University of New Mexico
UNM Digital Repository

Biology ETDs | Electronic Theses and Dissertations

Spring 3-25-2019

GENOMICS BASED APPROACHES TO FUNGAL EVOLUTION

Aaron J. Robinson
University of New Mexico

Donald O. Natvig
University of New Mexico

Follow this and additional works at: https://digitalrepository.unm.edu/biol_etds

Part of the Bioinformatics Commons, Biology Commons, Desert Ecology Commons, Environmental Microbiology and Microbial Ecology Commons, Evolution Commons, and the Genomics Commons

Recommended Citation
Robinson, Aaron J. and Donald O. Natvig. "GENOMICS BASED APPROACHES TO FUNGAL EVOLUTION." (2019). https://digitalrepository.unm.edu/biol_etds/313

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Biology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact amywinter@unm.edu.

Aaron Robinson
Candidate

Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Dr. Donald O. Natvig, Chairperson

Dr. Donald Lee Taylor

Dr. Scott L. Collins

Dr. Jason Stajich

GENOMICS BASED APPROACHES TO FUNGAL EVOLUTION

BY

AARON ROBINSON

B.S., Biology, University of New Mexico, 2012
M.S., Biology, University of New Mexico, 2016

Ph.D. DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy in Biology

The University of New Mexico
Albuquerque, New Mexico

May, 2019

In memory of Uncle Pat

Patrick Norris
(1965-2010)

ACKNOWLEDGMENTS

I would like to thank Dr. Donald Natvig for serving as a mentor to me during my time at the University of New Mexico and aiding in my development as both a scientist and individual. Don sparked my interest in mycology years ago and I thank him for introducing me to such a unique and vibrant scientific community. I would also like to thank my committee members for being a great support structure to me these last few years.

I would like to thank John Urbane for being a great friend. Graduate school is a very stressful process and I owe my mental stability and fortitude to John and his ability to brighten any situation.

Finally, I would like to acknowledge my mother and father, Melissa and John Robinson, for being amazing parents and providing me with amazing support throughout my life. I am incredibly lucky to have such an amazing family and am eternally grateful for all the guidance they have provided.

GENOMICS BASED APPROACHES TO FUNGAL EVOLUTION

by

Aaron J. Robinson

B.S., Biology, University of New Mexico, 2012
M.S., Biology, University of New Mexico, 2016
Ph.D., Biology, University of New Mexico 2019

ABSTRACT

Advances in DNA sequencing and data analysis make it possible to address questions in population genetics and evolution at the genomic level. Fungi are excellent subjects for such studies, because they are found in diverse environments, have short generation times, can be maintained in culture and have relatively small genomes. My research employed genetic approaches using a variety of sequencing technologies and methods of analysis to explore questions in fungal evolution.

In one study, I explored the genetics behind differences in thermotolerance between isolates of *Neurospora discreta* from Alaska and New Mexico. Isolates from the two states exhibited differences in maximal growth temperature, with New Mexico isolates being substantially more thermotolerant than isolates from Alaska. Genomic scale comparisons of progeny from crosses between isolates from New Mexico and Alaska indicated that two regions, one on chromosome III and another on chromosome I, are responsible for differences in thermotolerance. Examination of these regions revealed numerous differences between the New Mexico and Alaska isolates at nucleotide and amino-acid levels; and it identified candidate genes for being important for differences in maximal growth temperatures.

In a second study, I explored the genomic differences between pathogenic and endophytic isolates in the genus *Monosporascus*. Culture and sequence-based surveys of root associating fungi at the Sevilleta National Wildlife Refuge (SNWR) revealed the ubiquitous presence of members of this genus. Although *M. cannonballus* is known as a severe pathogen of melon roots in agricultural settings, all of the host plants associating with *Monosporascus* species in natural settings appeared to be disease free. Complete genome sequences were obtained from three *M. cannonballus* isolates, an *M. ibericus* isolate and six SNWR isolates. Comparative genome analyses revealed that 1) isolates of *Monosporascus* possess genomes that are more than twice the size of those typical for members of the Sordariomycetes, while having typical numbers of protein-coding genes; 2) isolates from diverse grasses, tree and forbs include lineages closely-related to previously described species including *M. cannonballus*, in addition to novel lineages; and 3) species of *Monosporascus* and other Xylariales lack mating-type gene regions typical of other members of the Pezizomycotina.

TABLE OF CONTENTS

LIST OF FIGURES ix

LIST OF TABLES x

INTRODUCTION 1

References 6

CHAPTER 1

DIFFERENCES IN THERMOTOLERANCE BETWEEN ECOTYPES OF NEUROSPORA DISCRETA ARE DUE PRIMARILY TO ONLY TWO GENOMIC REGIONS

Abstract 7
Introduction 8
Methodology 10
Crossing procedure and progeny collection 10
Phenotyping with racetubes 11
Molecular methods and sequencing 12
Bioinformatics 13
Bulked-segregant analysis 14
Results 15
Identification of genomic regions important for temperature growth phenotypes 15
Outlier progeny 17
Distribution and substitution patterns of encoded amino-acids across genomes and within regions important for thermotolerance 18
Identification of genes potentially involved in thermotolerance 19
Discussion 20
Acknowledgments 23
References 24

CHAPTER 2

EVOLUTION OF MONOSPORASCUS IN ARID ENVIRONMENTS: GENOMIC-BASED APPROACHES

Abstract 39
Introduction 40

Methodology 42
Sample collection 42
Molecular methods and sequencing 43
Genome assembly 43
Bioinformatics 44
Molecular alignments and phylogenetic analyses 44
Results and discussion 45
Acknowledgments 50
References 50

CHAPTER 3

DIVERSE MEMBERS OF THE XYLARIALES LACK CANONICAL MATING-TYPE REGIONS

Abstract 66
Introduction 66
Materials and methods 67
Genomes examined 67
Mating-type gene searches 68
Molecular alignments and phylogenetic analyses 68
Results and discussion 68
Acknowledgments 70
Appendix A. Supplementary material 70
References 70

SUMMARY 72

LIST OF FIGURES

CHAPTER 1.

Figure 1. Comparison of *N. discreta* growth rates 31

Figure 2. Summary analysis of contig 10 from FGSC strain 10343 (Alaska) 32

Figure 3. Summary analysis of contig 1 from FGSC strain 10343 (Alaska) 34

Figure 4. Summary analysis of an example genomic region with no apparent contribution to thermotolerance.. 36

Figure 5. Number of SNPs found per 100 Kb window across contig 10 from FGSC strain 10343 assembly based on Illumina data.. 37

Figure 6. Amino-acid differences in the 1.1 Mb region on linkage group III.................. 38

CHAPTER 2.

Figure 1. Maximum likelihood tree for strains of *Monosporascus* based on analysis of ribosomal RNA ITS sequences 62

Figure 2. Relationship between GC content and genome size for several genomes from the Sordariales and Xylariales 63

Figure 3. Five-gene maximum-likelihood tree 64

Figure 4. Synteny comparisons of *M. cannonballus* isolates 65

CHAPTER 3.

Figure 1. Comparison of *sla2/apn2* gene regions from three Xylariales genomes with the mating-type regions from *Neurospora* crassa 69

LIST OF TABLES

CHAPTER 1.

Table 1. Reexamination of progeny exhibiting either a discrepancy between intial phenotypic assignment and genotype or mixed LG III and LG I genotypes 28

Table 2. Comparison of amino-acid compositions between parental assemblies 29

Table 3. Comparison of amino-acid compositions between parental assemblies in 1.1 Mb high separation region of linkage group III 30

CHAPTER 2.

Table 1. Metadata of *Monosporascus* isolates selected for whole genome sequencing .. 55

Table 2. Summary *Monosporascus* assembly statistics 56

Table 3. Xylariales annotation and assembly size comparison 57

Table 4. dbCAN2 results 58

Table 5. PHI-base results 59

Table 6. BUSCO results 60

Table 7. Sordariales and Xylariales sequences used as queries in multi-gene phylogenetic analysis.. 61

CHAPTER 3.

Table 1. Xylariales genomes examined for mating-type genes 67

Table 2. Sordariomycete mating-type sequences used as queries in alignment based searches 68

INTRODUCTION

The field of comparative genomics has vastly improved the capabilities and resolution of evolutionary studies. Advancements in sequencing technologies have made it possible to obtain and assemble complete nuclear and mitochondrial genomes rapidly and relatively inexpensively. Sequencing whole nuclear genomes may seem excessive when compared with more conserved methods of targeted sequencing that have been popular in the past, but there are many benefits to having whole genomes available for comparative analysis. Targeted sequencing analyses require prior knowledge or insight as to what regions of the genome may be involved with a particular adaptation being examined. Having access to the complete genome provides an opportunity to examine regions of interest independent of their identification prior to the initial sequencing. Complete genome assemblies also allow for the expansion of analyses to include comparisons of overall genome structure and organization. Genetic changes associated with adaptation could be highly complex and genomic-based studies have the capability to capture the full effects of any changes associated with this evolutionary process.

Fungi have been used in previous genomic studies of adaptation because of the many advantages they provide. Many fungi have small haploid genomes, which greatly simplify the genome assembly and annotation processes while also eliminating the need to examine effects of heterozygosity. Fungal isolates can often be maintained in culture and sometimes produce resilient asexual spores that allow for long-term storage of stock cultures. Short generation times are another valuable aspect of fungi as some isolates are capable of generating large numbers of progeny via sexual reproduction in short intervals.

Finally, isolates from natural fungal populations can be grown and maintained under diverse environmental conditions in the laboratory, making them great for studying environmental adaptation.

Members of the fungal genus *Neurospora* have long served as a model organisms in studies of genetics and biochemistry. The most widely studied member of this genus is *Neurospora crassa*, which has been utilized in biological research since the early 20th century. Most famous are the biochemical mutant studies of Beadle and Tatum (Beadle and Tatum 1941) for which they won the Nobel Prize in. The complete genome of *Neurospora crassa* was sequenced and assembled before the appearance of next-generation technologies. Functional annotations for nearly all of the approximately 10,000 genes were completed following knock-out experiments. This nearly-chromosomal level annotated assembly has served as a reference for high-resolution genome level comparisons within *N. crassa* and across the multiple recently diverged species found within the genus.

As an example of the power of genomic studies for studies in ecology and evolution, Ellison et al. (2011) employed comparative genomes in a 'reverse-ecology' approach to identify two genomic 'islands' of divergence between populations from Louisiana and the Caribbean that suggested a selective sweep. Gene deletion experiments demonstrated a link between genes found within these genomic islands and their ability to grow at low temperatures.

To explore further the links between genomic evolution and differences in environmental temperature, I employed *Neurospora discreta* isolates from wild populations in the western United States. *Neurospora discreta* is commonly described as

a primary colonizer of burnt plant tissue following forest fire events in the western United States. Isolates of *N. discreta* have been collected from various latitudes in the western United States, Europe and other locations. Phylogenomic analyses of these isolates revealed multiple distinct populations (Gladieux et al. 2015). These closely related populations of *N. discreta* present a unique opportunity to examine the effects of environmental conditions on genome evolution. Isolates from New Mexico and Alaska are of particular interest given the substantial differences in climate at these two locations. New Mexico also experiences very warm seasonal periods, which could select for more thermotolerant lineages of *N. discreta* in these locations.

Thermotolerant adaptation is of particular interest to both the industrial and evolutionary research sectors. The industrial sector is interested in utilizing thermotolerant organisms to obtain stable, efficient enzymes. Using thermotolerant organisms to produce commonly used enzymes also reduces the risk of environmental contamination due to the fact that warmer incubation temperatures make it more difficult for contaminants to grow. Evolutionary biologists and ecologists are interested in how organisms adapt to various temperatures in order to better predict the impacts of climate change on susceptible populations.

The primary goals of the research presented in the first chapter of this dissertation were to determine if *N. discreta* isolates from New Mexico are better adapted to high ambient temperatures than are isolates from Alaska, and, if so, to use genomic comparisons to identify regions of the genome that are associated with differences in thermotolerance. Gladieux et al. (2015) demonstrated that despite being closely related, *N. discreta* isolates from New Mexico and Alaska were different enough at the genome

level to make genome-level comparisons feasible. The acquisition of genomic data for this work employed both Illumina short-read and Oxford Nanopore long-read sequencing technologies. This hybrid sequencing approach, coupled with computer-assisted genome assembly methods, minimized the disadvantages associated with using either sequencing platform separately and produced very high quality *N. discreta* genome assemblies. High quality genome assemblies simplify comparative analyses because they are less fragmented, and thereby provide greater confidence when comparing genomic regions that are not highly conserved.

The second chapter of this dissertation focuses on the evolution of the fungal genus *Monosporascus* in arid environments. *Monosporascus* is comprised of three recognized species, two of which cause disease in members of the cucurbitaceae in agricultural settings. The majority of research conducted with members of this genus has been performed with strains associated with disease in agricultural settings, and as a result little is known about the evolution and ecology of species. Molecular surveys of root endophytes in central New Mexico revealed the near ubiquitous presence of *Monosporascus* sequence reads across highly diverse hosts including grasses, shrubs and forbs. Comparisons of the internal transcribed spacer (ITS) regions from these isolates indicated substantial diversity among *Monosporascus* isolates. All of the hosts containing *Monosporascus* reads also appeared to be disease free, indicating potential differences in primary ecological function between the agricultural and New Mexico isolates. In order to better understand the evolutionary diversity and history of *Monosporascus*, whole genome comparisons were employed.

Unlike *Neurospora*, genetic based studies of *Monosporascus* are limited, and as a result the genomic resources available are also very limited. The work presented in Chapter 2 resulted in the first complete genome sequences and assemblies for three *M. cannonballus* isolates, one *M. ibericus* isolate, and six *Monosporascus* sp. isolates from New Mexico, towards the goal of whole genome comparisons. Phylogenetic comparisons conducted using these assemblies also revealed that some of the New Mexico *Monosporascus* isolates were closely related to previously described species, while other isolates appeared to represent novel lineages. Genomes for these *Monosporascus* isolates were larger than what is typical of members of the Xylariales and indicate numerous differences in genome structure and organization.

The research presented in the final chapter of this dissertation examined the evolution of mating-type regions in members of the Xylariales. Mating-type genes in fungi control sexual reproduction and have been widely studied in the Pezizomycotina. The high level of conservation typically observed in these genes across the Pezizomycotina simplifies their identification via common alignment based methods such as BLAST. Attempts to identify mating-type genes within multiple Xylariales genomes indicated the absence of genes (MAT1-1-1 and MAT1-1-2) known to be important for mating in diverse members of the Pezizomycotina. While I was able to identify distant homologs of high-mobility group (HMG) domain mating-type genes (MAT1-1-3 and MAT1-2-1), because filamentous fungi contain members of this gene family that are not involved in determination of mating-type, it was not possible to say with certainty that any of these genes functions as a mating-type gene. It appears that mating-type regions in

members of the Xylariales are either absent or are highly divergent relative to other members of the Pezizomycotina.

References

Beadle, G.W. and Tatum, E.L. 1941. Genetic Control of Biochemical Reactions in *Neurospora*. PNAS 27, 499-506.

Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L. and Taylor, J. W. 2011. Population genomics and local adaption in wild isolates of a model microbial eukaryote. PNAS 108, 2831-2836.

Gladieux, P., Wilson, B. A., Perraudeau, F., Montoya, L. A., Kowbel, D., Hann-Soden, C., Fischer, M., Sylvain, I., Jacobson, D. J. and Taylor, J. W. 2015. Genomic sequencing reveals historical, demographic and selective factors associated with the diversification of the fire-associated fungus *Neurospora discreta*. Mol Ecol, 24: 5657-5675.

CHAPTER 1

Differences in thermotolerance between ecotypes of *Neurospora discreta* are due primarily to only two genomic regions

Abstract

Differences in maximal growth temperature among *Neurospora discreta* isolates from the western United States correlate with differences in mean annual environmental temperature. Isolates from New Mexico and Alaska exhibit comparable growth rates below 35°C, but isolates from New Mexico grow much better near and above 40°C. Individual progeny from crosses between isolates from New Mexico and Alaska either possess one of the two parental temperature phenotypes or have an intermediate phenotype. The range of progeny phenotypes suggests the involvement of multiple gene regions. With support from the DOE Joint Genome Institute (JGI) Community Science Program (CSP), we obtained complete genome sequences for 82 progeny from crosses with parents from NM and AK. Progeny were selected to exhibit either the New Mexico parental temperature phenotype or the Alaska parental phenotype (39 NM-like and 43 AK-like progeny). High-quality genome assemblies of the parental strains were obtained utilizing sequence data from both Illumina (JGI) and Oxford Nanopore MinION platforms. Bulked-segregant analysis was also performed using the MinION platform and combined genomic DNA from 22 New Mexico-like progeny and 29 Alaska-like progeny. Comparative analyses of genomes from these two progeny pools demonstrated two regions associated with thermotolerance above 40ºC. Differences between the parental isolates in a region of linkage group III demonstrated a strong link between genotype and phenotype and indicated amino-acid modifications that could result in more

thermotolerant proteins. It is striking that this region in *N. discreta* overlaps a region previously identified in *N. crassa* that may be under selection in adaptation to cold. A region on linkage group I plays a secondary role in determining the thermotolerant phenotype.

Keywords: Neurospora, thermotolerance, genomics, New Mexico, Alaska

Introduction

Understanding how organisms adapt to diverse environments is a complex question, but it is crucial to the comprehension of the evolutionary processes that result in biological diversity. Every environment has numerous biotic and abiotic factors that contribute to selection and evolutionary divergence on varying scales. Understanding the effects of individual environmental factors is an important first step in the development of evolutionary models regarding adaptation. For example, in the case of adaptation to low or high temperature, it is not clear the extent to which such adaptation requires global genetic changes that alter the GC content of DNA or the amino-acid content of proteins, versus less global changes involving changes in gene expression or the thermostability of a few proteins.

Genetic changes that result in heritable thermotolerance are of particular interest given that current global climate change models project significant warming over the next few decades (Pachauri and Meyer 2014). Thermotolerance is often studied in a comparative manner where differences between closely related thermophilic and mesophilic organisms are examined at the genomic scale to identify genetic components that are correlated with phenotype. Haploid fungi simplify these comparisons due to their

small genome size and ability to be maintained in culture. Specific types of amino-acid substitutions are reported to correlate with thermophiliy in some fungal lineages, and may contribute to protein thermostability. It remains unclear, however, if these changes are solely responsible for conferring thermotolerance (van Noort et al. 2013).

The goals of the present study were to determine if the genetic basis for differences in temperature growth phenotypes can benefit from comparisons within a single fungal species capable of inhabiting environments with substantially different climates. These intraspecies comparisons may provide a clearer resolution to this question by enhancing signals that may be masked by evolutionary noise in comparisons across species.

Neurospora discreta is a filamentous ascomycete commonly found colonizing burnt vegetation following forest fires in western North America (Jacobson et al. 2004). *N. discreta* isolates have been collected across states throughout the western United States and in Europe. Phylogenetic studies conducted on these isolates indicate several distinct *N. discreta* populations with strong evidence of endemism at the state level (Gladieux et al. 2015). *N. discreta* occurs at very diverse latitudes, ranging in the United States from New Mexico to Alaska. Therefore, populations of this species provide a compelling opportunity to examine the role of climate-related adaption in the divergence of closely related fungal lineages.

