12-11-2015

Towards a Genome Sequence of the Brown Spot Needle Blight Pathogen (Mycosphaerella Dearnessii) Infecting Longleaf Pine

Benjamin Douglas Bartlett

Towards a genome sequence of the brown spot needle blight pathogen (*Mycosphaerella dearnessii*) infecting longleaf pine

By

Benjamin Douglas Bartlett

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Agronomy
in the Department of Plant and Soil Sciences

Mississippi State, Mississippi

December 2015

Towards a genome sequence of the brown spot needle

blight pathogen (*Mycosphaerella dearnessii*)

infecting longleaf pine

By

Benjamin Douglas Bartlett

Approved:

_____

Daniel G. Peterson
(Major Professor)

_____

Susan V. Diehl
(Committee Member)

_____

C. Dana Nelson
(Committee Member)

_____

Zenaida Magbanua
(Committee Member)

_____

Michael S. Cox
(Graduate Coordinator)

_____

J. Mike Phillips
Department Head

_____

George Hopper
Dean
College of Agriculture and Life Sciences

Name: Benjamin Douglas Bartlett

Date of Degree: December 11, 2015

Institution: Mississippi State University

Major Field: Agronomy

Major Professor: Daniel G. Peterson

Title of Study: Towards a genome sequence of the brown spot needle blight pathogen (*Mycosphaerella dearnessii*) infecting longleaf pine

Pages in Study: 105

Candidate for Degree of Master of Science

A major disease damaging seedlings of *Pinus palustris* is caused by the fungus *Mycosphaerella dearnessii*. Population structure of this pathogen was studied in a population in Mississippi. High genetic diversity (0.65) was measured using microsatellite markers and coincides with the high number of vegetative compatibly groups observed. A 30 Mb genome sequence for a single isolate of *M. dearnessii* was assembled, representing 65% of the estimated genome size. Nearly all (93%) of the core set of genes present in eukaryotes were detected from a total of 10,996 predicted genes. A total of 853 enzymatic associations were identified along with several genes homologous to pathogenic genes in other fungal pathogens. These results provide insights into the infection process and host-pathogen interactions. Further investigating this pathosystem will lead to effective disease management strategies.

DEDICATION

I would like to dedicate this work to my family, Richard, Sally, & Alice.

ACKNOWLEDGEMENTS

Clearly I could not have accomplished this alone. Many people pushed me to finish my thesis after several years. Kurt Shoemaker and Jim Roberds have contributed more than anyone for the progress I made on this project. Daniel Peterson and Dana Nelson provided me with the resources that allowed me to achieve this goal. I would also like to thank Susan Diehl and Zeniaida Magbanua for remaining on as committee members.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE ECOSYSTEM OF *PINUS PALUSTRIS*

Rationale

Every species relies on natural selection to survive and adapt to their environment. Humans have the additional ability to engineer, which provides solutions to problems on a much shorter time scale; experimental time versus ecological or geological time (Norton, 1995). Humans are unusual, in that we are able to assert a degree of control over our environment, health, and resources. We can clear land for agriculture that allows many people to embark on alternative fruitful endeavors. We can eradicate many diseases allowing an individual's contribution to society to be prolonged. We can mine the earth for energy reserves that allow us to mimic daylight conditions and extend productive operation hours. Most importantly, though, we are able to predict changes, recognize the impact of our actions, determine what risks we are willing to accept, and pass knowledge on to future generations. These decisions we make, based upon reasonable predictions, are influenced by the moral framework of a society.

One of the most influential premises for moral philosophy in modern western society comes from Immanuel Kant. His Categorical Imperative – "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" - creates the groundwork on which society characterizes the treatment of fellow peoples, and by extension how the environment must be treated. It is important to

distinguish this framework regarding environmental preservation in two ways. First, integral ecological habitats must be preserved for future generations to appreciate. Previous generations were unknowingly altering ecosystems irrevocably. Contemporary generations recognize that their behavior can depreciate the quality of life for future generations and limit their ability to understand the interactions of nature. Second, (and I think more importantly) natural habitats must be conserved for their own intrinsic value. These two notions are codified in *Restoring Nature* with "a desired future condition managed by humans that produces socially valued goods and services" and "a natural, authentic, or pristine condition existing prior to or with minimum human modification", respectively (Hull & Robertson, 2000).

Life on earth has endured many hardships and will inevitably endure more. *Homo sapiens* are a relatively recent species, yet they have arguably been the most influential in shaping the environment. If life on earth is to persist, the conditions that foster life should be protected. Human society should respect the natural forces that have permitted all life on earth to prevail, including humans themselves. Humans do not have the competence to define what species are fit to survive and which are not. It is not our place to create unnecessary hardships for species that may not directly benefit us. Every species has a right to exist. Evolutionary processes have selected all living organisms to be suited for their environment. We should continue to reward those organisms by preserving the environment they evolved within, while emphasizing as little interference as possible regarding that environment. Kant may have designed his morality for the rational being, however the total impact of human activities was unrecognized at the time. While one may be able to justify that preserving ecologically significant habitats

for future generations as a substantial reason for maintaining a habitat; the primary justification rests on a higher moral plane than benefit for human society.

Longleaf Pine

Background

An important legacy of the human species will be determined by how we are able to manage our environment in a harmonious manner. Key conditions that will influence this legacy will include ecosystem reparation, sustainable forestry practices, maintaining species diversity, and watershed management. One of the most devastated forest ecosystems resides in the Southeastern United States; specifically, the longleaf pine ecosystem. Prior to European settlement, longleaf pine (*Pinus palustris*) spanned an estimated 37 million hectares. It was the dominant tree species on 30 million hectares and existed in mixed-species forests on an additional 7 million hectares (Frost, 1993). This established ecosystem extended from southeast Virginia traversing the South Atlantic Coastal Plain (North Carolina, South Carolina, Georgia) on into the Gulf Coast region (Florida, Alabama, Mississippi, Louisiana) reaching as far west as East Texas (Frost, 1993). Acreage now devoted to growing longleaf pine is fragmented and has diminished to less than 3% of its historic range; currently, 1.9 million hectares (Oswalt *et al.*, 2012; Ware *et al.,* 1993). Old-growth forests of longleaf pine are all but absent from the southeastern United States, with only 5095 hectares remaining (Varner & Kush, 2004).

It is instructive to compare the outcome from the decline of another vast ecosystem to the state of the longleaf pine ecosystem. Wetland degradation in the US progressed over roughly the same time period as the decline of the longleaf pine forests. Wetland acreage has been reduced to 53% of its historic range; at present 44.6 million

hectares, down from 89 million hectares before European settlement (Dahl, 1990; Dahl, 2011). Such dramatic losses stimulated legislation designed to protect against further devastation. Conservation programs established as a result of legislation, such as the Wetland Conservation Provisions ("Swampbuster" 1985), North American Wetlands Conservation Act (1989), and the Wetlands Reserve Program (1993), have been effective in controlling loss of this vital ecosystem. With the passage of these legislative measures, society has set a precedent to preserve and maintain ecosystems that are vulnerable to destruction. Considering the relative extent of depletion, compared to wetlands, a similar level of commitment is appropriate to support a substantial increase in land area dedicated to longleaf pine as the characteristic species of the ecosystem. As a pragmatic matter, wetland conservation centers on preserving North American waterfowl breeding grounds. These migratory birds also depend on the uplands for food and shelter. It would be unwise to preserve breeding habitat while ignoring habitat responsible for food and shelter.

The decline of the longleaf forest type has been attributed to a number of causes, beginning with the arrival of European settlers, the advent of steam engines, fire suppression, and human preference for fast-growing loblolly pine (*Pinus taeda*). European settlers were the first to exploit longleaf pine resources in such a way that kept the species from adequately regenerating itself. These settlers were motivated to produce naval stores (pitch, tar, rosin, and turpentine) which were essential commodities for ships, wagons, and lanterns. The production value of these commodities were enough to give North Carolina its nickname; the 'Tar Heel' state. Much of the forests were spared due to the requirement of nearby navigable waterways for efficient transportation of logs to the

mills. However, the invention of the steam engine and subsequent spread of railroads allowed commercial logging of forests previously protected by geographical restraints. The desolation of virgin forests was amplified by longleaf's inability to regenerate sufficiently. Longleaf is intolerant to competition for resources and seed production is capricious from year to year (Palik, *et al.,* 1997; Croker & Boyer, 1975). These attributes make regeneration problematic after clearing a forest. Hogs were one of the first species to be brought from the old world and quickly reached carrying capacity (Frost, 1993). Grazing animals, especially imported swine, fed on longleaf seeds and seedlings. The nutritious and vulnerable longleaf seedlings provided hogs with year round forage, particularly during winter months. Finally, the adoption of fire suppression practices in the 20th century led to the establishment of loblolly pine, slash pine, and various hardwood species to the detriment of longleaf. More than 13 million hectares of loblolly pine surpass all other southern pine species in acreage devoted to timber (Schultz, 1997).

Fire

Longleaf pine has evolved within a natural disturbance regime of intermittent fire, encompassing large swaths of the landscape. This pattern of disturbance has culminated in the development of a particular seedling adaptation for longleaf pine to endure spatiotemporally, called the "grass stage." The apical meristem is protected from fire injury by dense needle foliage that resembles grass in appearance (Brockway *et al.,* 2006). The longleaf seedling remains adjacent to the soil for 3-5 years with little above ground height growth. Height initiation creates a vulnerability to fire damage for the terminal bud (Maple, 1975; Bruce, 1951). However, once the tree reaches a height above

two meters, mortality is drastically reduced (Maple, 1975; Brockway *et al.,* 2006; Bruce, 1951).

The primary structure of the longleaf pine ecosystem consists of a grass savanna understory with longleaf pine as the prevailing tree species (Varner, *et al.* 2003a). Like all savannas, fire plays a central role in sustaining the openness that defines a savanna. For centuries 'pine savannas' were maintained by frequent low-intensity fires that excluded mid-story tree species and kept fire-tolerant ground cover dominant. Fire has a unique role in shaping the landscape as the only environmental force (glacial recession, tornado, earthquake, volcano, flood, etc…) that typically relies on organisms (living or dead) as a fuel. Vegetative abundance and variety regulates fire intensity; likewise, fire determines species composition (Varner, *et al.* 2003b). This unique relationship preserves the longleaf pine ecosystem as an apex community.

Fire sustains all of the major influences that contribute to the health of the longleaf pine ecosystem (Hunter, 1990; Pyne *et al.,* 1996). It creates conditions favorable to perennial grasses and prevents hardwood succession (Brockway *et al*. 2006). Moreover, it prepares a suitable seedbed by exposing mineral soil, increasing soil organic matter, and releasing nutrients; these actions result in improved seed germination (Gonzalez-Perez *et al*. 2004; Smith *et al.,* 1997). It also removes forest debris (including leaf litter and downed trees damaged by wind, disease, or insects) that would otherwise initiate more intense wildfires with greater destructive consequences (Smith *et al.*, 1997). Additionally, it preserves an open canopy by pruning low limbs and limiting proliferation of short woody perennials (Provencher *et al*., 2001; Engstrom *et al.,* 1984; Russell *et al.,* 1999; Litt *et al.,* 2001). An open canopy increases solar radiation which, in turn,

regulates the thermal and moisture content of the soil (Smith *et al.,* 1997). In these ways, fire is responsible for maintaining the structure and function of the forest.

Current practices in longleaf pine stand management include herbicide application, mechanical thinning, and prescribed fire. These techniques are employed to mimic the natural disturbance regime that fire provided. However, prescribed fire is less favorable to land managers, due to the hazards associated with smoke obstruction, high fuel loads, and the unpredictable nature of fire control (Smith *et al.,* 1997; Brennan *et al.,* 1998; Brennan & Hermann, 1994; Varner *et al.* 2005; Ottmar *et al.*, 1996). The fragmented condition of the longleaf ecosystem often puts it in close proximity to residential homes, which further complicates use of prescribed fire management.

Diversity

The longleaf pine ecosystem is one of the most species diverse ecosystems in the temperate zones (Peet & Allard, 1993). This diversity can be attributed to the diverse ecoregions that constitute the large historic range of longleaf. The longleaf pine ecosystem has been divided into six ecoregions that differ with regard to soil texture, moisture content, climate, and physiography. These are the Atlantic Coastal Plain, Fall-line Sand-hills, Southern Coastal Plain, Eastern Gulf Coastal Plain, Piedmont, and Montane Uplands (Peet, 2006). The vast geographical range of the longleaf ecosystem has revealed 135 different plant associations, 72 understory plant species, and 400+ associated species (Peet, 2006; Varner *et al.*, 2003b; Brennan *et al.*, 1998), while retaining a consistent bi-layered park-like structure throughout its territory.

The conventional structure of this forest ecosystem consists primarily of herbaceous ground cover, most importantly graminoids, including forbs and a canopy of

widely spaced uneven-aged longleaf pines (Oswalt *et al*., 2012, Varner, *et al.* 2003a). Much of the biodiversity is retained in the understory vegetation, herpetofauna, and arthropods in the ecosystem. Reports of species richness can exceed 40 species of plants per square meter (Walker & Peet, 1983) and 170 per square kilometer (Peet, Carr & Gramling, 2006). Herpetofauna within the longleaf ecosystem includes 73 species of amphibians and 95 species of reptiles (Guyer & Bailey 1993). Of particular interest, the Endangered Species Act lists several species that rely heavily on habitat that only longleaf pine provides; specifically, gopher tortoise (*Gopherus polyphemus*), gopher frog (*Rana sevosa*), Mississippi sandhill crane (*Grus canadensis pulla*), red-cockaded woodpecker (*Picoides borealis*), and the eastern indigo snake (*Drymarchon corais couperi*). The fate of the flora and fauna that have evolved within the longleaf pine ecosystem is inextricably linked to the longleaf pine itself.

Brown Spot Needle Blight

A severe needle disease of longleaf pine known as brown spot needle blight occurs throughout the southeastern United States. This disease results from infection by the fungus *Mycosphaerella dearnessii* and can disrupt development and vigor of longleaf seedlings. Brown spot needle blight can slow seedling growth, delay stand development, inflate fire damage, and cause an increase in seedling mortality (Siggers, 1934; Maple, 1975). The symptoms associated with the disease first appear as small, circular spots of yellowish to light grayish-green lesions (Wolf *et al.,* 1941, Hedgcock, 1929). Within a few days, the lesions become necrotic, turn brown, and encircle the needle in a narrow band. Typically, the brown necrotic spots are surrounded by a yellow halo and occur as multiple infections throughout the needle length. Eventually, the girdling of the needle

kills the tissue distal to the infection. Necrotic needles can remain attached to the seedling for more than a year and provide a source of inoculum. Each successive season's growth is commonly infected, unless measures are taken to reduce inoculum. The consequence of failing to reduce inoculum will retard growth of the seedling and allow the seedling to remain in the "grass stage" for 8 to 10 years longer than normal and/or the seedling will succumb to the disease. Disease fatality is observed mostly before the onset of height initiation and generally noticed within the third successive season with the disease (Wolf *et al.* 1941, Hedgcock, 1929).

The correlation of seedling height and fatality seems to be rooted in the dispersal method of spores. Infection from conidia spores occurs more often than infection from ascospores (Verrall, 1936). Conidia from acervuli erupt from the infected lesions during rainfall. The "grass stage" is therefore the most susceptible life stage of longleaf pine, due to the proximity to rain splash and the close proximity of all needles of the tree. The delayed onset of stem elongation characteristic of the "grass stage" increases the time and needle proximity required for thorough establishment of infection. Once height initiation begins, the fungal conidia spores are effectively out of range of the host tissue. Conidia spores can be released throughout the year, yet the greatest quantities are present beginning in April and peak in August (Kais, 1974). Ascospores are wind disseminated and are also released in the greatest quantities in April through August (Kais, 1974). These periods of greatest spore dispersal correlates with the growing season for newly emerging needles, which are especially vulnerable to infection.

Problem

America's Longleaf Restoration Initiative is currently coordinating an effort involving a number of agencies to increase the area occupied by longleaf pine forests to 8 million acres (8% of the original range). The longleaf pine ecosystem has an advantage for restoration efforts that will ease the transition back to historic ecological norms. Longleaf pine has not been wholly substituted for other agricultural systems, but rather has been selectively removed and replaced by other conifers (loblolly and slash). Reversion back to longleaf pine will not drastically alter the silviculture practices currently being implemented. However, due to the fragmented distribution of longleaf pine forests, natural regeneration will be insufficient to achieve substantial progress. Therefore, planting containerized longleaf pine seedlings will play an essential role in achieving objectives set forth by America's Longleaf Restoration Initiative. Poor nursery stock will undermine the performance of any regeneration program. Nursery operations will therefore become an important tool in providing quality planting material. The most important element to control in a nursery setting is the management of pests and disease. Minimizing costs from pest and diseases will hasten the adoption of longleaf pine forests for land managers.

Widespread incidence of brown-spot needle blight is a potential threat that could hinder progress in a restoration program. While prescribed fire is recommended management practice for minimizing disease incidence, its use is problematic for many land managers (Brennan & Hermann, 1994; Ottmar *et al.*, 1996). The absence of a proper fire regime will result in inadequate management for this needle disease on many sites, suggesting that approaches other than prescribed fire are needed for establishing

longleaf pine stands (Varner *et al.*, 2005). It has been roughly 40 years since there has been any major study of the biology of this fungus. Recent advances in host-pathogen interactions have provided an opportunity to apply molecular approaches to explore the infection process. At the same time, few resources exist for elucidating the mechanisms by which this fungus is able to cause disease. Genome sequencing projects have yielded abundant information for model and non-model organisms. The decreased costs of genome scale investigations have created a means to attain copious data from which to mine biomolecular pathways associated with pathogenicity and virulence. Moreover, approaching the disease from a population standpoint is relevant to any disease management program. Monitoring changes in pest populations can help assess the efficacy of those programs. Pathogens will respond to selective pressures imposed by disease management practices and lead to the pathogen overcoming plant defenses. Prompt realization to fluctuations in pathogen populations can guide continuous recommendations. Obtaining an understanding of the systems involved in the infection process should prove valuable in efforts to select and breed trees resistant to brown-spot needle blight. Here, I describe research conducted to address these shortcomings, namely the development and analysis of wild collected isolates from a local population (Chapter II) of longleaf pine and development and analysis of a genome sequence of two isolates (Chapter III).

CHAPTER II

DIVERISTY OF MYCOSPHAERELLA DEARNESSII IN A SMALL POPULATION IN SOUTH MISSISSIPPI

Introduction

America's Longleaf Restoration Initiative is encouraging the development of 8 million acres of longleaf pine forests within the present decade. Achieving those restoration objectives will rely upon the availability of healthy, disease free planting material. A major disease damaging seedlings of *P. palustris* is caused by the fungus *Mycosphaerella dearnessii*. This fungus produces necrotic lesions on the needles of seedlings and saplings of longleaf pine, resulting in a disease known as brown spot needle blight (Wolf *et al.,* 1941, Hedgcock, 1929). Young trees of this species in the grass stage of development, although protected from fire damage, are particularly susceptible to this disease (Siggers, 1934; Maple, 1975). Dissemination of conidial spores by rain allows recurrent infection of *M. dearnessii* (Verrall, 1936). Severe infection results in delayed tree growth and can result in seedling mortality (Wolf *et al.* 1941, Hedgcock, 1929). Without proper disease management, stand development will be delayed and ultimately impede restoration goals.

Investigations aimed at obtaining information about the behavior of pathogen populations are critical for the development of strategies that are effective in reducing disease incidence. Research focused on gaining an understanding of the biological

properties of *M. dearnessii* has been halted for nearly four decades. Since then, molecular techniques have progressed to become an effective tool for investigating biological questions. Employing these techniques, we studied genetic diversity using microsatellite loci in a local population of *M. dearnessii* infecting longleaf pine seedlings growing in the Harrison Experimental Forest located in Southeast Mississippi. Furthermore, haplotypes were assigned based upon an initial three microsatellite marker system. Haplotype assignment allowed us to evaluate this fungal population for variation within vegetative compatibility groups (VCGs). Successful disease management programs depend on predicting how diseases respond to control measures. Recognizing changes in pathogen population structure will assist in guiding effective disease management programs.

Materials and Methods

Fungal Isolation

Needles from seedlings of *Pinus palustris* were collected in the Harrison Experimental Forest near Saucier, MS on October 15, 2012. Seedlings were in the grass stage of development, meaning that their height was less than four inches and diameter was less than one inch (Wahlenberg, 1946). Needles were harvested from 37 different *P. palustris* seedlings that displayed brown spot needle blight symptoms (Wahlenberg, 1946).

Needles with visible lesions were surface sterilized with a 10% v/v bleach solution, and the lesions were excised with a razor blade. Approximately 30 lesions were excised from each *P. palustris* seedling (~1,110 total). Lesions were embedded in acidic potato dextrose agar (APDA), spaced roughly 1 cm apart with 15 lesions per petri dish.

Lab prepared APDA consisted of these ingredients: dextrose (10 g/L), potato flakes (14 g/L), yeast extract (2 g/L), agar (18 g/L), hydrochloric acid (250 µL/L) and chloramphenicol (12.5 mg/L). All procedures were carried out under a laminar flow cabinet using aseptic techniques. A sterile toothpick was used to capture conidia spores for microscopic identification (400X magnification), and positively identified spores were subsequently transferred to fresh media. Following successful capture of an isolate, half-strength media (except agar, hydrochloric acid, and chloramphenicol were left at full strength) was found to be sufficient to manage cultures. Every 14 days, cultures were examine for visible contamination and then maintained by transferring conidia to fresh media.

Genomic DNA Isolation

Isolates were grown on APDA at 21°C for 21 days prior to DNA extraction. This time period was selected to allow the cultures to grow as much as possible before contaminating fungi became apparent in the cultures based on our prior observations. Only cultures without noticeable contamination were selected for DNA isolation. Genomic DNA was isolated from each of the cultures using a CTAB protocol (Telfer *et al.*, 2013). In brief, conidia from isolates were removed with a sterile toothpick and placed in extraction buffer [0.5 M Tris-HCl (pH 8.0), 0.05 M EDTA, 5 mM polyethylene glycol (PEG), 0.35 mM sorbitol, 1.25 mM spermine, 1.75 mM spermidine, 15 µM bovine serum albumin (BSA), and 10 mM 2-mercaptoethanol]. Liquid nitrogen was used in conjunction with a pestle and mortar to macerate the conidia tissue. For each sample, 750 µL of the tissue mixture was transferred to a 1.5 mL microfuge tube and centrifuged at 8,000 x g for 20 minutes. The supernatant was decanted and the pellet was

resuspended in 200 µL organelle wash buffer [1 M Tris-HCl (pH 8.0), 0.1 M EDTA, 5 mM PEG, 0.175 mM sorbitol, 0.625 mM spermine, 0.875 mM spermidine, 4 µL RNase A (10 mg/mL), and 80 µL sarkosyl (5% w/v)] which then was incubated at room temperature for 30 minutes. The tubes were again centrifuged at 8,000 x g for 2 minutes, before 80 µL CTAB (0.7 M NaCl, 236 µM hexadecyltrimethylammonium bromide), 72 µL 5 M NaCl, and 3 µL proteinase K (20 mg/mL) were added to the samples. The mixtures were incubated at 65°C for 15 minutes. Afterwards, the samples were placed into a -20°C freezer for 5 minutes, and then centrifuged at 8,000 x g for 2 minutes. Each sample was then mixed gently with 400 µL of 24:1 chloroform/isoamyl alcohol, and samples were then centrifuged at 8,000 x g for 20 minutes. The aqueous phase was transferred to a new 1.5 mL microfuge tube and an equal volume of isopropyl was added. Samples were then centrifuged at 8,000 x g for 20 minutes, and afterwards the supernatant was decanted without disturbing the pellet. Pellets were subsequently washed with 70% v/v ethanol and allowed to dry overnight. Each pellet was then resuspended in 100 µL LTE (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0).

Amplification of Internal Transcribed Spacer and Microsatellites

Microsatellite primers designed from *M. dearnessii* were provided by Josef Janoušek, Irene Barnes, and Michael Wingfield (Janoušek *et al.,* 2013). Microsatellites are short repeated DNA sequences that are thought to be selectively neutral. These sequences can vary among individuals, which permits investigations into the genetic diversity within and among populations. Universal ITS1 and ITS4 primers were used in this study to identify fungal species. The ITS region resides between the ribosomal RNA gene sequences in eukaryotic organisms. These genes are highly conserved, which

allows for species identification and phylogenetic analysis. Primer sequences used in this research are listed in Table 2.1. The forward primers were modified with the addition of a fluorescent M_{13} tail sequence (CACGACGTTGTAAAACGAC) to the 5' end.

PCR reactions were performed in 6 µL volumes, consisting of 1X buffer (Kleartaq Buffer B, KBiosciences), 1.8 mM $MgCl_2$, 66 µM dNTPs, 0.08 µM of the forward primer, 0.32 µM of reverse primer (Integrated DNA Technologies) and M_{13}-dye, 0.5 U Taq polymerase (Kleartaq, KBiosciences), and 2 µL of genomic DNA. PCR cycling conditions used for both the microsatellite primers and ITS primers were as follows; denaturation at 94 °C for 2 minutes, 35 cycles of 94 °C for 30 seconds, 48 – 60 °C (annealing temps – see below) for 45 seconds, and 72 °C for 60 seconds, followed by final extension at 72 °C for 15 minutes. The annealing temperature for a particular primer was based on values in Janoušek *et al.,* (2013). An annealing temperature of 51 °C was used for the ITS primers. Capillary electrophoresis was carried out on an ABI 3730xl DNA Analyzer (Applied Biosystems) to evaluate the PCR products. GeneMapper v3.7 (Applied Biosystems) performed the fragment analysis of the PCR reactions.

