard, A., and Keller, H. (2011). An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant, Cell & Environment *34*, 1944-1957.

Holub, E.B., Beynon, J.L., and Crute, I.R. (1994). Phenotypic and genotypic characterization of interactions between isolates of Peronospora parasitica and accessions of Arabidopsis thaliana. Molecular Plant-Microbe Interactions *7*, 223-239.

Homa, K., Barney, W.P., Ward, D.L., Wyenandt, C.A., and Simon, J.E. (2014). Evaluation of Fungicides for the Control of Peronospora belbahrii on Sweet Basil in New Jersey. Plant Disease *98*, 1561-1566.

J Flatauer, L., F Zadeh, S., and Bardwell, L. (2005). Mitogen-Activated Protein Kinases with Distinct Requirements for Ste5 Scaffolding Influence Signaling Specificity in Saccharomyces cerevisiae, Vol 25.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. The EMBO journal *19*, 4004-4014.

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

Kanetis, L., Vasiliou, A., Neophytou, G., Samouel, S., and Tsaltas, D. (2014). First report of downy mildew caused by Peronospora belbahrii on sweet basil (Ocimum basilicum) in Cyprus. Plant Disease *98*, 283-283.

Kang, E.-H., Kim, J.-a., Oh, H.-W., and Park, H.-M. (2013). LAMMER kinase LkhA plays multiple roles in the vegetative growth and asexual and sexual development of Aspergillus nidulans. PLoS ONE *8*, e58762.

Kapteyn, J., Qualley, A.V., Xie, Z., Fridman, E., Dudareva, N., and Gang, D.R. (2007). Evolution of Cinnamate p-Coumarate Carboxyl Methyltransferases and Their Role in the Biosynthesis of Methylcinnamate. Plant Cell *19*, 3212-3229.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., and Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLoS ONE *7*, e49423.

Khateri, H., Calmin, G., Moarrefzadeh, N., Belbahri, L., and Lefort, F. (2007). First report of downy mildew caused by Peronospora sp. on basil in Northern Iran. Journal of Plant Pathology *89*.

Kim, S.Y., Shang, Y., Joo, S.H., Kim, S.K., and Nam, K.H. (2017). Overexpression of BAK1 causes salicylic acid accumulation and deregulation of cell death control genes. Biochem Biophys Res Commun *484*, 781-786.

Kinkema, M., Fan, W., and Dong, X. (2000). Nuclear Localization of NPR1 Is Required for Activation of PR Gene Expression. Plant Cell *12*, 2339-2350.

Kistler, H.C. (1997). Genetic Diversity in the Plant-Pathogenic Fungus Fusarium oxysporum. Phytopathology *87*, 474-479.

Knoth, C., and Eulgem, T. (2008). The oomycete response gene LURP1 is required for defense against Hyaloperonospora parasitica in Arabidopsis thaliana. Plant J *55*, 53-64.

Kosti, I., Mandel-Gutfreund, Y., Glaser, F., and Horwitz, B.A. (2010). Comparative analysis of fungal protein kinases and associated domains. BMC Genomics *11*, 1-12.

Krems, B., Charizanis, C., and Entian, K.D. (1996). The response regulator-like protein Pos9/Skn7 of Saccharomyces cerevisiae is involved in oxidative stress resistance. Curr Genet *29*, 327-334.

Laplante, M., and Sabatini, D.M. (2009). mTOR signaling at a glance. Journal of Cell Science *122*, 3589-3594.

Lee, J., Moir, Robyn D., McIntosh, Kerri B., and Willis, Ian M. (2012). TOR signaling regulates ribosome and tRNA synthesis via LAMMER/Clk and GSK-3 family kinases. Molecular Cell *45*, 836-843.

Leslie, J.F., and Summerell, B.A. (2006). The fusarium laboratory manual. In The fusarium laboratory manual, J.F. Leslie, and B.A. Summerell, eds. (Blackwell Publishing), pp. i-xii.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, D., Agrellos, O.A., and Calderone, R. (2010). Histidine kinases keep fungi safe and vigorous. Current Opinion in Microbiology *13*, 424-430.

Li, J., Brader, G., and Palva, E.T. (2004). The WRKY70 Transcription Factor: A Node of Convergence for Jasmonate-Mediated and Salicylate-Mediated Signals in Plant Defense. Plant Cell *16*, 319-331.