As presented here, an examination of the temperature growth profiles of *N. discreta* isolates from New Mexico and Alaska revealed substantial differences. New Mexican isolates had consistently higher growth rates than did Alaskan isolates when incubated at temperatures above 40°C. To explore the genetic basis for differences in

thermotolerance observed for the AK and NM strains, we obtained progeny from NM x AK crosses and sequenced progeny with either strong AK or strong NM parental phenotypes. In a second experiment, a bulked-segregant analysis was performed using separate sets of progeny, again with either strong AK-like or NM-like phenotypes. Genome-wide analyses were employed to determine which regions of the genome are associated with the differences in thermotolerance observed in the parental strains. Our results indicate that a region on chromosome III has a major role in determining differences in thermotolerance, while a region on chromosome I plays a secondary role. These results suggest that relatively few genetic changes are required to account for the differences in thermotolerance between AK and NM ecotypes. Our results also reinforce a previous study with *N. crassa* (Ellison et al. 2011) that suggested this region on chromosome III contributes to adaptation to local environment conditions.

Methodology

Crossing procedure and progeny collection

Progeny were obtained from a cross between FGSC strains 10342 (Bernalillo, NM) and 10343 (Tok, Alaska). Parental strains were crossed by inoculating opposite sides of plates containing Vogel’s N medium (1.5% sucrose, 1.5% agar; Vogel 1956) with conidial suspensions, followed by incubation at 25°C until mature perithecia appeared. Perithecia were collected from the plates and rinsed with sterile water to aid in the removal of hyphal fragments and conidia. Perithecia were ruptured in sterile water and ascospores were collected with a micropipette and transferred to a 1.5-mL microcentrifuge tube containing sterile water. These tubes were placed in a hot water bath

set at 60°C for one hour to induce ascospore germination. A few (10-20) microliters of heat-shocked ascospore suspensions were spread on plates of Vogel's N medium and incubated for 24 hours. Individual germlings derived from ascospores were identified using a compound microscope and transferred to fresh N medium plates. Conidia from these cultures were streaked onto N medium plates containing sorbose (1% sorbose, 0.05% glucose, 0.05% fructose, 1.5% agar) to foster colonial growth (Davis and deSerres 1970), followed by subculturing on Vogel's N medium to ensure that each progeny strain was derived from a single spore.

Phenotyping with racetubes

Conidial suspensions of progeny and parental isolates were created in sterile distilled water, and 5 μL aliquots were transferred to the ends of 40 cm glass racetubes containing Vogel's N medium and capped with sterile cotton plugs. In preliminary experiments the growth rates of several wild type isolates from each location (AK and NM) were assessed at multiple temperatures ranging from 10°C to 45°C. Based on the differences observed in these preliminary tests, subsequent routine scoring experiments were performed at 42°C and 12°C. In progeny scoring, two racetubes were inoculated for each strain, and racetubes were incubated for 24 hours at 30°C. After 24 hours, mycelial growth was marked, and the tubes were separated into incubators set at 42°C and 12°C. For each such experiment, the two parental strains were included as controls. Mycelial growth rate (millimeter per hour) was recorded daily for a week. The highest growth rate recorded, excluding the first, last and control measurements, was used to generate the final phenotype by dividing the 42°C growth rate by the 12°C growth rate for each

isolate. Ratios were used to account for differences in growth rates among the progeny that could have been independent of incubation temperature (see Ellison et al. 2011). The resulting parental and progeny ratios from each experiment were used to assign individual progeny phenotypes as NM-like, AK-like, or intermediate.

Molecular methods and sequencing

We selected 43 progeny with strong Alaska-like phenotypes and 39 progeny with strong New Mexico-like phenotypes to undergo complete genome sequencing. These progeny and the parental isolates were grown for 48 hours in 25 mL of liquid N medium, and DNA was extracted following a CTAB protocol (Hutchinson et al. 2015). Genomic sequencing was performed at the DOE Joint Genome Institute (JGI) employing 101 base-pair paired-end sequencing on the Illumina HiSeq-2500 platform. Sequencing data for this project has been deposited at NCBI under accession numbers PRJNA441644 – PRJNA441729. In addition to Illumina sequencing, genomic libraries for the parental isolates were generated using the SQK-LSK108 kit and sequenced for 48 hours on separate FLO-MIN106 flowcells on a MinION sequencer from Oxford Nanopore Technologies (https://nanoporetech.com/products/minion).

The progeny selected for sequencing have been deposited at the Fungal Genetics Stock Center (FGSC) at Kansas State University along with the parental isolates. Progeny with an Alaskan phenotype are represented by strains ARPr1-ARPr43, and progeny with a New Mexican phenotype are represented by strains ARPr44-ARPr83. One progeny is represented by two strain identifiers, ARPr75 and ARPr78. Only ARPr75 was used for the analyses described below.

Bioinformatics

Genome assemblies for the parental isolates were initially generated with raw Nanopore long-reads (>500 bp) using the Canu assembler v1.8 (Koren et al. 2017). Racon (https://github.com/isovic/racon) and Pilon (Walker et al. 2014) were used to further improve these parental assemblies in an iterative manner using alignments of Illumina data generated with the Burrows-Wheeler Aligner (BWA) MEM algorithm (Li 2013). Assembly statistics were generated using Quast (Gurevich et al., 2013). Based on the number of contigs produced, the genome assembly for FGSC strain 10343 (Alaska parent) assembly was of higher-quality than the FGSC strain 10342 (New Mexico parent) and was selected to serve as the reference for downstream analyses.

Illumina reads from all 82 progeny and both parental isolates were mapped to the final FGSC strain 10343 assembly using BWA-MEM, and variant calling was performed using BCFtools (https://github.com/samtools/bcftools). MinION long-reads from the bulked segregant analysis (described below) were mapped back to the assembly for FGSC strain 10343 using Minimap2 (Li 2018). A pileup file for each Minimap2 alignment was generated using the Samtools mpileup script (https://github.com/samtools/samtools). Variant calling for each pooled sample was completed using VarScan 2 (Koboldt et al. 2012). Synonymous and missense mutations were identified with SnpEff (Cingolani et al. 2012).

Genomes were annotated using the gene prediction program AUGUSTUS with the species parameter set to *Neurospora crassa*, limited to no alternative transcripts and only complete gene predictions (Stanke and Morgenstern, 2005; Hoff and Stanke, 2013). Nucmer was used for synteny comparisons and these comparisons were visualized using

mummerplot (MUMmer4 package, Kurtz et al. 2004). *Neurospora crassa* strain 74-OR23-1VA (aka OR74A) (GCF_000182925.2) was used as a reference in synteny comparisons because the assembly is near chromosomal resolution, while the *N. discreta* reference genome is more fragmented.

Bulked-segregant analysis

In addition to the progeny described above, progeny obtained from crosses with the FGSC 10342 (NM) and 10343 (AK) parents were employed in a separate bulked-segregant analysis (Pomraning et al., 2011) as an alternative to the progeny-sequencing approach described above. An alternative phenotyping procedure was also employed. Conidia from 206 individual progeny were suspended in sterile water in 1.5-mL microcentrifuge tubes and 5 µL of each suspension was spotted onto each of three different N Medium plates with sorbose (see above). Each 10-cm plate was arrayed with 25 progeny, and these plates were incubated at 25°C, 42°C and 45°C. After 24 and 48 hours each colony was scored for growth. As a second check of phenotype, 29 progeny with an apparent Alaska-like phenotype (no growth at 45°C) and 46 progeny with a New Mexico-like phenotype (growth at 45°C) were re-screened individually at 25ºC, 42ºC and 45ºC on N medium plates with 1.5% sucrose. From this second screen, 22 New Mexico-like and 29 Alaska-like progeny were selected for further analysis. Progeny were grown for 48 hours in 25 mL liquid N medium, and DNA was extracted following a CTAB protocol (Hutchinson et al. 2015). DNA pools were created for each of the two phenotypic classes by combining 300 ng of DNA from each progeny strain. The pooled DNA preparations were further purified using an equal volume of AMPure XP bead

suspension following instructions provided by the manufacturer (Beckman Coulter). Genomic libraries for each of the DNA pools were generated using the SQK-LSK109 kit and sequenced for 24 hours on separate FLO-MIN106 flowcells on a MinION sequencer from Oxford Nanopore Technologies.

Results

Identification of genomic regions important for temperature growth phenotypes

At temperatures above 40°C, isolates from Bernallilo, New Mexico, (mean annual high 21°C, mean annual low 5°C) had higher mycelial growth rates than isolates from Tok, Alaska, (mean annual high 2°C, mean annual low -10.67°C), while having similar growth rates at 12°C. In order to quantify these differences and score progeny we used ratios of maximal growth rates (mm/hour) at 42°C and 12°C (Fig. 1). Two *N. discreta* isolates, FGSC strain 10343 (Tok, AK) and FGSC strain 10342 (Bernalillo, NM), consistently showed the greatest growth-rate difference and lowest variance in terms of this phenotype across multiple experiments. Analysis of progeny obtained from sexual crosses of these two parental isolates demonstrated the heritability of the thermotolerance phenotype (Fig. 1). The occurrence and frequency of progeny with intermediate phenotypes indicated the involvement of multiple regions of the genome in conferring thermotolerance.

The genome assembly for the Alaskan parent (FGSC 10343) consisted of 28 contigs covering 37.68 Mb. This assembly was used as a reference in initial genetic analyses. A total of 223,106 SNPs across the 28 contigs were identified between this genome and that of the NM parent (FGSC 10342) using high-coverage Illumina reads.

The high similarity observed between the parental isolates (~6 SNPs every thousand bases) was expected, given the known relationship between the New Mexico and Alaska *N. discreta* lineages (Gladieux et al. 2015).

To determine which regions of the *N. discreta* genome are associated with thermotolerance, we sequenced the genomes of 43 progeny with strong Alaska phenotypes and 39 progeny with strong New Mexico phenotypes. Relationships between progeny genotype and phenotype were examined through variant calling. Comparisons of phenotypic class and SNPs were performed for both the Illumina and bulked-segregant data using a sliding window technique. In an effort to eliminate noise created by sequencing errors, only positions that showed variation between the two parental isolates were examined. Positions that showed disagreement between the Illumina and MinION data from the Alaska parental isolate were also excluded. These variant analyses focused exclusively on SNPs and excluded insertion/deletion (indel) mutations. Only two genomic regions, represented by portions of contigs 10 and 1 from FGSC 10343 (Alaska), displayed substantial correlation between genotype and phenotype (Fig. 2 and Fig. 3). Contig 14, which displayed no clear separation between the two phenotypic classes for both datasets, reflected the results observed for the other 26 contigs and was used to represent regions not influenced by our phenotypic selections (Fig. 4).

Contig 10 showed the most substantial separation between the two phenotypic classes (Fig. 2). The Illumina-based variant analysis for contig 10 demonstrated a high proportion of the New Mexico-like progeny contained New Mexico SNPs, which were largely absent from the Alaska-like progeny (Fig. 2A). This strong relationship between genotype and phenotypic class was observed throughout the majority of this large contig

(3.88 Mb) and was supported by a similar analysis performed on the bulked-segregant data (Fig. 2B). Examinations of synteny indicated this contig resembles a large part of linkage group III in *Neurospora crassa* OR74A (NC_026503.1) (Fig. 2C).

The region of highest differentiation between the two phenotypic classes on linkage group III mapped to a 1.1 Mb region (Fig. 2A) which contained 2,746 parental SNPs and 258 putative gene annotations. A total of 245 missense mutations contained within 112 predicted genes were found between the parental isolates in this region. This region is also very conserved between the two parental isolates in terms of SNP content relative to the rest of contig 10 (Fig. 5).

*C*ontig 1 also showed a strong relationship between genotype and phenotype for the New Mexico-like progeny, but a similar relationship was not observed for the Alaska-like progeny (Fig. 3A). Synteny comparisons with *N. crassa* indicated this contig represents a large region of linkage group I and demonstrated a large chromosomal inversion event between the Alaska parent and *N. crassa* (Fig. 3C). This chromosomal inversion was not found in synteny comparisons between the two parental strains. The importance of this region was confirmed by bulked-segregant analysis (Fig. 3B).

Outlier progeny

With respect to contig 10 and contig 1, several individual progeny exhibited a discrepancy between our initial phenotype assignments and observed genotype in coding regions (CDS). Accordingly, we re-examined the phenotypes of these progeny. In three instances, progeny with a strong AK genotype in the linkage group III region but with at least a partial NM genotype in the linkage group I region were found to have intermediate

temperature growth phenotypes (Table 1). In contrast, progeny with a strong NM genotype in the linkage group III region but an AK genotype in the linkage group I region were observed to possess an NM phenotype consistent with our original phenotype assignment. Two progeny possessed NM genotypes on linkage group III but exhibited AK phenotypes in both initial and follow-up tests. Progeny that displayed a strong AK phenotype despite having a strong NM genotype in the linkage group III region could aid in highlighting loci most important to the phenotype.

Distribution and substitution patterns of endcoded amino-acids across genomes and within regions important for thermotolerance

Substantial evidence exists that certain amino-acids are over-represented in the proteins of thermotolerant and thermophilic microorganisms relative to mesophilic organisms (van Noort et al. 2013, MacDonald et al. 1999, Zeldovich et al. 2007). In particular, it has been reported that the occurrences of Ile, Val, Tyr, Trp, Arg, Glu, and Leu (the IVYWREL group) increase in a nearly linear fashion with an increase in optimal growth temperature (Zeldovich et al. 2007), an observation supported by comparisons of related mesophilic and thermophilic Bacteria and Archaea (McDonald et al. 1999). Accordingly, we sought to determine whether the frequencies of these amino-acids differed at the global level between the proteomes of AK and NM isolates, or differed within the regions observed to be important for differences in thermotolerance between AK and NM isolates. While no individual amino-acid was significantly more frequent in the predicted proteins from the NM parental isolate, five of the seven amino-acids from the IVYWREL group (IVYW and L) were more frequent in the NM proteins. Proline,

which is also reported to be overrepresented in thermophilic organisms (Zeldovich et al. 2007), was also more abundant in NM proteins (Table 2). No significant differences were observed between the linkage-group III region of interest and the parental genomes as a whole.

The 1.1 Mb region on linkage group III showing the highest degree of differentiation between NM-like and AK-like progeny contained 258 predicted protein-coding genes, of which 112 contained at least one missense difference between the isolates. Among the differences in this region involving the IVYWREL group, IY and L were more highly represented in the NM parent (Table 3).

Identification of genes potentially involved in thermotolerance

The 1.1 Mb region on linkage group III possessed 245 amino-acid differences between the AK and NM parents. There were an additional 562 SNPs observed between the two parents within 1000 bp of start codons. While recognizing that differences in thermotolerance could involve factors other than differences in protein amino-acid sequences, e.g. differences in gene regulation, we examined the coding sequences of the genes in this region for signatures that could flag differences in thermotolerance between the two parental strains. Given that *Neurospora crassa* is thermotolerant, in one comparison we looked for amino-acid acid positions that were shared between *N. crassa* and our NM *N. discreta* parent but not between the NM and AK parents. Of the 245 missense differences between the NM and AK parental isolates, 72 resulted in amino-acid positions shared between *N. crassa* and the NM parental isolate, but not between the AK and NM parents. These amino-acid positions were distributed across 59 protein-coding

genes (Fig. 6). Many of these genes are involved in common cellular processes (for example, RNA processing and cell-wall biosynthesis), and in several cases they appear to be clustered. The region of highest separation between the parental isolates, represented by the position from peroxin 8 to the position in BEAK-1, contained the highest density of shared amino-acid positions between *N. crassa* and the NM parental isolate (Fig. 6).

Discussion

Genomic comparisons of individual progeny combined with bulked-segregant analysis revealed two distinct regions of the *N. discreta* genome to be important in conferring differences in thermotolerance between closely related *N. discreta* isolates. Variant analysis performed with the two parental isolates and 82 progeny identified regions on linkage group III and linkage group I that correlated strongly with genotype and phenotype class. These findings were confirmed by a bulked-segregant analysis performed with a separate pool of progeny. The most surprising aspect of these results is the fact that one region of the genome, a portion of linkage group III, accounts for major differences in thermotolerance among the *N. discreta* ecotypes from AK and NM. This was surprising given the frequency of progeny with intermediate phenotypes and the predicted genetic complexity of thermotolerance. A second region, on linkage group I, contributes substantially to the differences between the two ecotypes. While it is entirely possible that other regions of the genome contribute to the temperature growth phenotypes of the parental strains employed here, any such contributions would appear to be minor in comparison.

The regions implicated in this study are quite large, which could be a result of co-adapted linked loci being involved directly or indirectly in the thermotolerant phenotype. It is also possible that these portions of the genome are subject to reduced recombination, adding to the difficulty in narrowing down the precise genes most responsible for phenotypic differences. It is also clear that the most important region, the region on linkage group III, is more conserved than surrounding regions in terms of both synonymous and non-synonymous nucleotide divergences both within and between species. In fact, typical approaches to evaluating levels of selection, such as Tajima's D approach, were ineffective due to low levels of divergence in the region.

In an attempt to eliminate noise created by linkage disequilibrium, we focused our efforts to identify linkage group III genes involved in thermotolerance on amino-acid variant sites. If we make the simplistic assumption that *N. crassa* and our NM *N. discreta* parent share amino-acid variants that are not shared with the AK parent, there are still at least 59 thermotolerance relevant gene candidates on linkage group III.

Ellison et al. (2011) identified two genomic "islands" in the *N. crassa* genome that might be involved in adaptation to cold temperatures. One of these regions was on linkage group VII, where we did not identify differences correlated with thermotolerance. Remarkably, however, the other genomic island identified by Ellison et al. is within the region on linkage group III that we identified as being most highly correlated with thermotolerance (Fig. 6.) One of the genes in this region identified as potentially important was MRH4-like RNA helicase, knockout mutants of which were found to be cold sensitive (Ellison et al. 2011). Interestingly, the predicted amino acid sequence for this protein was identical for the two *N. discreta* parental strains.

Ellison et al. identified two other *N. crassa* genes in this region on linkage group III, a phospholipase C and an unidentified gene designated NCU06247. This phospholipase C was one of the proteins that we found to have amino-acid positions that are shared between *N. crassa* and our *N. discreta* NM parent, but not with the AK parent. The predicted NCU06247 proteins from the NM and AK parents differed at several amino-acid positions, but the NM parent and *N. crassa* did not share variants that were not present in the AK parent.

While the importance of individual genes in this region to temperature growth phenotypes remains speculative, the fact remains that this region was identified as important in this context in two different species of *Neurospora* using three different experimental approaches. It is therefore tempting to speculate that this region is particularly important in helping to tune populations to local climatic conditions.

Separate from the question of which specific genes are responsible for the observed segregation of temperature growth phenotypes between NM and AK strains, there is the question of whether NM and AK lineages have undergone global shifts in amino-acid coding across the genome. While there were hints that the amino-acids encoded by the genome of the NM parent are shifted slightly in favor of amino-acids associated with thermotolerance and thermophiliy (the IVYWREL group), the differences were not statistically significant. In any case, any such global difference between the two genomes could not account for the segregation of progeny with different levels of thermotolerance in crossing experiments.