Three isolates were chosen for ITS verification (Figure 2.1) by sequence analysis. DNA from the ITS region was amplified by PCR and then sequenced using an ABI 3730xl DNA Analyzer (Applied Biosystems). Sequence data was aligned to the NCBI non-redundant database using the BLAST algorithm to confirm species identification. Fragment size analysis from the ITS region for the three isolates were compared to all other isolates to determine if all isolates shared the same amplicon length.

All isolates were initially examined for allelic differences at three microsatellite marker loci (MD1, MD2, & MD9). These three loci were obtained before publication

and facilitated the grouping of isolates into haplotype groups. Subsequently, all isolates were genotyped for eight additional microsatellite loci and genotypic differences were calculated for the eleven total loci. Gene diversity (Nei, 1973) was estimated for each microsatellite locus using PopGen 1.31 (http://www.ualberta.ca/~fyeh/popgene.html), as well as unbiased gene diversity (DeGiorgio & Rosenberg, 2008). Values obtained along with observed allele number and allele range were compared to those previously reported from another sample of infections collected from the same population (Janoušek *et al.,* 2013). In addition, estimates were obtained for the effective number of alleles (Kimura & Crow, 1964) at each locus as well as Shannon's information index (Lewontin, 1972).

Vegetative Incompatibility

Vegetative incompatibility was investigated for seventeen unique haplotypes identified from preliminary analysis of the microsatellite markers MD1, MD2, and MD9. A sample isolate representing each haplotype was selected for testing vegetative incompatibility against every other haplotype. Each pairing was tested in three replicates and scored as being either compatible or incompatible. To form the pairings, conidia from one haplotype were placed on APDA media at a distance of 1 cm from that of a second haplotype and allowed to grow at 21°C for 14 days prior to scoring. Complete merging of mycelia into a confluent mass was the criterion used to indicate that pairings were compatible (Figure 2.2). Incompatibility was interpreted to have occurred when a barrage interaction was observed (Figure 2.3).

Scoring of the pairings took place in two stages, first by direct observation of the cultures and then afterwards from pictures that could be compared and contrasted much easier. Tallies were taken during each scoring phase, so that each pairing was scored 6

times (i.e. – two tallies for three replicates). Any ambiguities that needed to be settled were resolved by totaling the results for the individual tallies. If the majority (>3) of the tallies were scored as compatible, the pairing was interpreted to be compatible. If not, the pairing was deemed incompatible.

Mating Types

Mating type primer sequences were obtained from Josef Janoušek, Irene Barnes, and Michael Wingfield (Janoušek *et al.,* 2013). The identical M_{13} tag was added to the 5' forward primer sequence and is listed in Table 2.1. Several attempts were made to amplify the mating type gene sequences in our samples. The first attempt was conducted along with the microsatellite primers with the same PCR conditions; annealing temperature 56°C for MAT1-1 and 58°C for MAT1-2. Analysis of the PCR product was carried out by slab electrophoresis on a 2% w/v agarose gel and UV visualization. Amplification of the mating type genes were not detected in our analyses. Since the primers were slightly altered with the addition of the M_{13} tail sequence, the PCR conditions were altered to determine the optimal annealing temperature for the primers. A temperature gradient (52°C - 62°C) was applied during the annealing cycle for a subset of the samples; A2_MdHEF, A3_MdHEF, B1_MdHEF, B5_MdHEF, B7_MdHEF, D2_MdHEF, D4_MdHEF, D5_MdHEF. In addition, cycling conditions were strictly followed as reported in Janoušek *et al.* (2013). After optimal annealing temperatures were identified, another attempt was made to amplify the mating type gene sequences for the entire population. After that attempt failed, a final attempt was made with new reagents and additional positive controls. DNA from *Cronartium fusiforme* was included as another positive control for amplifying the ITS gene sequence with the ITS primers

described previously. Microsatellite primers known to consistently amplify *Cronartium fusiforme* DNA were included in the final attempt.

Results

Fungal Isolation

Pinus palustris needles were collected that exhibited symptoms consistent with the disease brown-spot needle blight (Wolf *et al.,* 1941, Hedgcock, 1929; Figure 2.4). Of the approximately 1,110 lesions excised, only 41 lesions produced isolates for further experiments. Of those 41 isolates, 22 shared a parent seedling sample. Isolates originated from two separate collections. The first collection of isolates were taken from seedlings that were located in close proximity (i.e. – within 50 meters) to each other. This collection was labeled alphanumerically (e.g. - A1_MdHEF & A2_MdHEF). The letter prefix designates the seedling it was collected from and the number represents the individual isolate. Therefore, seedling A_MdHEF produced isolates A1_MdHEF and A2_MdHEF. The goal was to capture many isolates from a single seedling source. The second collection of isolates were taken from seedlings located further apart (i.e. – within 8 kilometers). These isolates were labeled numerically (e.g. - 1_MdHEF). The objective was to capture one isolate from each of those seedlings. However for some seedlings, the lesions did not produce any visible spores for isolation and/or the isolation failed, primarily because of contaminating organisms. Of the 30 seedlings designated numerically, 11 did not produce any isolates for further analysis. Seedling sample B_MdHEF generated six isolates from a parent source, the most of any seedling. Seedling samples C_MdHEF, F_MdHEF, G_MdHEF only produced one isolate. Isolates

18-1_MdHEF and 18-2_MdHEF were collected from the same seedling, due to contamination creating the assumption that the first isolate could not be recovered.

Visual inspection of the embedded needles took place daily to ensure acquisition of any fungal isolate. Examination of conidia from embedded lesions required careful observation of the cultures, to avoid transferring contaminating fungi. Acervuli produced from the lesions appeared as a black speck when observed on the lesion and were typically visible within 10-14 days. After 14 days, contaminating fungi frequently dominated the media.

Plastic storage containers were adopted for organizational purposes. Even though all Petri dishes were stored upside down, the plastic storage containers appeared to provide extra protection against contaminating fungi, or at least seemed to prolong the interval before contamination appeared.

Conidia morphology under 400X magnification was consistent with previous descriptions (Wolf *et al.,* 1941; Figure 2.5). Microscopic visualization was determined to be a sufficient method for identification of *Mycosphaerella dearnessii*. Spores produced from the initial transfer were identical to the spores produced afterwards, indicating successful isolation. Growth rate was not measured explicitly; however, cultures roughly grew to a diameter of approximately 18 mm in 14 days and 25 mm in 21 days.

Several distinctive culture morphologies were apparent in the collections. Coloration ranged from light grey (almost white) to dark grey (almost black). Surface texture ranged from isolates with a dry fibrous texture resulting from dense mycelium to a mucoid texture produced by a confluence of conidia (Figure 2.2 & 2.3).

All isolates presented irregular shape with entire margins. Also, all dry fibrous isolates displayed raised colony elevation, however distinct forms were apparent. These forms included irregular undulations across the surface, and more uniform undulations radiating out from the center. All mucoid isolates displayed either convex or umbonate elevation. Dry fibrous isolates eventually produced conidia with recognizable forms. Some isolates produced conidia only in the center of the culture, while other isolates produced conidia that encircled the culture in a ring.

Internal Transcribed Spacer Verification

Three isolates (C1_MdHEF, 7_MdHEF, 19_MdHEF) were selected for DNA sequencing within the internal transcribed spacer (ITS) region. Sequencing yielded 501, 495, and 508 bp amplicons, respectively (Figure 2.1). When these sequences were compared to the NCBI nucleotide database using the BLAST algorithm, two sequences (C1_MdHEF, 7_MdHEF) generated alignments with 100% identity to *Mycosphaerella dearnessii* in the database, while the third (19_MdHEF) aligned with 99.8% identity to *Mycosphaerella dearnessii* with only one nucleotide difference at base pair 45 (Figure 2.1). Sequencing the ITS region verified the identity of the cultures as *Mycosphaerella dearnessii* and confirmed the microscopic evaluation of conidia.

All isolates were investigated by PCR fragment analysis in the ITS region. The amplicon length of the ITS region for all isolates were demonstrated to be 580 bp, as determined by GeneMapper v3.7 (Figure 2.6). Considering that all ITS amplification lengths were identical (specifically C1_MdHEF, 7_MdHEF, and 19_MdHEF), the sequences were interpreted to be equivalent, thus implying all cultures were of the same species.

Three Marker Haplotype

Three microsatellite primer sequences were obtained in advance of the eventual total eleven microsatellite primer sequences eventually acquired. Initial analysis of all 41 isolates revealed 17 haplotypes (haplotype 7 was placed into another group after further examination), based upon the three microsatellite markers (MD1, MD2, & MD9). Several isolates were appointed to the same haplotype group (Table 2.2). Assigning isolates to haplotype groups simplified vegetative compatibility experiments on a much more manageable scale. These microsatellite primers were developed specifically for *Mycosphaerella dearnessii* (Janoušek *et al.,* 2013), and we were able to successfully amplify DNA from all 41 isolates studied in this experiment. Amplification of at least one primer for each isolate provided additional evidence to support the previous visual and molecular identification of the species.

The frequency of haplotypes is illustrated in Figure 2.7; dark shaded bars indicate the isolate was captured from a distinct seedling, while the light shaded bars indicate isolates were captured from a shared seedling source. The haplotype of eight isolates were not determined, chiefly due to failure of the primer MD9 to amplify DNA in that isolate. Haplotypes represented by multiple isolates are largely comprised from isolates originating from a single seedling source. Only five haplotypes are represented by isolates originating from more than one seedling source. It is noteworthy that isolates E1_MdHEF & E2_MdHEF are from the same seedling source, but were assigned to different haplotype groups.

Vegetative Incompatibility Analysis

A vegetative incompatibility (VIC) experiment was conducted between representatives from each of the 17 haplotypes detected from our collection (Table 2.3). Each haplotype was paired against every other haplotype, including itself. Each of the self-pairings were used as a guide for assessing compatibility. Of the 135 pairings evaluated, 80 pairings were identified as compatible (59%) while 55 parings were identified as incompatible (41%).

There seems to be no discernable pattern of compatibility that would allow us to place these isolates into vegetative compatibility groups (VCGs). However, ten haplotypes have a propensity for compatibility in our sample group, five haplotypes have a propensity for incompatibility, and two haplotypes have an equal compatibility/incompatibility ratio (Figure 2.8).

There is the possibility that there exists an intermediate level of compatibility between certain pairings of haplotypes. However, this study did not examine the various degrees of compatibility, but rather made a distinction between compatibility and incompatibility. The limits of human perception also played a role in determining the significance of each pairing. While every effort was made to be unbiased, errors in perception/interpretation may have provided some inaccuracies.

Microsatellite Analysis

All microsatellites, except MD5, were found to be polymorphic (Table 2.4). With a population of 27 isolates, 93 alleles were found across all 11 microsatellites. Gene diversity of the polymorphic loci ranged from 0.3786 (MD2) to 0.93 (MD8) (see Table 2.5). Unbiased gene diversity was similar to gene diversity, as indicated by their

averages (0.6463 and 0.6748, respectively) indicating the sample size used for estimating gene diversity did not result in a large bias. Most allele ranges were consistently higher than values reported in Janoušek *et al.* (2013), but the observed number of alleles were similar with those investigators' findings. The notable exception being allele number and range for markers MD4 and MD11. The results from these two markers are vastly different between this study and Janoušek *et al*. (2013).

Results from marker MD4 in our study seem to be more similar to results from MD11 in Janoušek *et al*. (2013); and *vice versa*. Marker MD4 in this study was observed to have ten alleles, whereas Janoušek *et al*. (2013) reported only three alleles for MD4. Moreover, Janoušek *et al*. (2013) reported twelve alleles for locus MD11, which is closer to the reported number of alleles for MD4 in this study. Furthermore, six alleles were detected for locus MD11 in this study, a number closer to that reported in Janoušek *et al*. (2013) for locus MD4. In the absence of these two differences, allele range averaged 11.82 base pairs larger than from Janoušek *et al*. (2013).

Mating Type Analysis

Problems arose with the mating type primer sequences (MAT1-1 & MAT1-2) published in Janoušek *et al.* (2013). Amplification of the mating type gene proved unsuccessful in five attempts to achieve satisfactory results. In the first attempt at amplifying fungal DNA using the published MAT1-1 and MAT1-2 primer sequences, an unused portion of a PCR plate was used at the same annealing temperature as that specified for the mating type primers, i.e., 56°C and 58°C, respectively. Cycling conditions used for microsatellite amplification were slightly different from those published for mating type primer sequences. The resulting product was analyzed on a

2% w/v agarose gel, but successful amplification was not indicated. The second attempt followed the same cycling conditions as reported in Janoušek *et al.* (2013), but again without success in amplification. Next an experiment was conducted to determine the optimal annealing temperature for each primer. PCR conditions followed the same criteria as the second attempt, except that a gradient of annealing temperatures (52°C to 62°C) was employed. Not every isolate was used in testing the temperature gradient; only a subset of DNA samples were needed, but were still from the same extraction set. Successful amplification was detected for all MAT1-1 samples, but for only two MAT1-2 samples (Figure 2.9). These results indicate that an optimal annealing temperature for the MAT1-1 primer is 60°C, while the MAT1-2 primer is optimal at 53°C. In a third attempt, optimal annealing temperatures for each primer was implemented and DNA from all isolates were included in the trial. Again, the amplification was unsuccessful. New primer stocks were created for the fourth attempt and still amplification of the mating type sequences failed (Figure 2.10). A fifth and final attempt included use of fresh PCR ingredients (Taq, dNTPs, buffer) while using optimal annealing temperatures for each primer. Moreover, additional positive controls were used to verify the efficacy of the PCR ingredients and the isolated DNA. Positive controls on the final attempt indicated that the PCR ingredients and DNA from *M. dearnessii* were adequate for amplification for the ITS primer sequences (Figure 2.11 & 2.12), thus implicating that other factors were responsible for the unsuccessful attempts to amplify the mating type gene sequences.

Discussion

Fungal Isolation and Culture

Isolating *M. dearnessii* from infected tissue and producing a pure culture proceeded with little difficulty. Several attempts were made using previously designed methods (Kais, 1971) without success, due to contaminating fungi that prevailed on the media. The next isolation technique was successful (described in Materials and Methods) and required surface sterilization of the needle tissue, embedding the excised lesion in nutrient rich media, and persistent observation of the culture before contaminating organisms were observed. Even then, not all needle samples yielded fungal spores for transfer. The low success rate of isolation is not wholly troublesome, but needs to be accounted for in future studies that might include a rangewide collection of the fungus. Needle samples were stored in plastic sandwich bags at 4°C for several months and were still able to produce spores for isolation. Long term storage of needle samples is beneficial to researchers for coordinating field collections.

In order to combat rampant contamination, attempts were made to reduce exposure to contaminants and for long term storage of pure cultures. Efforts to produce mycelium in liquid culture, as illustrated in Huang *et al.* (1995), failed. The previous recipe of APDA was made without the agar, to produce liquid media. Conidia were placed in the liquid media at 26°C in a shaking incubator, but no growth was ever noticed. However, no contaminating organisms were noticed either, suggesting our liquid potato dextrose media is inadequate for *M. dearnessii* cultivation. Efforts were also made to curtail the need of transferring the culture to fresh media to maintain the isolate. Again, liquid media was produced, but with the addition of 1/3 volume of

glycerol. An entire fungal culture (2 weeks old) was placed in the liquid media and stored at -80°C. Eight months later the culture was thawed and transferred to fresh APDA media. No growth was observed. Long term storage and effortless preservation of pure cultures is necessary for future projects to build a continuity of data.

The colony morphologies were consistent with those previously reported for *M. dearnessii* (Wolf & Barbour, 1941; Huang *et al.*, 1995). However, in contrast to descriptions of isolates occurring in Huang *et al.,* (1995), cultures appearing as black mucoid in texture were detected in southern Mississippi. This could suggest greater diversity in the southern states than in the northern states, or that the populations have merged in the last two decades. The wide variety of cultural morphology conforms to the broad genetic diversity from microsatellite markers and vegetative incompatibility from our study. Accurate identification was accomplished through microscopic identification. However, given the limited distinguishing morphological characteristics and shared amplification of microsatellite and mating type primers with other species (Huang *et al.*, 1995; Janousek *et al.*, 2013), then sequencing the ITS region is currently the only approach for unambiguous species recognition by molecular methods.

Haplotypes

Haplotypes were assigned based upon a three microsatellite classification system. The primers for markers MD1, MD2, and MD9 were obtained before the complete eleven marker set was published. Hence, the constraint on the three marker classification system. In doing so, many of the isolates that were assigned to the same three marker haplotype group do not share a complete set of identical alleles from all eleven markers (Table 2.4). Nevertheless, many of the isolates from the same seedling sample are

consistently found to belong to the same haplotype group (Table 2.2). Isolates collected from the same seedling sample that did not amplify at all loci were assumed to share identical alleles at the remaining loci.

Most haplotypes (71%) were represented by only one isolate (Figure 2.7), revealing genetic diversity from a small sample size using only three markers. *M. graminicola* isolates studied in France were found to have similar results to our findings (Chartouni *et al.,* 2011). In their population 84% of the isolates belong to unique haplotype groups based upon 6 markers. Moreover, Banke *et al.* (2004) reported 68% of isolates from *M. graminicola* belong to unique multi-locus haplotypes in a worldwide population, while *M. fijiensis* has been reported as having 78% of the haplotypes represented by a single isolate (Hayden *et al.,* 2003). Similarly, 85.5% of haplotypes of *M. musicola* from populations in Brazil were singletons (Peixouto *et al*., 2015). However, there are examples of *Mycosphaerella sp.* that do not match our findings. *M. musicola* populations in Australia were reported as having only 50% of the haplotypes represented by a single genotype (Hayden *et al.,* 2005), and *M. cryptica* only 18% (Milgate *et al.,* 2005). Our data is not as robust as these other experiments; nonetheless, we demonstrate that *M. dearnessii* is highly variable within our population.

Genetic Diversity

Microsatellite markers are a genetic tool for biologist to study population structure. Since the microsatellites are not thought to be under selection, any differences can be attributed to lineage provenance (Schlotterer, 2003; Andolfatto, 2001; Li *et al.*, 2002). Population geneticists use this tool to study genetic diversity within and among populations. Janousek *et al.* (2013) developed polymorphic markers for *M. dearnessii*

using isolates from seedlings collected from the same population of seedlings as used in this study. Predictably, allele number and range were quite similar (Table 2.5). The small differences in allelic range are probably the result of aberrations in techniques between labs. The addition of M_{13} tail sequence could have contributed to the increase in allelic range; while the disparities in observed allele numbers could be explained by scoring techniques and parameters set within the GeneMapper software. The differences in the results of the markers MD4 and MD11 seem to be substantial. This is striking because the isolates were derived from the same population. Furthermore, our results for marker MD4 seem to be more similar to MD11 in Janousek *et al.* (2013), than they do for MD4 from that study, and *vice versa*. Such a large disparity between results from the same population has developed questions regarding errors in reporting accurate microsatellite sequences.

In the absence of marker MD5, genetic diversity (0.71) in our population of *M. dearnessii* displays high variation at ten microsatellite markers. Even in the presence of MD5 genetic diversity (0.65) is relatively high. High genetic diversity appears to be common in this genus. Isolates of *M. graminicola* reveal genetic diversity of 0.70 in French populations (Chartouni *et al.,* 2011). However, global populations reveal a lower gene diversity, near 0.50, for *M. graminicola* (Linde *et al.,* 2002). Likewise, a single population of *M. pinodes* has documented gene diversity of 0.67 (May *et al.*, 2012). In contrast, *M. fijiensis* and *M. musicola* are reported as having moderate gene diversities; 0.31 and 0.22-0.44, respectively (Rivas *et al,* 2004; Hayden *et al.,* 2005). Investigations among various populations of *M. dearnessii* are required for a more complete understanding of the genetic structure. However, we have determined in this

investigation that high genetic diversity can be observed at a very small spatial scale. We conclude that this is unlikely to change at larger spatial scales.

Only one seedling sample yielded infections from two different isolates (E1 & E2_MdHEF), demonstrating that adjacent lesions could originate from separate fungal strains. Still, most isolates from a particular seedling originated from a single source, indicating clonal propagation from conidial infection as the most likely source of inoculum. However, the high genetic diversity demonstrates that sexual recombination is likely occurring frequently. Dissemination of conidia by rain is limited in range and illustrates that further investigations need to be carried out to show outcrossing frequency.

Inspecting population genetics will address concerns including rates of infection, virulence, host specificity, and the ability to overcome plant defenses (Anagnostakis & Waggoner, 1981; Caten & Newton, 2000; Mehrabi *et al.,* 2011). Selective pressures that maintain a high level of diversity indicate the pathogen's potential to overcome host resistance (Morran *et al.,* 2011). Any disease prevention strategy (breeding resistant trees, fungicide application, etc.) will apply selective pressure to the pathogen population, which could inadvertently select for a more virulent pathogen (Thrall & Burdon, 2003; Zhan *et al.*, 2002b). Surveying the pathogen population will remain an important diagnostic tool to measure the effectiveness of programs designed to reduce disease incidence.

Vegetative Incompatibility

Heterokaryon and vegetative incompatibility (VIC) are interchangeable terms that reflect one aspect of filamentous fungal biology (see reviews: Worrall, 1998; Leslie,

1993; Saupe, 2000; Glass & Kaneko, 2003; Bégueret *et al.,* 1994). VIC is another tool biologist use to study population structure. The incompatibility reaction regulates intra-species asexual connections, whereby genetically distinct individuals will reject hyphal fusion. A major function of the VIC response is the recognition of self/non-self for the purposes of anastomoses, heterkaryosis, and transmittance of other cytoplasmic elements (e.g., dsRNA, mitochondria, plasmids). Vegetative compatibility groups (VCGs) are strains within a species that are successful at heterokaryon formation. Typically, any allelic differences, at common loci, detected after cell fusion will result in cell lysis of incompatible strains (Glass & Dementhon, 2006). However, alleles interacting at multiple loci have been reported to control heterokaryon stability (Glass & Kaneko, 2003). In addition, mating type genes have been implicated in determining VCGs (Shiu & Glass, 2000). Pleiotropic and epistatic associations further contribute to the complexity of VIC gene regulation (Loubradou & Turcq, 2000; Saupe, 2000). It is evident that genetic control of heterokaryon interactions vary across species, sometimes within the same genus (Glass & Kaneko, 2003).

The evolutionary benefit of heterokaryotic interactions is not fully understood, nor is it resolved whether genes that control these processes are under selection or are selectively neutral (Saupe, 2000). It may be possible that each species shares little in common, with respect to VIC genes; consequently, evolutionary forces may shape genetic environments independently. Continued research into VCGs among separate species of fungi will add to the literature of any communal genetic elements. The genetic background of VCGs in a limited number of plant pathogenic ascomycetes have been

characterized (Cortesi & Milgroom, 1998; Loubradou & Turcq, 2000, Saupe *et al.,* 2000; Leslie, 1993). No study of VCGs have been described for *M. dearnessii.*

Here, we characterize a small population of *M. dearnessii* on longleaf pine with respect to VIC in the Harrison Experimental Forest. For the sake of simplicity, we assume members of the same VCG share a common progenitor (Caten & Newton, 2000). We were unable to determine any common VCGs among the isolates studied (i.e., each isolate belonging to its own group), presumably, due to the high genetic diversity as detected from the SSR data. The failure to detect discernable VCGs is not uncommon in natural populations of other fungi (Milgroom & Cortesi, 1999; Mylyk, 1979; Cai & Schneider, 2005; Caten & Newton, 2000). The apparent number of incompatible genotypes suggests that *M. dearnessii* could possess many unlinked loci to regulate incompatibility, as seen in other species (Glass & Kaneko, 2003). Genetic regulation of vegetative incompatibility in *M. dearnessii* will need to be examined further. In addition, the frequencies of VCGs need to be investigated to describe any selective pressure acting on the genes controlling incompatibility (Figure 2.8).

All isolates appeared to be self-compatible, even though additional investigations may discover self-incompatible reactions (Cai & Schneider, 2005). The self-compatible reaction was used as a guide for all compatible reactions, while the appearance of a barrage formation was used as a guide to indicate an incompatible reaction (Anagnostakis, 1977; Leslie, 1993). Barrage formation results from cell lysis, and it inhibits two isolates from merging. Investigations into barrage development and the method of cell lysis (apoptosis vs. necrosis vs. autophagy) will rely on discerning the signaling pathway initiated after self/non-self recognition (Saupe, 2000; Glass &

Dementhon, 2006; Saelens *et al.*, 2005). The signaling pathway involved in self/non-self recognition could also be involved in host recognition in the fungus. Understanding the signaling pathway for host recognition will be useful in accelerating host resistance research. The examination of VCGs in plant pathogens will reveal insights into the genetic control of development, cell lysis, pathogenicity, and signaling pathways (Glass & Kuldau, 1992; Glass & Kaneko, 2003; Loubradou & Turcq, 2000; Caten & Newton, 2000; Hillman *et al.*, 1990; Anagnostakis & Waggoner, 1981; Glass & Dementhon, 2006). In this study, we begin the process to understand these aspects of *M. dearnessii* biology.