Li, Z., Hao, Y., Wang, L., Xiang, H., and Zhou, Z. (2015). Genome-wide identification and comprehensive analyses of the kinomes in four pathogenic Microsporidia species. PLoS ONE *9*, e115890.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research *41*, e108.

Loewith, R., and Hall, M.N. (2011). Target of rapamycin (TOR) in nutrient signaling and growth control. Genetics *189*, 1177-1201.

López-Berges, M.S., Rispail, N., Prados-Rosales, R.C., and Di Pietro, A. (2010). A nitrogen response pathway regulates virulence functions in Fusarium oxysporum via the protein kinase TOR and the bZIP protein MeaB. Plant Cell *22*, 2459-2475.

Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., and Ton, J. (2011). Callose deposition: a multifaceted plant defense response. Molecular Plant-Microbe Interactions *24*, 183-193.

Lundgren, K., Walworth, N., Booher, R., Dembski, M., Kirschner, M., and Beach, D. (1991). mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2. Cell *64*, 1111-1122.

Ma, L.-J., Does, H.C.v.d., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., e, Pietro, A.D., Dufresne, M., Freitag, M., Grabherr, M., *et al.* (2010). Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature *464*, 367-373.

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

Madden, K., Sheu, Y.J., Baetz, K., Andrews, B., and Snyder, M. (1997). SBF cell cycle regulator as a target of the yeast PKC-MAP kinase pathway. Science *275*, 1781-1784.

Malumbres, M. (2014). Cyclin-dependent kinases. Genome Biol *15*, 122.

Manning, G., Plowman, G.D., Hunter, T., and Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. Trends in Biochemical Sciences *27*, 514-520.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002b). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., *et al.* (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res *43*, D222-226.

Martin, H., Shales, M., Fernandez-Piñar, P., Wei, P., Molina, M., Fiedler, D., Shokat, K.M., Beltrao, P., Lim, W., and Krogan, N.J. (2015). Differential genetic interactions of yeast stress response MAPK pathways. Molecular Systems Biology *11*.

Martinez de La Parte, E., Pérez-Vicente, L., Bernal, B., and Garcia, D. (2010). First report of Peronospora sp. on sweet basil (Ocimum basilicum) in Cuba. Plant Pathology *59*, 800-800.

Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev *70*, 910-938.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G.M., Goff, S., Holub, E.B., and Dangl, J.L. (1998). Intragenic Recombination and Diversifying Selection Contribute to the Evolution of Downy Mildew Resistance at the RPP8 Locus of Arabidopsis. Plant Cell *10*, 1861-1874.

McHale, L., Tan, X., Koehl, P., and Michelmore, R.W. (2006). Plant NBS-LRR proteins: adaptable guards. Genome Biology *7*, 212-212.

McLeod, A., Coertze, S., and Mostert, L. (2006). First Report of a Peronospora Species on Sweet Basil in South Africa. Plant Disease *90*, 1115-1115.

Mendy, B., Wang'ombe, M.W., Radakovic, Z.S., Holbein, J., Ilyas, M., Chopra, D., Holton, N., Zipfel, C., Grundler, F.M.W., and Siddique, S. (2017). Arabidopsis leucine-rich repeat receptor–like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. PLOS Pathogens *13*, e1006284.

Meyer, F.E., Shuey, L.S., Naidoo, S., Mamni, T., Berger, D.K., Myburg, A.A., Van den Berg, N., and Naidoo, S. (2016). Dual RNA-sequencing of Eucalyptus nitens during Phytophthora cinnamomi challenge reveals pathogen and host factors influencing compatibility. Frontiers in plant science *7*, 191.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell *15*, 809-834.

Miranda-Saavedra, D., Stark, M.J., Packer, J.C., Vivares, C.P., Doerig, C., and Barton, G.J. (2007). The complement of protein kinases of the microsporidium Encephalitozoon cuniculi in relation to those of Saccharomyces cerevisiae and Schizosaccharomyces pombe. BMC Genomics *8*, 1-21.

Mohr, T.J., Mammarella, N.D., Hoff, T., Woffenden, B.J., Jelesko, J.G., and McDowell, J.M. (2010). The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. Mol Plant Microbe Interact *23*, 1303-1315.

Mok, J., Kim, P.M., Lam, H.Y.K., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., *et al.* (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci STKE *3*, ra12-ra12.

Morgan, D.O. (1995). Principles of CDK regulation. Nature *374*, 131-134.