A clearer understanding of the ancestral state of thermotolerance in *N. discreta* would aid in clarifying if the New Mexico linage experienced selection towards

thermotolerance or if there was an evolutionary trade-off between warm and cold adaptation in the Alaska lineage. The inclusion of isolates from other warm or cold environments in the future may make it possible to infer the ancestral state of thermotolerance in *N. discreta*.

The parental genomes assembled in this study are considerably less fragmented than other *N. discreta* assemblies currently available. Examinations of synteny revealed a large chromosomal inversion between the Alaska parental and *N. crassa* assembly on linkage group III. This chromosomal inversion was not present in synteny comparisons between the two parental isolates, but was also present in comparisons between the NM parent and *N. crassa*. The disappearance of this chromosomal inversion in intraspecies comparisons indicates it is specific to interspecies comparisons between *N. crassa* and *N. discreta*.

The value of *N. discreta* for examining natural evolution as the result of varying environmental factors is made clear in this study. Isolates of *N. discreta* have been collected from numerous geographical locations representing a multitude of environmental conditions. This study also demonstrates the ability to comparatively analyze phenotypic differences guided by parental variance using populations of progeny and artificial selection.

Acknowledgments

Sequencing support for this work was provided through the Department of Energy Joint Genome Institute Community Science Program (proposal ID: 1462). I thank the UNM Center for Advanced Research Computing, supported in part by the National

Science Foundation, for providing the high performance systems and resources used in this work. I thank the Fungal Genetics Stock Center (Kansas State University) for curating parental strains and progeny.

References

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80-92.

Davis, R. H., and F. J. deSerres. 1970. Genetic and microbiological research techniques for *Neurospora crassa*. Methods Enzymol 71A, 79–143.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100.

Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L. and Taylor, J. W. 2011. Population genomics and local adaption in wild isolates of a model microbial eukaryote. PNAS 108, 2831-2836.

Gladieux, P. , Wilson, B. A., Perraudeau, F. , Montoya, L. A., Kowbel, D. , Hann-Soden, C., Fischer, M. , Sylvain, I. , Jacobson, D. J. and Taylor, J. W. 2015. Genomic sequencing reveals historical, demographic and selective factors associated with the diversification of the fire-associated fungus *Neurospora discreta*. Mol Ecol. 24, 5657-5675.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.

Hoff, K.J., Stanke, M., 2013. WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 41, W123–W128.

Jacobson, D.J., Powell, A.J., Dettman, J.R., Saenz, G.S., Barton, M.M., Hiltz, M.D., Dvorachek Jr., W.H., Glass, N.L., Taylor, J.W. and Natvig, D.O. 2004. *Neurospora* in temperate forests of western North America. Mycologia 96, 66-74.

Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L., & Wilson, R. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research 22, 568-576.

Koren, S., Walenz, B.P., Berlin, K., Miller. J.R. and Phillippy, A.M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27, 722-736.

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

McDonald, J.H., Grasso, A.M. and Rejto, L.K. 1999. Patterns of temperature adaptation in proteins from *Methanococcus* and *Bacillus*. Mol Biol Evol. 16, 1785-1790.

Pachauri, R.K. and Meyer L.A. 2014. IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. Geneva, Switzerland, 151 pp.

Pomraning, K.R., Smith, K.M. and Freitag, M. 2011. Bulk Segregant Analysis Followed by High-Throughput Sequencing Reveals the Neurospora Cell Cycle Gene, ndc-1, To Be Allelic with the Gene for Ornithine Decarboxylase, *spe-1*. Eukaryotic Cell 10, 724-733.

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467.

Vogel, H.J. 1956. A convenient growth medium for *Neurospora* (Medium N). Microbial Genetics Bulletin 13, 42-43.

van Noort, V., Bradatsch, B., Arumugam, M., Amlacher, S., Bange, G., Creevey, C., Falk, S., Mende, D.R., Sinning, I., Hurt, E. and Bork, P. 2013. Consistent mutational paths predict eukaryotic thermostability. BMC Evolutionary Biology 13, 7.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. and Earl, A.M. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9, e112963.

Zeldovich, K.B., Berezovsky I.N. and Shakhnovich EI. 2007. Protein and DNA sequence determinants of thermophilic adaptation. PLoS Comput Biol. 3, e5.

Table 1. Reexamination of progeny exhibiting either a discrepancy between initial phenotypic assignment and genotype or mixed LG III and LG I genotypes.

Progeny ID	Original phenotype	Percent of NM CDS SNPs in LG III region (genotype)	Percent of NM CDS SNPs in LG I (genotype)	Re-examined phenotype
ARPr61	NM	100% (NM)	72% (mixed)	NM
ARPr71	NM	99% (NM)	27% (AK)	NM
ARPr67	NM	93% (NM)	17% (AK)	NM
ARPr22	AK	99% (NM)	24% (AK)	NM
ARPr21	AK	90% (NM)	98% (NM)	Intermediate
ARPr24	AK	0% (AK)	62% (mixed)	Intermediate
ARPr13	AK	0% (AK)	58% (mixed)	Intermediate
ARPr59	NM	0% (AK)	35% (mixed)	Intermediate
ARPr32	AK	0% (AK)	0% (AK)	AK
ARPr66	NM	0% (AK)	64% (mixed)	AK
ARPr27	AK	99% (NM)	11% (AK)	AK
ARPr11	AK	99% (NM)	98% (NM)	AK

Table 2. Comparison of amino-acid compositions between parental assemblies.

Amino-acid	AK parental isolate frequency	NM parental isolate frequency	Difference between AK and NM
A	0.087856388410906	0.087949938363997	-9.35499530909933E-05
L	0.084179274912215	0.084279652626265	-0.00010037771405
S	0.082087154278759	0.082041222087632	4.59321911270016E-05
G	0.071560755655349	0.071283241964574	0.000277513690775
E	0.065212477791301	0.064807447698596	0.000405030092705
P	0.064454627803193	0.064614333448102	-0.000159705644909
T	0.061386067626362	0.061440213076823	-5.41454504610003E-05
R	0.060944079545614	0.060811811277872	0.000132268267742
V	0.060043053398805	0.060075479574489	-3.24261756840041E-05
D	0.056741476756061	0.056686168381355	5.53083747059993E-05
K	0.051311712203243	0.051256223809159	5.5488394083994E-05
I	0.044712712119305	0.044877432659768	-0.000164720540463
Q	0.042830437458905	0.042828771436574	1.66602233100249E-06
N	0.037114976357354	0.037162005462502	-4.70291051479985E-05
F	0.034048820665632	0.034214225696594	-0.000165405030962
Y	0.025780014969013	0.025822006639919	-4.19916709059991E-05
H	0.024440279584785	0.024486485570396	-4.62059856110031E-05
M	0.02136778479596	0.021390190688058	-2.24058920980028E-05
W	0.013399976916943	0.013410959612808	-1.0982695864999E-05
C	0.010527928750297	0.010562189924521	-3.42611742240005E-05

Table 3. Comparison of amino-acid compositions between parental assemblies in 1.1 Mb high separation region of linkage group III.

Amino-acid	AK parental isolate LG III region frequency	NM parental isolate LG III region frequency	Difference between AK and NM
A	0.09039741779302	0.090357128368785	4.02894242349933E-05
L	0.08408311478717	0.084441048802586	-0.000357934015416
S	0.086039943514222	0.085718921988885	0.000321021525337
G	0.067090309999328	0.066604914064719	0.000485395934609
E	0.064750184923677	0.064238482238239	0.000511702685438
P	0.066404411270258	0.066537301726819	-0.000132890456561
T	0.062181426938336	0.062257440737786	-7.6013799449999E-05
R	0.062013314504741	0.06176387067112	0.000249443833621
V	0.059128505144241	0.059039093453774	8.94116904670006E-05
D	0.05701701297828	0.056902543576152	0.000114469402128
K	0.051597068119158	0.051770767129586	-0.000173699010428
I	0.043978212628606	0.044062960609052	-8.47479804459983E-05
Q	0.044166498554233	0.044793173858366	-0.000626675304133
N	0.036836796449465	0.036814917986234	2.18784632310051E-05
F	0.033891466612871	0.034049573366148	-0.000158106753277
Y	0.024826844193397	0.024908385282146	-8.15410887490021E-05
H	0.023831618586511	0.023921245148815	-8.96265623040009E-05
M	0.020671104834914	0.020608240591743	6.2864243171E-05
W	0.011734247864972	0.011723979391759	1.02684732130008E-05
C	0.009360500302602	0.009486011007289	-0.000125510704687

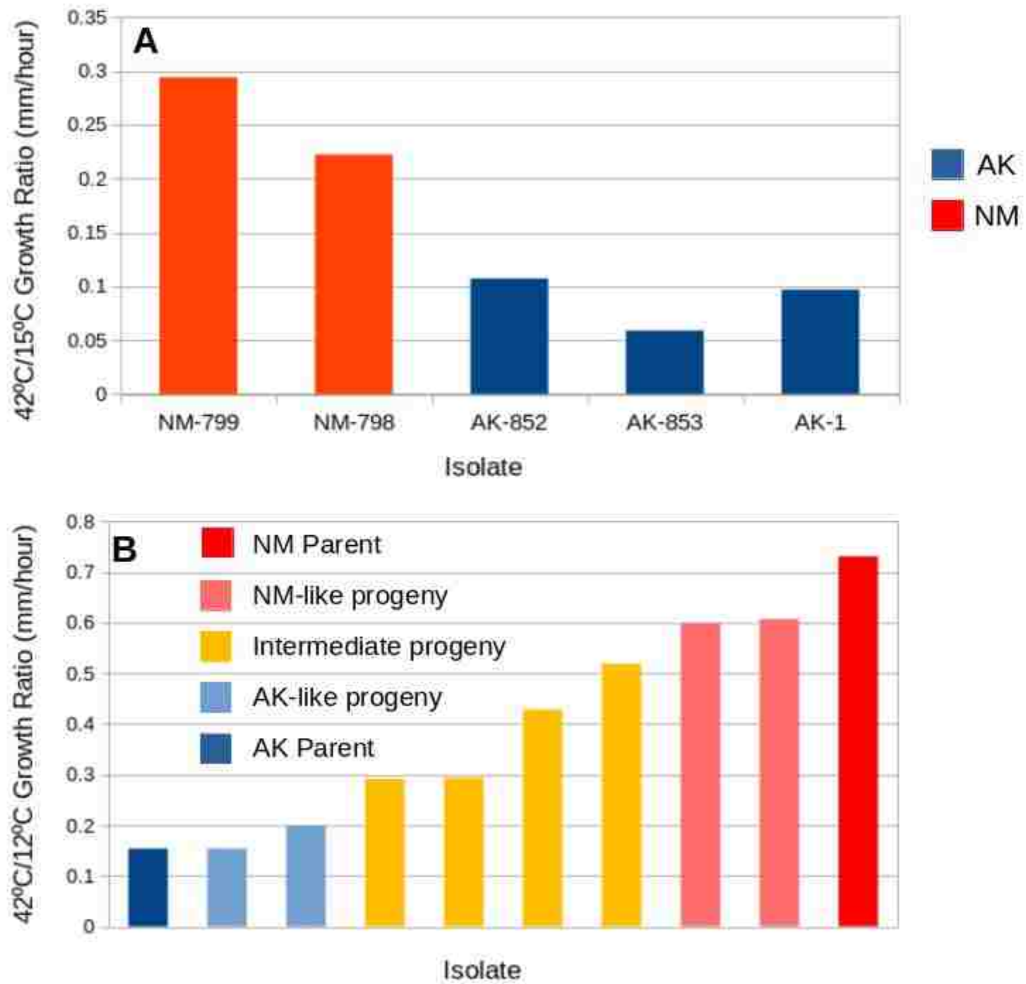

Figure 1. Comparison of *N. discreta* growth rates. (A) Comparisons of 42ºC/15ºC growth-rate ratios of wild isolates from New Mexico and Alaska. (B) Comparisons of 42ºC/12ºC growth-rate ratios for parental isolates and several progeny. FGSC strain 10343 (AK parent) is shown in dark blue and progeny with similar phenotypes are shown in light blue. FGSC strain 10342 (NM parental) is shown in dark red, and progeny with similar phenotypes are shown in light red. Progeny with intermediate phenotypes are displayed in orange.

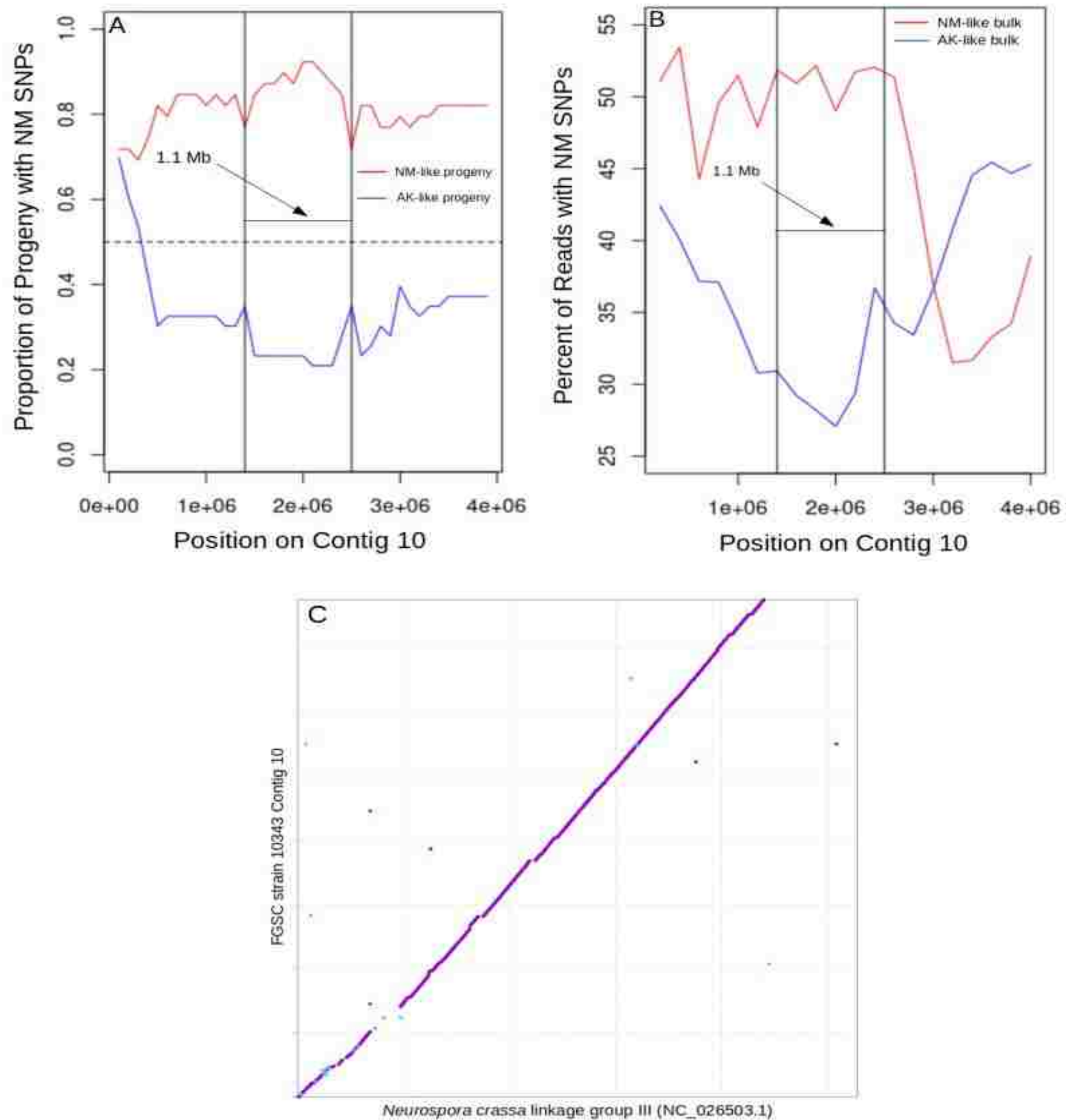

Figure 2. Summary analysis of contig 10 from FGSC strain 10343 (Alaska). (A) Results obtained by analysis of progeny based on Illumina data. The vertical axis represents the proportion of progeny with New Mexico SNPs in 100 Kb windows across the contig. New Mexico-like progeny are represented by a red line and the Alaska-like progeny are represented by a blue line. The dashed line at 0.5 indicates the expected proportion for regions not displaying selection between the phenotypic classes. At the point of highest separation (vertical black lines) 36 of 39 progeny with NM phenotypes possessed NM genotypes, and 37 of 43 progeny with AK phenotypes possessed AK genotypes. (B) Comparison of phenotypic bulks for contig 10 from FGSC strain 10343 based on bulked-segregant data. The vertical axis represents the percentage of reads containing New Mexico SNPs in 200 Kb windows across contig 10. The New Mexico-like bulk is

represented by a red line and the Alaska-like bulk is represented by a blue line. (C) Synteny comparison of linkage group III (NC_026503.1) from *Neurospora crassa* and contig 10 from the FGSC 10343 (Alaska parent) assembly generated with nucmer. Contig 10 represents a large portion of linkage group III.