Mating Types

Sexual reproduction in filamentous fungi involves interaction between mating types, and is analogous to male and female reproductive roles in other eukaryotic kingdoms. However, each mating type is not morphologically distinct from the other in most fungi species (Poggeler, 2001). Mating type systems are, typically, determined genetically by a single locus having two alternate sequences (Kronstad & Staben, 1997; Ni *et al.*, 2012). The two alternate mating type loci are called idiomorphs to illustrate non-homologous sequence similarity. Mating type reproductive strategies are highly varied across the fungal kingdom, and even within phyla. The sexual roles of individuals are not universal, with certain species being classified as self-sterile (heterothallic) and others as self-fertile (homothallic). Heterothallic species require different idiomorphs for sexual reproduction, while homothallic species maintain both idiomorphs within an individual (Coppin *et al.*, 1997).

M. dearnessii is described as a heterothallic fungus and two mating types were reported to occur in equal frequencies in the population investigated in this study (Janousek *et al.,* 2013). Equal mating type frequencies can result from sexual or frequency-dependent selection at the mating type loci (Nieuwenhius & Aanen, 2012; Zhan *et al.*, 2002a). To help explain the high genetic variation found in populations of this fungus, a more comprehensive study into the frequency of mating type idiomorphs will need to be further investigated. We can speculate that high genetic variation is maintained through sexual recombination; however, non-meiotic recombination (parasexual) and horizontal gene transfer cannot be ruled out (Ishikawa, 2012; Mehrabi *et al.*, 2011). The very presence of vegetative compatibility suggests other factors besides sexual recombination may be involved in maintaining genetic diversity within populations of this species. Although other methods of genetic resilience may exist in the absence of suitable environmental conditions for sexual reproduction, outcrossing could be the preferred method of reproductive behavior in natural populations (Wilson *et al.*, 2015).

Furthermore, *M. dearnessii* is reported to have ascospores disseminated by the wind (Kais, 1974), which would support heterothallism (Saupe, 2000). To accomplish this, the spore must be small enough to tolerate dispersal by the wind; creating the expectation that spores from the same ascus would be scattered farther apart and more likely to encounter a spore from another ascus, thus permitting genetic mixing. Homothallism would not require the same expectations, since all characteristics needed for sexual reproduction are maintained within a single individual. Vegetative incompatibility mediated by mating type genes can follow the same rationale as

heterothallism regarding dissemination methods. However, vegetative incompatibility mediated by two mating type genes would be apparent in the VIC analysis for *M. dearnessii*. Because no VCGs were established, it is unlikely that mating type idiomorphs influence VIC. Although, the assumption that only two mating types exist could be erroneous. Detecting other mating type idiomorphs could reveal a connection with mating types and vegetative incompatibility.

Summary and Conclusions

Research introduced in this thesis has begun to acknowledge the genetic diversity of the fungal plant pathogen *M. dearnessii.* Fungal isolates displayed various cultural morphologies and textures. Isolates displaying a mucoid texture were not previously reported in cultures from the southern United States (Huang *et al.*, 1995). We have identified several isolates displaying mucoid texture as well as those with dry, fibrous textures. All isolates were verified by microscopic visualization, ITS size comparison, and microsatellite amplification. We can speculate that the cultural morphologies are predicated on genetic factors. Future research will need to examine if the various isolates are linked to pathogenicity and virulence in the fungus.

Haplotypes were assigned based upon three microsatellite primers. Most haplotypes were represented by one isolate. The haplotype assignments demonstrate genetic diversity within a small sample of the population. Accordingly, when all microsatellite primers were applied to the isolates, a high degree of genetic diversity was established. This finding is in agreement with a sample set from the same population (Janoušek *et al.,* 2013). It is not unusual for fungi in this genus to possess high genetic diversity. Likewise, we suspect the high genetic diversity is maintained by

recombination and sexual outcrossing. However, we were unable to confirm an equal distribution of mating type frequencies that would indicate outcrossing frequency.

High genetic diversity is implicated in the pathogens potential to overcome host plant defenses. It will be essential to survey the fungal pathogen population for changes in genetic frequencies. Any disease management technique will apply selective pressure to the pathogen population. Inspecting changes in pathogen population will help reveal the effectiveness of any disease management program. Vegetative incompatibility experiments provided evidence for many different VCGs established in this population. It is unlikely that the mating type genes are responsible for determining vegetative incompatibility. Understanding the genetic background that controls vegetative incompatibility may expose other genetic factors that govern other cellular processes involved in pathogenicity. Here we describe, for the first time, vegetative incompatibility in *M. dearnessii*.

In order to meet the objectives proposed to increase longleaf pine acreage, it will be essential to reduce disease incidence. Providing resources to better understand of the genetic background of this fungal pathogen will supply knowledge into the infection process. Population genetic studies will rely on surveying changes in the pathogen imposed by disease management programs. We have initiated research to address these issues.

Table 2.1 Mycosphaerella dearnessii Primer Sequences.

Primer ID	Primer Sequence
ITS1.f_m13	CACGACGTTGTAAAACGACTCCGTAGGTGAACCTGCGG
ITS4.r	TCCTCCGCTTATTGATATGC
16S.f	CACGACGTTGTAAAACGACAGAGTTTGATCCTGGCTCAG
16S.r	ACGGCTACCTTGTTACGACTT
MD1.f_m13	CACGACGTTGTAAAACGACGATGACCGCTCTGCTTTG
MD1.r	ACAAACATCCCACCCACA
MD2.f_m13	CACGACGTTGTAAAACGACTTACTCCCGAGACTGGATTG
MD2.r	CCAGACCAAGAACGAAGAAA
MD9.f_m13	CACGACGTTGTAAAACGACGGGAACACACGCTCTTTG
MD9.r	GGGCAAGAAATCCAGGAC
MD4.f_m13	CACGACGTTGTAAAACGACATCCGGATCTTGACCTCCT
MD4.r	CGGTAACTTCTCGCAACCT
MD5.f_m13	CACGACGTTGTAAAACGACCAGGCACAAGGAGAAAGAGA
MD5.r	TCCTCAAGACTCCTCACCTG
MD6.f_m13	CACGACGTTGTAAAACGACAGAGTAAGGGAAAGGAAGAGA
MD6.r	CGGCTACCGTCCTAATCTAAC
MD7.f_m13	CACGACGTTGTAAAACGACCCAACCCGTCAATCAGAA
MD7.r	CGAGAGCGCGAGAAAGTA
MD8.f_m13	CACGACGTTGTAAAACGACCACAGCACGGAAGACACGAG
MD8.r	TCTGTTTCTGAGCGGTAGGAG
MD10.f_m13	CACGACGTTGTAAAACGACCCTACCTACTTCCCTTTATATCTCC
MD10.r	TTAGGACGGTAGCCGTAGAG
MD11.f_m13	CACGACGTTGTAAAACGACGTGGGATGTTTGTTGGGTAG
MD11.r	GCCACCACAGATTGGATAAC
MD12.f_m13	CACGACGTTGTAAAACGACAGTCATAAAGAACCAGGA
MD12.r	GCTATCTAGGCCATTGAA
MAT1-1.f_m13	CACGACGTTGTAAAACGACCGCATTCGCACATCCCTTTGT
MAT1-1.r	ATGACGCCGATGAGTGGTGCG
MAT1-2.f_m13	CACGACGTTGTAAAACGACGCATTCCTGATCTACCGTCT
MAT1-2.r	TTCTTCTCGGATGGCTTGCG

Mycosphaerella dearnessii primer sequences originated from Janoušek *et al.* (2013) and modified with an M13 tail sequence added to the 5' end of the forward primer.

Table 2.2 Designation of 17 haplotypes from 41 isolates of *Mycosphaerella dearnessii.*

Samples	MD1	MD2	MD9	Haplotype
B1_MdHEF	144	114	237	1
B3_MdHEF	144	114	237	1
B4_MdHEF	144	114	237	1
B5_MdHEF	144	114	237	1
B7_MdHEF	144	114	237	1
B2_MdHEF	144	114	-	-
B6_MdHEF	144	114	-	-
21_MdHEF	144	116	237	2
8_MdHEF	144	118	235	3
16_MdHEF	144	118	235	3
20_MdHEF	144	118	235	3
C1_MdHEF	144	118	237	4
E2_MdHEF	144	118	237	4
26_MdHEF	144	118	237	4
7_MdHEF	144	118	239	5
15_MdHEF	144	118	239	5
6_MdHEF	144	118	247	6
4_MdHEF	144	118	252	-
19_MdHEF	144	117	245	8
10_MdHEF	145	118	237	9
25_MdHEF	145	118	241	10
11_MdHEF	145	118	245	-
E1_MdHEF	146	116	237	11
18-1_MdHEF	146	118	251	12
18-2_MdHEF	146	118	251	12
G_MdHEF	147	116	235	13
1_MdHEF	147	118	237	14
A1_MdHEF	147	118	241	15
A2_MdHEF	147	118	-	-
A3_MdHEF	147	118	-	-
D1_MdHEF	147	118	-	-
D2_MdHEF	147	118	-	-
D3_MdHEF	147	118	243	16
D4_MdHEF	147	118	243	16
D5_MdHEF	147	118	243	16
D6_MdHEF	147	118	243	16
F_MdHEF	147	118	243	16
9_MdHEF	147	118	243	16
23_MdHEF	147	118	245	17
13_MdHEF	147	118	245	17
12_MdHEF	147	120	243	18

All 41 isolates were originally examined and arranged into haplotype groups based upon three microsatellite markers. Criteria for placement into separate haplotype groups included a unique combination of all microsatellite markers. Haplotype 7 is missing due to further examination placing it with haplotype 1.

Table 2.3 Compatibility matrix for all haplotype combinations.

Haplotype	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18
1	C	C	C	I	C	I	I	I	C	C	I	C	C	I	I	C	I
2		C	C	I	C	C	C	C	I	C	C	I	C	I	C	C	C
3			C	I	C	I	I	C	C	C	I	I	C	C	C	C	C
4				C	C	I	C	C	C	C	I	I	C	C	C	C	C
5					C	C	C	C	C	C	I	I	C	C	C	C	C
6						C	I	I	I	C	I	I	C	C	C	I	I
8							C	C	C	I	I	I	I	I	C	I	I
9								C	C	C	I	C	C	C	C	C	C
10									C	C	C	C	C	I	I	C	C
11										C	I	I	C	I	I	C	C
12											C	I	C	I	I	I	C
13												C	C	I	C	I	I
14													C	C	C	I	C
15														C	I	I	C
16															C	I	C
17																C	I
18																	C

This table represents the designation of each haplotype pair combination. Compatibility is represented by a C (light). Incompatibility is represented by an I (dark). All self-pairings were designated as compatible.

Table 2.4 Microsatellite results from a representative of all samples.

Haplo (2013)	Sample	Md1 (2013)	Md2 (2013)	Md4	Md5	Md6	Md7	Md8	Md9 (2013)	Md10	Md11	Md12
1	B3_MdHEF	144	114	237	304	152	321	375	237	-	174	135
2	21_MdHEF	144	116	233	304	220	325	347	237	184	181	139
3	8_MdHEF	144	118	239	304	150	331	-	235	198	174	137
3	16_MdHEF	144	118	236	304	-	321	-	235	-	174	141
3	20_MdHEF	144	118	237	304	175	319	356	235	184 191	173	139
4	26_MdHEF	144	118	241	304	156	313	322	237	191	174	137
4	C1_MdHEF	144	118	239	304	156	319	362	237	185	174	135
4	E2_MdHEF	144	118	237	304	152	319	377	237	-	195	141
5	15_MdHEF	144	118	236	304	197	334	-	239	216	174	139
5	7_MdHEF	144	118	241	304	165	325	345	239	185	174	137
6	6_MdHEF	144	118	237	304	168	313	323	247	198	174	137
8	19_MdHEF	144	117	237	304	153	314	-	245	-	188	139
9	10_MdHEF	145	118	231	304	158	317	369 371	237	191	173	143
10	25_MdHEF	145	118	-	304	220	329	-	241	184	174	141
11	E1_MdHEF	146	116	258	304	153	319	379	237	204	174	139
12	18_MdHEF	146	118	239	304	156	321	344	252	191	174	139
13	G_MdHEF	147	116	236	304	153	314	342	235	185	188	141
14	1_MdHEF	147	118	236	304	-	313	352	237	-	173	141
15	A1_MdHEF	147	118	231 237	304	179	346	340	241	185	174	137
16	F_MdHEF	147	118	240	304	159	319	346	243	191	174	139
16	D3_MdHEF	147	118	240	304	159	319	326	243	191	174	139
16	9_MdHEF	147	118	-	304	-	319	-	243	-	173	139
17	13_MdHEF	147	118	249	304	161	319	337	245	185	173	143
17	23_MdHEF	147	118	236	304	192 195	319	340	245	200	174	137
18	12_MdHEF	147	120	237	304	195	314	356	243	185	209	143
-	4_MdHEF	144	118	228 237	304	170	325	344	252	185	174	139
-	11_MdHEF	145	118	239	304	-	327	352	245	210	174	141

Table of all allele sizes for 27 samples. Microsatellites MD1, MD2, and MD9 were performed in the previous experiment to establish haplotypes (Table 2.2). Some of the microsatellites produced two alleles for certain samples, indicated as the second number in the column.

Table 2.5 Comparison of microsatellite results.

		J. Janoušek *et al.* 2013									
Microsatellite	Size of cloned allele	Repeat motif	Allele range	Observed alleles	Gene diversity	Allele range	Observed alleles	Gene diversity	Unbiased gene diversity	Shannon index	Effective number of alleles
MD1	148	(GA)9	149-153	3	0.5212	144-149	4	0.6392	0.6155	1.1550	2.7719
MD2	103	(TC)8	97-105	4	0.2663	114-120	5	0.3786	0.3646	0.8085	1.6093
MD4	169	(CT)14	155-169	3	0.3038	228-258	10	0.8073	0.7689	1.8193	5.1882
MD5	290	(CT)2(TC)8T(TC)3T(TC)4TT(TC)2	286-288	2	0.095	304	1	0.0000	0.0000	0.0000	1.0000
MD6	169	(GA)7AA(GA)9(GAA)13	129-205	19	0.927	150-220	16	0.9174	0.8757	2.6024	12.1000
MD7	298	(CT)12	296-328	11	0.835	313-346	11	0.8313	0.8005	2.0752	5.9268
MD8	337	(GA)20	303-366	17	0.9307	322-379	18	0.9300	0.8835	2.7185	14.2857
MD9	220	(GT)9	218-236	8	0.8213	235-252	8	0.8285	0.7978	1.9095	5.8320
MD10	224	(CT)3(TATAAC)13	162-197	7	0.7622	184-216	8	0.7850	0.7458	1.7737	4.6512
MD11	195	(TGG)3(GGGAAAT)10(GTT)3	209-232	12	0.8638	173-209	6	0.5597	0.5390	1.1626	2.2710
MD12	124	(GA)14	119-133	7	0.7812	135-143	5	0.7462	0.7186	1.4733	3.9405
Average					0.6461			0.6748	0.6463	1.5905	5.4161

Results reported in Janoušek *et al.* (2013) are displayed in light color, while results from this study are displayed in dark color. Results from Janoušek *et al.* were taken from the same population in Mississippi. Results from this study only include samples from the Harrison Experimental Forest, Mississippi USA.

Sample C1_MdHEF - ITS

GGCCGGGGGAGTGATTTTCAAACCCTTGTGAACTACAACTCTGTTGCTTCGGGGGCG
ACCCCGCCGTCTCGGCGGTGGTGCTCCCGGTGGCCATCTATCAAACTCTGCATTACC
TTGCGTCGGAGTCTTATAAAGAATTAAACAAAACTTTCAACAACGGATCTCTTGGTT
CTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTC
AGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCGTGGTATTCCGCGGGGCATG
CCTGTTCGAGCGTCATTTCACCACTCAAGCCTGGCTTGGTATTGGGCGTCGCGGCCT
CCGCGCGCCTCAAAGTCTCCGGCTGAGCAGTCCGTCTCCGAGCGTTGTGACATTTC
GCTAGGGAGTTCGCGTCTGCCGCGGCCGTTAAATCATTAACACCAAAGGTTGACCTC
GGATCAGGTAGGGATACCCGCTGAACTTAAGCATATCAATAAGCGG

Sample 7_MdHEF - ITS

CCGGGGGAGTGATTTTCAAACCCTTGTGAACTACAACTCTGTTGCTTCGGGGGCGAC
CCCGCCGTCTCGGCGGTGGTGCTCCCGGTGGCCATCTATCAAACTCTGCATTACCTT
GCGTCGGAGTCTTATAAAGAATTAAACAAAACTTTCAACAACGGATCTCTTGGTTCT
GGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAG
TGAATCATCGAATCTTTGAACGCACATTGCGCCCCGTGGTATTCCGCGGGGCATGCC
TGTTCGAGCGTCATTTCACCACTCAAGCCTGGCTTGGTATTGGGCGTCGCGGCCTCC
GCGCGCCTCAAAGTCTCCGGCTGAGCAGTCCGTCTCCGAGCGTTGTGACATTTTCGC
TAGGGAGTTCGCGTCTGCCGCGGCCGTTAAATCATTAACACCAAAGGTTGACCTCGG
ATCAGGTAGGGATACCCGCTGAACTTAAGCATATCAATAA

Sample 19_MdHEF - ITS

GGTCCCGGGCCGGGGGAGTGATTTTCAAACCCTTGTGAACTACCACTCTGTTGCTTC
GGGGGCGACCCCGCCGTCTCGGCGGTGGTGCTCCCGGTGGCCATCTATCAAACTCTG
CATTACCTTGCGTCGGAGTCTTATAAAGAATTAAACAAAACTTTCAACAACGGATCT
CTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGC
AGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCGTGGTATTCCGCG
GGGCATGCCTGTTCGAGCGTCATTTCACCACTCAAGCCTGGCTTGGTATTGGGCGTC
GCGGCCTCCGCGCGCCTCAAAGTCTCCGGCTGAGCAGTCCGTCTCCGAGCGTTGTGA
CATTTTCGCTAGGGAGTTCGCGTCTGCCGCGGCCGTTAAATCATTAACACCAAAGGT
TGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAAGCATATCAATAAGCGG

Figure 2.1 Internal Transcribed Spacer (ITS) sequences of samples C1_MdHEF, 7_MdHEF, and 19_MdHEF.

DNA sequences of the ITS region from three isolates were aligned to *Mycosphaerella dearnessii* as the top match using BLAST in the NCBI nucleotide database. The first two samples aligned with 100% identity, while the last sample exhibited one nucleotide difference at base pair 45 (red highlight shows mismatch to green highlight) yielding 99.8% identity.

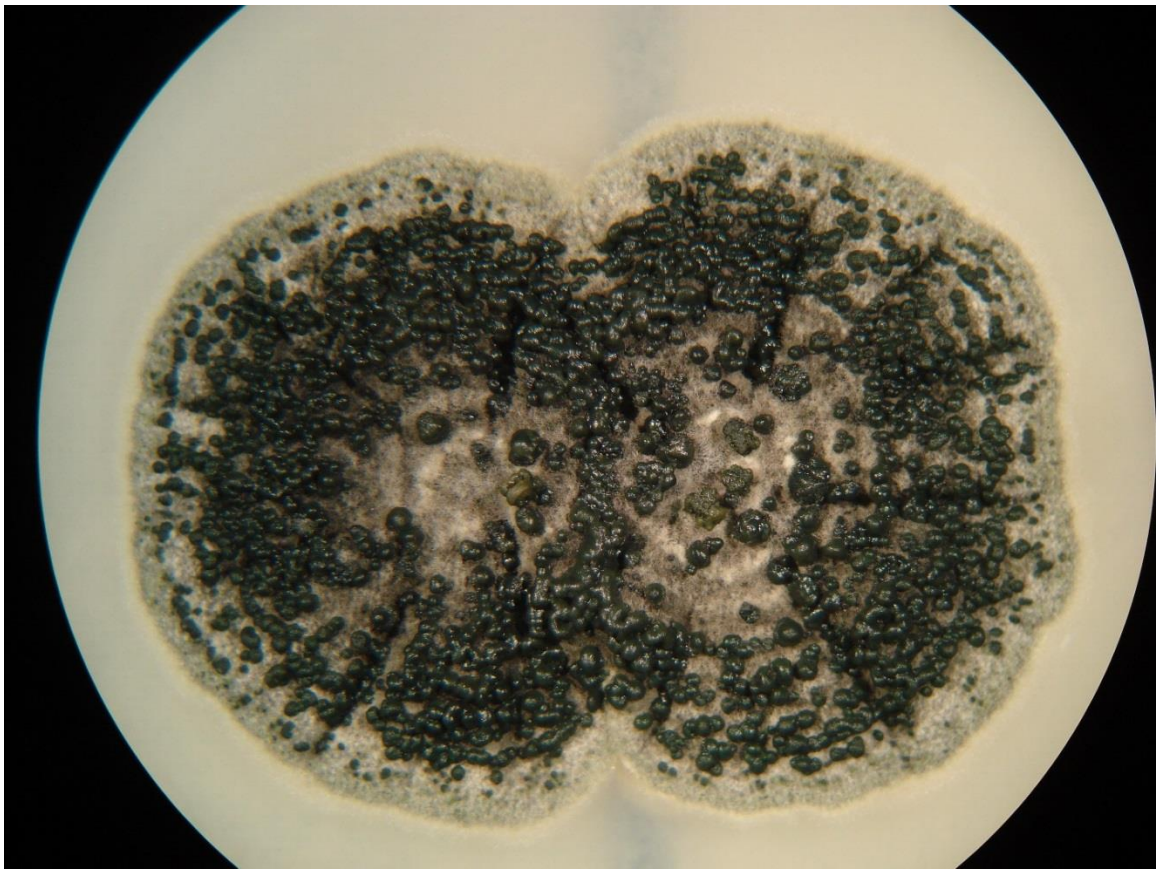

Figure 2.2 Compatible reaction of vegetative cultures.

Haplotype 1 and haplotype 11 exhibit compatibility in the vegetative incompatibility experiment. Notice the merging of mycelia and conidia. The texture/coloration of both haplotypes are similar.

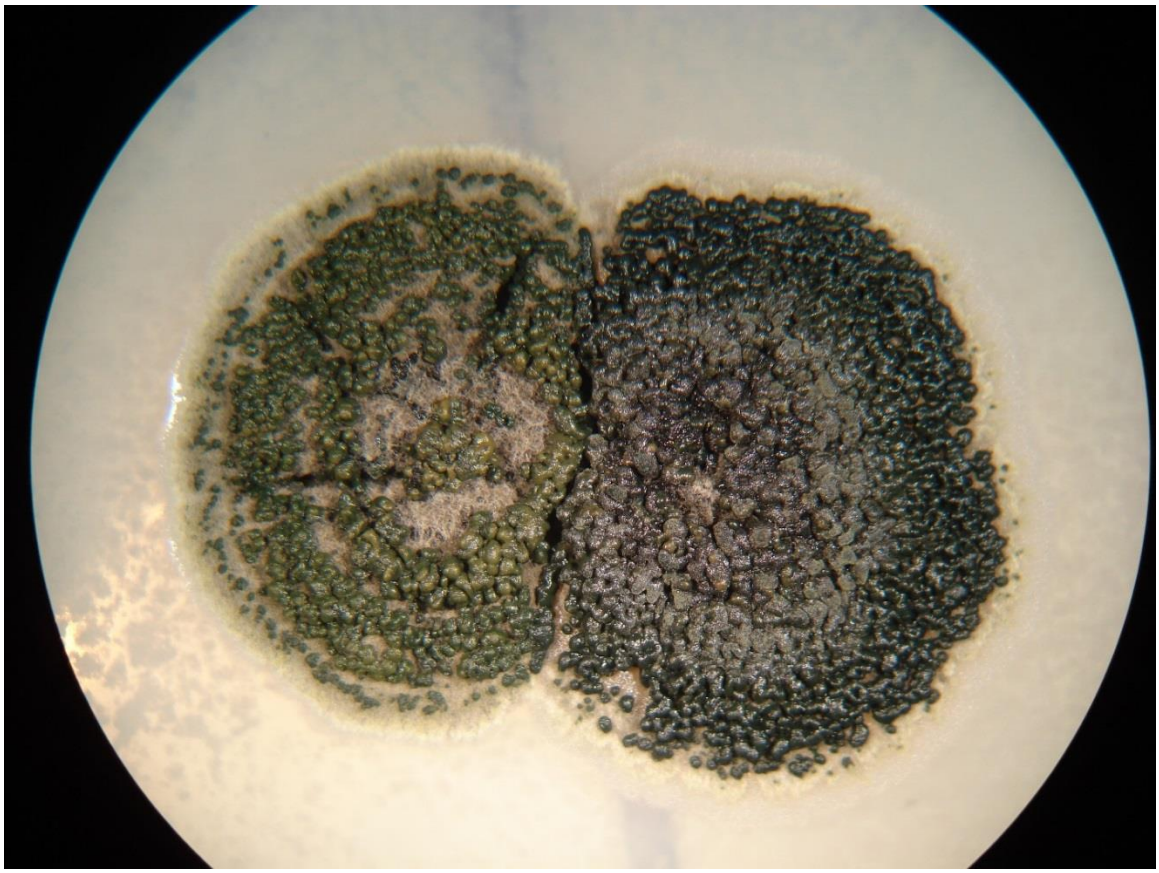

Figure 2.3 Incompatible reaction of vegetative cultures.

Haplotype 12 and haplotype 8 exhibit incompatibility in the vegetative incompatibility experiment. Notice the barrage formation of conidia at the point of convergence with mycelia and conidia. The texture/coloration of both haplotypes are dissimilar.

Figure 2.4 Longleaf pine needles displaying symptoms of Brown Spot Needle Blight.

Diseased needles exhibit symptoms from infection by *Mycosphaerella dearnessii.* Dark brown lesions encircle the needle tissue, surrounded by a yellow halo.