Morgan, W., and Kamoun, S. (2007). RXLR effectors of plant pathogenic oomycetes. Curr Opin Microbiol *10*, 332-338.

Mulet, J.M., Leube, M.P., Kron, S.J., Rios, G., Fink, G.R., and Serrano, R. (1999). A novel mechanism of ion homeostasis and salt tolerance in yeast: the Hal4 and Hal5 protein kinases modulate the Trk1-Trk2 potassium transporter. Mol Cell Biol *19*, 3328-3337.

Nagy, G., and Horvath, A. (2011). Occurrence of downy mildew caused by Peronospora belbahrii on sweet basil in Hungary. Plant Disease *95*, 1034-1034.

Nianiou-Obeidat, I., Madesis, P., Kissoudis, C., Voulgari, G., Chronopoulou, E., Tsaftaris, A., and Labrou, N.E. (2017). Plant glutathione transferase-mediated stress tolerance: functions and biotechnological applications. Plant Cell Reports *36*, 791-805.

Nucci, M., and Anaissie, E. (2007). Fusarium Infections in Immunocompromised Patients. Clinical Microbiology Reviews *20*, 695-704.

Olsen, K.M., Lea, U.S., Slimestad, R., Verheul, M., and Lillo, C. (2008). Differential expression of four Arabidopsis PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. Journal of Plant Physiology *165*, 1491-1499.

Palma-Guerrero, J., Ma, X., Torriani, S.F.F., Zala, M., Francisco, C.S., Hartmann, F.E., Croll, D., and McDonald, B.A. (2017). Comparative Transcriptome Analyses in Zymoseptoria tritici Reveal Significant Differences in Gene Expression Among Strains During Plant Infection. Molecular Plant-Microbe Interactions *30*, 231-244.

Pancaldi, V., Saraç, Ö.S., Rallis, C., McLean, J.R., Převorovský, M., Gould, K., Beyer, A., and Bähler, J. (2012). Predicting the fission yeast protein interaction network. G3: Genes|Genomes|Genetics *2*, 453-467.

Park, S.-W., Kaimoyo, E., Kumar, D., Mosher, S., and Klessig, D.F. (2007). Methyl Salicylate Is a Critical Mobile Signal for Plant Systemic Acquired Resistance. Science *318*, 113-116.

Pearce, L.R., Komander, D., and Alessi, D.R. (2010). The nuts and bolts of AGC protein kinases. Nat Rev Mol Cell Biol *11*, 9-22.

Perrin, R.M., Fedorova, N.D., Bok, J.W., Cramer Jr, R.A., Wortman, J.R., Kim, H.S., Nierman, W.C., and Keller, N.P. (2007). Transcriptional regulation of chemical diversity in Aspergillus fumigatus by LaeA. PLoS pathogens *3*, e50.

Pogány, M., von Rad, U., Grün, S., Dongó, A., Pintye, A., Simoneau, P., Bahnweg, G., Kiss, L., Barna, B., and Durner, J. (2009). Dual Roles of Reactive Oxygen Species and NADPH Oxidase RBOHD in an Arabidopsis-Alternaria Pathosystem. Plant Physiology *151*, 1459-1475.

Posas, F., and Saito, H. (1998). Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. The EMBO Journal *17*, 1385-1394.

Pyne, R., Honig, J., Vaiciunas, J., Koroch, A., Wyenandt, C., Bonos, S., and Simon, J. (2017). A first linkage map and downy mildew resistance QTL discovery for sweet basil (Ocimum basilicum) facilitated by double digestion restriction site associated DNA sequencing (ddRADseq). PLoS ONE *12*, e0184319.

Pyne, R.M., Koroch, A.R., Wyenandt, C.A., and Simon, J.E. (2015). Inheritance of Resistance to Downy Mildew in Sweet Basil. Journal of the American Soceity for Horticultural Science, 396-403.

Qiu, J.L., Zhou, L., Yun, B.W., Nielsen, H.B., Fiil, B.K., Petersen, K., Mackinlay, J., Loake, G.J., Mundy, J., and Morris, P.C. (2008). Arabidopsis mitogen-activated protein kinase kinases MKK1 and MKK2 have overlapping functions in defense signaling mediated by MEKK1, MPK4, and MKS1. Plant Physiol *148*, 212-222.

Raid, R.N., Zhang, S. (2015). The importance of Drench Fungicie Treatments for Successful Basil Downy Mildew Control. Basil Downy Mildew Workshop Atlantic City, NJ.