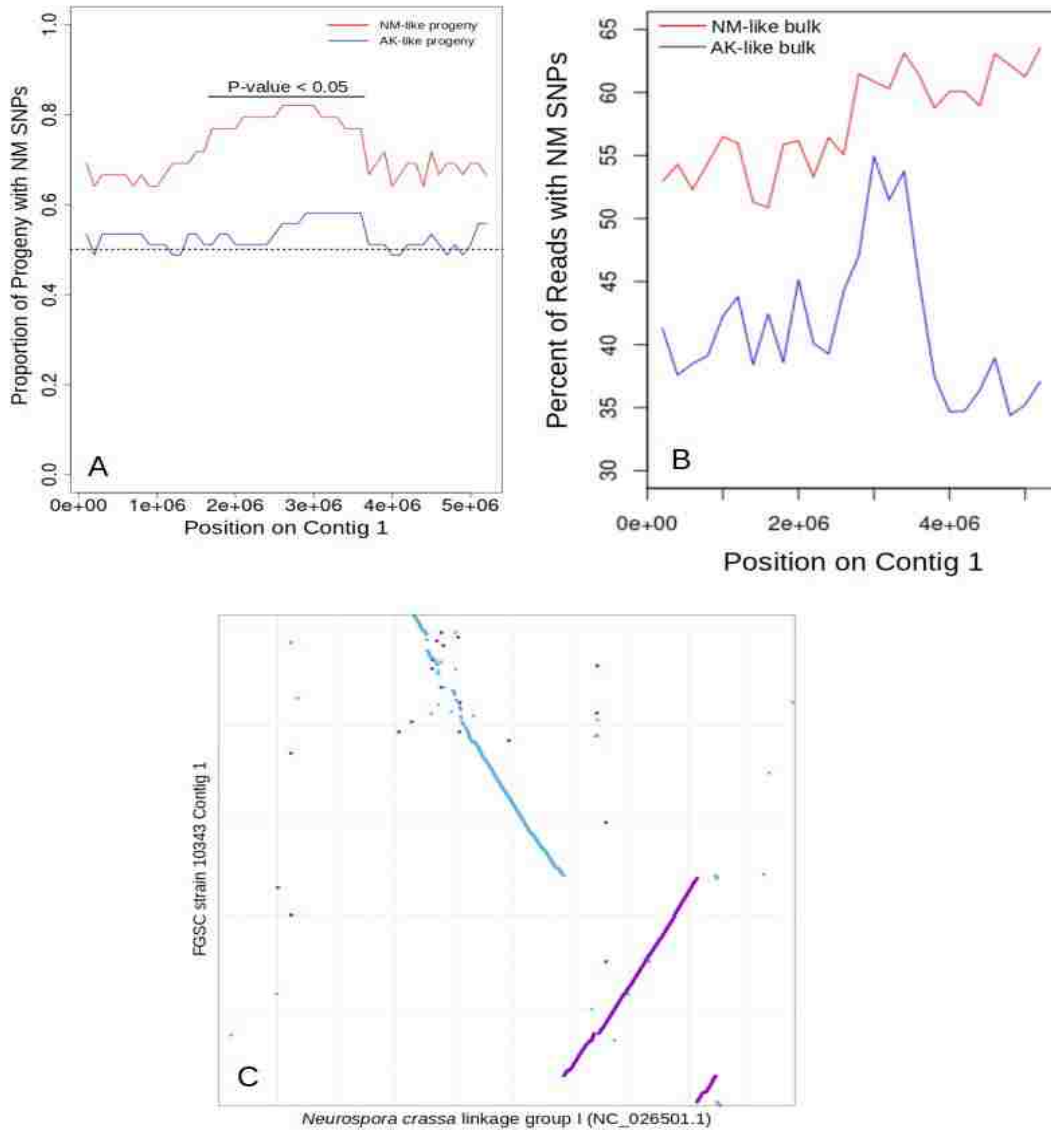

Figure 3. Summary analysis of contig 1 from FGSC strain 10343 (Alaska). (A) Comparison of progeny by phenotypic class based on Illumina data. The vertical axis represents the proportion of progeny with New Mexico SNPs in 100 Kb windows across the contig. New Mexico-like progeny are represented by a red line and the Alaska-like progeny are represented by a blue line. The dashed line at 0.5 indicates the expected proportion for regions not displaying selection between the phenotypic classes. Significant regions resulting from a chi-squared test using this proportion of 0.5 as the expected value are represented with a black bar. At the point of highest separation 32 of 39 progeny with NM phenotypes possessed NM genotypes. Genotype and phenotype were not significantly correlated for progeny with AK phenotypes. (B) Comparison of phenotypic bulks for contig 1 from FGSC strain 10343 based on bulked-segregant data. This figure represents the average percentage of reads containing New Mexico SNPs in

200 Kb windows across contig 10. The bulk for progeny with NM-like phenotypes is represented by a red line, and the AK-like bulk is represented by a blue line. (C) Synteny comparison of linkage group I (NC_026501.1) from *Neurospora crassa* and contig 1 from the FGSC 10343 (Alaska parent) assembly generated with nucmer. Contig 1 represents a large portion of linkage group I.

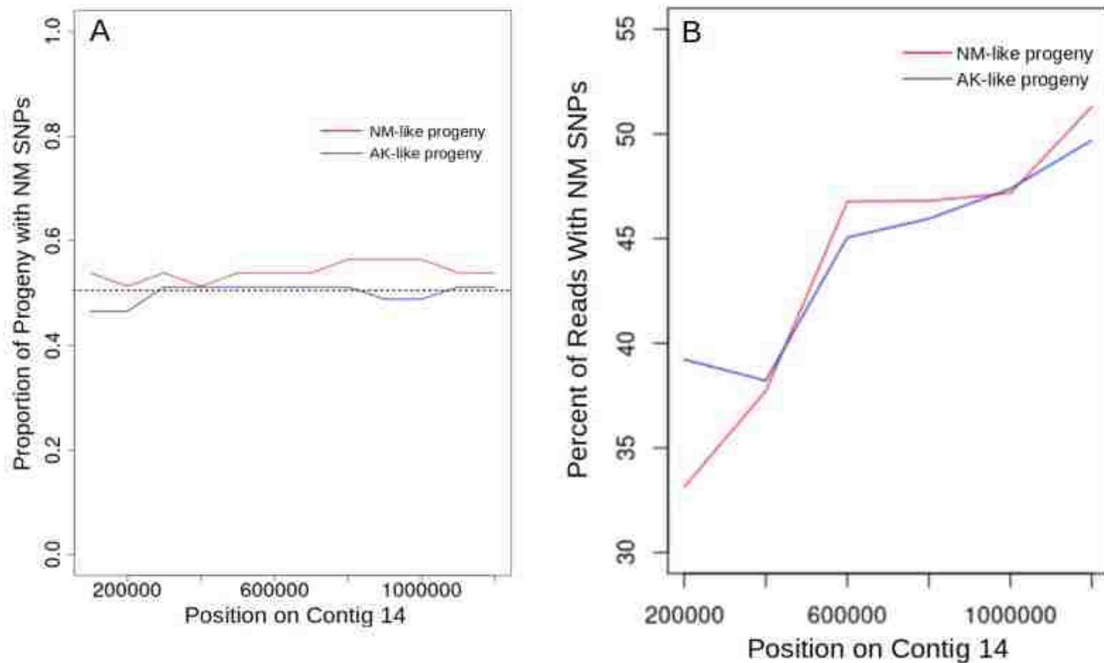

Figure 4. Summary analysis of an example genomic region with no apparent contribution to thermotolerance. (A) Comparison of progeny by phenotypic class for contig 14 from FGSC strain 10343 (Alaska) based on Illumina data. NM-like progeny are represented by a red line and Alaska-like progeny are represented by a blue line. The dotted line at 0.5 indicates the expected proportion of progeny for regions not contributing to temperature growth phenotypes The vertical axis represents the proportion of progeny from each class containing New Mexico SNPs in 100 Kb windows. (B) Results obtained for NM-like and AK-like phenotypic bulks for contig 14 using FGSC strain 10343 (Alaska) genome as a reference. The vertical axis indicates the average percentage of reads containing New Mexico SNPs in 200 Kb windows across the contig. The NM-like bulk is represented by a red line and the AK-like bulk is represented by a blue line. Both figures A and B demonstrate no significant separation between the two phenotypic classes, suggesting no relationship between genotype and temperature growth phenotype in this genomic region. This result is representative of what was observed for all 22 contigs other than contigs 10 and 1 in the FGSC 10343 (AK parent) genome assembly.

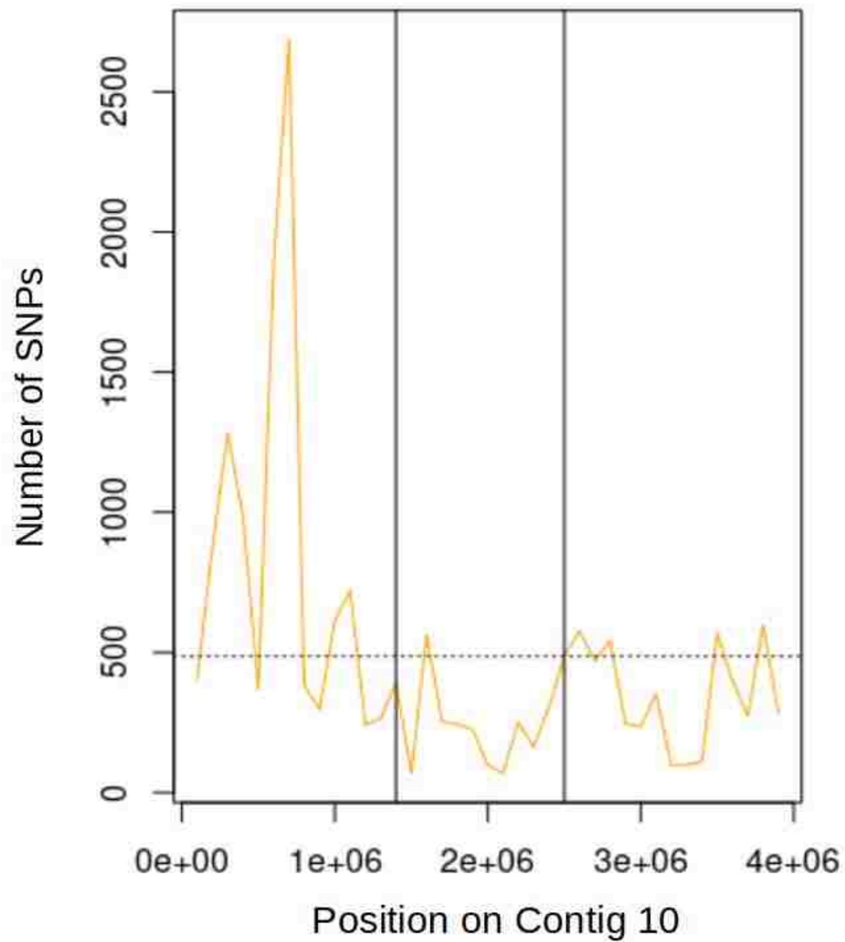

Figure 5. Number of SNPs found per 100 Kb window across contig 10 from FGSC strain 10343 assembly based on Illumina data. The 1.1 Mb region of high separation shown in Figure 3 is represented with black vertical lines. The dotted horizontal line represents the mean number of SNPs found across these 100 Kb windows (486.49).

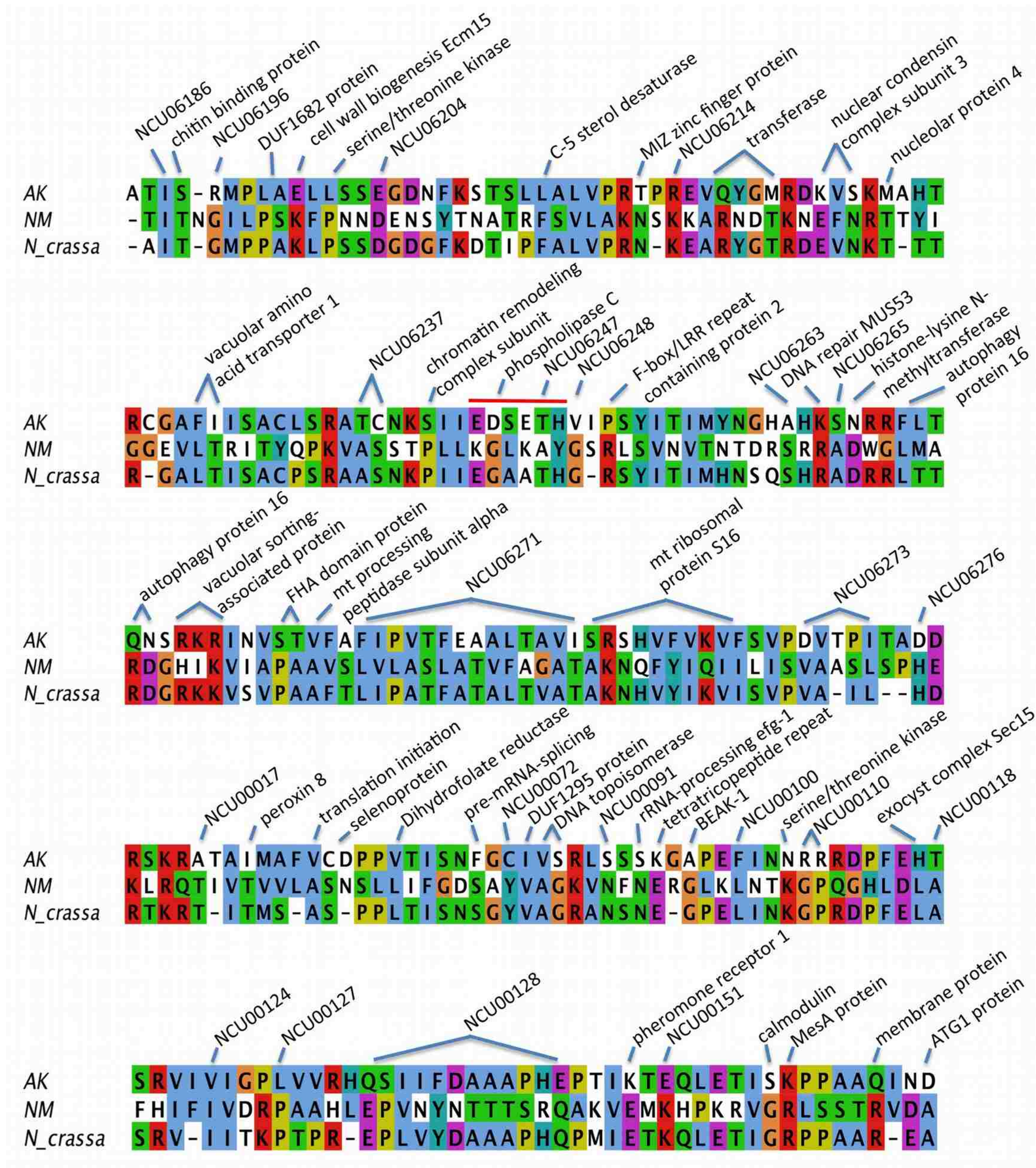

Figure 6. Amino-acid differences in the 1.1 Mb region on linkage group III. The amino-acid positions shown represent those that differed between *N. discreta* parents from NM and AK. Gene names are presented for variations that are shared between *N. crassa* and the *N. discreta* NM parental isolate. The red line above the second alignment panel corresponds to a region identified by Ellison et al. (2011) as possibly being under selection in the context of adaptation to cold in a population of *N. crassa* (see Discussion). This alignment figure was generated with Jalview and the coloration is a representation of the chemical classification of the amino-acid side chain (Clustal setting).

CHAPTER 2

Evolution of *Monosporascus* in arid environments: genomic-based approaches

Abstract

The genus *Monosporascus* represents an enigmatic group of fungi important in agriculture and widely distributed in natural arid ecosystems. Of the three recognized species of *Monosporascus*, two (*M. cannonballus* and *M. eutypoides*) are devastating pathogens on the roots of members of Cucurbitaceae in agricultural settings. A third species, *M. ibericus*, was described as an endophyte. Recent molecular and culture studies have shown that members of the genus are nearly ubiquitous as root endophytes in arid environments of the southwestern United States; isolates have been obtained from apparently healthy roots of grasses, shrubs and herbaceous plants located in central New Mexico and other regions of the Southwest. Phylogenetic and genomic analyses reveal substantial diversity among these isolates. The New Mexico isolates include close relatives of the three recognized species of *Monosporascus,* as well as isolates that represent previously unrecognized lineages. To resolve the evolutionary relationships within the genus and gain insights into potential ecological functions, we sequenced and assembled the genomes of three *M. cannonballus* isolates, one *M. ibericus* isolate and several New Mexico isolates. The assembled genomes were significantly larger than what is typical for the Xylariales despite having predicted gene numbers similar to other members of the order. Variation in the *M. cannonballus* genomes indicated substantial diversity in genome size and gene content within the species. While regions of the *Monosporascus* genomes exhibit synteny with genomes of diverse members of the

Xylariales, certain regions do not. Genomic regions lacking clear synteny with other members of the Xylariales are in general AT-rich and are in some cases enriched for genes involved in secondary metabolism. Comparisons of predicted carbohydrate-active enzymes and enzymes involved in pathogenicity suggest that endophytic *Monosporascus* isolates possess higher numbers of genes for both groups of enzymes than do pathogenic lineages. Several *Monosporascus* isolates appear to harbor bacterial endosymbionts from the genus *Ralstonia*.

Introduction

Members of the fungal genus *Monosporascus* (Ascomycota, Sordariomycetes, Xylariales) are known for their association with plant roots in agricultural and natural arid environments. Most research focused on the genus has addressed aspects of pathogenicity in agricultural settings, with the result that little is known about the functional and evolutionary diversity of members of this genus in other environments. In order to gain deeper insight into the diversity and evolutionary history of this group, we have obtained the first complete genome sequences for members of the genus *Monosporascus,* targeting isolates from diverse environments and locations. One goal was to evaluate whether lineages that possess agricultural pathogens differ from those that appear to be benign endophytes in natural environments.

There are currently three recognized *Monosporascus* species, including the closely related *M. cannonballus and M. eutypoides,* which have been studied for their ability to cause disease in the roots of many agriculturally important members of the Cucurbitaceae (Ben Salem et al. 2013). The third species, *M. ibericus*, was described as an endophyte

occurring in the roots and stems of several plant species in Spain (Collado et al. 2002). Diverse members of the genus are also common in root endophyte surveys conducted in the southwestern United States (Dean et al. 2015, Herrera et al. 2010, Porras-Alfaro et al. 2008) using both molecular and culturing methods. The near ubiquity of *Monosporascus* species on the roots of plants in arid southwestern ecosystems raises questions about their evolutionary diversity and primary ecological roles. There is no evidence that isolates of *Monosporascus* from the roots of plants in natural ecosystems are capable of causing disease despite their broad host association, which raises the question: are the southwest isolates genetically different from agricultural isolates, or are cucurbits especially susceptible hosts?

As reported here, preliminary comparisons of internal transcribed spacer (ITS) DNA regions indicated that certain *Monosporascus* isolates from arid environments were closely related to previously described species, while other isolates appeared to represent novel lineages. In order to explore this diversity more deeply, we sequenced and assembled the genomes of ten *Monosporascus* isolates including representatives of two described species and several isolates from New Mexico. This work represents the first genome sequences and assemblies for the genus. The resulting genome sizes revealed by these assemblies are significantly larger than what is commonly found in the Xylariales and other Ascomycota. Despite their larger size, the *Monosporascus* genomes have gene numbers and genome organization similar to other members of the Xylariales. Phylogenetic analyses confirmed substantial diversity across the genus while aiding in understanding the relationship of *Monosporascus* to other members of the Xylariales.

Comparisons of synteny between the *Monosporascus* assemblies and other Xylariales genomes indicated substantial differences in genome organization. Certain regions in the *Monosporascus* genomes did not exhibit clear synteny with other Xylariales genomes. Comparisons among *M. cannonballus* genomes revealed differences in genome organization and content between closely related isolates. Within the genus *Monosporascus*, endophytic strains possessed higher numbers of genes for carbohydrate-active and pathogenesis enzymes than did pathogenic isolates. Across the Xylariales, however, there was no clear correlation between gene groups and primary ecological function. Members of the genus *Monosporascus* appear to have the capability to accommodate bacterial endosymbionts from the Burkholderiaceae.

Methodology

Sample collection

Monosporascus cannonballus (CBS 586.93 & CBS 609.92) and *Monosporascus ibericus* (CBS 110550) strains were obtained from the CBS-KNAW fungal collection (http://www.westerdijkinstitute.nl/Collections/). Six additional isolates were obtained from surface sterilized roots collected in previous endophyte surveys conducted at the Sevilleta National Wildlife Refuge (SNWR) in New Mexico (Dean et al. 2015, Porras-Alfaro et al. 2014). *Monosporascus cannonballus* strain MC13-8B was provided courtesy of Michael Stanghellini, University of California, Riverside. Geographic and host information is provided in Table 1.