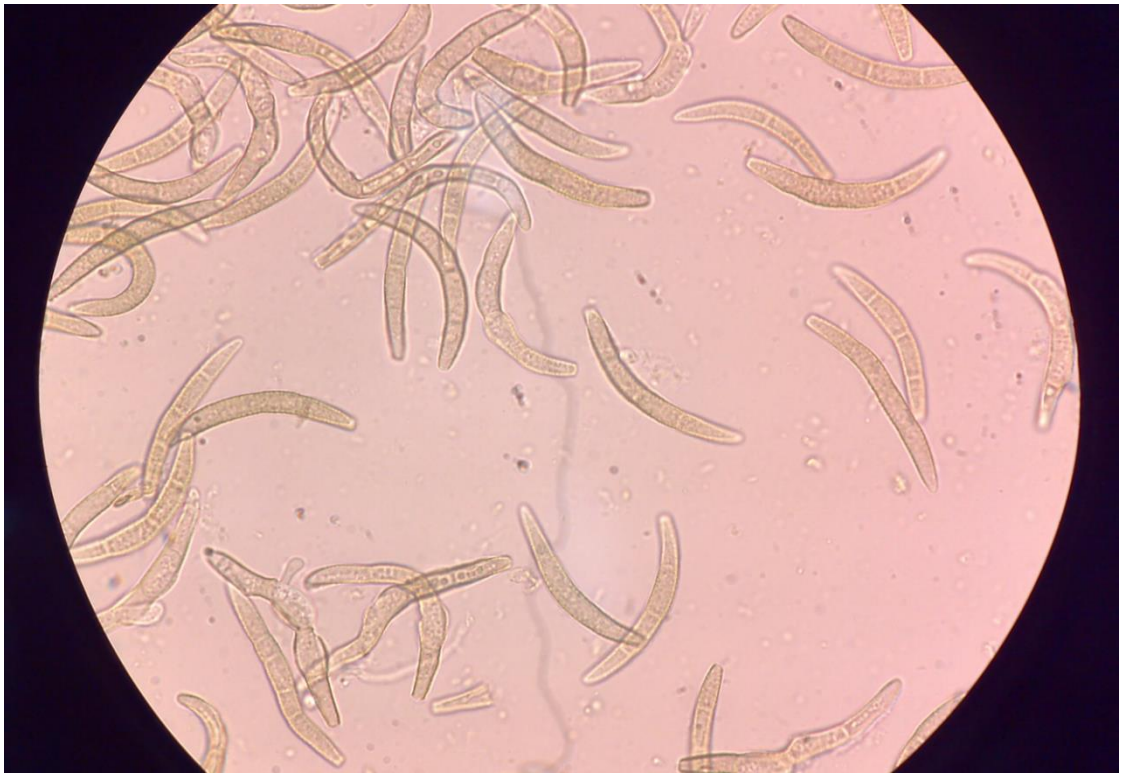

Figure 2.5 Conidia from excised lesions on longleaf pine needles (400X magnification).

Mycosphaerella dearnessii spores were harvested from diseased longleaf pine needles and placed on nutrient media. Forty-one isolates were collected and developed into pure cultures. Microscopic evaluation provided sufficient evidence of species identification.

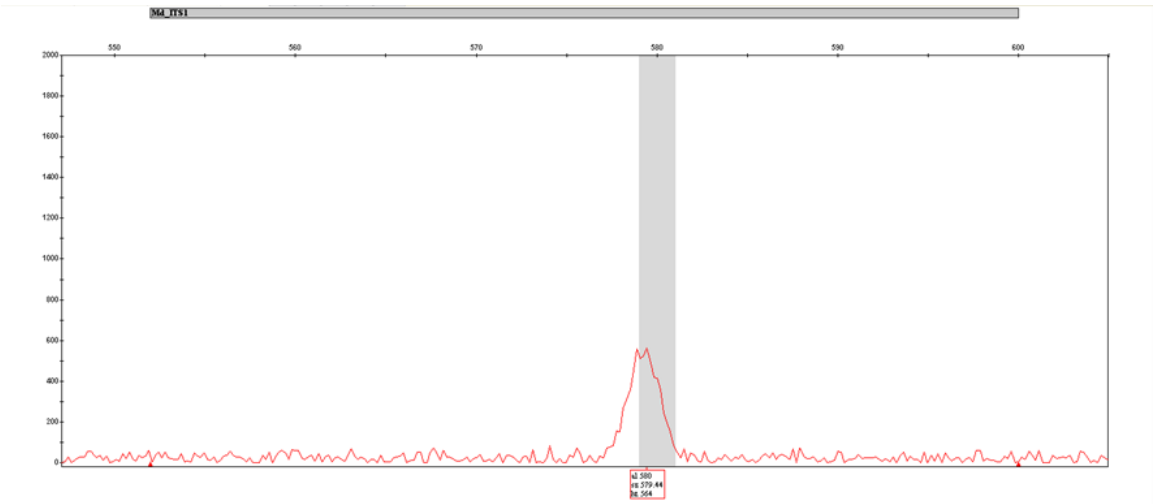

Figure 2.6 Electropherogram from Internal Transcribed Spacer amplification (ITS) amplification.

The graph displays excitation from successful amplification of the ITS region around 500 bp. Identical peaks were detected in all samples, including the samples selected for ITS sequence analysis.

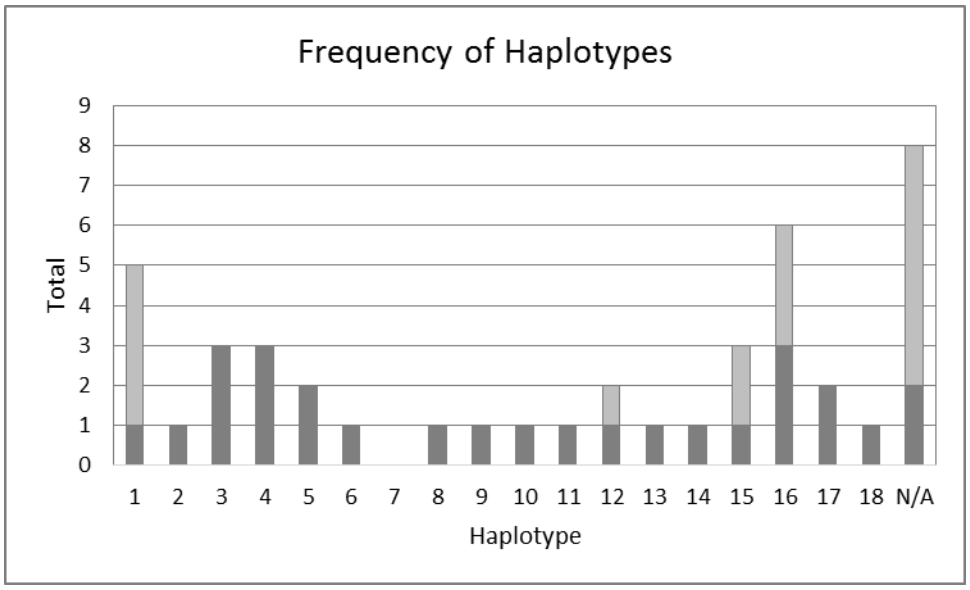

Figure 2.7 The frequency of each haplotype.

This bar graph displays the number of isolates that were designated to each haplotype group. The dark bar shows the number of isolates in each haplotype after removing isolates taken from the same needle samples (except E1 & E2_MdHEF). Most haplotypes are represented by only one isolate. Eight isolates were not designated haplotypes (N/A), of which only two were from unique samples.

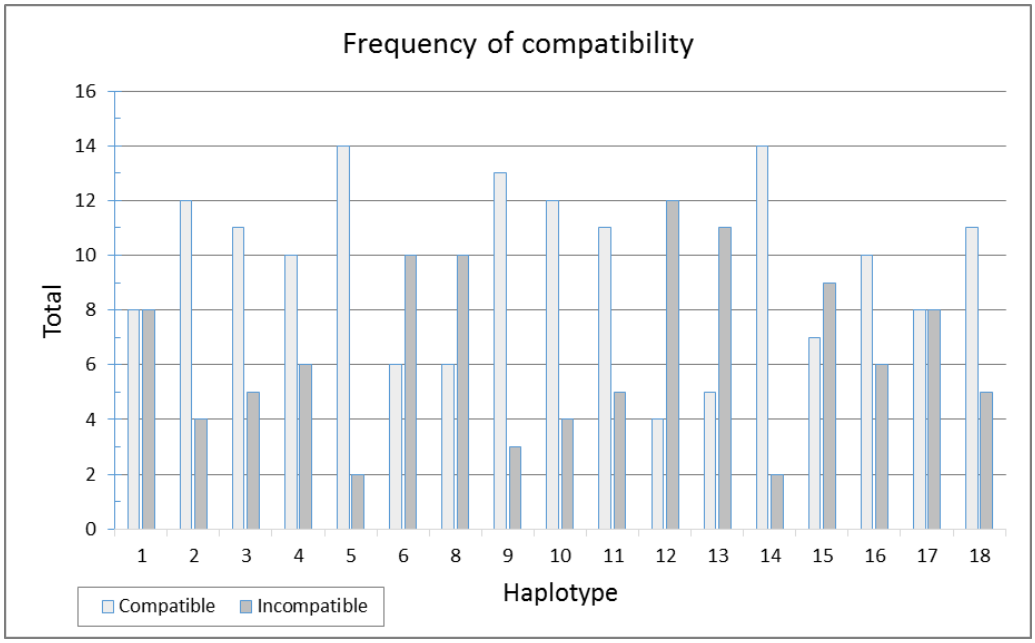

Figure 2.8 Frequency of compatibility of each haplotype.

This graph shows the number of times each haplotype was designated compatible (light) or incompatible (dark). These results exclude the self-pairings data.

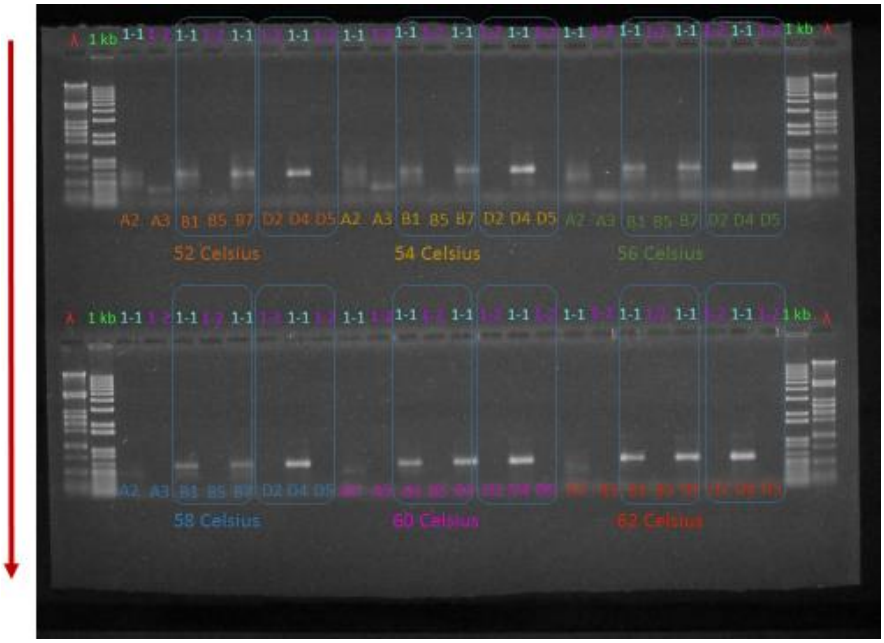

Figure 2.9 Temperature gradient for mating type primers.

A temperature gradient (52 °C to 62 °C) was applied to a subset of isolates to determine the optimal annealing temperature for the mating type primers (MAT1-1 & MAT1-2).

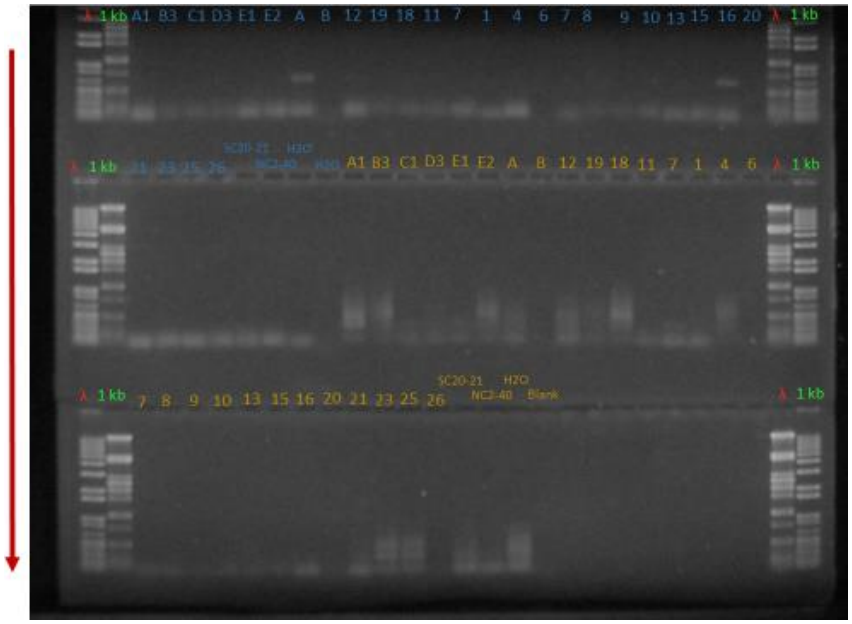

Figure 2.10 Mating type PCR amplification.

Optimal annealing temperatures were applied to all isolates for both sets of mating type primers (MAT1-1 &MAT1-2).

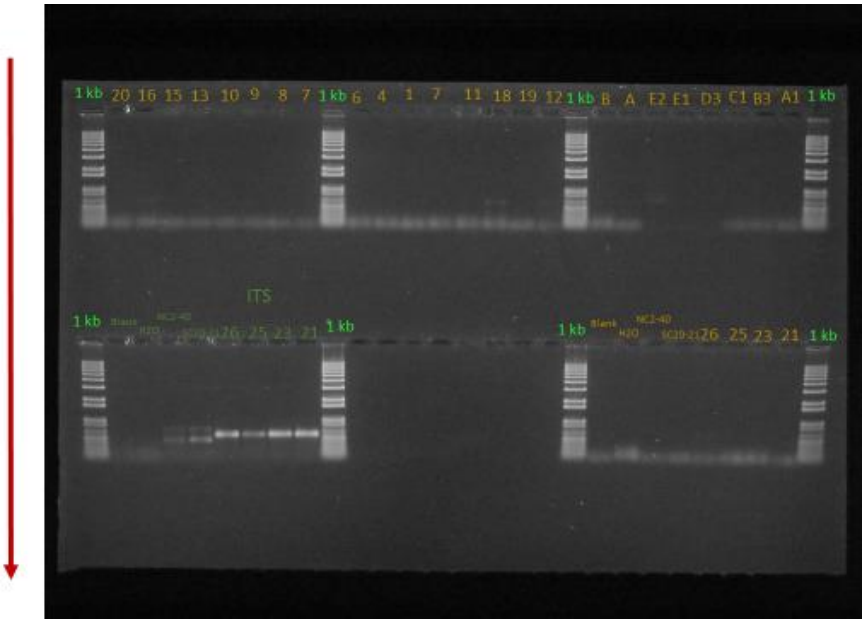

Figure 2.11 MAT1-1 PCR amplification with ITS positive controls.

Mating type primer (MAT1-1) was applied to all isolates with *Cronartium fusiforme* (NC2-40 & SC20-21) as negative controls. Universal ITS fungal primers were applied as positive controls.

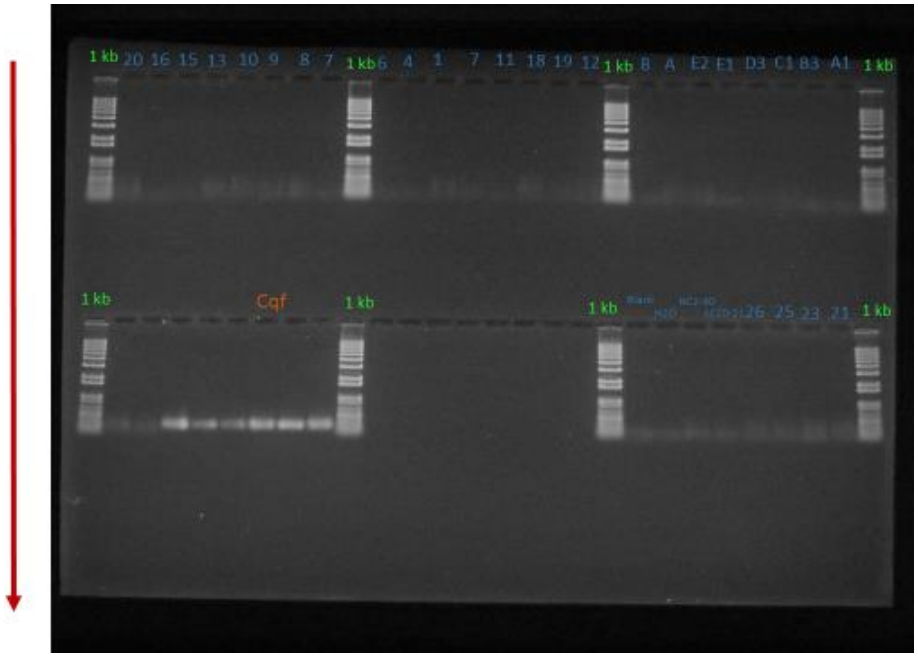

Figure 2.12 MAT1-2 PCR amplification with Cqf positive controls.

Mating type primer (MAT1-2) was applied to all isolates with *Cronartium fusiforme* (NC2-40 & SC20-21) as negative controls. *C. fusiforme* fungal primers (Cqf) were applied as positive controls to *C. fusiforme* DNA.

CHAPTER III

MYCOSPHAERELLA DEARNESSII DRAFT GENOME SEQUENCE

Introduction

The fungus *Mycosphaerella dearnessii* (class: *Dothideomycetes,* order: *Capnodiales*) is the teleomorph of the ascomycete *Lecanosticta acicula*, which causes the disease brown spot needle blight. This phytopathogen infects many pine species and has been observed in North America, South America, Europe, and Asia (Wolf & Barbour, 1941; Huang *et al.* 1995). *Pinus palustris* (longleaf pine) is particularly susceptible to this fungal pathogen. *P. palustris* has adapted to frequent environmental disturbance from wildfire by postponing above ground growth in the early stages of development. The primary source of inoculum for *M. dearnessii* is believed to be conidia, which are disseminated by rain splash. Persistent secondary infections of *M. dearnessii* onto longleaf seedlings delays height initiation and can result in mortality. The fungal pathogen is capable of entering its host through stomatal openings and within 10 days visible symptoms can be detected on the host (Snow, 1961; Setliff & Patton, 1974).

The majority of biological investigations into this organism were conducted before molecular techniques became commonplace. Consequently, little is known about the genetics of this organism and how it is able to infect its host. Despite this drawback, many other fungal phytopathogens have considerable information pertaining to the genomics and pathogenicity (Amselem *et al.*, 2011; Comeau *et al.*, 2014; Ohm *et al.*,

2012; de Wit *et al.*, 2012). Comparing gene similarities between these pathogens will deliver insights into how *M. dearnessii* is able to cause disease and provide a framework for future investigations.

Generally, a pathogen must accomplish a series of steps to infect and sustain a plant host. Broadly, these steps include the formation of an infection structure to locate and penetrate the host, the ability to degrade the cuticle and plant cell wall, evade host detection and respond to host defenses, produce toxins for nutrient assimilation, and signal changes either within the host to act as a carbon sink or the pathogen for preparing a feeding structure (Idnurm & Howlett, 2001). Disruption to any one of these processes can prevent disease, although not every process is present in all pathogens. Recognizing genes that are responsible for these processes will contribute to a better understanding to the lifecycle of the pathogen.

Materials and Methods

Genomic DNA Preparation, Library Construction, and Sequencing

Fungal cultures (*Mycosphaerella dearnessii*) were isolated from lesions on *Pinus palustris* at the Harrison Experimental Forest in southern Mississippi as described in Chapter 2. Also, genomic DNA was extracted as previously described in Chapter 2. Two isolates (B3_MdHEF & D3_MdHEF) were selected for genome sequencing. Paired-end library preparation from genomic DNA was performed with TruSeq DNA Library Prep Kit, following the manufacturer's instructions (Illumina, San Diego, CA). Library construction attempted to capture a target fragment size selection of 500 bp. The DNA libraries of both isolates used TruSeq Index Adapters (6-nt indices) to distinguish between the each sample; CGATGT for B3_MdHEF and TGACCA for D3_MdHEF. All

samples utilized short-read sequencing of our libraries on a single flow-cell of an Illumina MiSeq with a 500 cycle MiSeq Reagent Kit v2. Raw read data (251-nt read length) was processed by the MiSeq System and the sequence data was provided in FASTQ format for further processing.

Assembly and Annotation

Sequence data depicted in FASTQ format were trimmed to remove adapter sequences and filtered to remove low quality reads from the data using the program *cutadapt* (Martin, 2011). Strict stringency measures were used to filter out low quality reads. Reads with a PHRED quality score below 28 were considered low quality. Only complete and perfect reads were kept for sequence assembly. Incomplete reads, less than 250-nt in length, were removed from the data. Reads that contained any ambiguous base calls (i.e. – N) were treated as imperfect and discarded, along with the corresponding paired-end read.

Sequence data was further filtered to remove any sequences with similarity to bacterial or archaeal genomes. Parameters within the program *DeConSeq* (Schieder & Edwards, 2011) were designated as the most rigorous. Any sequence that contained 60% identity at a 1% threshold was regarded as contamination and rejected from dataset.

The short read assembly tool *ABySS* (Simpson *et al.,* 2009) was employed to obtain assembly metrics. Assembly metrics of interest were N50 and contigs of significant length for various k-mer's. Implementing the *ABySS* program to evaluate the various k-mer lengths, from 41-mer to 249-mer in 2-mer increments, we were able to determine the optimal k-mer parameter.

The sequence assembly program *Velvet* (Zerbino & Briney, 2008) utilized the optimal k-mer (51-mer) parameter to assemble the genome sequence. This program makes use of algorithms manipulating de Bruijn graphs from specified k-mer lengths to efficiently assemble sequences from short-read data. Paired-end datasets are treated within the program to classify repeat sequences, but not resolve the complete structure of the repeat sequence region. Assembly sequences less than 100 nt were removed from the data. The resulting sequence assembly was used for downstream analyses.

In order to assess the assembly coverage, a genome estimate was required. The program *jellyfish* (Marcais & Kingsford, 2010) identified the number of identical sequences of the size 17-mer and calculated every instance of that sequence. Counting the coverage of the highest identified 17-mer sequence and then dividing that count into the total number of reads produced an estimate of genome size of the organism. A 17-mer sequence size is small enough to generate abundant sequences with the same identity, while large enough to remain infrequent for approximate size evaluation.

Bowtie (Langmead *et al.,* 2009) functioned to discern sequence similarity from two closely related *Mycosphaerella sp.* This program aligned sequence reads from *M. dearnessii* to the assemblies of *M. fijiensis* and *M. graminicola*. This measure was taken to decide whether to include experimental data from these two species when annotating genes from *M. dearnessii*. Only perfect alignments were considered for analysis which enabled a judgment call for concluding how to include experimental data.

Genome completeness was tested by the software program *CEGMA* (Parra *et al*, 2007). In addition, 52 other genome assemblies within the Class Dothideomycetes were tested for completeness. A core set of 458 genes represented conserved proteins present

in a wide range of taxa. This core set of conserved genes served to illustrate the thoroughness of the genome assembly. Novel microsatellite identification was accomplished through SciRoKo (Kofler *et al.,* 2007) software.

Gene prediction was carried out by *GeneMark-ES v2* (Ter-Hovhannisyan *et al.,* 2008) from all contigs to generate coding sequence and protein translation. This program was chosen due to a lack of experimental data for this organism to use as a training set. Only genes longer than 35 amino acids (a.a.) were retained for successive analysis. Protein translations of predicted genes were then processed into *SignalP* (Emanuelsson *et al.,* 2007) for recognition of signal peptides or membrane anchors as putative secretory proteins. These secretory proteins are most likely responsible for host recognition and infection.

Functional annotation was automated and prepared using AgBase pipeline to segregate genes by GO terms (McCarthy *et al.*, 2006). Analysis was performed using only the predicted proteins from *GeneMark-ES*. The algorithm BLAST was implemented to search AgBase's fungal database with only experimentally curated entries. Default BLAST parameters, except for the e-value which was changed to 4e-70, were kept for inquiry. Also, BLAST hits were excluded that had a percent identity less than 60% or a query coverage of less than 60%. AgBase's GOanna tool was operated to assign GO associations to similar sequences from the BLAST hits. Then the tool GOSlimViewer assisted in summarizing the GO terms assigned.

A reciprocal best BLAST hit (RBBH) strategy was employed to identify orthologous genes from 38 fungal species. Sequences for 44 fungal genome assemblies were obtained from the Joint Genome Institute's (JGI) MycoCosm database. Only

sequences from the Class Dothideomycetes were retrieved. The RBBH strategy applies the BLAST algorithm to detect the most similar sequences from the query genome against the genome assembly of *M. dearnessii*. Then the reciprocal strategy is applied; detecting the most similar sequences from *M. dearnessii* against the genomes from the JGI database. If the most similar sequences from each BLAST analysis match, then that sequence is considered orthologous.