Rastogi, S., Meena, S., Bhattacharya, A., Ghosh, S., Shukla, R.K., Sangwan, N.S., Lal, R.K., Gupta, M.M., Lavania, U.C., Gupta, V., *et al.* (2014). De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. BMC Genomics *15*, 588.

Rispail, N., and Di Pietro, A. (2009). Fusarium oxysporum Ste12 controls invasive growth and virulence downstream of the Fmk1 MAPK cascade. Mol Plant Microbe Interact *22*, 830-839.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research.

Roberts, P.D., Raid, R.N., Harmon, P.F., Jordan, S.A., and Palmateer, A.J. (2009). First Report of Downy Mildew Caused by a Peronospora sp. on Basil in Florida and the United States. Plant Disease *93*, 199-199.

Robinson, L.C., Hubbard, E.J., Graves, P.R., DePaoli-Roach, A.A., Roach, P.J., Kung, C., Haas, D.W., Hagedorn, C.H., Goebl, M., Culbertson, M.R., *et al.* (1992). Yeast casein kinase I homologues: an essential gene pair. Proc Natl Acad Sci U S A *89*, 28-32.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010a). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010b). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*.

Ronald, P.C., Salmeron, J.M., Carland, F.M., and Staskawicz, B.J. (1992). The cloned avirulence gene avrPto induces disease resistance in tomato cultivars containing the Pto resistance gene. J Bacteriol *174*, 1604-1611.

Ronco, L., Rollán, C., Choi, Y., and Shin, H. (2009). Downy mildew of sweet basil (Ocimum basilicum) caused by Peronospora sp. in Argentina. Plant Pathology *58*, 395-395.

Rudd, J.J., Kanyuka, K., Hassani-Pak, K., Derbyshire, M., Andongabo, A., Devonshire, J., Lysenko, A., Saqi, M., Desai, N.M., and Powers, S.J. (2015). Transcriptome and metabolite profiling of the infection cycle of Zymoseptoria tritici on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. Plant physiology *167*, 1158-1185.

Šafránková, I., and Holková, L. (2014). The First Report of Downy Mildew Caused by Peronospora belbahrii on Sweet Basil in Greenhouses in the Czech Republic. Plant Disease *98*, 1579-1579.

Santos, J.L., and Shiozaki, K. (2001). Fungal histidine kinases. Sci STKE *2001*, re1.

Sauer, E., Imseng, S., Maier, T., and Hall, M.N. (2013). Conserved sequence motifs and the structure of the mTOR kinase domain (Portland Press Limited).

Sawano, Y., Miyakawa, T., Yamazaki, H., Tanokura, M., and Hatano, K. (2007). Purification, characterization, and molecular gene cloning of an antifungal protein from Ginkgo biloba seeds. Biol Chem *388*, 273-280.

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F.L.W., and Rep, M. (2013). MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum. BMC Genomics *14*, 1-21.

Seger, R., and Krebs, E.G. (1995). The MAPK signaling cascade. FASEB J *9*, 726-735.

Serrano, M., Wang, B., Aryal, B., Garcion, C., Abou-Mansour, E., Heck, S., Geisler, M., Mauch, F., Nawrath, C., and Métraux, J.-P. (2013). Export of Salicylic Acid from the Chloroplast Requires the Multidrug and Toxin Extrusion-Like Transporter EDS5. Plant Physiology *162*, 1815-1821.

Shapiro, A.D., and Zhang, C. (2001). The role of NDR1 in avirulence gene-directed signaling and control of programmed cell death in Arabidopsis. Plant Physiol *127*, 1089-1101.

Sharifpoor, S., van Dyk, D., Costanzo, M., Baryshnikova, A., Friesen, H., Douglas, A.C., Youn, J.-Y., VanderSluis, B., Myers, C.L., Papp, B., *et al.* (2012). Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. Genome Res *22*, 791-801.

Shertz, C.A., Bastidas, R.J., Li, W., Heitman, J., and Cardenas, M.E. (2010). Conservation, duplication, and loss of the Tor signaling pathway in the fungal kingdom. BMC Genomics *11*, 1-14.

Shiu, S.-H., and Bleecker, A.B. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proceedings of the National Academy of Sciences *98*, 10763-10768.

Shor, E., and Chauhan, N. (2015). A Case for Two-Component Signaling Systems As Antifungal Drug Targets. PLOS Pathogens *11*, e1004632.