Molecular methods and sequencing

Genomic DNA was obtained from pure cultures following a CTAB extraction protocol (Robinson and Natvig, 2019). The KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, Massachusetts) was used to prepare genomic libraries for each isolate. Genome sequencing was performed using the Illumina NextSeq 500 platform configured for 150 base-pair read lengths. A total yield of 52.99 Gbp was generated from this sequencing run.

Genome Assembly

Quality control of the resulting Illumina reads was performed using Trimmomatic (Bolger et al. 2014) and three separate software packages designed for short read de-novo microbial genome assemblies were independently optimized to assemble these QC-filtered reads (Robinson and Natvig, 2019). Quast (Gurevich et al., 2013) was used to generate assembly statistics and further assembly quality assessment was performed using the BUSCO software package and Ascomycota dataset (Simão et al., 2015). Assemblies produced in SPAdes (Bankevich et al., 2012) consistently showed the highest quality and were annotated using AUGUSTUS with *Neurospora crassa* selected for the species parameter and limited to few alternative transcripts and only complete gene predictions (Stanke and Morgenstern, 2005; Hoff and Stanke, 2013). These assemblies have been deposited at GenBank with the following accession numbers: QJNS00000000, QJNT00000000, QJNU00000000, QJNV00000000, QJNW00000000, QJNX00000000, QJNY00000000, QJNZ00000000, QJOA00000000, and QJOB00000000.

Bioinformatics

CD-HIT (Li and Godzik 2006) was used to examine our assemblies for gene duplications and to compare protein similarities among assemblies. Examinations of synteny were conducted using nucmer with default settings and visualized using mummerplot, which are both part of the MUMmer4 package (Kurtz et al. 2004). Predicted protein sequences generated by AUGUSTUS were annotated for carbohydrate-active enzymes using the dbCAN2 meta server (http://cys.bios.niu.edu/dbCAN2/). For the dbCAN2 annotation we utilized HMMER (E-Value < 1e-15, coverage > 0.35), DIAMOND (E-Value < 1e-102) and Hotpep (Frequency > 2.6, Hits > 6) and only candidates agreed upon by at least two of these methods were retained. A local version of the PHI-base (Urban et al. 2017) database was created using BLAST, and predicted protein sequences were queried against this database using blastp. Only the top hit produced by blastp was retained. InterProScan was used to infer the function of protein sequences from non-syntenic regions of *Monosporascus* and other members of the Xylariales using default settings.

Molecular alignments and phylogenetic analyses

ITS sequences were obtained from the National Center for Biotechnology Information (NCBI) databases and aligned with Clustal Omega (Sievers et al. 2011). Sequences from our *Monosporascus* assemblies and genome assemblies of other members of the Xylariales for beta-tubulin, minichromosome maintenance complex component 7 (MCM7), translation elongation factor 1 alpha (EF1α), DNA-directed RNA polymerase II subunit RPB1 and DNA-directed RNA polymerase II subunit RPB2 were

used in a five-gene phylogenetic analysis. These sequences were identified based on blastn and blastp alignments generated using queries from members of the Xylariales and Sordariales (Table 7). Sequences were aligned separately using Clustal Omega (Sievers et al. 2011) and the aligned sequences were concatenated in alphabetical order. These concatenated sequences were then used in phylogenetic tree construction. RaxML (Stamatakis 2014) was used to construct maximum likelihood phylogenetic trees using the GTRCAT substitution model and 1000 boostrap replicates. Separate phylogenetic trees were generated for each gene individually to check for conflicts prior to performing the concatenated analysis. Final graphical representations of each tree were produced with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Results and discussion

We sequenced, assembled and annotated the genomes of ten *Monosporascus* isolates that spanned the diversity of isolates initially observed based on ITS sequence analyses. The sizes of these genome assemblies ranged from 70 to > 100 Mb, which is larger than what is typical within the Sordariomycetes. Three different software packages designed to assemble microbial genomes de-novo produced assemblies of similar size for each isolate. The consistency among these assembly methods supported the large genome estimates. Summary statistics for final SPAdes assemblies demonstrated no relationship between assembly fragmentation (number of contigs) and total assembly size (Table 2). A linear regression analysis employing *Monosporascus* genomes as well as genomes from the Sordariales and other members of the Xylariales revealed a negative relationship between GC content and genome size (Figure 2). BUSCO, which is commonly used to

assess genome assemblies quantitatively, was able to locate over 97% of the expected complete single-copy Ascomycota orthologs in every *Monosporascus* assembly (Table 6). These results helped substantiate the validity of the large predicted genome sizes within this genus and supported the conclusion that assemblies were of good quality.

During quality control filtering for NCBI submission we noticed several contigs had strong hits to the bacterial species *Ralstonia pickettii*. Two of the *Monosporascus* assemblies contained nearly complete genomes for *R. pickettii,* and numerous other *R. pickettii* sequences were found in nearly every assembly. Given that our *Monosporascus* isolates were originally cultured on media containing antibiotics and the lack of other bacterial sequences at such abundances, it is unlikely these sequences are the result of contamination in our laboratory. The genus *Ralstonia* belongs to the Burkholderiaceae, a family that contains known fungal endosymbionts (Lastovetsky et al. 2018).

To better understand diversity within the genus *Monosporascus* we performed a five-gene phylogenetic analysis using the genes beta-tubulin, MCM7, EF1α, RPB1 and RPB2. These genes are commonly used to examine phylogenetic relationships in fungi at various scales (James et al. 2006). This multi-gene analysis confirmed the high level of diversity initially observed with ribosomal RNA ITS analyses. There are close relatives of both *M. cannonballus* and *M. ibericus* among the New Mexican isolates. Several members of the Xylariales were included in this analysis to gain a better understanding of the placement of this genus within the order. The *Monosporascus* isolates were most closely related to *Eutypa lata*, which is commonly associated with disease in grape vines and has been placed in the family Diatrypaceae (Acero et al. 2004).

Annotation of the *Monosporascus* genome assemblies indicated they have predicted gene content similar to other members of the Xylariales, despite their inflated genome sizes. There was no consistent relationship between genome size and predicted gene number among the *Monosporascus* assemblies or across the Xylariales in general (Table 3). Synteny comparisons between *Monosporascus* genomes and other members of the Xylariales resulted in regions of synteny as well as large regions lacking synteny. The genome from the GIB2 strain was used to examine in some detail the regions lacking synteny. This genome was used as a reference because it was the smallest *Monosporascus* genome and would therefore minimize noise from *Monosporascus*-specific variation (Table 2). We identified 371 contigs in the GIB2 assembly that did not show synteny with any of the non-*Monosporascus* assemblies listed in Table 3, although the majority of these contigs possessed synteny with other *Monosporascus* genomes. The combined size of these contigs was 11.08 Mb. Of these, 207 (1.0 Mb) contained at least one gene annotation with a total of 975 gene annotations across all 207 contigs, and an average GC content of 40.08% (comparable to the average genome-wide GC content of 39.53% for all *Monosporascus* genomes). These contigs appear to be enriched for genes involved in secondary metabolism. The remaining 164 contigs (10.8 Mb) had no predicted gene content and possessed an average GC content of 21.96%, indicating the existence of AT-rich regions with no or obscured synteny with respect to members of the Xylariales outside the genus *Monosporascus*.

Substantial variation in predicted gene number existed among the three agricultural isolates from the single species *M. cannonballus*. *M. cannonballus* isolates from the southwestern United States, CBS609.92 and MC13-8B, were found to have very

similar predicted gene numbers (10,682 and 10,641, respectively), but the strain from Egypt, CBS586.93, had a larger predicted genome size (91.50 Mb) and number of genes (11,585). Comparisons of protein similarity using CD-HIT-2D indicated that the predicted amino-acid sequences in the southwestern *M. cannonballus* isolates were more similar, based on identity, than the predicted amino-acid sequences found in the Egyptian isolate. The number of predicted protein sequences from the Egyptian isolate that did not pair with sequences from either of the southwestern isolates based on identity was larger than the differences between southwesten isolates. To examine further this apparent divergence among the *M. cannonballus* isolates we examined synteny across the *M. cannonballus* genome assemblies. The southwestern isolates exhibited nearly complete synteny with each other, while comparisons between southwestern and Egyptian isolates indicated gaps in synteny (Figure 4).

Endophytic isolates of *Monosporascus* from New Mexico associate with a broad range of hosts, and evidence from molecular and culture studies suggest that this broad host range can apply to individual terminal clades within the genus. This raises questions regarding whether presumed endophytic lineages differ from pathogenic lineages with respect to specific genes or groups of genes. We compared the genomes of endophytic and pathogenic isolates within the genus *Monosporascus* and also across genera within the Xylariales with respect to genes for carbohydrate-active enzymes, genes known to be involved in pathogenicity in certain fungi, and genes for effector proteins that facilitate colonization of plant tissues.

Carbohydrate-active enzymes interact with a number of substrates that are common in plant cells, and it is predicted that some of these enzymes may facilitate

plant-fungal interactions (Knapp et al. 2018). The endophytic New Mexican *Monosporascus* isolates have a greater number of these enzymes than do the pathogenic isolates (Table 4). Differences in the number of enzymes present between the New Mexico and agricultural isolates were greatly reduced when examining the presence of unique enzymes from each gene family suggesting differences in gene copy number for certain enzymes (Table 4). Glycosidases comprised the majority of these enzymes and the genes for some types exhibited high copy numbers. The order Xylariales contains genera with species classified as pathogens as well as others classified as endophytes. Comparing profiles across these two ecological groups, there appears to be no observable relationship between either the total or unique number of genes for these enzymes and primary ecological function (Table 4).

The PHI-base database catalogs genes that have been experimentally verified to be involved in pathogenicity for both eukaryotic and prokaryotic pathogens that interact with plant hosts (Urban et al. 2017). The *Monosporascus* isolates from New Mexico contain a higher number of pathogenicity homologs than do the *M. cannonballus* isolates (Table 4). This result is somewhat surprising given that the New Mexican isolates have not been observed to cause disease in any of their hosts, while *M. cannonballus* is a well described pathogen. This also raises a question about the function of these homologs and if perhaps they could have been altered to allow access to a wider host range or possibly encouraging a more mutualistic lifestyle.

Finally, we examined effector-protein genes within the genomes of *Monosporascus* isolates. Effector proteins are secreted by fungi to facilitate colonization of plant cells. Again, the New Mexico isolates had a higher content of these effector proteins (Table 5),

further supporting the possibility that endophytic lineages have adapted to associate with a broad host range by facilitating a wide range of dynamic plant-fungal interactions. Our comparison also included other members of the Xylariales and once again this analysis showed no obvious relationship between primary ecological function and the content of these two categories of PHI-base genes (Table 5).

Acknowledgments

This research was partially supported by a National Science Foundation award to the University of New Mexico for the Sevilleta Long-term Ecological Research program. I thank George Rosenberg and Lijing Bu of the UNM Department of Biology's Molecular Biology Facility for help with sequencing and analysis and I acknowledge support from the UNM Center for Evolutionary & Theoretical Immunology (CETI) under National Institutes of Health grant P30GM110907. I thank the UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing access to high performance systems for data analyses.

References

Acero, F.J., González, V., Sánchez-Ballesteros, J., Rubio, V., Checa, J., Bills, G.F., Salazar, O., Platas, G. and Peláez, F. 2004. Molecular phylogenetic studies on the Diatrypaceae based on rDNA-ITS sequences. Mycologia 96, 249-259.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi,

N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19, 455-477.

Ben Salem, I., Correia, K.C., Boughalleb, N., Michereff, S.J., León, M., Abad-Campos, P., García-Jiménez, J. and Armengol, J. 2013. *Monosporascus eutypoides*, a cause of root rot and vine decline in Tunisia, and evidence that *M. cannonballus* and *M. eutypoides* are distinct species. Plant Disease 97, 737-743.

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

Collado, J., González, A., Platas, G. and Stchigel, A.M. 2002. *Monosporascus ibericus* sp. nov., an endophytic ascomycete from plants on saline soils, with observations on the position of the genus based on sequence analysis of the 18S rDNA. Mycological Research 106, 118-127.

Dean, S.L., Warnock, D.D., Litvak, M.E., Porras-Alfaro, A. and Sinsabaugh, R.L. 2015. Root-associated fungal community response to drought-associated changes in vegetation community. Mycologia 107, 1089-1104.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.

Herrera, J., Khidir, H.H., Eudy, D.M., Porras-Alfaro, A., Natvig, D.O. and Sinsabaugh, R.L. 2010. Shifting fungal endophyte communities colonize *Bouteloua gracilis*: effect of host tissue and geographical distribution. Mycologia 102, 1012-1026.

Hoff, K.J., Stanke, M. 2013. WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 41, W123–W128.

James, T.Y. et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443,818-822.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30,1236-1240.

Knapp, D.G., Németh, J.B., Barry, K., Hainaut, M., Henrissat, B., Johnson, J., Kuo, A., Lim, J.H.P., Lipzen, A., Nolan, M., Ohm, R.A., Tamás, L., Grigoriev, I.V., Spatafora, J.W., Nagy, L.G. and Kovács, G.M. 2017. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. Scientific Reports 8, 6321.

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

Lastovetsky, O.A., Ahn, E., Mondo, S.J., Toomer, K.H., Zhang, A., Johnson L.M. and Pawlowska, T.E. 2018. Distribution and population structure of endobacteria in arbuscular mycorrhizal fungi at North Atlantic dunes. The ISME Journal 12, 3001-3013.

Li, W. and Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22,1658-1659.

Porras-Alfaro, A., Herrera, J., Sinsabaugh, R.L., Odenbach, K.J., Lowrey, T. and Natvig, D.O. 2008. Novel Root Fungal Consortium Associated with a Dominant Desert Grass. Applied and Environmental Microbiology 74, 2805-2813.

Porras-Alfaro, A., Raghavan, S., Garcia, M., Sinsabaugh, R.L., Natvig, D.O. and Lowrey, T.K. 2014. Endophytic fungal symbionts associated with gysophilous plants. Botany 92, 295-301.

Robinson, A.J. and Natvig, D.O. 2019. Diveerse members of the Xylariales lack canonical mating-type regions. Fungal Genetics and Biology 122, 47-52.

Sievers, F., Wilm A., Dineen D.G., Gibson T.J., Karplus K., Li W., Lopez R, McWilliam H., Remmert M., Söding J., Thompson J.D. and Higgins D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7, 539.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with singly-copy orthologs. Bioinformatics 31, 3210–3212.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stanke, M., Morgenstern, B. 2005. AUGUSTUS: a web server for gene prediction in eu-karyotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467.

Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., Sadanadan, V., Khamari, L., Billal, S., Mohanty, S. and Hammond-Kosack, K.E. 2017. PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. Nucleic Acids Research 45, D604-D610.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y. and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46, W95-W101.

Table 1. Metadata of *Monosporascus* isolates selected for whole genome sequencing

Species (strain)	Location	Host	NCBI Accession
Monosporascus cannonballus (CBS 609.92)	Arizona	*Cucumis melo*	QJNS00000000
Monosporascus cannonballus (CBS 586.93)	Egypt	*Cucumis melo*	QJNT00000000
Monosporascus cannonballus (MC13-8B)	California	*Cucumis melo*	QJNW00000000
Monosporascus ibericus (CBS110550)	Spain	Undetermined	QJNU00000000
Monosporascus sp. (5C6A)	New Mexico	*Bouteloua eriopoda*	*QJOB00000000*
Monosporascus sp. (CRB-8-3)	New Mexico	*Larrea tridentata*	*QJNZ00000000*
Monosporascus sp. (CRB-9-2)	New Mexico	*Larrea tridentata*	*QJOA00000000*
Monosporascus sp. (MG133)	New Mexico	*Mentzelia perennis Wooton*	QJNX00000000
Monosporascus sp. (MG162)	New Mexico	*Nerisyrenia linearifolia*	*QJNY00000000*
Monosporascus sp. (GIB2)	New Mexico	*Bouteloua eriopoda*	*QJNV00000000*

Table 2. Summary *Monosporascus* assembly statistics

Species (strain)	Number of contigs	N50	Total length (Mb)
Monosporascus sp. (GIB2)	935	138,727	70.59
Monosporascus sp. (MG133)	768	166,637	71.87
Monosporascus sp. (5C6A)	823	246,621	75.25
Monosporascus sp. (CRB-9-2)	2,348	88,305	76.10
Monosporascus ibericus (CBS110550)	1,612	104,428	87.25
Monosporascus cannonballus (MC13-8B)	1,441	151,284	89.29
Monosporascus sp. (MG162)	1,320	134,088	89.46
Monosporascus cannonballus (CBS609.92)	753	256,764	89.62
Monosporascus cannonballus (CBS586.93)	758	248,352	91.50
Monosporascus sp. (CRB-8-3)	2,382	93,740	102.96

Table 3. Xylariales annotation and assembly size comparison

Species (strain)	Predicted number of genes	Total length (Mb)	NCBI Accession
Monosporascus sp. (GIB2)	11,783	70.59	QJNS00000000
Monosporascus sp. (MG133)	11,569	71.87	QJNT00000000
Monosporascus sp. (5C6A)	12,537	75.25	QJNW00000000
Monosporascus sp. (CRB-9-2)	12,001	76.10	QJNU00000000
Monosporascus ibericus (CBS110550)	11,063	87.25	*QJOB00000000*
Monosporascus cannonballus (MC13-8B)	10,641	89.29	*QJNZ00000000*
Monosporascus sp. (MG162)	11,881	89.46	*QJOA00000000*
Monosporascus cannonballus (CBS609.92)	10,682	89.62	QJNX00000000
Monosporascus cannonballus (CBS586.93)	11,585	91.50	*QJNY00000000*
Monosporascus sp. (CRB-8-3)	11,775	102.96	*QJNV00000000*
Eutypa lata (UCREL1)	11,685	54.01	AORF00000000.1
Pestalotiopsis fici (W106)	15,413	51.91	ARNU00000000.1
Microdochium bolleyi (J235TASD1)	13,338	38.84	LSSP00000000.1
Rosellinia necatrix (W97)	12,644	44.26	BBSO00000000.2
Daldinia eschscholtzii (UM 1400)	N/A	35.76	CCED00000000.1
Xylaria striata (RK1-1)	N/A	59.79	LOBO00000000.1

Table 4. dbCAN2 results

Isolate (strain)	Functional Guild	Carbohydrate-active enzymes	Unique
M. cannonballus (CBS 609.92)	Pathogen	427	153
M. cannonballus (CBS 586.93)	Pathogen	441	158
M. cannonballus (MC13-8B)	Pathogen	428	151
M. ibericus (CBS 110550)	Endophyte	448	153
5C6A	Endophyte	474	154
CRB-8-3	Endophyte	469	161
CRB-9-2	Endophyte	471	158
MG133	Endophyte	476	161
MG162	Endophyte	470	157
GIB2	Endophyte	471	155
Eutypa lata (UCREL1)	Pathogen	494	158
Daldinia sp. (EC12)	Endophyte	407	145
Rosellinia necatrix (W97)	Pathogen	444	166
Microdochium bolleyi (J235TASD1)	Endophyte	433	158
Pestalotiopsis fici (W106-1)	Endophyte	683	172

These numbers are the results of automated CAZyme annotations on protein sequences from each isolate presented above. HMMER (E-Value < 1e-15, coverage > 0.35), DIAMOND (E-Value < 1e-102) and Hotpep (Frequency > 2.6, Hits > 6) were all utilized and only candidates found by at least two of these methods are represented in the last column. The functional guild indicates the primary description found in GenBank.