A list of known effectors was compiled and surveyed for homology. Accession numbers from the NCBI database were probed for sequences corresponding to the list of effectors. BLAST was then utilized to evaluate the list of effectors against the *M. dearnessii* genome sequence. Only results with a BLAST bit score greater than 60 were considered important for discussion.

The CAZy database is curated to contain known carbohydrate-active enzymes (Lombard *et al.,* 2014). These enzymes are considered important to the digestion of polysaccharides; which are essential for fungal nutrient assimilation, particularly for plant pathogens. This database was used to search our genome for homology. The CAT web server (Park *et al.,* 2010) facilitated the search for homologous sequences for our data.

Our genome sequence was inspected for ribosomal RNA (rRNA) and transfer RNA (tRNA) genes. WebMGA (Wu *et al.*, 2011) mediated the service for tRNAscan-SE (Lowe & Eddy, 1997) to provide tRNA gene predictions. In addition, WebMGA also mediated a search for rRNA gene sequences. Using a hidden Markov model algorithm, rRNA gene sequences were scanned for in the assembly (Eddy, 1998). The PHYMYCO-DB (Mahe *et al.*, 2012) provided 18S rRNA gene sequences from 31 species in the class *Dothideomycetes*. These sequences were then compared to the rRNA genes of *M.*

dearnessii. The program MEGA5 (Tamura *et al.,* 2011) performed phylogenetic analysis on the 18S rDNA subunit using ClustalW (Larkin *et al*., 2007) multiple sequence alignment. Maximum likelihood was the chosen method to interpret evolutionary relationships.

Results

Genomic DNA Preparation, Library Construction, and Sequencing

Two isolates (B3_MdHEF and D3_MdHEF) of the plant pathogen *M. dearnessii* were selected for genome sequencing. The two isolates were chosen based on a three marker haplotype assignment (Chapter 2). None of the microsatellite markers were shared between the isolates (Chapter 2). Initial DNA quality as detected by the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) indicated genomic DNA extracted produced a peak size of 18,455 bp at 3.8 ng/µL concentration and 17,480 bp at 4.9 ng/µL; B3_MdHEF and D3_MdHEF, respectively (Figure 3.1 & 3.2). After size selection and enrichment, quality control as indicated by the Agilent 2100 Bioanalyzer resulted in a peak size of 601 bp at 9.8 ng/µL concentration and 627 bp at 13.2 ng/µL for B3_MdHEF and D3_MdHEF, respectively (Figure 3.3 & 3.4).

A total of 7,361,060 sequences were produced using the Illumina MiSeq sequencing platform for the isolate B3_MdHEF. Another 5,475,481 sequences were produced for the D3_MdHEF isolate. Isolate B3_MdHEF was selected for further analysis and processing due to the additional 25% sequence data. The length of each sequence was 245 bp after removing the adapter and index sequences; totaling 3.6 billion nucleotide reads for B3_MdHEF.

Assembly and Annotation

Fourteen percent of the sequence data (764,852,965 reads) was removed due to similarity with bacterial and/or archaeal sequences. The remaining sequence was analyzed for optimal assembly parameters, resulting in a chosen k-mer length of 51 as a compromise between N50 and longest contig length, as indicated in Figure 3.5. Genome assembly of the B3_MdHEF isolate yielded a total sequence length of 30,512,903 bp. The longest contig of the B3_MdHEF assembly was 144,441 bp; yet, the average contig length was only 2,547 bp (Table 3.1). Most (92%) of the assembled sequences are smaller than 10,000 bp.

In order to determine how much of the *M. dearnessii* genome was assembled, a genome estimate was required. Previous reports utilized sequence data to estimate genome size (Zhang, *et al*., 2013; Li, *et al*., 2010). Following these procedures, a 17-mer oligonucleotide sequence was counted a total of 66 times (Figure 3.6). Using this calculation, a genome size estimate of 46 Mb was determined with the assembly representing 65% of the genome sequence. A set of 458 core genes is used to check how well the sequencing and assembly of a genome captures important sequences. The B3_MdHEF isolate was evaluated to have 93% of the core genes present within the assembly. An average of 95% of core genes were present in the assemblies of 52 other genomes within the Class Dothideomycetes (Table 3.2). In addition, 130 tRNA gene sequences and 46 rRNA sequences were detected in the B3_MdHEF assembly.

Evidence-driven gene prediction software can utilize sequence data (ESTs, RNA-Seq, etc.) from other experiments or from closely related species. Since no data is available for *M. dearnessii*, an examination of completed genomes with two closely

related species was performed. Sequence reads from the B3_MdHEF isolate were aligned to the genome assemblies of *M gramnicola* and *M. fijiensis* (Table 3.3). Neither assembly provided sufficient similarity (0.3%) to warrant the use of evidence-driven gene prediction. In the absence of evidence, only *ab initio* gene prediction was suitable for discovering gene models. The assembly of B3_MdHEF produced a total of 10,996 gene models larger than 35 bp (Table 3.4). The average peptide length is only 455 a.a. and only 7% of the peptides are larger than 1,000 a.a. In addition, 145,977 open reading frames were identified in all 6 reading frames. A total of 719 signal peptides were revealed within the genome assembly as well.

Microsatellite analysis unveiled 4,824 putative single sequence repeats (Table 3.5). Trinucleotide motifs were observed the most frequent, composing nearly half (2,268) of the putative sequences. A total of 7,839 peptides analyzed using the BLAST algorithm were associated with KOG database classes (Table 3.6). GO annotations also revealed 250,471 associations to a fungal database maintained by AgBase (McCarthy et al., 2006) (Figure 3.7 – 3.9). The majority of GO annotations belonged to the biological process category (133,534); with the cellular component with the next most (76,790) and molecular function last (40,147).

A total of 284,917 reciprocal best BLAST hits were observed between the B3_MdHEF assembly and 44 other genome assemblies representing 38 species within the Class Dothideomycetes. Twelve species within the same order (Capnodiales) clustered at the high end of the RBBH analysis (Figure 3.10). This matches a phylogenetic analysis of the 18S rRNA sequences from 31 species within the Class

Dothideomycetes (Figure 3.11). A basidiomycete (*Tritirachium oryzae*) was set as an outgroup in the phylogeny.

There were 151 different CAZy families identified in the genome assembly, totaling 853 associations (Table 3.7). Most were in the glycoside hydrolases (334) and the glycosyl transferases (223). A subset of CAZy families is shown in Table 3.8, and displays the substrate of each associate family. Particular genes of interest, mostly effectors, were selected from other fungal plant pathogen genome sequences and compared to the B3_MdHEF genome sequence (Table 3.9).

Discussion

Sequencing and Assembly

Mycosphaerella dearnessii belongs to a diverse genera of fungal pathogens. Only two species within the genera have completed genome sequences published as of this writing (Orhon *et al.*, 2011; Churchill, 2011). Comparing the genome sequences for these species will provide an opportunity to study the fundamental genetic components of this group of plant pathogens. In this study, we present an initial description of the genome sequence of *M. dearnessii* and begin the first steps to uncover links between these pathogenic filamentous fungi.

A popular sequencing platform (Illumina MiSeq, San Diego, CA) was chosen in this investigation for its ability to deliver sufficient data to assemble a credible genome sequence (Tang *et al.*, 2012). However, one of the drawbacks to this sequencing platform is the short length of the sequence reads. To compensate for this, a sequencing strategy utilizing paired-end reads was employed to improve genome assembly. This strategy assigns short DNA sequence tags to each end of the target DNA strand. The target DNA

is then sequenced and the orientation and boundaries of the strand is elucidated. By first creating a DNA sequencing library of a known nucleotide size, distances between any missing data can be interpreted and allow discrete contigs to be connected (Fullwood *et al*. 2009). For this sequencing project, a paired-end library size of approximately 500 bp was selected. The rationale behind this criterion was chosen for the purpose of creating artificially longer sequences. The MiSeq sequence data consisted of 250 bp read lengths from each paired-end sequence; therefore, each end of a paired-end sequence represents half of the total 500 bp sequence. The paired-end library sizes that resulted were found to be slightly larger than 500 bp (Figure 3.1 & 3.2).

Another drawback of next-generation sequencing technology is the increased error rate of 0.80% (Quail *et al.,* 2012; Nagarajan & Pop, 2013). Considering this drawback, a 50 fold coverage is desired to compensate for the error rate and short read length (Sims, 2014; Ekblom & Wolf, 2014). Initially the genome size was estimated based on the average fungal genome size and that of other closely related fungal species. Using data obtained from the Joint Genome Institutes MyoCosm database, the average genome size within the kingdom *Fungi*, phylum *Ascomycota*, and class *Dothideomycetes* are 38.3 Mb, 36.3 Mb, and 40.5 Mb; respectively. Given the highest average genome size (40.5 Mb) and MiSeq data totaling 4 Gb on the low end, a 98 fold coverage of sequence data was deemed sufficient to assemble the genome of *M. dearnessii*. In the end, data totaling 3.6 Gb was obtained for just one isolate (B3_MdHEF). The other isolate (D3_MdHEF) yielded 2.6 Gb of data. These outputs would have yielded 89 fold and 66 fold coverage, respectively. Sequence coverage needed to assemble a genome is based on the assumption that the organism being sequenced is a diploid individual. DNA

extracted from *M. dearnessii* was acquired from haploid isolates; presumably, the fold coverage is doubled for any given estimate.

Accounting for less than optimal sequence output and/or larger than average genome size proved to be advantageous. Preliminary analysis of the genome sequence detected the presence of bacterial 23S rDNA from the genus *Burkolderia*. Commissioning the most stringent criteria, similar sequences to any bacterial or archaeal sequences were removed from the data. Consequently 14% of the MiSeq data was removed, resulting in an estimated 76 fold coverage for the B3_MdHEF genome.

A more accurate estimate of the size of the genome was calculated after sequence data was obtained. Tallying the number of times a specific 17 nucleotide sequence is detected can reveal a better estimate of the genome size (Guo *et al.*, 2015; Zhang *et al.*, 2013; Li *et al.*, 2010). The most frequent 17 nucleotide sequence was detected 66 times in our genome (Figure 3.6). Dividing the most frequent count into the total number of sequence reads will generate a genome size estimate. In our case, the genome size is calculated to be 46.7 Mb (Table 3.1). Predictably, the fold coverage changes to 66 in light of the updated genome estimate. The calculated genome size estimate is larger than the average size of the class Dothidemycetes, but well within the extremities for that class (16 Mb for *Piedraia hortae* and 177 Mb for *Cenococcum geophilum*). Even within the genus, the two most studied species have vastly different genome sizes (*M. graminicola* with 39 Mb and *M. fijiensis* with 74 Mb), although the increase in genome size appears to be the result of an expansion of repeated sequences, presumably transposable elements (Churchill, 2011). Our estimated genome size for *M. dearnessii* is likely not exact, but accurate enough to determine the efficacy of our sequencing efforts.

Assembly and annotation was carried out with data from isolate B3_MdHEF only. The purpose of this document is to report the first description of a draft genome sequence for *M. dearnessii*. Future studies will utilize data from the D3_MdHEF isolate to make comparisons. There was 25% more data for B3_MdHEF, which provided the rationale to proceed with assembly for that isolate. To generate the best assembly with that data, a series of k-mer lengths were evaluated for optimization (Tang *et al.*, 2012). Metrics to evaluate the various k-mer lengths included the N50 value and the largest contig length (Figure 3.5). The N50 value is a widely used genomics statistic to measure assemblies. N50 is defined as the value at which half of all contigs are contained on contigs of that size or larger. A k-mer value of 51 was judged to produce the highest N50 value and reasonably large maximum contig length. Figure 3.5 demonstrates that a larger maximum contig length could have been achieved by sacrificing contig N50. Approximately 3.5 Kb of N50 value would have been relinquished for a gain of 15.6 Kb to the largest contig length. It was decided that a 17% increase in N50 took precedence to a 10% increase in longest contig length.

Assembly proceeded with the k-mer parameter set at 51 using the assembly program *Velvet*. In doing so, the B3_MdHEF isolate assembled a total 30.5 million base pairs. The total number of assembled sequences accounts for 65% of the estimated genome size. However, this does not indicate the accuracy nor the integrity of the genome sequence. One method to verify the thoroughness of an assembly is by searching for genes expected to be present. A core set of 458 genes are known to be present in a broad range of taxa (Parra *et al*, 2007). A collection of genome sequences compiled from the class Dothideomycetes (including *M. dearnessii*) were examined for presence of this

core set of genes (Table 3.2). Our *M. dearnessii* genome sequence is positioned below average for the percentage of core genes detected when compared to the other genome sequences from that set. Nonetheless, 93% of the core set of genes were found to be present in the B3_MdHEF assembly. While it may seem that the genome assembly is reasonably complete, the core set of 458 genes are presumably better represented in the sequence data. Genes vital to survival are highly conserved and likely to be present in multiple copies, due to mutations in those sequences that cause deleterious effects. As such, multiple copies of specific genes will be proportionally represented in sequence data. As stated before, short-read genome assembly relies upon a sufficient fold coverage to assemble a sequence. Assuming conserved genes are highly represented in the sequence data, it is expected that these genes are assembled to a greater degree as well. In light of these assumptions, it is not surprising that a highly conserved eukaryotic gene set is assembled in a newly drafted genome sequence.

Annotation

Annotation of any newly drafted sequence relies upon accurate gene/protein prediction. There are essentially two ways to predict protein models, evidence driven and *ab initio*. Evidence driven gene prediction relies upon previous studies of expressed sequence tags (ESTs) and/or RNA sequence (RNA-seq) projects. The *ab initio* methods utilize algorithms to predict protein sequences. Evidence driven gene prediction generates more accurate predictions, because portions of the sequences have already been annotated (Yandell & Ence, 2012). External evidence is not available for the fungus *M. dearnessii*; however, *M. fijiensis* and *M. graminicola* possess EST data (Orhon *et al.*, 2011; Churchill, 2011). Despite belonging to the same genus, evolutionary relationships

between these species remain polyphyletic (Crous, *et al*., 2009). In addition, fungi have low collinearity at the chromosomal level, even between closely related species (Oliver, 2012). In order to determine whether or not to include EST evidence for gene prediction, sequence data from *M. dearnessii* was aligned to the other *Mycosphaerella* genomes. Alignment to these genomes was meager, with only 0.3% of the reads aligning to the other genome sequences (Table 3.3). Because of the low alignment percentage, it was reasoned that the EST data would not contribute significantly to gene predictions. The consequence of this verdict resulted in *ab initio* as the method of choice for gene prediction.

Gene prediction identified 10,996 protein models that were longer than 35 a.a. (Table 3.4). Again, using data obtained from the MycoCosm database, an average of 12,199 proteins were identified within the kingdom *Fungi*, 11,508 from the phylum *Ascomycota*, and 12,836 from the class *Dothideomycetes*. Results are comparable to data in the MycoCosm database, but are vastly overrepresented by peptides shorter than 1,000 a.a. Only 770 peptides are longer than 1,000 a.a., which indicates that more data is needed to join shorter sequences together for better annotation.

Gene ontology (GO) is a data driven initiative to provide accurate gene product attributes to similar sequences across all species (McCarthy et al., 2006). GO terms were assigned to orthologous sequences in the *M. dearnessii* genome. The molecular function category (Figure 3.9) will be important for future investigations. This category contains enzymatic classes that may be linked to pathogenic factors for the fungus. Particularly, the high number of oxidoreductases is comparable to another pine pathogen, *Dothistroma septosporum* (Bradshaw *et al.*, 2015). Oxidoreductases in the *D. septosporum* are

upregulated during the early stages of infection. The GO annotation category transmembrane transporter activity and transferase activity likely contain candidate genes involved in host penetration and secretion. Likewise, several KOG designations for secretion and vesicular transport presumably play a similar role in host infection (Table 3.6).

The class *Dothideomycetes* is a highly diverse fungal taxonomic clade. Some members of this taxon are economically important plant pathogenic fungi, including forest pathogenic species belonging to the genera *Botryosphaeria* and *Mycosphaerella*. This taxon class is classically defined by the morphological characteristics of the ascus, which comprises fissitunicate asci and ascolocular development (Schoch *et al.*, 2009). Furthermore, taxonomic relationships within the order *Capnodiales* are not fully resolved. The order may include taxa having distant evolutionary relationships. Specifically, Schoch *et al.* (2009) recommends continued sampling and reorganization based on molecular data to resolve questions of phylogenetic relationships within the order. More directly, *M. dearnessii* clusters with members of the genus *Phaeophleospora* and may represent a clade within that genus (Crous *et al.*, 2009a; Crous *et al.*, 2009b). However, it has been demonstrated that the previously closest taxonomic relative to *M. dearnessii* (i.e. – *M. pini*) clusters outside of that clade (Crous *et al.*, 2009a; Crous *et al.*, 2009b). In this study we provide supporting evidence to conclude that *M. dearnessii* should belong to a clade separate from *M. pini*.

The 18S phylogenetic analysis (Figure 3.11) shows *M. dearnessii* within a clade detached from other members of the same genus. However, *M. dearnessii* does cluster nearer to *M. pini* and *M. graminicola* than it does with *M. fijiensis* or *M. populorum*.

These relationships are consistent with other published phylogenies (Crous *et al.*, 2009a; Crous *et al.*, 2009b). However, Figure 3.11 displays a few inconsistencies that need to be resolved with a more robust dataset. In particular, the species *M. zeae-maydis* should congregate closer to the other *Mycosphaerella* species. Also, different instances of *Zasmidium cellare* have segregated across many nodes. These peculiarities are most likely the result of insufficient data included in the phylogeny, namely the use of only one 18S sequence. It is well established that phylogenetic relationships are improved by incorporating additional DNA sequences for analysis. The addition of data from the genome of *M. dearnessii* will help to address this shortcoming regarding phylogenetic relationships among species within the taxa.

Furthermore, Figure 3.10 illustrates a pattern of relationships similar to that indicated by the phylogenetic analysis. Orthologous gene sequences can be infered from a reciprocal best blast hit investigation, which often outperforms other methods of grouping orthologs (Altenhoff & Dessimoz, 2009). The principle behind this investigation relies on the premise that genes derived from the same ancestral sequence are homologous. Accordingly, pairs of genes between different organisms that are more similar to each other than to any other gene between organisms are considered orthologous (Dalquen *et al.*, 2013). From this assertion, it is reasonable that organisms with higher number of orthologous sequences are more closely related. The pattern of relationships illustrated in Figure 3.10 largely matches that found in the 18S phylogenetic analysis, as well as that reported in other published phylogenies (Schoch *et al.*, 2009; Crous *et al.*, 2009a; Crous *et al.*, 2009b). Results indicate *Z. cellare* is the closest evolutionary relative to *M. dearnessii*; however, such a conclusion is not supported by

results described in Crous *et al.* (2009a & 2009b). Neither is the finding that *M. fijiensis* is more closely related to *M. dearnesii* than *M. graminicola*. Nonetheless, this investigation provides a resource for more thorough evolutionary scrutiny. Following this approach, each one of the several thousand reciprocal best blast hits can be individually analyzed phylogenetically. Moreover, common hits observed in every species are presumably conserved within that taxa and can be utilized for extensive evolutionary connections.

The ability of fungi to infect plants requires enzymes capable of modifying cell walls and metabolizing the carbohydrates of their host (Zhao *et al.*, 2013). Plant cell walls are constructed of the structural polymers: lignin, cellulose, hemicellulose, and pectin. Phytopathogenic fungi produce enzymes to degrade these polymers and use the metabolites for their own growth and development. To the same degree, a fungus must also be able to modify its own cell wall (composed of chitin), specifically the development of appressoria and hausorita which are specialized structures that invade plant cells in the process of infection (Kubicek *et al.*, 2014). Consequently, enzymatic proteins that degrade these particular carbohydrate substrates are targets for pathogenic activity. The composition of the enzymatic toolbox available to a pathogen reveals the nutritional strategy employed by that species (Perlin *et al.*, 2015). Of particular importance are enzymes classified as carbohydrate esterases (CE), glycoside hydrolases (GH), and polysaccharide lyases (PL), many of which are cell wall degrading enzymes found in phytopathogenic fungi (Kubicek *et al.*, 2014; Zhao *et al.*, 2013). The assignment of proteins to these classes is accomplished through a curated database

(CAZy, Lombard *et al.,* 2014) to discern particular enzymatic activities that cause disease.

A total of 853 associations with the CAZy database were found to occur in the *M. dearnessii* genome assembly described in this study. This number of associations is notably high, although not beyond the scope found in other necrotrophic species. Members within the class *Dothideomycetes*, and plant pathogens explicitly, are known to possess sequences that exhibit a high number of CAZy associations (Ohm *et al.,* 2012). The number of CAZy associations detected in the current genome assembly of *M. dearnessii* is approaching the upper limit found in fungal species, with only *Nectria haematococca* and *Fusarium verticilloides* possessing more than 800 associations (Zhao *et al.*, 2013). Somewhat different results have been reported for *M. graminicola* and *M. pini,* for which approximately half the number of CAZy associations have been detected as those found for *M. dearnessii* (Zhao *et al.*, 2013). The fragmented nature of the current draft assembly of *M. dearnessii* could be a leading cause of the high number of associations detected. This result may not accurately represent the organism at this stage of the genome assembly process.

Nonetheless, 151 different CAZy families were identified in the *M. dearnessii* genome assembly, compared to 187 CAZy families identified throughout the kingdom *Fungi* (Zhao *et al.*, 2013). Notably, Zhao *et al.* (2013) reported the CAZy family GH74 occurs in every species studied, yet it is missing from the *M. dearnessii* CAZy results. Again, a more thorough assembly could uncover this missing family. Zhao *et al.* (2013) also noted that family CE6 is phylum specific to *Zygomycota*, yet it is detected in *M. dearnessii*. The presence of CE6 may demonstrate a novel way for this fungus to procure

nutrients from its host. Additionally, no cutinases (CAZy family CE5) were identified in the *M. dearnessii* genome. As the name suggests, cutinases digest cutin (plant cuticle component). *M. dearnessii* is reported to enter its host through the stomata (Setliff & Patton, 1974; Snow, 1961; Wolf & Barbour, 1941) and seemingly circumvents the plant cuticle without the need to digest cutin. Investigating how the fungus responds to nutrient limited media will confirm these biological aspects of the fungus.

The CAZy associations have demonstrated certain aspects of *M. dearnessii* biology that are most likely involved in pathogenicity. Appressoria are fungal cell types that function to infect host tissue. These cells form a germ tube, aided by enzymes that modify chitin and β-1,3 glucan for development (Kubicek *et al.*, 2014). Not surprisingly, enzymes that act upon substrates (chitin and β-1,3 glucan) in fungal cell walls have been identified 156 times in *M. dearnessii*. Furthermore, enzymes with substrates (lignocellulose, pectin, PCW, starch, hemicellulose) found in plant cells were identified 316 times. This pathogen presumably uses these enzymes for nutrient assimilation accomplished through haustoria cells (Kubicek *et al.*, 2014). All of the enzymes involved in cell wall modification should be considered potential targets for pathogenicity and thus studied to guide exploration for the processes that are involved in plant resistance (Vleeshouwers & Oliver, 2014).

Other genes that could also be responsible for pathogenicity were investigated in the *M. dearnessii* genome (Table 3.9). A partial list of genes encoding CWDEs, effectors, toxins, and secreted proteins were assembled to help determine systems involved in the development of the disease brown spot needle blight. Effectors are proteins that bind to plant cell signals and act to subdue plant defenses. Most CWDEs

and effectors have signal peptides that configure proteins for secretion (Kubicek *et al.*, 2014). Hemibiotrophic fungi first establish a biotrophic relationship with the host plant, before switching to a necrotrophic relationship. This dual aspect of hemibiotrophs relies on effectors to evade the plant immune system and also the production of toxic molecules to its host. Many hemibiotrophic and necrotrophic species develop haustorial cells that enable nutrient acquisition (Kemen & Jones, 2012). Several *Mycosphaerella* species are hemibiotrophic plant pathogenic fungi; but the distinction between necrotroph and hemibiotroph has not been described explicitly for *M. dearnessii*. Snow (1961) briefly described a latent period after inoculation until the visualization of disease symptoms. The partial list of homologous genes described below, together with Snow (1961), indicates that *M. dearnessii* is a hemibiotrophic fungus. Likewise, this concludes that longleaf pine possesses pathogen recognition systems and likely genes corresponding to avirulent genes in the pathogen.

Biosynthesis of secondary metabolites that produce host specific toxins usually exist in gene clusters that contain polyketide synthase (PKS) enzymes (Stergiopoulos *et al.*, 2013). A total of 76 PKS homologous sequences were identified in the genome of *M. dearnessii*. In *Cochliobus heterostrophus*, a gene cluster containing two PKS genes is responsible for controlling T-toxin synthesis. Likewise, *M. zeae-maydis* contains PKS genes controlling PM-toxin synthesis; a toxin structurally related to T-toxin. The gene *Ace1* in *Magnaporthe oryzae* is expressed exclusively in appressoia although the function is currently unknown. The *Ace1* gene product contains a putative PKS class of enzymes (Stergiopoulos & de Wit, 2009). The PKS genes in *M. dearnessii* likely play a similar role as they do in other fungi.

The production and transport of HC-toxin in *C. carbonum* is regulated by a gene cluster containing *Hts1*, *ToxA*, *ToxC*, *ToxE*, *ToxF*, and *ToxG* (Stergiopoulos *et al.*, 2013). Homologs for the entire gene cluster responsible for pathogenicity were found in *M. dearnessii*. Additionally, *ToxB* is well characterized in the wheat pathogen *Pyrenophora tritici-repentis* and produces a toxin that induces chlorosis in susceptible hosts. The ACT-toxin gene in *Alternaria alternaria* prevents the pathogen from succumbing to reactive oxygen species initiated by plant defenses. Necrotrophic fungi require these toxins to metabolize the host cell into constituent molecules (Oliver & Solomon, 2010). The particular toxin repertoire of each pathogen species provides an observation into how the pathogen obtains nourishment from its host. Similar genes for these toxins were discovered in *M. dearnessii* and reveal the enzymatic toolbox available to the pathogen.