Shwab, E.K., and Keller, N.P. (2008). Regulation of secondary metabolite production in filamentous ascomycetes. Mycological Research *112*, 225-230.

Siebel, C.W., Feng, L., Guthrie, C., and Fu, X.D. (1999). Conservation in budding yeast of a kinase specific for SR splicing factors. Proc Natl Acad Sci U S A *96*, 5440-5445.

Smith, D.A., Morgan, B.A., Quinn, J., and Sullivan, D. (2010). Stress signalling to fungal stress-activated protein kinase pathways. Fems Microbiology Letters *306*, 1-8.

Sobko, A. (2006). Systems biology of AGC kinases in fungi. Sci STKE *2006*, re9-re9.

Stajich, J.E., Wilke, S.K., Ahren, D., Au, C.H., Birren, B.W., Borodovsky, M., Burns, C., Canback, B., Casselton, L.A., Cheng, C.K., *et al.* (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom Coprinopsis cinerea (Coprinus cinereus). Proc Natl Acad Sci U S A *107*, 11889-11894.

Tamuli, R., Kumar, R., and Deka, R. (2011). Cellular roles of neuronal calcium sensor-1 and calcium/calmodulin-dependent kinases in fungi. Journal of Basic Microbiology *51*, 120-128.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol *30*.

Tan, G., Liu, K., Kang, J., Xu, K., Zhang, Y., Hu, L., Zhang, J., and Li, C. (2015). Transcriptome analysis of the compatible interaction of tomato with Verticillium dahliae using RNA-sequencing. Frontiers in plant science *6*, 428.

Taylor, B.L., and Zhulin, I.B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev *63*, 479-506.

Taylor, S.S., and McKeon, F. (1997). Kinetochore localization of murine Bub1 is required for normal mitotic timing and checkpoint response to spindle damage. Cell *89*, 727-735.

Teichert, S., Wottawa, M., Schönig, B., and Tudzynski, B. (2006). Role of the Fusarium fujikuroi TOR kinase in nitrogen regulation and secondary metabolism. Eukaryotic Cell *5*, 1807-1819.

Thatcher, L.F., Gardiner, D.M., Kazan, K., and Manners, J.M. (2012). A Highly Conserved Effector in Fusarium oxysporum Is Required for Full Virulence on Arabidopsis. Molecular Plant-Microbe Interactions *25*, 180-190.

Thomason, P., and Kay, R. (2000). Eukaryotic signal transduction via histidine-aspartate phosphorelay. Journal of cell science *113*, 3141-3150.

Tieman, D., Zeigler, M., Schmelz, E., Taylor, M.G., Rushing, S., Jones, J.B., and Klee, H.J. (2010). Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. Plant J *62*, 113-123.

Torres, M.A., Dangl, J.L., and Jones, J.D.G. (2002). Arabidopsis gp91phox homologues AtrbohD and AtrbohF are required for accumulation of reactive oxygen intermediates in the plant defense response. Proceedings of the National Academy of Sciences *99*, 517-522.

Turrà, D., Segorbe, D., and Di Pietro, A. (2014). Protein kinases in plant-pathogenic fungi: conserved regulators of infection. Annual Review of Phytopathology *52*, 267-288.

Uppalapati, S.R., Ishiga, Y., Doraiswamy, V., Bedair, M., Mittal, S., Chen, J., Nakashima, J., Tang, Y., Tadege, M., and Ratet, P. (2012). Loss of abaxial leaf epicuticular wax in Medicago truncatula irg1/palm1 mutants results in reduced spore differentiation of anthracnose and nonhost rust pathogens. Plant Cell *24*, 353-370.

USDA (2014). National Agricultural Statistics Service.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016). Effector profiles distinguish formae speciales of Fusarium oxysporum. Environmental Microbiology, n/a-n/a.

van Esse, H.P., Bolton, M.D., Stergiopoulos, I., de Wit, P.J., and Thomma, B.P. (2007). The chitin-binding Cladosporium fulvum effector protein Avr4 is a virulence factor. Mol Plant Microbe Interact *20*, 1092-1101.

Veneault-Fourrey, C., and Talbot, N.J. (2007). Autophagic cell death and its importance for fungal developmental biology and pathogenesis. Autophagy *3*, 126-127.

Vieira, R.F., and Simon, J.E. (2006). Chemical characterization of basil (Ocimum spp.) based on volatile oils. Flavour and Fragrance Journal *21*, 214-221.