Table 5. PHI-base results

Isolate (strain)	Genes Associated with Pathogenicity	Effector Genes	Functional Guild
CBS609.92	1000	276	Pathogen
CBS586.93	1081	316	Pathogen
MC13-8B	1008	282	Pathogen
CBS110550	1042	346	Endophyte
GIB2	1108	350	Endophyte
5C6A	1175	394	Endophyte
MG133	1079	364	Endophyte
MG162	1148	376	Endophyte
CRB-8-3	1129	363	Endophyte
CRB-9-2	1154	385	Endophyte
Rosellinia necatrix (W97)	1191	384	Pathogen
Microdochium bolleyi (J235TASD1)	1217	407	Endophyte
Eutypa lata (UCREL1)	1089	317	Pathogen
Daldinia sp. (EC12)	1094	334	Endophyte

Table 6. BUSCO results

Isolate	BUSCO results
CBS 609.92	C:98.2%[S:98.0%,D:0.2%],F:0.8%,M:1.0%,n:1315
CBS 586.93	C:97.8%[S:97.4%,D:0.4%],F:1.1%,M:1.1%,n:1315
CBS 110550	C:98.1%[S:97.9%,D:0.2%],F:0.8%,M:1.1%,n:1315
5C6A	C:98.1%[S:97.6%,D:0.5%],F:0.8%,M:1.1%,n:1315
CRB-8-3	C:98.4%[S:97.9%,D:0.5%],F:0.8%,M:0.8%,n:1315
CRB-9-2	C:98.3%[S:97.9%,D:0.4%],F:0.7%,M:1.0%,n:1315
MG133	C:98.4%[S:97.9%,D:0.5%],F:0.9%,M:0.7%,n:1315
MG162	C:98.4%[S:98.1%,D:0.3%],F:0.5%,M:1.1%,n:1315
GIB2	C:98.1%[S:97.6%,D:0.5%],F:0.8%,M:1.1%,n:1315

Given 1315 (n) total BUSCO groups searched this table represents the percentage of complete (C) , complete and single-copy (S), complete and duplicated (D), fragmented (F) and missing (M) BUSCOs.

Table 7. Sordariales and Xylariales sequences used as queries in multi-gene phylogenetic analysis

Gene	Species (strain)	Sequence type	NCBI Accession
Beta-tubulin	*Pestalotiopsis fici* (W106-1)	Nucleotide	NW_006917102.1: 1335824-1337715
EF1α	Neurospora crassa (OR74A)	Protein	EAA35632.2
MCM7	Neurospora crassa (OR74A)	Protein	EAA34642.1
RPB1	*Pestalotiopsis fici* (W106-1)	Protein	ETS81956.1
RPB2	*Biscogniauxia marginata* (CBS 124505)	Protein	AMQ10338.1

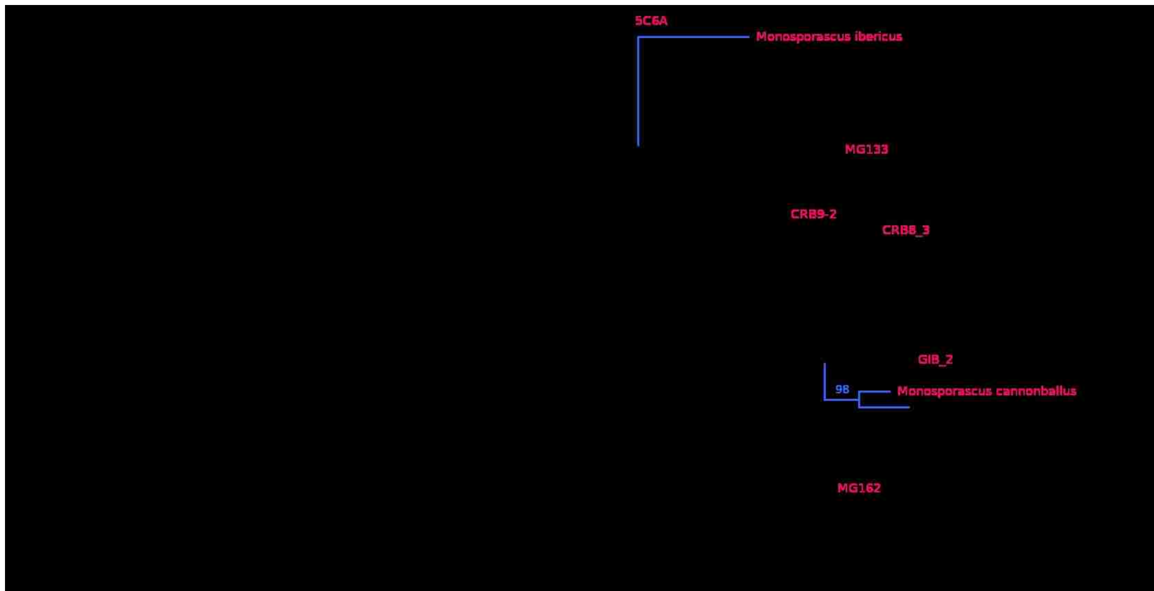

Figure 1. Maximum likelihood tree for strains of *Monosporascus* based on analysis of ribosomal RNA ITS sequences. Previously described isolates of *Monosporascus* are indicated by branches shown in blue. Isolates selected for whole genome sequencing are highlighted in red. Other isolates are from culture surveys of root endophytes at the SNWR in cental New Mexico. Bootstrap values (percent of 1000 replicates) are shown for branches with greater than 65% support.

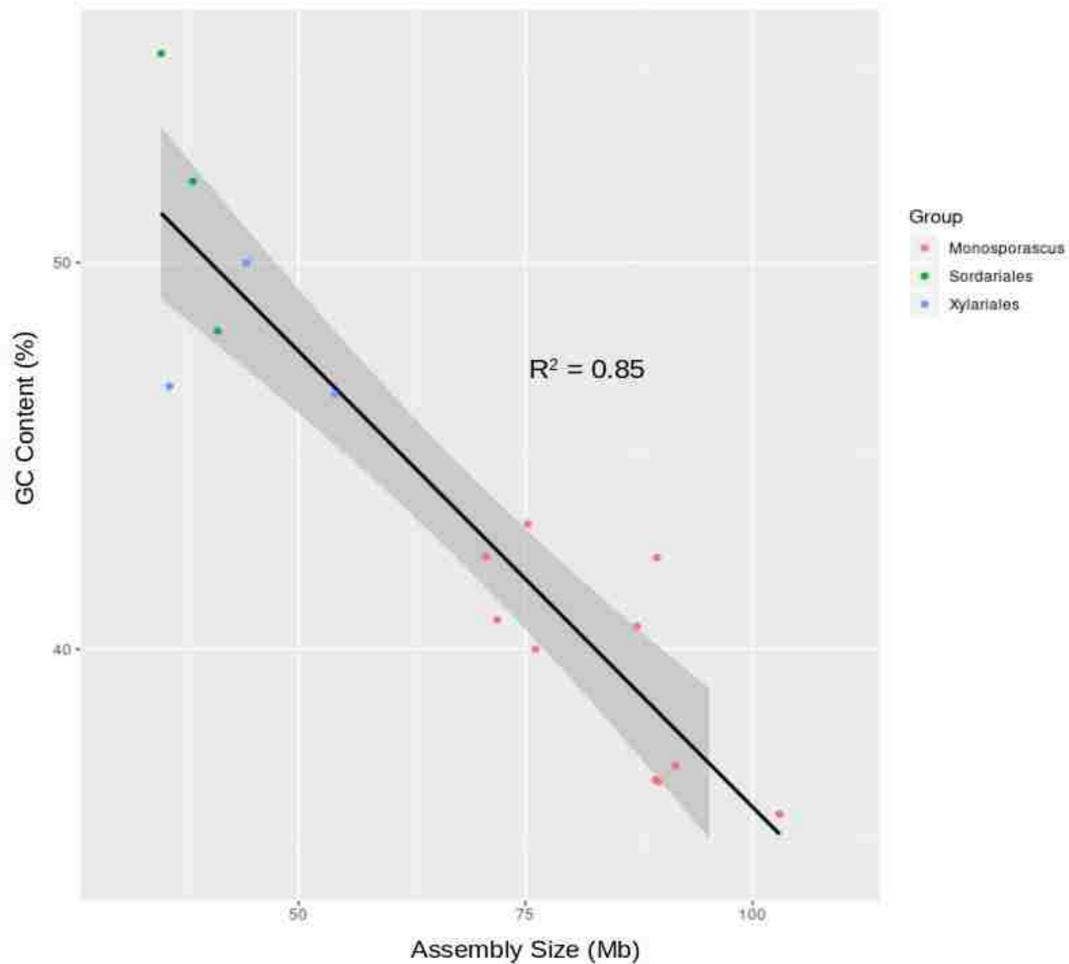

Figure 2. Relationship between GC content and genome size for several genomes from the Sordariales and Xylariales. Linear regression analysis demonstrating the strong negative relationship between GC content and assembly size (p-value = 2.701e-07). The ten *Monosporascus* assemblies are represented in red. The green circles indicate assemblies for the following members of the Sordariales: *Neurospora crassa* (41.1 Mb), *Sordaria macrospora* (38.39 Mb) and *Chaetomium globosum* (34.89 Mb). The blue circles indicate assemblies for the following members of the Xylariales: *Eutypa lata* (54.01 Mb), *Rosellinia necatrix* (44.26 Mb) and *Daldinia eschscholtzii* (35.81 Mb).

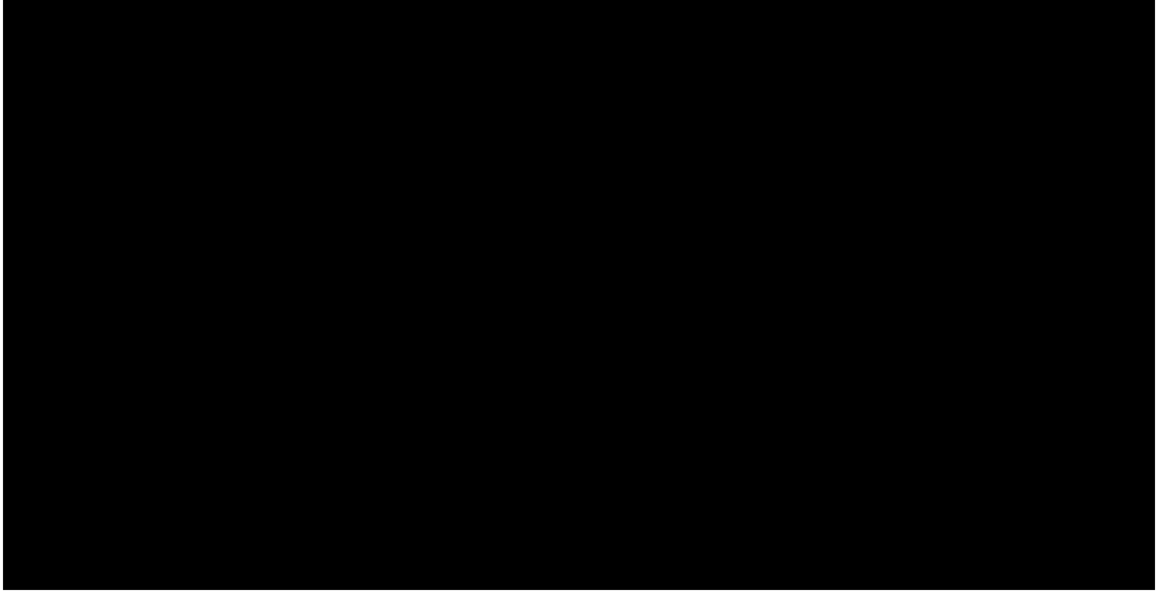

Figure 3. Five-gene maximum-likelihood tree. Bootstrap values (percent of 1000 replicates) are shown for branches with greater than 65% support.

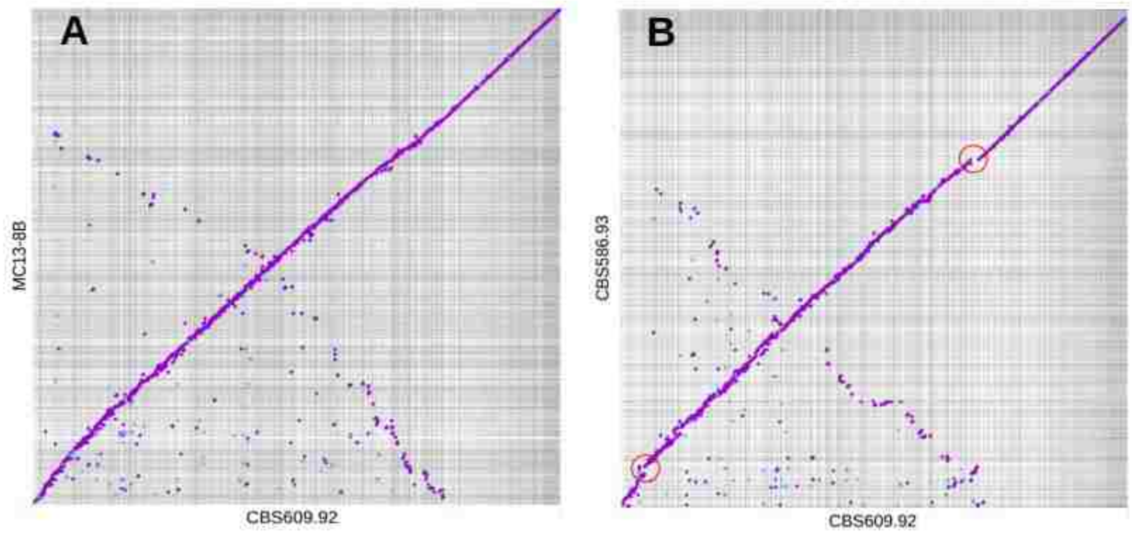

Figure 4. Synteny comparisons of *M. cannonballus* isolates. These plots were generated using nucmer and visualized with mummerplot. Figure A is a comparison of *M. cannonballus* isolates from the southwestern United states, which demonstrates strong synteny. Figure B is a comparison of synteny between M. cannonballus isolates from the southwestern United States and Egypt with gaps in synteny shown in red circles.

Fungal Genetics and Biology 122 (2019) 47–52

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Diverse members of the Xylariales lack canonical mating-type regions

Aaron J. Robinson, Donald O. Natvig*

Department of Biology, University of New Mexico, Albuquerque, NM 87131 USA

ARTICLE INFO

Keywords:
Xylariales
Monosporascus
Xylaria
Microdochium
Mating type
Genomics

ABSTRACT

A survey of genomes reported here for 10 isolates of *Monosporascus* species and an additional 25 genomes from other members of the Xylariales (representing 15 genera) available in public databases indicated that genes typically associated with *MAT1-1* (*mat A*) or *MAT1-2* (*mat a*) mating types are absent or have diverged greatly relative to counterparts in other Pezizomycotina. This was particularly surprising for isolates known to be homothallic, given that homothallic members of the Pezizomycotina typically possess a *MAT1-1-1* (*mat A-1*) gene and one or both of two other closely-linked mating-type genes, *MAT1-1-2* (*mat A-2*) and *MAT1-1-3* (*mat A-3*), in addition to *MAT1-2-1* (*mat a-1*). We failed to detect candidate genes for either *MAT1-1-1* or *MAT1-1-2* in any member of the Xylariales. Genes related to *MAT1-2-1* and *MAT1-1-3* are present in the genomes examined, but most appear to be orthologs of MATA_HMG (high-mobility group) genes with non-mating-type functions rather than orthologs of mating-type genes. Several MATA_HMG genes were found in genome positions that suggest they are derived from mating-type genes, but these genes are highly divergent relative to known *MAT1-2-1* and *MAT1-1-3* genes. The genomes examined represent substantial diversity within the order and include *M. cannonballus*, *M. ibericus*, *Xylaria hypoxylon*, *X. striata*, *Daldinia eschscholzii, Eutypa lata*, *Rosellinia necatrix*, *Microdochium bolleyi* and several others. We employed a number of avenues to search for homologs, including multiple BLAST approaches and examination of annotated genes adjacent to genes known to flank mating regions in other members of the Ascomycota. The results suggest that the mating regions have been lost from, or altered dramatically in, the Xylariales genomes examined and that mating and sexual development in these fungi are controlled differently than has been reported for members of the Pezizomycotina studied to date.

1. Introduction

The mating-type genes that control sexual reproduction have been studied in diverse homothallic and heterothallic members of the Ascomycota. Within the Sordariales, Hypocreales, Ophiostomatales and other orders of Ascomycota, individuals from heterothallic species possess one of two mating idiomorphs. One of these idiomorphs (referred to as *mat A* in *Neurospora*, but *MAT1-1* in other groups) typically has three different linked genes: *MAT1-1-1* (*mat A-1*), *MAT1-1-2* (*mat A-2*), and *MAT1-1-3* (*mat A-3*). The *MAT1-1-1* gene shares a region of homology with the yeast *mat α1* gene. The other idiomorph (designated *mat a* in *Neurospora*, *MAT1-2* in other groups) typically has one gene, designated *MAT1-2-1* (aka *mat a-1*) (reviewed by Martin et al. (2010)).

The mating-gene arrangement in homothallic species varies, but when it is known, it usually falls within one of three types. In one type, both mating-type regions are present in the same haploid nucleus but are in different parts of the genome. A second type has *MAT1-1* and *MAT1-2* genes present within a single mating-type region. A third type has the *MAT1-1* (*mat α1*) region but lacks an identifiable *MAT1-2* region. In the genus *Neurospora*, these three homothallic types are represented by *N. sublineolata*, *N. pannonica*, and *N. africana*, respectively (Gioti et al., 2012).

Relevant to the results presented below, several mating-type genes have domains that are characteristic of the high mobility group (HMG) of transcriptional activator genes. Among the mating-type genes with HMG domains, *MAT1-1-1* genes are the most distantly related. *MAT1-1-1* genes encode proteins that are part of a family referred to as MATα_HMG, which includes the *Saccharomyces cerevisiae* MATα1 protein, whereas *MAT1-2-1* and *MAT1-1-3* genes encode proteins in the MATA_HMG family (Jackson et al., 2013; Martin et al., 2010). Although the MAT1-2-1 and MAT1-1-3 proteins possess clear regions of homology, phylogenetic analyses employing sordariomycete versions of these proteins suggested they form two distinct clades (Debuchy et al., 2010). The proteins encoded by the fourth group of mating-type genes present in diverse members of the Pezizomycotina, *MAT1-1-2* (*mat A-2*), have not been linked to the HMG superfamily, but homologs of this gene group are conserved widely among members of the Sordariomycetes and have a conserved domain (PFAM PF17043). There is

* Corresponding author.
E-mail address: dnatvig@gmail.com (D.O. Natvig).

https://doi.org/10.1016/j.fgb.2018.12.004
Received 19 July 2018; Received in revised form 12 December 2018; Accepted 12 December 2018
Available online 14 December 2018
1087-1845/ © 2018 Elsevier Inc. All rights reserved.

evidence that the MAT1-1-1 protein interacts with other mating-type proteins and can be required for ascocarp development (Dyer et al., 2016).