The gene *Avr4* in *Cladosporium fulvum* encodes a protein with a chitin binding domain that protects chitin from degradation by chitinases elicit in plant defenses (Stergiopoulos & de Wit, 2009). *Avr2* encodes a protease inhibitor of plant defenses in a similar manner to *Avr4* in *C. fulvum*. These genes were identified in the genome of *M. dearnessii,* as well as two other *C. fulvum* homologs for the genes *Avr9* and *Nrf1*. These genes are considered to be involved in nitrogen metabolism, with *Avr9* being expressed under nitrogen-limiting conditions. *Nrf1* mutants do not express *Avr9* and virulence is compromised in susceptible tomato plants (Stergiopoulos & de Wit, 2009). All of the genes described above inhibit plant defenses from destroying the pathogen. The observation that similar sequences to these genes are found in *M. dearnessii* genome suggests that similar gene products may be active in this fungal pathogen.

Initiation of the infection process by *Leptosphaeria maculans* induces expression of the genes *AvrLm1* and *AvrLm6* (Stergiopoulos & de Wit, 2009). Their function remains unclear, but they are recognized as avirulence genes having corresponding host R genes. Additionally, *AvrLm4*-7 is upregulated during the infection process in *L. maculans* (Stergiopoulos & de Wit, 2009). Deletion of this gene creates susceptible rapeseed plants that are otherwise resistant. Likewise, *Avr3* in *Fusarium oxysporum* f. sp. *lycopersici* is expressed only during colonization and is required for full virulence in susceptible individuals (Stergiopoulos & de Wit, 2009). Moreover, *Six2* and *Avr3* are genes that reside on the same chromosome and neither are found in nonpathogenic *Fusarium* strains (Stergiopoulos *et al.*, 2013). However, the gene products of *Six2* are found to localize in the host plant xylem. The locus *Avr1* functions to suppress plant mediated resistance, but is not required for full virulence in *F. oxysporum*. Homologous sequences to these genes were identified in *M. dearnessii.* Discerning genes that are expressed during colonization will help elucidate mechanisms responsible for infection. Such genes may help protect pathogens from plant defenses or may help pathogenic fungi to recognize a susceptible host. Further inquiry is needed to obtain an understanding of the biochemical processes that result when these genes are expressed.

Several sequences detected in the assembly for *M. dearnessii* are homologous to avirulent genes found in the rice pathogen *M. oryzae.* Cultivars of wild rice with R genes *Pi-ta, Pia, Piz* correspond to the avirulent genes *Avr-Pita*, *Avr-pia*, and *Avr-piz* for resistance to *M. oryzae* (Giraldo & Valent, 2013). Additionally, the avirulence gene *Avra10* is found in the pathogen *Blumeria graminis* f. sp. *hordei. Melampsora lini* genes *AvrM* and *AvrL* are considered to be effectors, with *AvrL* producing a protein that is

secreted in haustoria and *AvrM* expressing a compound that is involved in plasma membrane binding (Stergiopoulos & de Wit, 2009; Presti *et al.*, 2015). There is also an R gene that counteracts the effects of *AvrM* by triggering programed cell death in the plant (Presti *et al.*, 2015). These avirulence genes may play a similar role in *M. dearnesii*. Avirulent genes have been employed to probe host genomes for R genes and have been used to breed resistance in potato cultivars (Vleeshouwers & Oliver, 2014). Moreover, fungal strains presumably vary in virulence due to specific repertoires of effector genes. The particular pathogen strains present at certain geographical locations can be identified and this information can be applied to select cultivars with appropriate R genes for deploying in those locations. The genome sequence of *M. dearnessii* will promote the application of these techniques to accelerate for breeding resistance to brown spot needle blight in longleaf pine.

Genes involved in protein secretion were identified in draft genome sequence of *M. dearnessii*. Plasma membrane proton pump genes (Pma1 & Pma2) help secrete proteins in *Saccharomycetes cerevisiae*. The SNARE (soluble NSF attachment protein receptors) complex is also involved in transporting secreted proteins, but does so by bypassing the endoplasmic reticulum and Golgi apparatus (Stergiopoulos & de Wit, 2009). The genes *Syn2* and *Syn8* encode SNARE complex components. These effector proteins are effective in protecting the pathogen from host detection (Giraldo & Valent, 2013). Linking the genes responsible for protein secretion to effector proteins is likely to be an effective method for identifying genes that are involved in host resistance.

Several other homologous genes involved in pathogenicity were found in the *M. dearnessii* genome and are discussed below. *Nip* gene products in the fungus

Rhynchosporium secalis trigger host defense responses that ultimately lead to necrosis. The *Nep1* gene performs a similar function in the necrotrophic phase of hemibiotrophic fungi. Both the *Nip* and *Nep1* genes are responsible for causing necrosis, which essentially elicits disease. If these genes are required for virulence, then committing research to determine how to incapacitate these genes will lead to nonpathogenic fungi on resistant hosts. The *Rid* gene is a locus found in other fungi that is responsible for disabling the repeat induced point mutation (RIP) function in that particular fungal strain. Repeat induced point mutation is a fungal characteristic that preserves genome integrity by disabling transposable elements (Perlin *et al.*, 2015). It operates by producing a mutation in repeated sequences that works to inactivate transposons. It also produced accelerated gene evolution by provoking a higher mutation rate. Many effector genes are species specific and are located near transposable elements (Perlin *et al.*, 2015). It is theorized that this mechanism operates in fungi to generate novel mutations that function to defeat plant resistance. Detection of any gene associated with this phenomenon in fungi will shed light on genetic diversity in pathogen populations, as well as provide insights into the pathogens ability to overcome resistant host genotypes.

Summary and Conclusion

An initial investigation into the genome of *Mycosphaerella dearnessii* has revealed many previously undisclosed aspects into the biology of the fungus. The genome assembly of the isolate B3_MdHEF needs refinement to create longer contigs and connect scaffolds. However, assembly of approximately 65% of the genome and the recognition of about 11,000 predicted genes suggests that most of the protein coding genes have been identified. The remaining 35% of the genome is likely repeat sequences

(i.e. – transposable elements, centromeres, telomeres and microsatellites), which are known to be problematic for short read sequencing platforms (Tang *et al.*, 2012). Validation of the genome assembly can be accomplished through physical mapping, but may not represent the best appropriation of resources at this phase.

Additional sequence data is available for another isolate (D3_MdHEF) which will allow comparisons between individuals in a highly diverse population. A diverse population ordinarily displays variation in pathogenicity and virulence. Extensive sampling of *M. dearnessii* across the geographic range, coupled with an examination of host virulence will collect the necessary information to probe the genome for malicious genes. These genes will become the targets for any control measure to reduce disease incidence. The genome sequence of this organism has also revealed additional microsatellite markers to analyze population structure and possibly markers for virulence or other traits of interest. However, mating type genes still need to be identified in the genome sequence. Characterizing the mating type gene sequences will provide genetic markers that allow the frequency of each mating type to be determined in a population. The frequency of outcrossing corresponds to the frequency of each mating type. Novel genetic combinations increase with the rate of outcrossing and the ability of a pathogen population to respond to control measures.

Phylogenetic analysis performed in this study confirms previous research to suggest *M. dearnessii* should be placed in with *Phaeophleospora* in a separate clade. These findings highlight the need to better characterize the *Fungi* kingdom with sequence data and evolutionary analysis. The genome sequence of *M. dearnessii* will join a list of growing datasets to address this problem. Furthermore, orthologous sequences from

several closely related species were identified in this investigation. The distribution of species with a higher number of orthologs with *M. dearnessii* closely resembles the phylogenetic analysis. These orthologous sequences can deliver a set of sequences that are evolutionarily conserved and serve as a dataset for in depth phylogenies.

This study was the first endeavor to investigate the genetics, of *M. dearnessii* strains that cause disease, and the first study in several decades to answer biological questions about this fungus. We have presented evidence for *M. dearnessii* to be considered a hemibiotrophic pathogen. This requires the pathogen to infect the host initially without detection, then to induce necrosis and disease symptoms. We detected homologous sequences to known effectors in other fungi, demonstrating a probable method of evading host detection. We also did not detect any cutin degrading enzymes, suggesting the pathogen uses alternative methods to enter the host cell. Effectors are seemingly secreted by pathogens and several signal peptides were identified from the genome sequence of *M. dearnessii*. Moreover, homologous genes involved in secretory pathways were identified.

We have also detected genes associated with the formation of appressorium and haustoria. The appressorium presumably perceive the host through mechanical or chemical signals. Once these signals are discerned, ways to disrupt them can be examined. Haustoria are feeding structures that supply the pathogen with nourishment. Phytotoxins are generally released to decompose the plant material for fungal digestion. Several putative genes encoding toxin molecules were identified in the genome sequence of *M. dearnessii*. Typically, toxins are able to bind to molecules in the host to induce

programmed cell death or apoptosis. Altering those molecules, either by breeding or genetic engineering, can impede the signaling pathway that stimulates cell death.

Similarities to avirulent genes were recognized in *M. dearnessii* and likely have corresponding R genes in the host. These genes produce effectors and have been utilized as functional markers to help breed resistance. Employing similar techniques to *M. dearnessii* will accelerate the breeding process to select resistant trees. The identification of R genes with this technique relies on a genome sequence to discover avirulent genes. Furthermore, cultivars with different R genes corresponding to specific pathogen strains can be planted where those strains are located.

A gene involved in a defective RIP pathway was present in the *M. dearnessii* genome. The RIP pathway induces point mutations in transposable elements that inactivate those elements. Effector genes have been located near transposons and may be exposed to point mutations from the RIP pathway. These point mutations could accelerate the mutation rate in those genes and explain why fungal pathogens can overcome host resistance quickly. An allele resulting in a defective RIP pathway permits transposable elements to prosper, which could cause effector genes to duplicate.

The genome sequence of the fungal phytopathogen *M. dearnessii* provides a structure to further investigate molecular processes in the fungus. This genome can contribute to understanding the host-pathogen system and eventually to disease management strategies. Comparing genomic sequences with other fungal pathogens can link various fungal lifestyles with particular genomic structures. Additional genomic data across species will allow for a more thorough investigation into evolutionary histories of those species.

Table 3.1 Genome assembly statistics of *Mycosphaerella dearnessi* isolate B3_MdHEF.

Species	*M. dearnessii*
Genome assembly size	30,512,903
Number of predicted gene models	10,996
Sequencing coverage depth	66
% GC content	53%
Est Genome size (Mb)	46.7
% of genome assembled	65%
Number of contigs	12,317
Contig N50	19,120
Average contig length	2,547
Longest contig length	144,441
% of Genome < 10,000 bp	92%
% of Genome > 10,000 bp	7%

Table 3.2 Genome completeness measured by CEGMA.

Genome Assembly	Complete
Acidomyces richmondensis	95.16
Aplosporella prunicola	97.58
Aulographum hederae	95.97
Aureobasidium pullulans var. melanogenum	95.97
Aureobasidium pullulans var. namibiae	96.77
Aureobasidium pullulans var. pullulans	96.77
Aureobasidium pullulans var. subglaciale	95.56
Baudoinia compniacensis	95.97
Botryosphaeria dothidea	97.18
Cenococcum geophilum	95.16
Cercospora zeae-maydis	95.16
Cladosporium fulvum	95.16
Cochliblus berberidis	97.18
Cochliblus carbonum	94.35
Cochliblus heterostrophus	95.56
Cochliblus heterostrophus	95.56
Cochliblus lunatus	96.37
Cochliblus miyabeanus	96.37
Cochliblus sativus	94.76
Cochliblus victoriae	95.16
Didymella exigua	96.77
Dissoconium aciculare	95.56
Dothidotthia symphoricarpi	96.37
Dothistroma septosporum	96.37
Hysterium pulicare	97.18
Lentithecium fluviatile	96.37
Lepidopterella palustris	95.97
Leptosphaeria maculans	96.37
Lophiostoma macrostomum	96.37
Macrophomina phaseolina	96.77
Melanomma pulvis-pyrius	96.37
Mycosphaerella dearnessii	**93.55**
Mycosphaerella fijiensis	94.35
Mycosphaerella graminicola	94.76
Myriangium duriaei	96.37
Neofusicoccum parvum	97.58
Patellaria atrata	87.5
Piedraia hortae	95.16
Pleaomassaria siparia	97.58
Polychaeton citri	95.16
Pyrenophora teres f. teres	94.35
Rhytidhysteron rufulum	95.16
Septoria musivia	95.97
Septoria populicola	94.35
Setosphaeria turcica	95.56
Sporomia fimetaria	97.18
Stagonospora nodorum	97.18
Trypethelium eluteriae	96.37
Zasmidium cellare	97.18
Zopfia rhizophila	95.56
Average	**95.7812**

A database of 458 core genes is utilized to test the completeness of 52 genomes. The *Mycosphaerella dearnessii* genome sequence is below average, but still 93% complete.

Table 3.3 Read alignment with *Bowtie.*

Bowtie			
	B3_MdHEF	*M. graminicola*	*M. fijiensis*
B3_MdHEF	-	0.32%	0.36%
D3_MdHEF	75%	0.34%	0.38%

Sequence reads from two isolates of *Mycosphaerella dearnessi* were aligned to two other *Mycosphaerella sp.* genomes. Isolate D3_MdHEF was aligned to the genome assembly of the isolate B3_MdHEF.

Table 3.4 Predicted peptides from the genome assembly.

GeneMark ES	
Total	10,996
Average	455 a.a.
Max	4,901 a.a.
% of Peptides < 1000	93%
% of Peptides > 1000	7%

Proteins sequences were predicted with *GeneMark ES* and summarized.

Table 3.5 Microsatellite analysis.

Motif	**Counts**	**Average Length**
mononucleotide	136	17.14
dinucleotide	666	19.53
trinucleotide	2268	19.22
tetranucleotide	860	20.07
pentanucleotide	609	23.66
hexanucleotide	285	26.44

Predicted microsatellite motifs from the *Mycosphaerella dearnessii* genome sequence.

Table 3.6 Eukaryote orthologous group of proteins annotation.

KOG Class	Count	Description
A	316	RNA processing and modification
B	123	Chromatin structure and dynamics
C	395	Energy production and conversion
D	245	Cell cycle control, cell division, chromosome partitioning
E	361	Amino acid transport and metabolism
F	94	Nucleotide transport and metabolism
G	337	Carbohydrate transport and metabolism
H	121	Coenzyme transport and metabolism
I	419	Lipid transport and metabolism
J	409	Translation, ribosomal structure and biogenesis
K	501	Transcription
L	226	Replication, recombination and repair
M	91	Cell wall/membrane/envelope biogenesis
N	4	Cell motility
O	597	Posttranslational modification, protein turnover, chaperones
P	174	Inorganic ion transport and metabolism
Q	297	Secondary metabolites biosynthesis, transport and catabolism
R	1136	General function prediction only
S	371	Function unknown
T	592	Signal transduction mechanisms
U	587	Intracellular trafficking, secretion, and vesicular transport
V	67	Defense mechanisms
W	23	Extracellular structures
X	1	Multiple functions
Y	37	Nuclear structure
Z	315	Cytoskeleton

The database euKaryote Orthologous Groups (KOG) was used to annotate proteins from the *Mycosphaerella dearnessii* genome sequence.

Table 3.7 Summary of carbohydrate activated enzymes database association.

CAZy Enzyme Class	Count
Auxiliary Activities	73
Carbohydrate Binding Module	140
Carbohydrate Esterases	80
Glycoside Hydrolases	334
Glycosyl Transferases	223
Polysaccharide Lyases	3

Homologous protein sequences of *Mycosphaerella dearnessii* to enzymes in the CAzY database.

Table 3.8 Carbohydrate activated enzymes database association.

Substrate	CAZy Family	Count
β-1,3 Glucan	GH16	13
	GH25	1
	GH32	4
	GH64	3
	GH72	4
	GH81	1
	GH55	4
	GH128	2
Chitin	CBM2	13
	CBM1	4
	CBM12	1
	CBM14	6
	CBM18	5
	CBM50	9
	GH18	84
	GH18|CBM19	1
	GH75	1
LignoCellulose	CBM6	4
	GH12	2
	AA9	1
	CBM1	4
	CBM10|CBM2	1
	CBM2	13
	CBM63	1
	GH1	1
	GH5	14
	GH7|CBM1	2
	GH35	1
	GH43	14
	GH31	7
	GH27	4
	GH3	16
	GH38	1
	GH47	11
	GH67	1

Substrate	CAZy Family	Count
PCW	CBM32	2
	CBM6	4
	CBM52	1
Pectin	GH2	7
	GH43	14
	GH35	1
	CE12	2
	CE8	3
	GH28	7
	GH51	2
	GH53	1
	GH54|CBM42	7
	GH78	2
	GH93	3
	PL3	1
	PL4	2
	CBM13	18
Starch	GH31	7
	GH13	4
	CBM20	35
	CBM21	1
	CBM48	11
Hemicellulose	CBM6	4
	CBM2	13
	GH43	14
	GH3	16
	GH67	1
	CBM13	18
	CE1	6
	GH10	1
	GH12	2
	GH31	7
	GH29	1

Homologous protein sequence of *Mycosphaerella dearnessii* to enzymes involved in carbohydrate metabolism are assigned to various classes and their substrate.

Table 3.9 Genes of interest from other plant pathogenic fungi.

Number of M.dearnessii BLASTx Hits	Gene of Interest	Function
32	*Ace1*	Hybrid polyketide synthase/nonribosomal peptide synthetase
12	*Act*	*ACT-Toxin*
2	*Avr1*	Suppresses plant mediated resistance
1	*Avr2*	Protease inhibitor
1	*Avr3*	Expressed during colonization
1	*Avr4*	Chitin binding
1	*Avr9*	Charboxypeptidase
2	*Avra10*	Avirulent gene
2	*AvrL*	Expressed in hausoria
52	*AvrLm1*	Expressed in colonization
5	*AvrLm6*	Expressed in colonization
6	*AvrLm4-7*	Expressed in colonization
2	*AvrM*	Avirulent gene
6	*Avr-pia*	Avirulent to pia rice
6	*Avr-pita1*	Metalloporteases, avirulent to pi-ta rice
2	*Avr-pita2*	Metalloporteases, avirulent to pi-ta rice
1	*Avr-pita3*	Metalloporteases, avirulent to pi-ta rice
2	*Avr-piz*	Avirulent to piz gene
27	*Hts1*	Histidyl-tRNA synthase
6	*Nep1*	Pseudourine methyltransferase
6	*Nip1*	Non-specific toxin/induces necrosis and plasma membrane H+Atpase
1	*Nip3*	Non-specific toxin/induces necrosis
2	*Nrf1*	Nitrogen reponse
76	*PKS*	Polyketide synthase
44	*Pma1*	Plasma membrane proton pump
5	*Pma2*	Plasma membrane proton pump
5	*Rid*	RIP (receptor interacting serine threonine kinase) defective
2	*Six2*	Required for virulence
1	*Syn2*	SNARE complex
12	*Syn8*	SNARE complex
1	*ToxA*	Putative HC-toxin efflux pump
2	*ToxB*	HC-toxin biosynthesis
3	*ToxC*	Glutathione S-transferase
1	*ToxE*	Leucine zipper transcription
9	*ToxF*	Alkane-inducible cytochrome P450
4	*ToxG*	Alanine racemase
2	*ToxH*	HC-toxin biosynthesis

Particular genes of interest were selected from other plant pathogenic fungi for protein sequence similarity to the genome sequence of *Mycosphaerella dearnessii*.

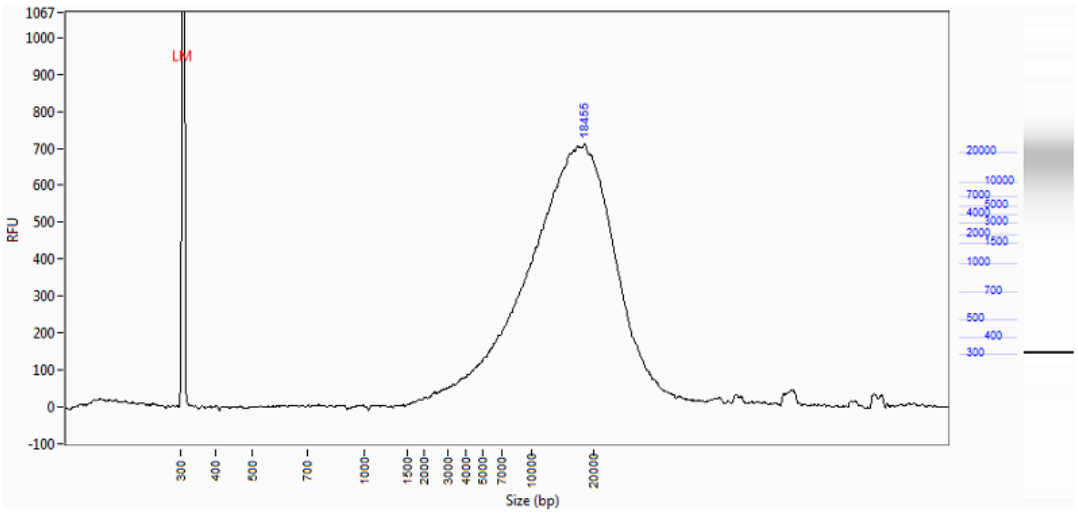

Figure 3.1 B3_MdHEF genomic DNA quality.

Genomic DNA was extracted from sample B3_MdHEF and analyzed on the Aglilent 2100 Bioanalyzer, which indicated a concentration of 3.8 ng/µL.

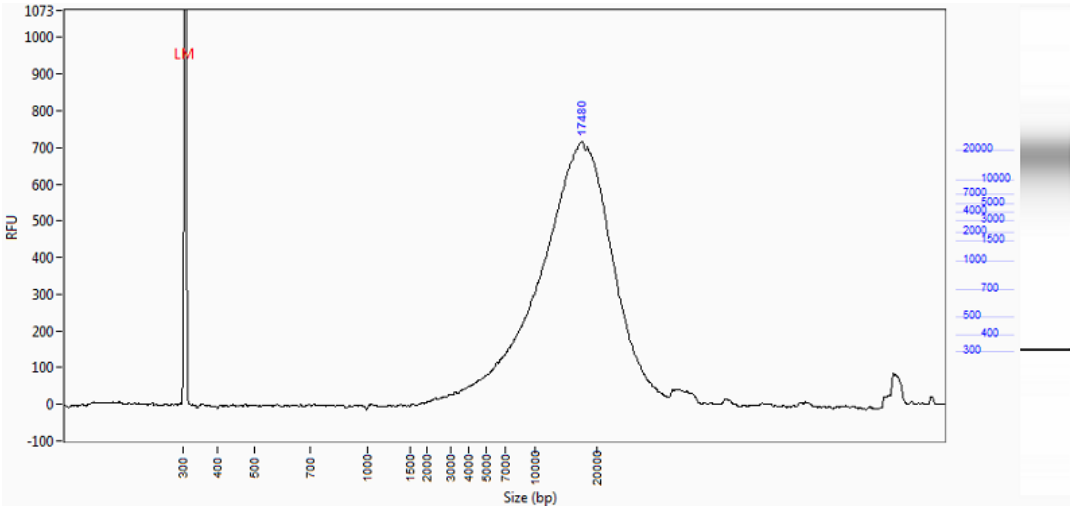

Figure 3.2 D3_MdHEF genomic DNA quality.

Genomic DNA was extracted from sample D3_MdHEF and analyzed on the Aglilent 2100 Bioanalyzer, which indicated a concentration of 4.9 ng/µL.

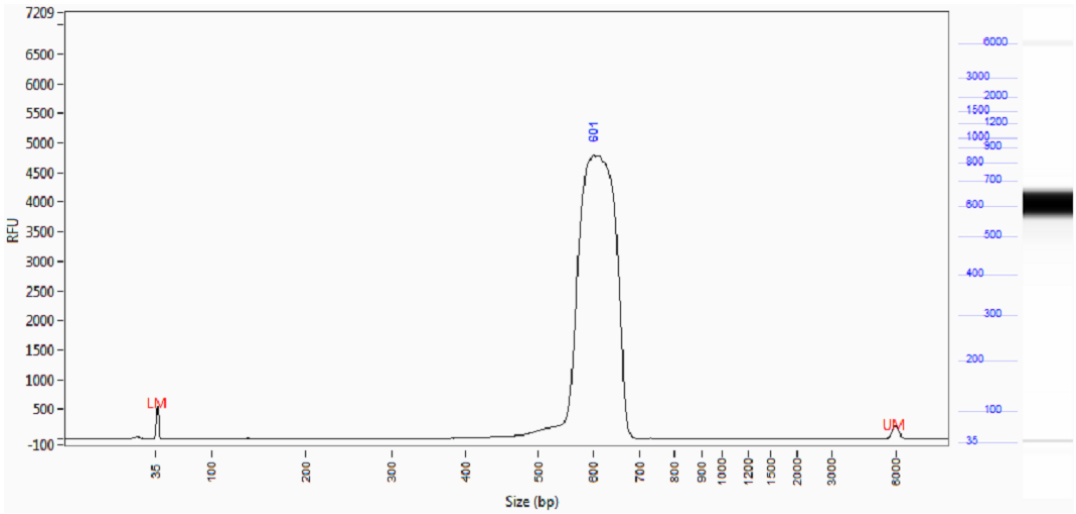

Figure 3.3 B3_MdHEF DNA library fragment size.