Vlaardingerbroek, I., Beerens, B., Schmidt, S.M., Cornelissen, B.J.C., and Rep, M. (2016). Dispensable chromosomes in Fusarium oxysporum f.sp lycopersici. Molecular Plant Pathology, n/a-n/a.

von Röpenack, E., Parr, A., and Schulze-Lefert, P. (1998). Structural analyses and dynamics of soluble and cell wall-bound phenolics in a broad spectrum resistance to the powdery mildew fungus in barley. J Biol Chem *273*, 9013-9022.

Wang, C., Zhang, S., Hou, R., Zhao, Z., Zheng, Q., Xu, Q., Zheng, D., Wang, G., Liu, H., Gao, X., *et al.* (2011). Functional analysis of the kinome of the wheat scab fungus Fusarium graminearum. PLoS Pathog *7*, e1002460.

Wawra, S., Trusch, F., Matena, A., Apostolakis, K., Linne, U., Zhukov, I., Stanek, J., Koźmiński, W., Davidson, I., Secombes, C.J., *et al.* (2017). The RxLR Motif of the Host Targeting Effector AVR3a of Phytophthora infestans Is Cleaved before Secretion. Plant Cell *29*, 1184-1195.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., Armstrong, M.R., Grouffaud, S., van West, P., Chapman, S., *et al.* (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature *450*, 115-118.

Widmann, C., Gibson, S., Jarpe, M.B., and Johnson, G.L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiological Reviews *79*, 143-180.

Wildermuth, M.C., Dewdney, J., Wu, G., and Ausubel, F.M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature *414*, 562-565.

Wyenandt, C.A., Simon, J.E., Pyne, R.M., Homa, K., McGrath, M.T., Zhang, S., Raid, R.N., Ma, L.J., Wick, R., Guo, L., *et al.* (2015). Basil Downy Mildew (Peronospora belbahrii): Discoveries and Challenges Relative to Its Control. Phytopathology *105*, 885-894.

Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., Li, Y., Tang, X., Zhu, L., and Chai, J. (2008). Pseudomonas syringae effector AvrPto blocks innate immunity by targeting receptor kinases. Current biology *18*, 74-80.

Xu, J.-R., and Hamer, J.E. (1996). MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus Magnaporthe grisea. Genes Dev *10*, 2696-2706.

Xu, J., and Zhang, S. (2015). Mitogen-activated protein kinase cascades in signaling plant growth and development. Trends in Plant Science *20*, 56-64.

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L., and Zhang, X. (2011). Lignin metabolism has a central role in the resistance of cotton to the wilt fungus Verticillium dahliae as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. Journal of experimental botany *62*, 5607-5621.

Yu, E.Y., Lee, J.H., Kang, W.H., Park, Y.H., Kim, L., and Park, H.M. (2013). Fission yeast LAMMER kinase Lkh1 regulates the cell cycle by phosphorylating the CDK-inhibitor Rum1. Biochem Biophys Res Commun *432*, 80-85.

Zhang, L., Tian, L.H., Zhao, J.F., Song, Y., Zhang, C.J., and Guo, Y. (2009). Identification of an apoplastic protein involved in the initial phase of salt stress response in rice root by two-dimensional electrophoresis. Plant Physiol *149*, 916-928.

Zhang, S., and Xu, J.-R. (2014). Effectors and Effector Delivery in Magnaporthe oryzae. PLOS Pathogens *10*, e1003826.

Zhang, Y., Xu, S., Ding, P., Wang, D., Cheng, Y.T., He, J., Gao, M., Xu, F., Li, Y., Zhu, Z., *et al.* (2010). Control of salicylic acid synthesis and systemic acquired resistance by two members of a plant-specific family of transcription factors. Proceedings of the National Academy of Sciences *107*, 18220-18225.

Zhao, X., Kim, Y., Park, G., and Xu, J.-R. (2005). A mitogen-activated protein kinase cascade regulating infection-related morphogenesis in Magnaporthe grisea. Plant Cell *17*, 1317-1329.

Zhu, Q.-H., Stephen, S., Kazan, K., Jin, G., Fan, L., Taylor, J., Dennis, E.S., Helliwell, C.A., and Wang, M.-B. (2013). Characterization of the defense transcriptome responsive to Fusarium oxysporum-infection in Arabidopsis using RNA-seq. Gene *512*, 259-266.