Species of *Neurospora* and other Sordariomycetes possess other genes in the MATA_HMG transcription-factor family, including NCU03481 and *fmf-1* (NCU09387), that have been implicated in aspects of sexual development but which are not specifically mating-type genes. They are not located within the mating-type region and therefore in heterothallic species are found in both *MAT1-1* and *MAT1-2* strains. The *fmf-1* gene encodes the homolog of *Schizosaccharomyces pombe* Ste11p, a protein that activates multiple genes required for sexual development (Iyer et al., 2009). *N. crassa fmf-1* mutants are defective in ascocarp development, and the encoded protein appears to control other genes involved in sexual development (Johnson, 1979; Wang et al., 2014). The *N. crassa* gene designated NCU03481 is the ortholog of the *Podospora anserina* gene *PaHMG8*. Deletion of *PaHMG8* in *P. anserina* results in failed ascocarp development, and similar to *fmf-1,* its gene product has been implicated in the regulation of genes involved in sexual development (Ait Benkhali et al., 2013; Gautier et al., 2018).

The results reported here began with an attempt to identify mating-type genes in isolates of *Monosporascus*. The described species of *Monosporascus* include *M. cannonballus*, strains of which are important agricultural root pathogens on Cucurbitaceae (Edelstein et al., 1999), and *M. ibericus,* which was described as an endophyte occurring on diverse plants in Spain (Collado et al., 2002). Strains of a third described species, *M. eutypoides*, are also pathogenic on cucurbits and are closely related to *M. cannonballus* (Ben Salem et al., 2013). In the case of all three species, certain strains have been observed to produce ascocarps, and although publications do not specifically address the issue of homothallism versus homothallism, it is possible to infer that many strains are homothallic. Direct evidence for homothallism among strains of *M. cannonballus* and *M. eutypoides* exists in the fact that hyphal-tip derived cultures (Ben Salem et al., 2013) and ascospore-derived progeny (Michael Stanghellini, personal communication) produce perithecia and ascospores. In the publication describing *M. ibericus* it was noted that two strains were observed to produce ascocarps in culture although eight others did not (Collado et al., 2002).

Our interest in species of *Monosporascus* derives in part from the fact that they are common root endophytes of diverse plants in arid ecosystems of the western United States (Dean et al., 2015; Porras-Alfaro et al., 2008, 2014). The identification of isolates from our fungal endophyte surveys have relied on molecular characterization, however, and these isolates have not been observed to produce ascocarps. Although as mentioned above homothallism has been documented for certain isolates in the genus, heterothallism has not been reported. Our failure to obtain ascocarps, along with reports of similar failures for certain strains in studies cited above, suggested the possibility that members of certain *Monosporascus* lineages might be heterothallic. We have obtained genome sequences from 10 isolates that include known pathogenic strains as well as endophytic strains that span the range of known diversity in the genus. We employed alignment-based searches with diverse mating-type genes to examine assembled genomes, including genomes from known homothallic strains, for the mating-type genes typically associated with *MAT1-1* and *MAT1-2* regions. The *Monosporascus* genomes contained genes distantly related to *MAT1-2-1* (*mat a-1*) *and MAT1-1-3 (mat A-3)*, but homologs of *MAT1-1-1* (*mat A-1*) and *MAT1-1-2* (*mat A-2*) were not found in either the agricultural or endophytic isolates.

To determine if our results were unique to the genus *Monosporascus*, we performed similar searches with diverse members of the Xylariales. In no instance were we able to identify genes typically associated with the *MAT1-1* region. Moreover, although homologs of the *MAT1-2-1* and *MAT1-1-3* genes were identified, these genes were not clear orthologs of MATA_HMG-encoding mating-type genes based on sequence similarity. Nevertheless, the positions of several of these genes relative to genes that commonly flank mating-type regions suggest they might well be

Table 1
Xylariales genomes examined for mating-type genes.[1]

Species (strain)	NCBI accession or JGI[2] ID
Daldinia eschscholtzii (UM1400)	CCED00000000.1
Xylaria striata (RK1-1)	LOBO00000000.1
Xylaria sp. (MSU SB201401)	NPFG00000000.1
Xylaria hypoxylon (OSC100004) v1.0	JGI Project ID: 1050987
Xylaria sp. (JS573)	JWIU00000000.1
Pestalotiopsis fici (W106-1)	ARNU00000000.1
Pestalotiopsis sp. (JCM 9685)	BCGF00000000.1
Microdochium bolleyi (J235TASD1)	LSSP00000000.1
Rosellinia necatrix (W97)	BBSO00000000.2
Anthostoma avocetta (NRRL 3190) v1.0	JGI Project ID: 1006061
Apiospora montagnei (NRRL 25634) v1.0	JGI Project ID: 1006423
Biscogniauxia nummularia (BnCUCC2015) v1.0	JGI Project ID: 1106943
Daldinia eschscholzii (EC12)	MDGZ00000000.1
Entoleuca mammata (CFL468) v1.0	JGI Project ID: 1117716
Eutypa lata (UCREL1)	AORF00000000.1
Hypoxylon sp. (EC38)	MDCK00000000.1
Hypoxylon sp. (CO27-5)	MDCL00000000.1
Hypoxylon sp. (CI-4A)	MDGY00000000.1
Hypoxylon sp. (E7406B)	JYCQ00000000.1
Hypoxylon pulicicidum (MF5954)	PDUJ00000000.1
Annulohypoxylon stygium (MG137)	QLPL00000000.1
Kretzschmaria deusta (DSM 104547)	MLHU00000000.3
Microdochium trichocladiopsis (MPI-CAGE-CH-0230) v1.0	JGI Project ID: 1103673
Pseudomassariella vexata (CBS 129021)	MCFJ00000000.1
Truncatella angustata (HP017) v1.0	JGI Project ID: 1103645
Monosporascus cannonballus (CBS609.92)	QJNS00000000
Monosporascus cannonballus (CBS586.93)	QJNT00000000
Monosporascus sp. (GIB2)	QJNV00000000
Monosporascus sp. (MC13-8B)	QJNW00000000
Monosporascus sp. (MG133)	QJNX00000000
Monosporascus sp. (MG162)	QJNY00000000
Monosporascus sp. (CRB-8-3)	QJNZ00000000
Monosporascus sp. (CRB-9-2)	QJOA00000000
Monosporascus sp. (5C6A)	QJOB00000000
Monosporascus ibericus (CBS110550)	QJNU00000000

[1] Life cycle information: Strains listed as *Monosporascus cannonballus* and *M. ibericus* are known to be homothallic. As discussed in the text, ascocarps have not been observed for the strains listed as *Monosporascus* sp. Other strains listed in this table are assumed to be homothallic or are anamorphic stages of unknown life cycle.
[2] JGI = US Department of Energy Joint Genome Institute.

derived from mating-type genes. Our results suggest that species of *Monosporascus* and other Xylariales lack typical *MAT1-1* and *MAT1-2* regions, and although our sample size is limited by available Xylariales genomes, the results imply that the genetic mechanisms that control sexual reproduction in the Xylariales are different from those of other members of the Pezizomycotina.

2. Materials and methods

2.1. Genomes examined

Genome assemblies were available from public databases for the Xylariales members presented in Table 1 with the exception of 10 isolates from the genus *Monosporascus*. The results presented here for *Monosporascus* isolates were obtained using genome assemblies created from Illumina sequences. Six of the isolates examined came from root endophyte surveys conducted in New Mexico (eg. Dean et al., 2015). *Monosporascus cannonballus* strains CBS 609.92 and 586.93, as well as *Monosporascus ibericus* strain CBS 110550, were obtained from the CBS-KNAW fungal collection (http://www.westerdijkinstitute.nl/Collections/). *Monosporascus cannonballus* MC13-8B was a gift from Michael Stanghellini, University of California, Riverside. Genomic DNA was purified using a CTAB extraction protocol (Hutchinson et al. 2015). Genomic libraries for each isolate were prepared using the KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, Massachusetts). Paired-end

sequencing was conducted using an Illumina NextSeq 500 platform configured for mid-output (130 million reads) and 150 base-pair read lengths. The resulting Illumina reads were paired, trimmed and filtered using Trimmomatic (Bolger et al., 2014). The quality control filtered reads from each isolate were then assembled using three separate software packages designed for short read de-novo microbial genome assemblies: Velvet (Zerbino and Birney, 2008), SPAdes (Bankevich et al., 2012) and SOAPdenovo2 (Luo et al., 2012). The parameters for each assembly software package were optimized on a per-sample basis by generating multiple assemblies then comparing common assembly metrics and the overall quality of the assembly. Assembly statistics were generated using Quast (Gurevich et al., 2013), and assemblies that contained a small total contig number while maintaining a large N50 value were selected as representative assemblies to be used in subsequent steps. Quality assessment of these representative assemblies was performed using the BUSCO software package (Simão et al., 2015), which checks for the presence of fungal-specific single-copy orthologs. Assemblies produced in SPAdes showed the highest consistent results in both categories and were annotated using the gene prediction program AUGUSTUS (Stanke and Morgenstern, 2005; Hoff and Stanke, 2013). AUGUSTUS was run using *Neurospora crassa* for the species parameter, limited to few alternative transcripts and only complete gene predictions. These annotated genomes were used in the analysis described below. They have been deposited at GenBank with accession numbers QJNS00000000, QJNT00000000, QJNU00000000, QJNV00000000, QJNW00000000, QJNX00000000, QJNY00000000, QJNZ00000000, QJOA00000000, and QJOB00000000 (Table 1).

2.2. Mating-type gene searches

Given that genes required for mating and ascocarp development in the Xylariales have not been reported, we relied on genes from other orders in the Sordariomycetes to serve as references for identifying potential MAT homologs in members of the order. The NCBI database contains annotations for genes of both mating types for several members of the Sordariales, Hypocreales, Ophiostomatales and Magnaporthales. The protein sequences in Table 2 were selected based on sequence quality and completeness, and they included representatives from homothallic and heterothallic strains. These sequences were used as queries in alignment-based methods (BLAST) to identify potential homologs in the Xylariales. When nucleotide-based searches failed to find Xylariales mating-type gene homologs, we employed protein amino-acid sequences to search protein databases and translated nucleotide sequences. Those results are presented here.

Local BLAST protein and nucleotide databases were created using the annotation files for *Eutypa lata*, *Daldinia eschscholzii*, *Microdochium bolleyi*, *Pestalotiopsis fici, Rosellinia necatrix*, *Monosporascus cannonballus, Xylaria striata*, *Hypoxylon* sp., *Pseudomassariella vexata,* and several *Monosporascus* sp. isolates from our southwestern United States collections (Table 1). Diverse MAT protein sequences (Table 2) were used in BLASTp, tBLASTn and tBLASTx queries against the Xylariales databases. Potential homologs of MAT1-2-1 (Mat a-1) and MAT1-1-3 (Mat A-3) were identified as amino-acid and translated-nucleotide sequences based on best BLAST scores in the annotated Xylariales databases. Regions of putative homology were identified using the coordinates obtained from BLAST alignments. In cases where a BLAST alignment was shorter than the original query, sequence ranges were expanded to ensure that full regions of potential homology were captured. In cases where functional annotations were available they were also examined for hits to MAT-associated domains and superfamilies. Because proteins encoded by *MAT1-2-1* and *MAT1-1-3* genes belong to the MATA_HMG family and filamentous fungi typically possess distant homologs of these proteins not associated with the mating-type region (see, for example, Hutchinson et al. 2015), potential Xylariales MAT1-2-1 and MAT1-1-3 homologs were used in reciprocal searches against NCBI databases to assess the likelihood that these proteins were true orthologs of mating-type proteins.

Table 2
Sordariomycete mating-type sequences used as queries in alignment-based searches.

Species (strain)	Reproductive strategy & mating-type proteins	Protein accession #	Genomic region accession #
Hypocreales			
Fusarium graminearum (CBS 139514)	Homothallic		
	MAT1-1-1	AMP43945.1	KT855220.1
	MAT1-1-2	AMP43944.1	
	MAT1-1-3	AMP43943.1	
	MAT1-2-1	AMP43946.1	
Ophiostomatales			
Ophiostoma himal-ulmi (HP25)	Heterothallic		
	MAT1-1-1	AHL24887.1	KF961046.1
	MAT1-1-2	AHL24886.1	
	MAT1-1-3	AHL24885.1	
Ophiostoma himal-ulmi (HP62)	Heterothallic		
	MAT1-2-1	AAX83073.1	AY887030.1
Sordariales			
Neurospora crassa (74-ORS-A)	Heterothallic		
	MAT1-1-1	AAC37478.1	M33876.1
	MAT1-1-2	AAC37477.1	
	MAT1-1-3	AAC37476.1	
Neurospora crassa	Heterothallic		
	MAT1-2-1	AAA33598.2	M54787.1
Myceliophthora heterothallica (ThNM 146)	Heterothallic		
	MAT1-1-1	ALD16238.1	KR119057.1
	MAT1-1-2	ALD16239.1	
	MAT1-1-3	ALD16240.1	
Myceliophthora heterothallica (ThNM 053)	Heterothallic		
	MAT1-2-1	ALD16244.1	KR119058.1
Magnaporthales			
Magnaporthe grisea (Y93-164g-1)	Heterothallic		
	MAT1-1-1	BAC65091.1	AB080672.2
	MAT1-1-2	BAC65092.1	
	MAT1-1-3a	BAC65093.2	
	MAT1-1-3b	BAE66612.1	
Magnaporthe grisea (70-14)	Heterothallic		
	MAT1-2-1	BAC65090.1	AB080671.2
	MAT1-2-2a	BAE66610.1	
	MAT1-2-2b	BAE66611.1	

2.3. Molecular alignments and phylogenetic analyses

Amino-acid sequences for HMG homologs employed in phylogenetic analyses were aligned with Clustal Omega (Sievers et al., 2011). Trees were constructed using RaxML (Stamatakis, 2006) with the PROTCATDAYHOFF substitution model and 1000 bootstrap replicates.

In addition to BLAST searches against whole Xylariales genomes and protein databases, we carefully examined regions between and adjacent to homologs of genes that typically flank mating regions in other members of the Ascomycota (Butler et al., 2004, see below). These flanking genes were identified in all members of the Xylariales by employing searches with the Sla2 (XP_964240.1) and Apn2 (ESA43843.1) predicted-protein sequences from *Neurospora crassa*. Regions between and flanking these genes in the Xylariales genomes were employed in BLAST searches against GenBank databases to search for homologs of genes associated with mating type.

3. Results and discussion

We examined the genomes of 35 strains across 15 different Xylariales genera. Our searches employed known MAT1-1-1, 2, 3 (Mat A-1, 2, 3) and MAT1-2-1 (Mat a-1) amino-acid sequences from members of the Sordariales, Hypocreales, Magnaporthales and Ophiostomatales (Table 2). In no instance were we able to identify likely Xylariales gene

homologs (orthologs) for either *MAT1-1-1* or *MAT1-1-2*. While we detected homologs of the *MAT1-2-1* (*mat a-1*) and *MAT1-1-3* (*mat A-3*) mating-type genes, the interpretation of these results is complicated by the fact that proteins encoded by these genes are in the MATA_HMG family of transcriptional activators, and members of the Pezizomycotina typically have multiple genes for this family (discussed below).

Considering the possibility that genes of the mating-type region have undergone rapid change that would confound BLAST searches, we paid special attention to regions that possess homologs of two genes, *sla2* and *apn2*, that flank the mating-gene region in species of *Neurospora* and diverse Ascomycota (Butler et al., 2004; Gioti et al., 2012; Hutchinson et al., 2016). We used the *N. crassa* Sla2 and Apn2 predicted proteins to identify homologs in members of the Sordariomyetes, Leotiomycetes, Eurotiomycetes and Dothidiomycetes (Supplemental Table S1), as well as in the Xylariales (Supplemental Table S2). We then examined these regions for *MAT1-1-1* (*mat A-1*) and *MAT1-2-1* (*mat a-1*) genes. With the exception of the Xylariales, members of all groups examined possessed either one or both of these mating-type genes either between or adjacent to *apn2* and *sla2* genes. Again, none of the Xylariales genomes possessed genes with apparent homology to *MAT1-1-1* or *MAT1-1-2*.

Our results with respect to MATA_HMG genes are more difficult to interpret. While *sla2* and *apn2* homologous genes were found to be linked in most members of the Xylariales, in most cases, homologs of *MAT1-2-1* and *MAT1-1-3* were absent from *sla2/apn2* regions, further suggesting a loss or displacement of these genes. In some instances, however, MATA_HMG genes were found to be between or adjacent to *sla2/apn2* genes (Supplemental Tables S2 and S3). In cases of close linkage, the orientation and order of the three genes (*sla2, apn2,* MATA_HMG) varied among the genomes examined (Fig. 1).

In efforts to evaluate whether MATA_HMG-encoding genes were true orthologs of mating-type genes, we performed reciprocal BLAST searches against genome and protein sequences from species of *Neurospora* using the predicted Xylariales proteins identified in searches with known MAT1-2-1 and MAT1-1-3 sequences. We targeted sequences from species and strains of *Neurospora* because the mating-type and non-mating-type HMG genes in this genus have been characterized, and their chromosomal locations are known. The genome assemblies for *Monosporascus* strains and several strains from other Xylariales species each possessed multiple HMG genes. Almost invariably the Xylariales HMG proteins did not have top hits to either Mat a-1 (MAT1-2-1) or Mat A-3 (MAT1-1-3) proteins or their respective genes in *Neurospora* genomes. Instead, the top hits for the Xylariales sequences typically corresponded to one of the non-mating-type MATA_HMG genes discussed above, either the gene designated NCU03481 (GenBank accession XM_951277) or the gene designated *fmf-1* (GenBank accession 958740), both of which have been implicated in sexual development (Iyer et al., 2009).

Acknowledging that selection, differential rates of divergence or long divergence times could obscure true relationships and likewise make BLAST scores unreliable as indicators of relationships, we employed diverse Xylariales MATA_HMG-predicted proteins and those from other Pezizmycotina in phylogenetic analyses using alignments based on full-length proteins as well as alignments based on only the HMG core (residues 122–233 in the predicted *N. crassa* Mat a-1 protein, GenBank accession AAA33598; Supplemental Fig. S1). In these analyses, known MAT-1-2-1 proteins from diverse Sordariomycetes and a member of the Leotiomycetes (*Sclerotinia sclerotiorum*) consistently formed a distinct group. In contrast, similar to results obtained with BLAST results, several Xylariales proteins consistently grouped with either the NCU03481 gene product or FMF-1. Many Xylariales sequences, however, did not show clear affinities with any of these previous groups (Supplemental Tables S2 and S3, Supplemental Fig. S1, TreeBase submission ID 23036).