Pair-end library construction included size selection of 601 bp fragment length at a concentration of 9.8 ng/μL.

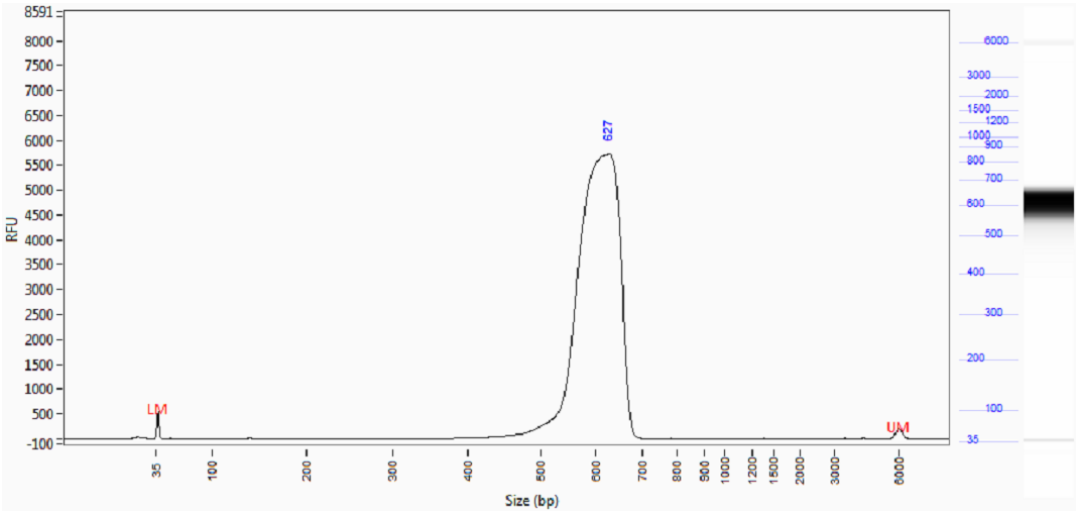

Figure 3.4 D3_MdHEF DNA library fragment size.

Pair-end library construction included size selection of 627 bp fragment length at a concentration of 13.2 ng/μL.

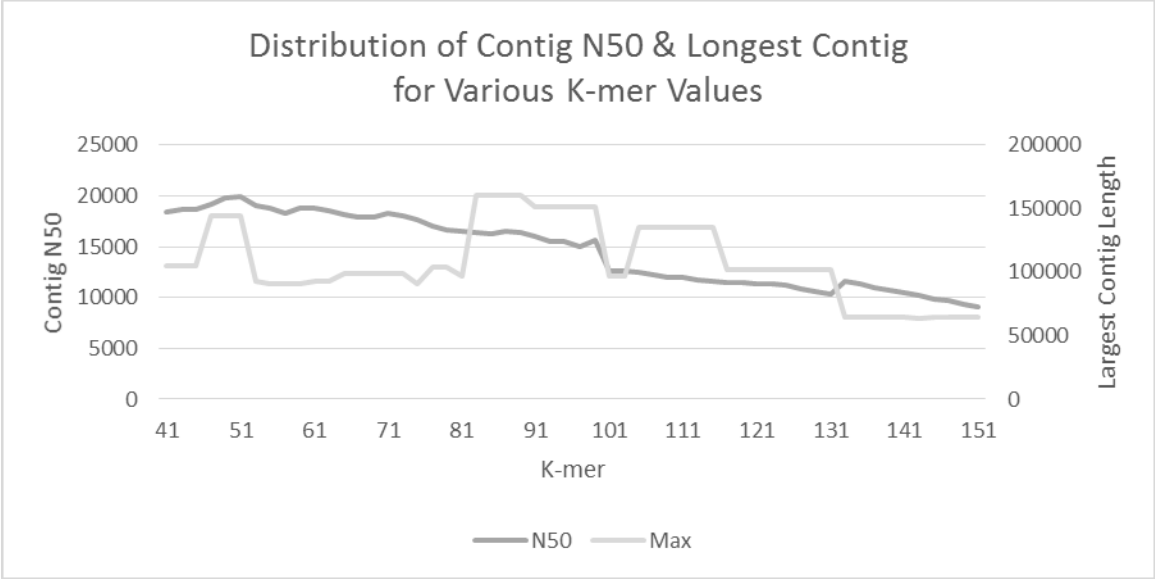

Figure 3.5 Distribution of contig N50 and longest contig length using Abyss.

Various k-mer lengths (41-151) were processed by the program Abyss to determine the optimal k-mer length for genome assembly. K-mer 51 produced the highest N50 length and reasonably long maximum contig length.

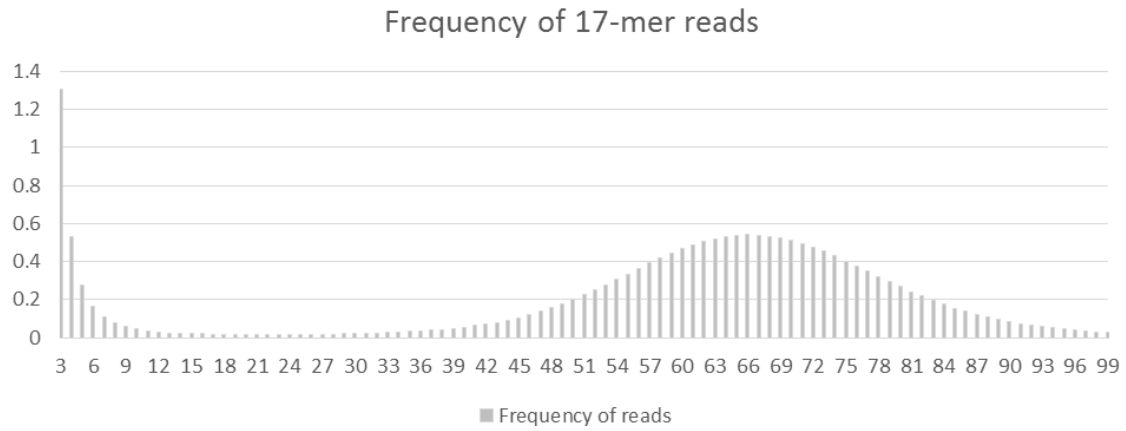

Figure 3.6 Frequency of 17-mer read length.

Read counter software (*Jellyfish*) counted every instance of a 17 nucleotide sequence found in the sequence data. The number of the most frequently found sequence (66) was used to calculate the genome size estimate.

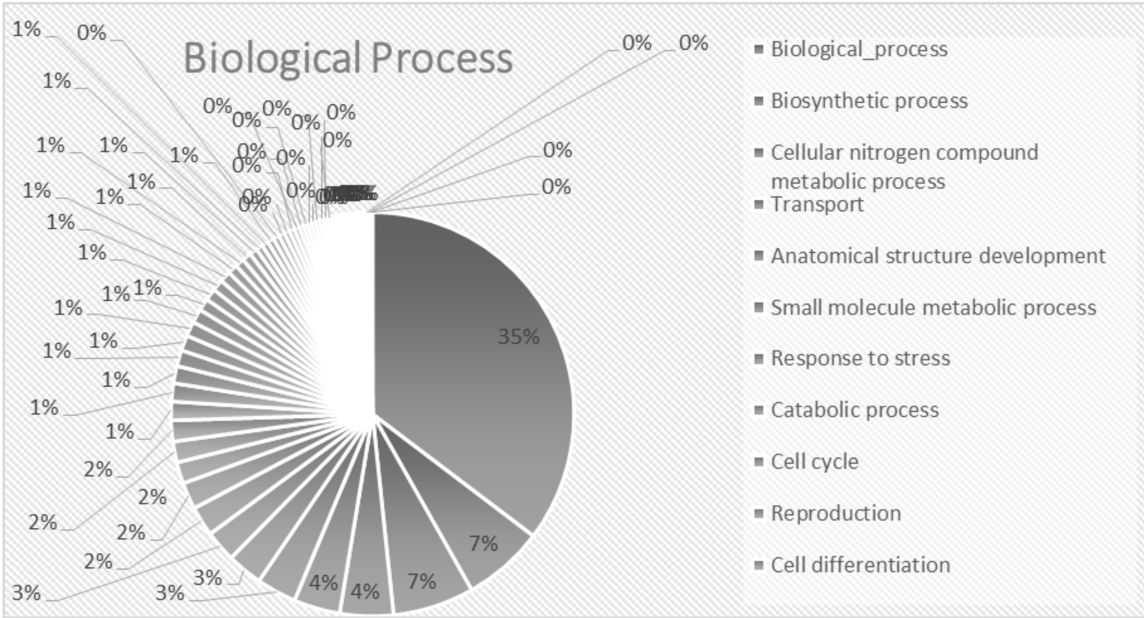

Figure 3.7 Gene ontology analysis of biological process.

Proteins predicted from the genome sequence of *Mycosphaerella dearnessii* were analyzed using AgBase Fungi database and assign functional annotation terms.

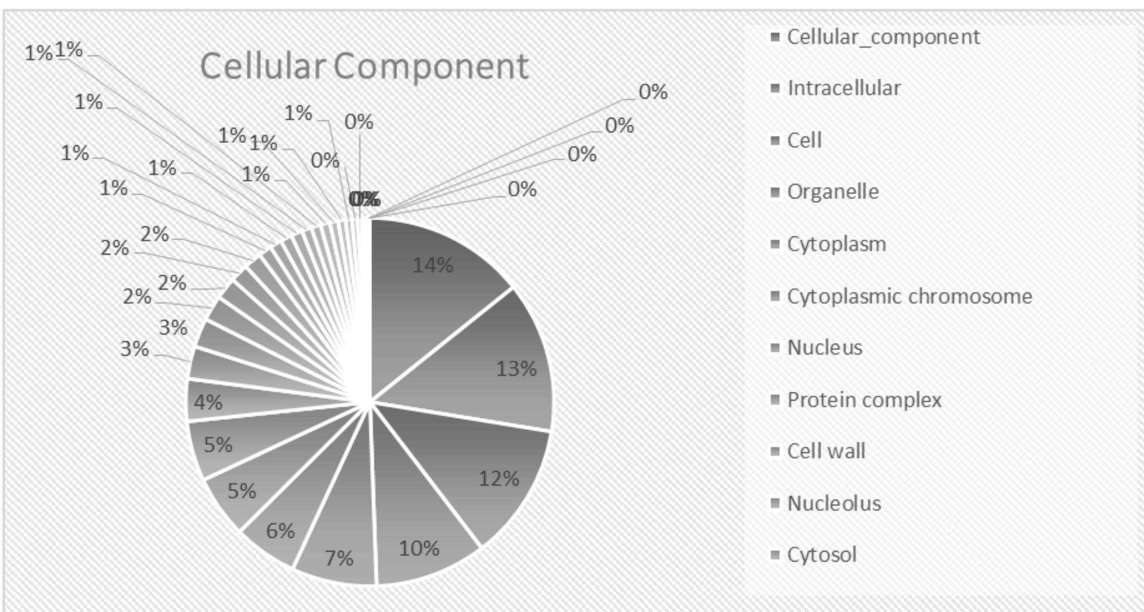

Figure 3.8 Gene ontology analysis of cellular component.

Proteins predicted from the genome sequence of *Mycosphaerella dearnessii* were analyzed using AgBase Fungi database and assign functional annotation terms.

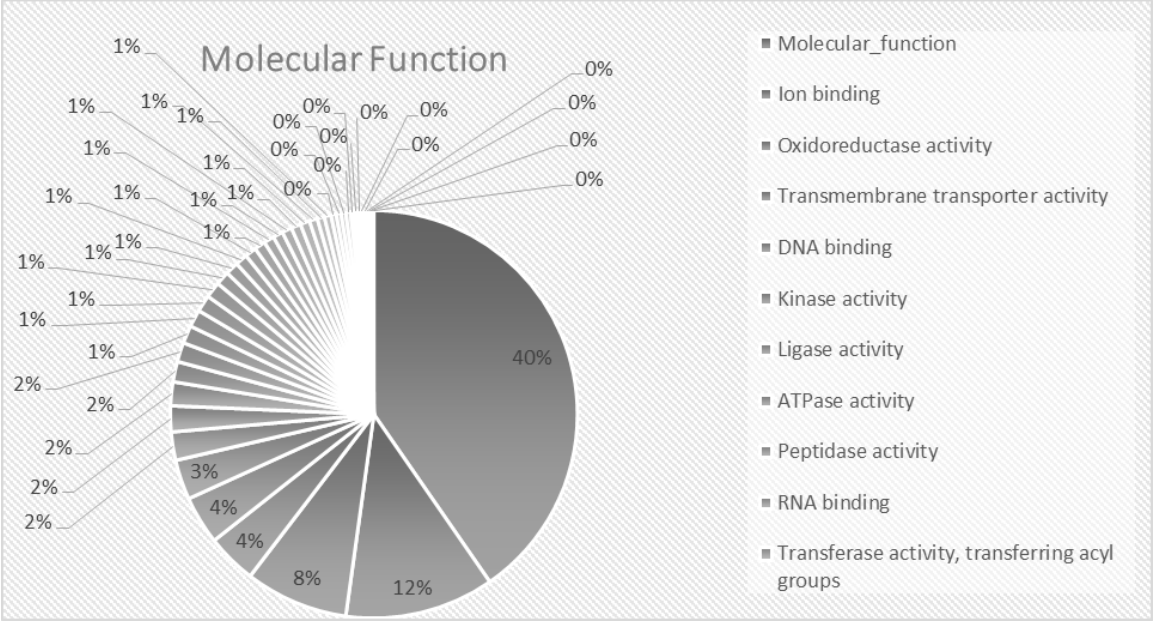

Figure 3.9 Gene ontology analysis of molecular function.

Proteins predicted from the genome sequence of *Mycosphaerella dearnessii* were analyzed using AgBase Fungi database and assign functional annotation terms.

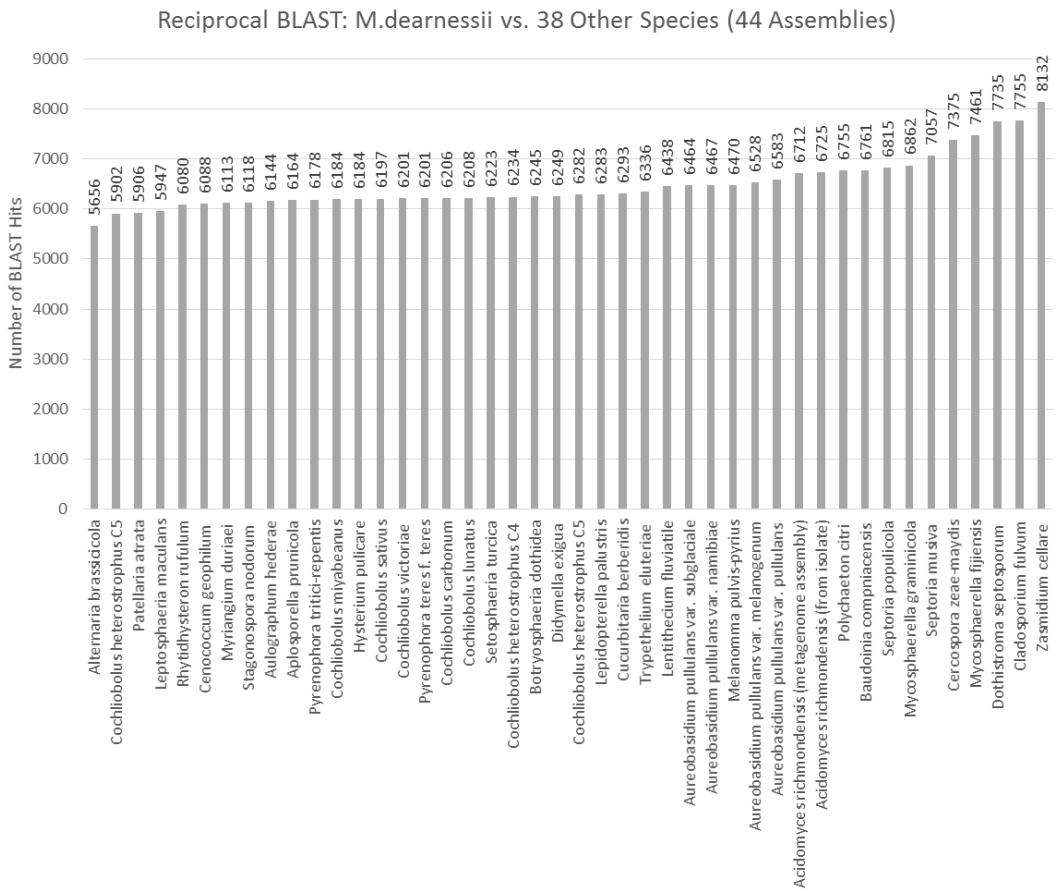

Figure 3.10 Reciprocal best BLAST hit with 38 *Dothideomycetes* species.

The top hit of a BLAST search between *Mycosphaerella dearnessii* and each of the 38 other species matches the top BLAST hit of the reciprocal search.

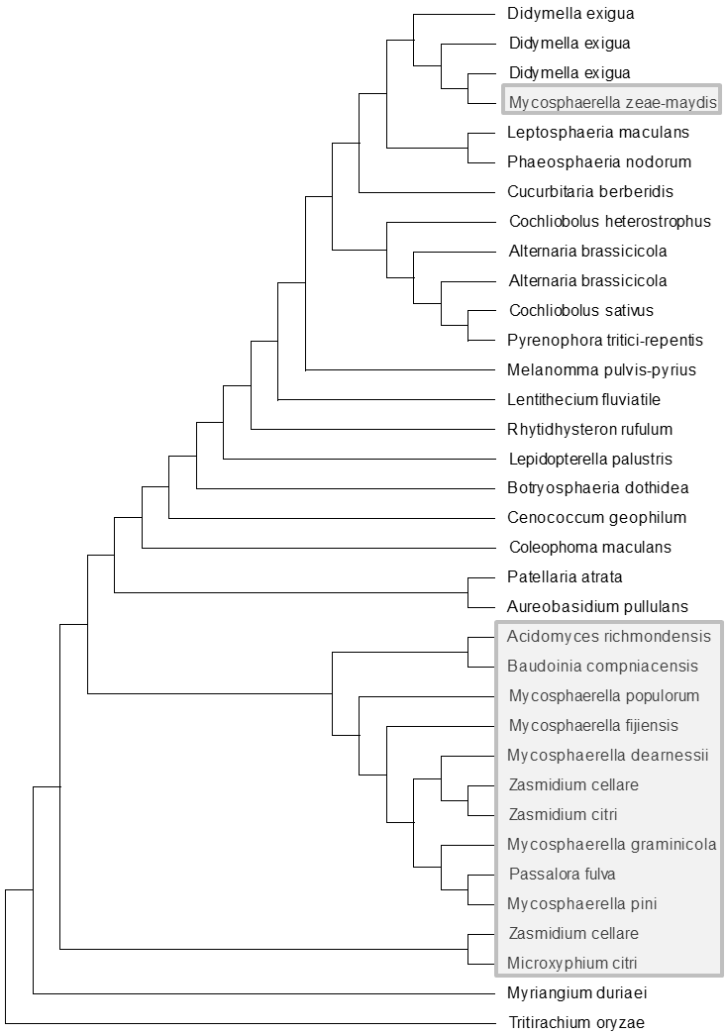

Figure 3.11 Phylogenetic analysis of 31 species within the Class *Dothideomycetes*.

All organisms belong to the Class *Dothideomycetes*, except *Tritirachium oryzae* which was selected as the outgroup belonging to the *Basidiomycota* phyla. Grey boxes indicate closely related species within the Order *Capnodiales*.

CHAPTER IV

CONCLUDING REMARKS

Mycosphaerella dearnessii is a filamentous fungal phytopathogen on many species of pine trees with a worldwide distribution. One host (*Pinus palustris*) is of particular importance, in which the disease brown spot needle blight is caused by *M. dearnessii*. Longleaf pine (*P. palustris*) is poised to increase in planted acreage across the southeastern United States. The promise of greater timber harvests with superior wildlife and ecosystem habitat encourage land managers to adopt longleaf pine. However, many obstacles confront land managers that decide to plant longleaf pine. The brown spot needle blight disease is the most devastating foliar fungal pathogen on longleaf pine. The grass stage of development is particularly susceptible to this disease and severe infection can lead to mortality. The rate of survival for planted seedlings will need to increase to persuade more land managers to grow longleaf pine. Brown spot needle blight has the potential to hinder progress towards adopting longleaf pine as the preferred pine species for planting in the Southeast U.S.

In order to combat the disease, resistant cultivars of longleaf pine will be developed. All diseases eventually overcome plant resistance; but determining how quickly pathogen resilience is acquired depends on knowledge of population structure. Observing changes within populations in which control measures have been implemented, will evaluate the effectiveness of the treatment and help predict when the

pathogen will overcome host resistance. Alternatively, highly virulent and infectious pathogen genotypes may be scarce in the population. Resistant host cultivars may only be resistant against the less virulent strains. Planting resistant varieties extensively could place selective pressure on the pathogen population; which might inadvertently select for the most virulent strains to increase in the population. Carefully sampling pathogen populations will guide decisions on where to plant the most resistant host varieties and where those varieties are not necessary. Distributing certain varieties based upon pathogen pressure can reduce the chances of highly virulent strains becoming widespread.

Polymorphic microsatellite markers for *M. dearnessii* have recently been developed to study population genetics. Results in our study show a genetically diverse population in a small sample set. It is conceivable that the numerous host species of this pathogen is a consequence of a highly variable genetic background within the species. Likewise, there is presumably variation of infection rates and virulence with different strains of the fungus. Future studies will need to distinguish which strains are more virulent and/or highly infectious. Comparing the different strains will identify the genetic environment that controls infection rate and virulence.

Additionally, the vegetative incompatibility experiments confirmed the conclusion that the population is highly variable. This result is a distinct verdict because microsatellite markers are selectively neutral, whereas vegetative incompatibility genes may have selective pressures. The variation observed in the population presumably comes from sexual outcrossing. However, other methods to share genetic information could predominate (i.e. – heterokaryosis). The processes involved in self/non-self

recognition during heterokaryon formation conceivably shares aspects with host recognition. Discovering those common aspects will lead to a better understanding of the infection process and ways to prevent infection.

Research devoted to this disease has diminished in the last several decades. Scarce resources exist for understanding how the fungus is able to cause infection and develop disease symptoms in the host. New technologies have emerged that enable enormous amounts of data to be generated with reduced effort. The genome sequence of *M. dearnessii* will contribute to the biological understanding of this organism. Preliminary genome analysis presented here has proposed certain fungal characteristics that have not been explicitly defined in this organism. The delayed onset of disease symptoms observed in *P. palustris* suggested a hemibiotrophic (biotrophic and necrotrophic) lifestyle for the *M. dearnessii*. Sequence similarity to known effector genes has provided evidence for a biotophic stage. Additionally, numerous plant cell wall degrading enzymes were identified, which are necessary for a necrotrophic pathogen. Effectors help the pathogen evade plant defenses and theoretically have analogous molecules in the host cell. Recently, experiments have been proposed to utilize pathogen effectors as probes for host resistant molecules; accelerating breeding performance. Eventually molecules can be designed to target effectors and genes encoding those molecules can be expressed in the host genome.

Infection begins with the formation of appressorium to locate a suitable host. Several genes were detected that play a role in modifying polymers within fungal cell walls, permitting appressorium development. Furthermore, appressorium presumably have ways to locate and recognize stomata. Understanding the mechanism behind host

recognition can lead to developments in plant resistance that disrupt those mechanisms. Haustoria are specialized fungal structures that conceivably secrete proteins and/or toxins for nutrient acquisition. Potential candidate toxins and proteins with signal peptides were identified in the genome sequence. Acquiring knowledge about these molecules has revealed how the pathogen is able to procure nutrients from its host.

The genome assembly is far from perfect and requires more sequence data to create scaffolds. Eventually physical mapping will be the only way to comprehensively determine the sequence orientation. Gene prediction will only improve with transcription and translational experiments. Yet, a draft genome sequence has provided a wealth of information about an organism that has not received sufficient attention in the past decades. In order to meet the demands of increased longleaf pine acreage, a disease management strategy will need to be developed. A better understanding of the pathogens biology will guide effective and efficient approaches to disease control. This study has extended biological information available for *Mycosphaerella dearnessii*.

REFERENCES

A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org

Altenhoff, A. M. & Dessimoz, C. (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Computational Biology*, 5 (1), e1000262.

Amselem, J., *et al.* (2011). Genomic analysis of the necrotrophic fungi pathogens *Sclerotinia sclerotiorum* and *Botryis cinerea*. *PLoS Genetics*, 7 (8), e1002230.

Anagnostakis, S. L. (1977). Vegetative incompatibility in *Endothia parasitica*. Experimental Mycology, 1, 306-16.

Anagnostakis, S. L., Waggoner, P. E. (1981). Hypovirulence, vegetative incompatibility, and the growth of cankers of chestnut blight. *Phytopathology,* 71, 1198-202.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics and Development*, 11, 635-641.

Banke, S., Peschon, A., McDonald, B. A., (2004). Phylogenetic analysis of globally distributed *Mycosphaerella graminicola* populations based on three DNA sequences. *Fungal Genetics*, 41, 226-238.

Barnett, J. P. Southern forest resource conditions and management practices: 1900-1950: Benefits from research. In H. M. Rauscher & K. Johnson (Eds.) *Southern Forest Science: Past, Present, and Future,* (pp. 15-22). Southern Research Station.

Bégueret J., Turcq B., Clavé C. (1994). Vegetative incompatibility in filamentous fungi: *het* genes begin to talk, *Trends in Genetics*, 10, 441–446.

Bistis, G. N. (1998). Physiological heterothallism and sexuality in Euascomycetes: a partial history. *Fungal Genetics and Biology*, 23, 213-222.

Bradshaw, R. E., Guo, Y., Sim, A. D., Kabir, M. S., Chettri, P., Ozturk, I. K., Hunziker, L., Ganley, R. J., Cox, M. P. (2015). Genome-wide gene expression dynamics of the fungal pathogen *Dothistroma septosporum* throughout its infection cycle of the gymnosperm host *Pinus radiate*. *Molecular Plant Pathology*

Brennan, L. A., Engstrom, R. T., Palmer, W. E., Hermann, S. M., Hurst, G. A., Burger, L. W., and Hardy, C. L. (1998). Whither wildlife without fire? *Proceedings of the Trans. 63rd North American Wildlife and Natural Resources Conference*, (pp. 402-414). Washington, D. C. Wildlife Management Institute.

Brennan, L. A., Hermann, S. M. (1994). Prescribed fire and forest pests: solutions for today and tomorrow. *Journal of Forestry,* 92(11), 34-37.

Brockway, D. G., Outcalt, K. W., Tomczak, D. J., & Johnson, E. E. (2006). Restoration of longleaf pine ecosystems. *General Technical Report SRS-83.* Southern Research Station.

Bruce, D. (1951). Fire, site, and longleaf height growth. *Journal of Forestry*, 49(1), 25-28.

Cai, G. & Schneider, R. W. (2005). Vegetative compatibility groups in *Cercospora kikuchii* the causal agent of cercospora leaf blight and purple seed stain in soybean. *Phytopathology,* 95(3), 257-261.

Caten, C. E., Newton, A. C. (2000). Variation in cultural characteristics, pathogenicity, vegetative compatibility and electrophoretic karyotype within field populations of *Stagonospora nodorum. Plant Pathology,* 49, 219-226.

Chartouni, L. E., Tisserant, B., Siah, A., Duyme, F., Leducq, J. B., Deweer, C., Roisin, C. F., Sanssene, J., Durand, R., Halama, P., Reignault, P. (2011). Genetic Diversity and population structure in French population of *Mycosphaerella graminicola. Mycologia*, 103 (4), 764-774.

Christensen, N. L. (2000). Vegetation of the Southeastern Coastal Plain. In M. G. Barbour and W. D. Billings (Eds.) *North American Terrestrial Vegetation, Second Edition*, (pp. 397-448). London. Cambridge University Press.

Churchill, A. C. L. (2011). *Mycosphaerella fijiensis,* the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and challenges of control. *Molecular Plant Pathology*, 12 (4), 307-328.

Comeau, A. M., Dufour, J., Bouvet, G. F., Jacobi, V., Nigg, M., Henrissat, B., Laroche, J., Levesque, R. C., Bernier, L. (2014). Functional annotation of the *Ophiostoma novo-ulmi* genome: insights into the phytopathogenicity of the fungal agent of dutch elm disease. *Genome Biology and Evolution,* 7 (2), 410-430.

Coppin, E., Debuchy, R., Arnaise, S., Picard, M. (1997). Mating types and sexual development in filamentous ascomycetes. Microbiology and *Molecular Biology Reviews*, 61(4), 411-428.

Cortesi, P., & M. G. Milgroom. (1998). Genetics of vegetative incompatibility in *Cryphonectria parasitica. Applied Environmental Microbiology,* 64, 2988–2994.

Croker, T. C., Boyer, W. D. (1975). Regenerating longleaf pine naturally. *U.S. Forest Service Research Note SO-105*. Southern Forest Experiment Station.

Crous, P. W., Schoch, C. L., Hyde, K. D., Wood, A. R., Gueidan, C., Hoog, G. S., Groenewald, J. Z. (2009a). Phylogenetic lineages in *Capnodiales*. *Studies in Mycology*, 64, 17-47.

Crous, P. W., Summerell, B. A., Carnegie, A. J., Wingfield, M. J., Hunter, G. C., Burgess, T. I., Andjic, V., Barber, P. A., Groenewald, J. Z. (2009b). Unraveling *Mycosphaerella*: do you believe in genera? *Persoonia*, 23, 99-118.

Dahl, T. E. (1990). Wetlands losses in the United States 1780's to1980's. Washington, D.C. *U.S. Department of the Interior, Fish and Wildlife Service*.

Dahl, T. E. (2011). Status and trends of wetlands in the conterminous United States 2004 to 2009. Washington, D.C. *U.S. Department of the Interior, Fish and Wildlife Service.*

Dalquen, D. A. & Dessimoz, C. (2013). Bidirectional best hits miss many orthologs in duplication-rich clades such as plant and animals. *Genome Biology and Evolution,* 5 (10), 1800-1806.

DeGiorgio, M., & Rosenberg, N. A. (2008). An unbiased estimator of gene diversity in samples containing related individuals. *Molecular Biology and Evolution*, 26 (3), 501-512.

de Wit, P. J., *et al.* (2012). The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaption to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genetics*, 8 (11), e1003088.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics Review*, 14(9), 755-763.

Ekblom, R. & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7, 1026-1042.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, N. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols*, 2(4), 953-971.

Engstrom, R. T., Crawford, R. L. & Baker, W. W. (1984). Breeding bird populations in relation to changing forest structure following fire exclusion: A 15-year study. *The Wilson Bulletin*. 96, 437-450.

Frost, C. C. (1993). Four centuries of changing landscape patterns in the longleaf pine ecosystem. *Proceedings of the Tall Timbers Fire Ecology Conference. The Longleaf Pine Ecosystem: Ecology, Restoration, and Management,* 18 (pp. 17-43). Tallahassee, Florida. Tall Timbers Research Station.

Fullwood, M. J., Wei, C. L., Liu, E. T., Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analysis. *Genoeme Research*, 19, 521-532

Garnjobst, L., & J. F. Wilson. (1956). Heterokaryosis and protoplasmic incompatibility in *Neurospora crassa. Proceedings of the National Academy of Sciences*, 2, 613–618.

Giraldo, M. C., & Valent, B. (2013). Filamentous plant pathogen effectors in action. *Nature Reviews Microbiology*, 11, 800-814.

Glass, N. L. & Dementhon, K. (2006). Non-self recognition and programmed cell death in filamentous fungi. *Current Opinion in Microbiology*, 9, 553-558.

Glass, N. L., and G. A. Kuldau. (1992). Mating type and vegetative incompatibility in filamentous ascomycetes. *Annual Review of Phytopathology*, 30, 201–224.

Gonzalez-Perez, J. A., Gonzalez-Vila, F. J., Almendros, G., & Knicker, H. (2004). The effect of fire on soil organic matter – a review. *Environment international,* 30 (pp. 855-870).

Guo, L. T., Wang, S. L., Wu, Q. J., Zhou, X. G., Xie, W., Zhang, Y. J. (2015). Flow cytometry and k-mer analysis estimates of the genome sizes of *Bemisia tabaci* B and Q (Hemiptera: Aleyrodidae). *Frontiers in Physiology*, 6 (144).

Guyer, C., & Bailey, M. A. (1993). Amphibians and reptiles of longleaf pine communities. *Proceedings of the Tall Timbers Fire Ecology Conference. The Longleaf Pine Ecosystem: Ecology, Restoration, and Management,* 18 (pp. 138-158). Tallahassee, Florida. Tall Timbers Research Station.

Harcombe, P. A., Glitzenstein, J. S., Knox, R. G., Orzell, S. L., and Bridges, E. L. (1993). Vegetation of longleaf pine region of the gulf west coast. *Proceedings of the Tall Timbers Fire Ecology Conference. The Longleaf Pine Ecosystem: Ecology, Restoration, and Management,* 18 (pp. 83-104). Tallahassee, Florida. Tall Timbers Research Station.

Hayden, H. L., Carlier, J., Aitken, E. A. B. (2003a). Genetic structure of *Mycosphaerella fijiensis* populations from Australia, Papua New Guinea, and the Pacific Islands. *Plant Pathology,* 52, 703-712.

Hayden, H. L., Carlier, J., Aitken, E. A. B. (2003b). Population differentiation in the banana leaf spot pathogen *Mycosphaerella musicola*, examined at a global scale. *Plant Pathology*, 52, 713-719.

Hayden, H. L., Carlier, J., Aitken, E. A. B. (2005). The genetic structure of Australian populations of *Mycosphaerella musicola* suggests restricted gene flow at the continental scale. *Population Biology,* 95 (5), 489-498.

Hillman, B. I., Shapira, R., Nuss, D. L. (1990). Hypovirulance-associated suppression of host functions in *Cryphonectria parasitica* can be partially revlieved by high light intensity. *Phytopathology*, 80(10), 950-956.

Huang, Z. Y., Smalley, E. B., Guries, R. P. (1995). Differentiation of *Mycosphaerella dearnessii* by cultural characters and RAPD analysis. *Phytopathology*, 85(5), 522-527.

Hull, R. B. & Robertson, D. P. (2000). The language of nature matters: we need more public ecology. In P. H. Gobster & R. B. Hull (Eds.) *Restoring nature: perspective from the social sciences and humanities,* (pp. 97-118). Washington, D. C. Island Press.

Hunter, M. L. (1990). *Wildlife, forests and forestry: principles of managing forests for biological diversity,* (pp. 370). New Jersey, NJ. Prentice Hall.

Idnurm, A. & Howlett, B. J. (2001). Pathogenicity genes of phytopathogenic fungi. *Molecular Plant Pathology*, 2 (4), 241-255.

Ishikawa, F. H., Souza, E. A., Shoji, J., Connolly, L., Freitag, M., Read, N. D., Roca, M. G. (2012). Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. *Public Library of Science One*, 7(2), e31175.

Janousek, J., Krumbock, S., Kirisits, T., Bradshaw, R. E., Barnes, I., Jankovsky, L., Stauffer, C. (2013). Development of microsatellite and mating type markers for the pine needle pathogen *Lecanosticta acicula*. *Australasian Plant Pathology*.

Kais, A. G. (1971). Dispersal of *Scirrhia acicula* spores in southern Mississippi. *Plant Disease Reporter*, 55(4), 309-311.

Kemen, E. & Jones, J. D. G. (2012). Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, 17 (8), 448-457.

Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49 (4), 725-738.

Kofler, R., Schlotterer, C., Lelley, T., (2007). SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, 23(13), 1683-1685.

Kronstad, J.W., Staben, C., (1997). Mating type in filamentous fungi. *Annual Review of Genetic*s, 31, 245–276.

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

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L., (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), 25.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948.

Leslie, J. F. (1993). Fungal vegetative compatibility. *Annual Reviews of Phytopathology*, 31, 127-150.

Lewontin, R. C. (1972). Testing the theory of natural selection. *Nature*, 236, 181-182.

Li, C. Y., Korol, A. B., Fahima, T., Beiles, A., Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11, 2453-2465.

Li, R., *et al.* (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature*, 463, 311-317.

Linde, C. C., Zhan, J., McDonald, B. A. (2002). Population structure of *Mycosphaerella graminicola* from lesions to continents. *Ecology and Population Biology*, 92 (9), 946-955.

Litt, A. R., Provencher, L., Tanner, G. W., Franz, R. (2001). Herpetofaunal responses to restoration treatments of longleaf pine sandhills in Florida. *Restoration Ecology,* 9(4), 462-474.

Liu, Y. C. & Milgroom, M. G. (2007). High diversity of vegetative compatibility types in *Cryphonectria parasitica* in Japan and China. *Mycologia,* 99(2), 279-284.

Lombard, V., Golaconda, R. H., Drula, E., Coutinho, P. M., Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42, D490-D495.

Loubradou, G. & Turcq, B. (2000). Vegetative incompatibility in filamentous fungi: a roundabout way of understanding the phenomenon. *Research in Microbiology*. 151, 239-245.

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

Mahe, S., Duhamel, M., Le Calvez, T., Guillot, L., Sarbu, L., Bretaudeau, O. C., Dufresne, A., Kiers, E. T., Vandenkoornhuyse, P. (2012). PHYMYCO-DB: a curated database for analyses of fungal diversity and evolution. *PLoS ONE*, 7(9), e43117.

Maple, W. R. (1975). Mortality of longleaf pine seedlings following a winter burn against brown-spot needle blight. *U.S. Forest Service Research Note SO-195*. Southern Forest Experiment Station.

Marcais, G. & Kingsford, C. (2010). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764-770.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10-12.

May, C., Guibert, M., Leclerc, A., Andrivon, D., Tivoli, B. (2012). A single, plastic population of *Mycosphaerella pinodes* causes ascochyta blight on winter and spring peas (*Pisum sativum*) in France. *Journal of Applied and Environmental Microbiology*, 78 (23), 8431-8440.

McCarthy, F. M., Wang, N., Magee, G. B., Nanduri, B., Lawrence, M. L., Camon, E. B., Barrell, D. G., Hill, D. P., Dolan, M. E., Williams, W. P., Luthe, D. S., Bridges, S. M., Burgess, S. C. (2006). AgBase: a functional genomics resource for agriculture. *BMC Genomics*, 7, 229.

Mehrabi, R., Bahkali, A. H., Abd-Elsalam, K. A., Moslem, M., M'Barek, S. B., Gohari, A. M., Jashni, M. K., Stergiopoulos, I., Kema, G. H. J., Wit, P. J. G. M. (2011). Horizontal gene and chromosome transfer in plant pathogenic fungi. *FEMS Microbiology Reviews,* 35, 542-554.

Milgate, A. W., Vaillancourt, R. E., Mohammed, C., Powell, M., Potts, B. M. (2005). Genetic structure of *Mycosphaerella cryptica* population. *Australasian Plant Pathology*, 34, 345-354.

Milgroom, M. G., & P. Cortesi. (1999). Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proceedings of the National Academy of Sciences*, 96, 10518–10523.

Morran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C., Lively, C. M. (2011). Running with the red queen: host-parasite coevolution selects for biparental sex. *Science,* 333, 216-218.

Mylyk, O. M. (1976). Heteromorphism for heterokaryon incompatibility genes in natural populations of *Neurospora crassa*. *Genetics,* 83, 275–284.

Nagarajan, N. & Pop, M. (2013). Sequence assembly demystified. *Nature Reviews Genetics*, 14, 157-167.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70, 3321-3323.

Neiuwenhuis, B. P. S. & Aanen, D. K. (2012). Sexual selectin in fungi. *Journal of Evolutionary Biology*, 25, 2397-2411.

Ni, M., Feretzaki, M., Sun, S., Wang, X., Heitman, J. (2012). Sex in fungi. *Annual Review in Genetics*, 45, 405-430.

Norton, B.G. (1995). Ecological Integrity and Social Values: At What Scale? *Ecosystem Health,* 1(4), 228-241.

Ohm, R. A., *et al.* (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathogens*, 8 (12), e1003037.

Oliver, R. (2012). Genomic tillage and the harvest of fungal phytopathogens. *New Phytologist*, 196 (4), 105-1023.

Oliver, R. P. & Solomon, P. S. (2010). New developments in pathogenicity and virulence of necrotrophs. *Current Opinions in Plant Biology*, 13, 415-419.

Orthon, E. S., Deller, S., Brown, J. K. M. (2011). *Mycosphaerella graminicola*: from genomics to disease control. *Molecular Plant Pathology*, 12 (5), 413-424.

Oswalt, C. M., Cooper, J. A., Brockway, D. G., Brooks, H. W., Walker, J. L., Connor, K. F., Oswalt, S. N., Conner, R. C. (2012). History and current condition of longleaf pine in the Southern United States. *General Technical Report SRS-166.* Southern Research Station.

Ottmar, R. D., Schaaf, M. D., Alvarado, E. (1996). Smoke considerations for using fire to maintain healthy forest ecosystems. *General Technical Report INT-GTR-341.* Intermountain Research Station.

Palik, B. J., Mitchell, R. J., Houseal, G., Pederson, N. (1997). Effects of canopy structure on resource availability and seedling responses in a longleaf pine ecosystem. *Canadian Journal of Forestry.,* 27, 1458-1464.

Park, B. H., Karpinets, T. V., Syed, M. D., Leuze M. R., Uberbacher, E. C. (2010). CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology*, 20(12), 1574-1584.

Parra, G., Bradnam, K., Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), 1061-1067.

Peet, R. K. (2006). Ecological classification of longleaf pine woodlands. In S. Jose, E. J. Jokela, & D. L. Miller (Eds.) *The longleaf pine ecosystem: ecology, silviculture, and restoration,* (pp. 51-94). New York, NY. Springer.

Peet, R. K. & Allard, D. J. (1993). Longleaf pine vegetation of the southern Atlantic and eastern gulf coast regions: a preliminary classification. *Proceedings of the Tall Timbers Fire Ecology Conference. The Longleaf Pine Ecosystem: Ecology, Restoration, and Management,* 18 (pp. 45-81). Tallahassee, Florida. Tall Timbers Research Station.

Peet, R. K., Carr, S., & Gramling, J. (2006). Fire adapted pineland vegetation of northern and central Florida: a framework for inventory, management, and restoration. *Florida Fish and Wildlife Commission.*

Peixouto, Y. S., Braganca, C. A. D., Andrade, W. B., Ferreira C. F., Haddad, F., Oliveira S. A. S., Brito, F. S. D., Miller, R. N. G., Amorim, E. P. (2015). Estimation of genetic structure of a *Mycosphaerella musicola* population using inter-simple sequence repeat markers. *Genetics and Molecular Research*, 14 (3), 8046-8057.

Perlin, M. H., Amselem, J., Fontanillas, E., Toh, S. S., Chen, Z., Goldberg, J., Duplessis, S., Henrissat, B., Young, S., Zeng, Q., Aguileta, G., Petit, E., Babouin, H., Andrews, J., Razeed, D., Gabaldon, T., Quesneville, H., Giraud, T., Hood, M. E., Schultz, D. J., Cuomo, C. A. (2015). Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus.

Poggeler, S. (2001). Mating-type genes for classical strain improvements of ascomycetes. *Applied Microbiology and Biotechnology*, 56, 58-601.

Presti, L. L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Phytopathology*, 66, 513-545.

Provencher, L., Herring, B. J., Gordon, D. R., Rodgers, H. L., Tanner, G. W., Hardesty, J. L., Brennan, L. A., Litt, A. R., (2001). Longleaf pine and oak responses to hardwood reduction in fire-suppressed sandhills in northwest Florida. *Forest Ecology and Management*, 148, 63-77.

Pyne, S. (1997). *Fire in America: a cultural history of Wildland and Rural fire 2nd ed.* (pp. 680) Seattle, WA. University of Washington Press.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences, and Illumina MiSeq sequencers. *BMC Genomics*, 13, 341.

Rivas, G. G., Zapater, M. F., Abadie, C., Carlier, J. (2004). Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis. Molecular Ecology,* 13, 471-482.

Russel, K. R., D. H. V. Lear, & J. D. C. Guynn. (1999). Prescribed fire effects on herpetofauna: review and management implications. *Wildlife Society Bulletin,* 30 (pp. 354-360).

Saelens, X., Festjens, N., Parthoens, E., Vanoverberghe, I., Kalai, M., Kuupevels, F., Vandenabeele, P. (2005). Protein synthesis persists during necrotic cell death. *The Journal of Cell Biology*, 169(4), 545-551.

Saupe, S. J. (2000). Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews*, 64(3), 489-502.

Saupe, S. J., Clave, C., Bégueret, J. (2000). Vegetative incompatibility in filamentous fungi: *Podospora* and *Neurospora* provide some clues. *Current Opinions in Microbiology,* 3, 608-612.

Schlotterer, C. (2003). Hitchhiking mapping – functional genomics from the population genetics perspective. *Trends in Genetics*, 19(1), 32-38.

Schmieder, R. & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomics datasets. *PLoS One,* 6(3), e17288.

Schoch, C. L. *et al.* (2009). A class-wide phylogenetic assessment of *Dothideomycetes*. *Studies in Mycology*, 64, 1-15.

Schultz, R. P. (1997). *Loblolly Pine: the ecology and culture of loblolly pine (Pinus taeda L.)* Washington, D. C. United States Forest Service.

Setliff, E. C. & Patton, R. F. (1974). Germination behavior of *Scirrhia acicula* conidia on pine needles. *Phytopathology*, 64, 1462-1464.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research*, 19, 117-1123.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., Ponting, C. P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics,* 15, 121-132.

Shiu, P. K. T. & Glass, N. L. (2000). Cell and nuclear recognition mechanisms mediated by mating type in filamentous ascomycetes. *Current Opinion in Microbiology*, 3, 183-188.

Smith, D. M., Larson, B. C., Kelty, M. J., Ashton, P. M. S. (1997). *The Practice of Silviculture: Applied Forest Ecology 9th ed.* New York, NY. John Wiley and Sons, Inc.

Snow, G. A. (1961). Artificial inoculation of longleaf pine with *Scirrhia acicula*. *Phytopathology,* 51, 186-188.

Stergiopoulos, I., Collermare, J., Mehrabi, R., de Wit, J. G. M. (2013). Phytotoxic secondary metabolites and peptides produced by plant pathogenic *Dothideomycete* fungi. *Federation of European Microbiological Review,* 37, 67-93.

Stergiopoulos, I. & de Wit, J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47, 233-263.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution,* 28, 2731-2739.

Tang, J. D., Perkins, A. D., Sonstegard, T. S>, Schroeder, S. G., Burgess, S. C., Diehl, S. V. (2012). Short-read sequencing for genomic analysis for the brown rot fungus *Fibroporia radiculosa*. *Applied Environmental Microbiology*, 78 (7), 2272-2281.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4 (1), 41.

Telfer, E., Graham, N., Stanbra, L., Manley, T., Wilcox, P. (2013). Extraction of high purity genomic DNA from pine for use in a high-throughput genotyping platform. *New Zealand Journal of Forestry Science*, 43, 3.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., Borodovsky, M. (2008). Gene prediction in novel fungal genomes using ab initio algorithm with unsupervised training. *Genome Research*, 18, 1979-1990.

Thrall, P. H., & Burdon, J. J. (2003). Evolution of virulence in a plant host-pathogen metapopulation. *Science.*, 299, 1735-1737.

Varner, J. M., Gordon, D. R., Putz, F. E., Hiers, J. K. (2005). Restoring fire to long-unburned *Pinus palustris* ecosystems: novel fire effects and consequences for long-unburned ecosystems. *Restoration Ecology,* 13(3), 536-544.

Varner, J. M. & Kush, J. S. (2004). Remnant old-growth longleaf pine (*Pinus palustris* Mill.) savannas and forests of the southeastern USA: status and threats. *Natural Areas Journal,* 24, 141-149.

Varner, J. M., Kush, J. S., Meldahl. (2003a). Structure of old-growth longleaf pine (*Pinus palustris* Mill.) forests in the mountains of Alabama, USA. *Castenea,* 68, 211-221.

Varner, J. M., Kush, J. S., Meldahl. (2003b). Vegetation of frequently-burned old-growth longleaf pine (*Pinus palustris* Mill.) savannas on Choccolocco Mountain, Alabama, USA. *Natural Areas Journal,* 23, 43-52.

Verrall, A. F. (1936). The dissemination of *Septoria acicula* and the effect of grass fires on it in pine needles. *Phytopathology*. 26, 1021-1024.

Vleeshouwers, V. G. & Oliver, R. P. (2014) Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular Plant-Microbe Interactions*, 27 (3), 196-206.

Wahlenberg, W. G. (1946). *Longleaf Pine: its use, ecology, regeneration, protection, growth, and management.* Washington D. C. Charles Lathrop Pack Forestry Foundation.

Walker, J., & Peet, R. K. (1984). Composition and species diversity of pine-wiregrass savannas of the Green Swamp, North Carolina. *Vegetatio,* 55, 163-179.

Ware, S., Frost, C., & Doerr, P. D. (1993). Southern mixed hardwood forests: the former longleaf pine forest. In W. H. Martin, S. G. Boyer, & A. C. Echternacht (Eds.) *Biodiversity of the Southern United States: Lowland Terrestrial Communities,* (pp. 447-493). New York, NY. John Wiley & Sons.

Wilson, A. M., Wilken, P. M., van der Nest, M. A., Steenkamp, E. T., Wingfield, M. J., Wingfield, B. D. (2015). Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus,* 6(1), 207-214.

Wolf, F. A. & Barbour, W. J. (1941). Brown-spot needle disease of pines. *Phytopathology.* 31, 61-74.

Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W. (2011). WebMGA: a customizable web server for fast metagenomics sequence analysis. *BMC Genomics*, 12, 444.

Worrall, J. J. (1997). Somatic incompatibility in basidiomycetes. *Mycologia,* 89(1), 24-36.

Yandell, M. & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature Review Genetics*, 13, 329-341.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 812-829.

Zhang, G. *et al.* (2013). Comparative analysis of bat genomes provides insights into the evolution of flight and immunity. *Science*, 339 (6118), 456-460.

Zhan, J., Kema, G. H. J., Waalwijk, C., McDonald, B. A. (2002a). Distribution of mating type alleles in the wheat pathogen *Mycosphaerella graminicola* over spatial scales from lesions to continents. *Fungal Genetics and Biology*, 36, 128-136.

Zhan, J., Mundt, C. C., Hoffer, M. E., McDonald, B. A. (2002b). Local adaptation and effect of host genotype on the evolution of the pathogen: an experimental test in a plant pathosystem. *Journal of Evolutionary Biology,* 15, 634-647.

Zhao, Z., Liu, H., Wang, C., Xu, J. R, (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, 14, 274.