Although we failed to identify obvious *MAT1-2-1* orthologs based on sequence similarity, it is safe to assume that some, if not most, of the MATA_HMG-family genes observed in these genomes serve directly or indirectly in sexual development, whether or not mating is involved. It is possible that members of Xylariales possess a mode of sexual reproduction similar to the "unisexual" reproduction reported for *Huntiella moniliformis,* which appears to rely on *MAT1-2-1* in the absence of the *MAT1-1* region (Wilson et al., 2015). Alternatively, given that the *P. anserina* orthologs of *fmf-1* (*PaHMG5*) and NCU03481 (*PaHMG8*) gene products have been proposed as upstream regulators of mating-type genes (Ait Benkhali et al., 2013), it seems plausible that regulatory pathways have been modified to bring mating and other aspects of sexual development under the control of these upstream regulators.

While none of the predicted Xylariales MATA_HMG proteins grouped with known MAT1-2-1 proteins in tree building analyses (Supplemental Fig. S1), the proximity of several of the MATA_HMG Xylariales genes to genes that typically flank mating genes in

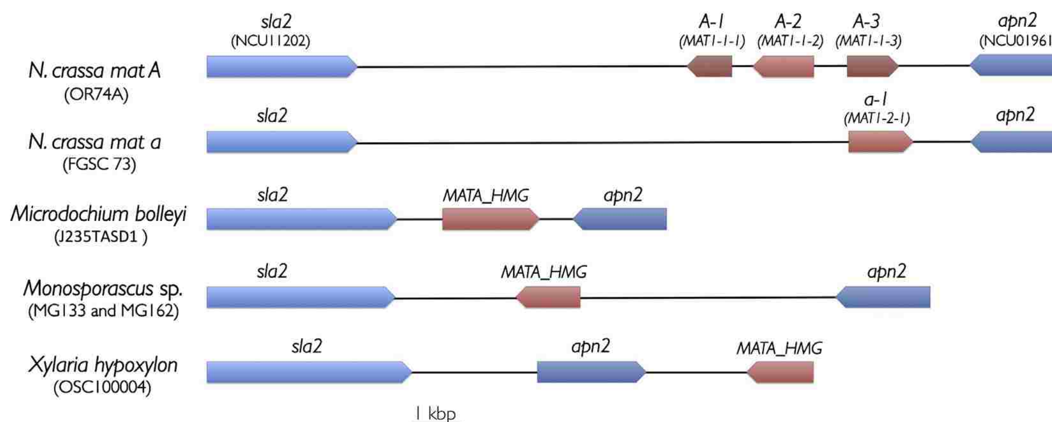

Fig. 1. Comparison of *sla2/apn2* gene regions from three Xylariales genomes with the mating-type regions from *Neurospora crassa*. Homologs of *sla2* and *apn2* are commonly found flanking mating-type genes in other Ascomycota and gene orientations are highly conserved (discussed in text). Note that the orientations of *sla2*, *apn2* and MATA_HMG genes in *Microdochium bolleyi* J235TASD1 are consistent with the orientations of *sla2*, *apn2* and MATA_HMG mating-type genes (*mat a-1* and *mat A-3*) in *N. crassa*. On the other hand, the orientation of the MATA_HMG gene in two strains of *Monosporascus* sp., MG133 and MG162, is reversed relative to homologs in *M. bolleyi* and *N. crassa*; and the order of the three genes is different in *Xylaria hypoxylon* OSC100004. The arrangement of these genes in other members of the Xylariales varies substantially. In contrast with the results shown here, in most cases MATA_HMG genes in members of the Xylariales were not observed to be linked with *sla2* and *apn2* homologs (Supplemental Table S2). Accession numbers for *N. crassa* mating-type proteins and gene regions are given in Table 2. Accession numbers for the Xylariales genomes are given in Table 1, and specific gene locations and protein IDs are presented in Supplemental Table S2.

Sordariomycetes (Fig. 1) suggests they are highly divergent *MAT1-2-1* genes. Evidence has been presented for the direct physical interaction of MAT1-2-1 with either MAT1-1-1 (Jacobsen et al., 2002) or other MAT transcription factors (Zheng et al., 2013). Our observations suggest the possibility that Xylariales genomes have lost the *MAT1-1* mating-type region. It is therefore possible that the loss of *MAT1-1* genes in the Xylariales could have resulted in reduced or altered constraint on *MAT1-2-1* genes, in turn resulting in rapid divergence.

Although the presence of MATA_HMG genes and their locations leave open the possibility that *MAT1-2* mating-type regions remain, if this is the case, genes and gene arrangements have been dramatically altered relative to other members of the Pezizomycotina. The order Xylariales represents a monophyletic group (Tang et al., 2009; Zhang et al., 2006) with a large number of species with diverse ecological roles across numerous genera and several families, as well as well-known iconic species such as *Xylaria polymorpha* ("dead man's fingers") and *Daldinia concentrica* ("coal fungus"). Despite the many studies that have focused on members of the order, information regarding life cycles is nearly non-existent in the literature. While homothallism appears to be the rule among those species for which single-ascospore cultures have been derived (Yu-Ming Ju, personal communication), circumstantial evidence for heterothallism has been presented for two species of *Hypoxylon.* Using RAPD markers, Vannini et al. (1999) presented evidence for genetic assortment among ascospore-derived progeny from single stromata of *H. mediterraneum*. Griffin et al. (1992) inferred heterothallism for *H. mammatum* after observing different genetically-based types of hyphal interactions, resembling compatible and incompatible somatic interactions, among ascospore progeny derived from the same ascus as well as from different stroma. Although these studies have not shed light on the genetics of mating or confirmed heterothallic mating in these species, they do suggest that some form of genetic assortment accompanies sexual reproduction.

Our results are consistent with a previous failed attempt to find mating-type genes in *Eutypa lata* with PCR using degenerate primers. In a report posted by the American Vineyard Foundation, Long and Bradshaw (2002) detailed efforts employing more than 100 different PCR experiments with multiple primer sets and 34 different *E. lata* isolates, targeting but failing to identify *MAT1-1-1* (*mat A-1*) and *MAT1-2-1* (*mat a-1*) genes.

We are therefore left with the possibility that diverse members of the Xylariales produce ascocarps and ascospores without the genes required for mating in other Pezizomycotina. Alternatively, it is possible that relative to other members of the Pezizomycotina mating-type genes in the Xylariales have diverged substantially in terms of both sequence and genome arrangement, with the result that inferences regarding function are not possible based on sequence analysis alone. This situation is made all the more puzzling by the fact that although the Xylariales are basal to the groups (particularly the Sordariales) within the Sordariomycetes for which mating-type genes have been best studied (Zhang et al., 2006), even more distantly related fungi, for example members of the Dothidiomycetes and Leotiomycetes, possess mating-type gene systems similar to those of the Sordariomycetes (Dyer et al., 2016; Martin et al., 2010).

We note that distantly-related members of the Ascomycota, notably the Saccharomycotina for which mating has been studied extensively, have mating systems that differ substantially from those of the Pezizomycotina discussed here. In the context of our results, it is interesting that a homothallic member of the Saccharomycetales, *Lodderomyces elongisporus* (a close relative of *Candida* species), appears to lack all four mating-type genes typically associated with members of this group (Butler et al., 2009). This result reinforces the many observations that despite the general themes that exist for mating systems across broad fungal groups evolution has resulted in a myriad of variations (Butler, 2010; Dyer et al., 2016; Gioti et al., 2012).

Acknowledgements

This research was partially supported by a National Science Foundation award to the University of New Mexico (UNM) for the Sevilleta Long-term Ecological Research program. We thank George Rosenberg and Lijing Bu of the UNM Department of Biology's Molecular Biology Facility for help with sequencing and analysis, and we acknowledge additional support for sequencing from this facility and from the UNM Center for Evolutionary & Theoretical Immunology (CETI) under National Institutes of Health grant P30GM110907. We also acknowledge the UNM CETI implementation of the Galaxy suite of computational tools. We thank the UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing high performance computing resources used in this work.

We thank Francis Martin for permission to use the genome sequence data for *Truncatella angustata* (HP017) and *Microdochium trichocladiopsis* (MPI-CAGE-CH-0230), which were produced by the US Department of Energy Joint Genome Institute (JGI) in collaboration within the framework of the CSP #1974 project "1KFG: Deep-sequencing of ecologically-relevant Dikarya." We also thank JGI and the following investigators for permission to use sequence data: Pedro Crous for *Apiospora montagnei* (NRRL 25634), Daniel Eastwood for *Biscogniauxia nummularia* (BnCUCC2015), Jon Magnuson for *Pseudomassariella vexata* (CBS 129021), Richard Hamelin and the JGI Pathobiome Project for *Entoleuca mammata* (CFL468), Joey Spatafora and Derek Johnson for *Xylaria hypoxylon* (OSC100004) obtained as part the 1000 Fungal Genome Project, and Kerry O'Donnell for *Anthostoma avocetta* (NRRL 3190).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fgb.2018.12.004.

References

Ait Benkhali, J., Coppin, E., Brun, S., Peraza-Reyes, L., Martin, T., Dixelius, C., Lazar, N., van Tilbeurgh, H., Debuchy, R., 2013. A network of HMG-box transcription factors regulates sexual cycle in the fungus *Podospora anserina*. PLoS Genet. 9, e1003642.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Ben Salem, I., Correia, K.C., Boughalleb, N., Michereff, S.J., Leon, M., Abad-Campos, P., Garcia-Jimenez, J., Armengol, J., 2013. *Monosporascus eutypoides*, a cause of root rot and vine decline in Tunisia, and evidence that *M. cannonballus* and *M. eutypoides* are distinct species. Plant Dis. 97, 737–743.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Butler, G., 2010. Fungal sex and pathogenesis. Clin. Microbiol. Rev. 23, 140–159.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., Wolfe, K.H., 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. USA 101, 1632–1637.

Butler, G., Rasmussen, M.D., Lin, M.F., Santos, M.A., Sakthikumar, S., Munro, C.A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J.L., Agrafioti, I., Arnaud, M.B., Bates, S., Brown, A.J.P., Brunke, S., Costanzo, M.C., Fitzpatrick, D.A., de Groot, P.W.J., Harris, D., Hoyer, L.L., Hube, B., Klis, F.M., Kodira, C., Lennard, N., Logue, M.E., Martin, R., Neiman, A.M., Nikolaou, E., Quail, M.A., Quinn, J., Santos, M.C., Schmitzberger, F.F., Sherlock, G., Shah, P., Silverstein, K.A.T., Skrzypek, M.S., Soll, D., Staggs, R., Stansfield, I., Stumpf, M.P.H., Sudbery, P.E., Srikantha, T., Zeng, Q., Berman, J., Berriman, M., Heitman, J., Gow, N.A.R., Lorenz, M.C., Birren, B.W., Kellis, M., Cuomo, C.A., 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459, 657–662.

Collado, J., Gonzalez, A., Platas, G., Stchigel, A.M., Guarro, J., Pelaez, F., 2002. *Monosporascus ibericus* sp. nov., an endophytic ascomycete from plants on saline soils, with observations on the positon of the genus based on sequence analysis of the 18S rDNA. Mycol. Res. 106, 118–127.

Debuchy, R., Berteaux-Lecellier, V., Silar, P., 2010. Mating systems and sexual morphogenesis in Ascomycetes. In: Borkovich, K.A., Ebbole, D.J. (Eds.), Cellular and Molecular Biology of Filamentous Fungi. ASM Press, Washington, DC, pp. 501–535.

Dean, S.L., Warnock, D.D., Litvak, M.E., Porras-Alfaro, A., Sinsabaugh, R.L., 2015. Root-associated fungal community response to drought-associated changes in vegetation community. Mycologia 107, 1089–1104.

Dyer, P.S., Inderbitzin, P., Debuchy, R., 2016. Mating-type structure, function, regulation

and evolution in the Pezizomycotina. In: Wendland, J. (Ed.), The Mycota I: Growth, Differentiation and Sexuality. Springer, Berlin Heidelberg New York, pp. 351–385.
Edelstein, M., Cohen, R., Burger, Y., Shriber, S., Pivonia, S., Shtienberg, D., 1999. Integrated management of sudden wilt of melons, caused by *Monosporascus cannonballus*, using grafting and reduced rate of methyl bromide. Plant Dis. 83, 1142–1145.
Gautier, V., Tong, L.C.H., Nguyen, T.S., Debuchy, R., Silar, P., 2018. *PaPro1* and *IDC4*, two genes controlling stationary phase, sexual development and cell degeneration in *Podospora anserina*. J. Fungi (Basel) 4, 85.
Gioti, A., Mushegian, A.A., Strandberg, R., Stajich, J.E., Johannesson, H., 2012. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. Mol. Biol. Evol. 29, 3215–3226.
Griffin, D.H., Qmm, K.E., Gilbert, G.S., Wang, C.J.K., Rosemarin, S., 1992. The role of ascospores and conidia as propagules in the disease cycle of *Hypoxylon mammatum*. Phytopathology 82, 114–119.
Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.
Hoff, K.J., Stanke, M., 2013. WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 41, W123–W128. https://doi.org/10.1093/nar/gkt418.
Hutchinson, M., Powell, A.J., Tsang, A., O'Toole, N., Berka, R.M., Barry, K., Grigoriev, I.V., Natvig, D.O., 2016. Genetics of mating in members of the Chaetomiaceae as revealed by experimental and genomic characterization of reproduction in *Myceliopthora heterothallica*. Fungal Genet. Biol. 86, 9–19.
Iyer, S.V., Ramakrishnan, M., Kasbekar, D.P., 2009. *Neurospora crassa fmf-1* encodes the homologue of the *Schizosaccharomyces pombe* Ste11p regulator of sexual development. J. Genet. 88, 33–39.
Jackson, D., Lawson, T., Villafane, R., Gary, L., 2013. Modeling the structure of yeast MATa1: an HMG-box motif with a C-terminal helical extension. Open J. Biophys. 3, 1–12.
Jacobsen, S., Wittig, M., Pöggeler, S., 2002. Interaction between mating-type proteins from the homothallic fungus *Sordaria macrospora*. Curr. Genet. 41, 150–158.
Johnson, T.E., 1979. A *Neurospora* mutation that arrests perithecial development as either male or female parent. Genetics 92, 1107–1120.
Long, P.G., Bradshaw, R.E., 2002. Understanding the sexual life-cycle of Eutypa lata. American Vineyard Association Research Report. < http://www.avf.org/wp-content/uploads/2017/10/962e3115f62ca51dd85af360f35efb95dfbdcb.pdf > .
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1, 18.
Martin, T., Lu, S.-W., van Tilbeurgh, H., Ripoll, D.R., Dixelius, C., Turgeon, B.G., Debuchy, R., 2010. Tracing the origin of the fungal α1 domain places its ancestor in the HMG-Box superfamily: implication for fungal mating-type evolution. PLoS ONE 5 (12), e15199. https://doi.org/10.1371/journal.pone.0015199.
Porras-Alfaro, A., Herrera, J., Sinsabaugh, R.L., Odenbach, K.J., Lowrey, T., Natvig, D.O., 2008. Novel root fungal consortium associated with a dominant desert grass. Appl. Environ. Microbiol. 74, 2805–2813. https://doi.org/10.1128/AEM.02769-07.
Porras-Alfaro, A., Raghavan, S., Garcia, M., Sinsabaugh, R.L., Natvig, D.O., Lowrey, T.K., 2014. Endophytic fungal symbionts associated with gypsophilous plants. Botany 92, 295–301. https://doi.org/10.1139/cjb-2013-0178.
Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with singly-copy orthologs. Bioinformatics 31, 3210–3212.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467.
Tang, A.M.C., Jeewon, R., Hyde, K.D., 2009. A re-evaluation of the evolutionary relationships within the Xylariaceae based on ribosomal and protein-coding gene sequences. Fungal Diversity 34, 127–155.
Vannini, A., Mazzaglia, A., Anselmi, N., 1999. Use of random amplified polymorphic DNA (RAPD) for detection of genetic variation and proof of the heterothallic mating system in *Hypoxylon mediterraneum*. Eur. J. For. Path. 29, 209–218.
Wang, Z., Lopez-Giraldez, F., Lehr, N., Farre, M., Common, R., Trail, F., Townsend, J.P., 2014. Global gene expression and focused knockout analysis reveals genes associated with fungal fruiting body development in *Neurospora crassa*. Eukaryot. Cell 13, 154–169.
Wilson, A.M., Godlonton, T., van der Nest, M.A., Wilken, P.M., Wingfield, M.J., Wingfield, B.D., 2015. Unisexual reproduction in *Huntiella moniliformis*. Fungal Genet. Biol. 80, 1–9.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhang, N., Castlebury, L.A., Miller, A.N., Huhndorf, S.M., Schoch, C.L., Seifert, K.A., Rossman, A.Y., Rogers, J.D., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Sung, G.H., 2006. An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. Mycologia 98, 1076–1087.
Zheng, Q., Hou, R., Juanyu, Z., Ma, J., Wu, Z., Wang, G., Wang, C., Xu, J.R., 2013. The *MAT* locus genes play different roles in sexual reproduction and pathogenesis in *Fusarium graminearum*. PLoS One 8, e66980.

Summary

Comparative genomic studies of fungi provide numerous advantages when studying environmental adaptation at the population level. This dissertation presented three separate studies of genomic evolution in fungi employing a variety of molecular and bioinformatics techniques. Below I summarize the main findings of my research:

Studies with *Neurospora discreta*

1. *Neurospora discreta* isolates from New Mexico and Alaska exhibit substantial differences in maximal growth temperature. Progeny that resulted from a cross between a New Mexico isolate and Alaska isolate had one of three possible phenotypes: New Mexico-like, Alaska-like or intermediate. These progeny phenotypes indicated that multiple regions of the genome are involved in conferring the thermotolerant phenotype.

2. Whole genome comparisons using either individual progeny or bulked samples indicated regions of two chromosomes, chromosome III and chromosome I, were associated with thermotolerance. These results were consistent across two separate experiments, which utilized both long and short read sequencing technologies. The region on chromosome I appeared to play a secondary role to the region on chromosome III, and this was confirmed upon reexamination of several progeny for which phenotypes and genotypes were mixed. Numerous differences between the parental isolates at both the nucleotide and amino-acid level existed in these regions, but further work will be required to implicate specific genes with thermotolerance.

Studies with the genus *Monosporascus*

1. *Monosporascus* isolates from the Sevilleta National Wildlife Refuge are phylogenetically diverse. None of the isolates recovered from the Sevilleta caused obvious disease in their hosts despite several isolates being closely related to the pathogenic *M. cannonballus*. *Monosporascus* genomes are large and are diverse in terms of both genome content and organization. The large size observed for genomes from *Monosporascus* isolates is not completely a result of non-coding regions, and several regions in the genomes are non-syntenic with close relatives from the Xylariales, despite the fact that non-syntenic regions contain putative gene annotations. It remains uncertain whether *M. cannonballus* strains isolated as pathogens and closely-related strains isolated as endophytes differ in pathogenicity or, instead, certain agriculturally important cultivars in the Cucurbitaceae are particularly susceptible to *M. cannonballus*.

2. MAT1-1-1 and MAT1-1-2 homologs were absent from all Xylariales examined including *M. cannonballus*, which has been observed to be homothallic. Distant homologs of HMG-domain mating-type genes were identified, but it is not possible to say with certainty that any of these function as mating-type genes. I was able to identify the genes that commonly flank the mating-type genes in other orders, which suggested that mating-type regions in members of the Xylariales are either absent or highly divergent relative to other members of the Pezizomycotina.