AN ABSTRACT OF THE DISSERTATION OF

C. Alisha Quandt for the degree of Doctor of Philosophy in Botany and Plant Pathology
presented on July 17, 2014
Title: Systematic, Genomic, and Transcriptomic Analysis of Mycoparasitic *Tolypocladium* Species and their *Elaphomyces* Hosts.

Abstract approved:_____
Joseph W. Spatafora

Fungi in the genus *Tolypocladium* are diverse in their host associations, but the predominant ecologies include parasites of the ectomycorrhizal genus *Elaphomyces* and pathogens of insects. The aim of this dissertation research is to examine the evolution of these fungi and their host associations. To accomplish this several lines of data collection, analyses and experimentation, including nomenclatural changes, genome sequencing, differential RNA expression, and metagenomic sequencing of a host sporocarp, were pursued in an integrated manner. The first chapter is an introduction to the study systems and a background of fungal genomics. Nomenclatural issues are addressed in family Ophiocordycipitaceae and *Tolypocladium* that were brought about by changes in the International Code of Nomenclature for algae, fungi, and plants, abolishing a system where a single fungal species could have different generic names at different parts of its life cycle. Proposals for names to be protected and suppressed are made for the family in addition to new combinations in *Tolypocladium*. The genome sequence of *T. ophioglossoides* reveals a great number of secondary metabolite genes and clusters, including three, large peptaibiotic genes. The evolution of these genes, which have only been identified in *Tolypocladium* and *Trichoderma* species, is different within these two genera. Phylogenomic analyses of Peptaibiotics reveal a pattern that is

consistent with speciation in the genus *Trichoderma*, while peptaibiotic diversity within *Tolypocladium* is inferred to be the product of lineage sorting and is inconsistent with the organismal phylogeny of the genus. To determine which genes are involved in the mycoparasitic infection in *Tolypocladium*, in Chapter 4 an RNA-Seq study identifies differentially expressed genes when *T. ophioglossoides* is grown on media containing host and non-host tissues. A chitinase, several G-protein coupled receptors hypothesized to play a role in host recognition, an adhesin involved in cellular attachment, and several secondary metabolite genes were found to be differentially expressed on the *Elaphomyces*-media created for the study. With the addition of two more *Tolypocladium* genome sequences, the *Elaphomyces*-parastite *T. capitatum* and the cicada pathogen *T. paradoxum*, Chapter 5 explores the evolution of host-jumping within *Tolypocladium* using genome scale datasets. A large number of genes within *Tolypocladium* have undergone lineage sorting, as the diversification of the genus happened relatively quickly and is inferred to be the product of host-mediated speciation events. The *Elaphomyces granulatus* metagenome as sequenced from sporocarp tissue is the subject of Chapter 6. Like other ectomycorrhizal Ascomycota, *E. granulatus* is expanded in genome size, but has a reduction in CAZymes compared to its closest sequenced relatives. The *E. granulatus* sporocarp microbiome is taxonomically diverse but the most common community component is Bradyrhizobiaceae bacteria. Despite the large amount of sequence data, the microbiome shows signatures of functional reduction. This dissertation uses advances in systematics, genomics, and transcriptomics to add to the understanding of how *Tolypocladium* and *Elaphomyces* utilize specific sources of nutrition, and what gene repertoires may tell us about these organisms' life histories.

Systematic, Genomic, and Transcriptomic Analysis of Mycoparasitic *Tolypocladium* Species and their *Elaphomyces* Hosts.

by
C. Alisha Quandt

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented July 17, 2014
Commencement June 2015

Doctor of Philosophy dissertation of C. Alisha Quandt presented on July 17, 2014

APPROVED:

_____

Major Professor, representing Botany and Plant Pathology

_____

Head of the Department of Botany and Plant Pathology

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

C. Alisha Quandt, Author

ACKNOWLEDGEMENTS

This thesis would not exist in its present form or maybe even at all without the guidance, advice, support, and encouragement of many people. Hopefully, I can accurately express my sincere appreciation to everyone who has been instrumental in this journey for me.

I would like to thank the Department of Botany and Plant Pathology (BPP) for providing a welcoming and supportive community in which to grow as a doctoral student. Dianne Simpson, Blaine Baker, Katie Remiyac, and Carleen Nutt graciously assisted in all manner of logistical matters, and I truly appreciate the time that they put into their jobs and helping me, which resulted in my education running smoothly. Linda Ciuffetti, our department chair, is amazingly supportive of graduate students and their experience in the department, and I have seen this and appreciated it personally. Since moving to Corvallis, BPP grads have been an integral part of my support system, and many, past and present, will be close friends for the rest of my life. BPP has been an incredibly positive and supportive environment with people knowledgeable in almost every aspect of biology, and I consider myself lucky to have worked in such a wonderful environment with a great group of people.

I want to thank Mark Desanko and Chris Sullivan of the Center for Genome Research and Biocomputing (CGRB) for all their help with sequencing and computational support throughout my dissertation.

My mother and father have always encouraged me to take my education seriously and to work diligently toward my goals. They may not have always understood why I chose the path that I did, but they have supported me nonetheless. Back during my

undergraduate degree, my sister, Jessica, was the first person to give me a mycology text book that I wanted, and she has always been supportive of my eccentric interests and inspiring to me in her pursuit of her own. My extended family has also been amazingly supportive, even though I saw them so rarely during this time working on my degree.

So many friends in Corvallis and beyond have helped me and encouraged me throughout my Ph.D. There were housemates who made Corvallis a real 'home' for me, BPP-ers who helped me relax after hard days in the lab, softball teammates, rock climbing buddies, La Conga Challenge conspirators, and so many others I dare not try to list them all by name here, lest I forget someone.

I want to thank all of my foreign collaborators for their hospitality when visiting their labs. In China, Professor Li and his group (including Mingjun) at Anhui Agricultural University planned several amazing collecting trips in different regions of the country, and it was there that I collected my first cordyceps s.l. (*Ophiocordyceps nutans*). In Korea, Gi-Ho Sung and his lab were so accommodating during my stay and collecting, and as a result I now have a younger Korean brother, fellow graduate student, Jun-Sang Oh. It was an honor to work for three months in the lab of Francis Martin at INRA in Nancy, France. Everyone was very helpful with the various aspects of my project, and Stéphan Uroz even tried to convince a bank manager to give me back my credit card that was eaten by an ATM machine, to no avail. The graduate students and other foreign researchers at INRA made my time in Nancy, away from home, really enjoyable and I appreciate their willingness to hang out with a 'loud American.'

I want to thank my committee members, Aaron Liston, Carol Mallory-Smith, Chris Marshall, and Pat Muir for their encouragement and willingness to ask tough, interesting questions, and point out ways in which I can grow as a scientist.

The lab of Joey Spatafora has been an amazing environment in which to become a mycologist. Former post-docs in the lab, including Conrad Schoch, Barbara Robbertse, Gi-Ho Sung, Mingjun Chen, and Kathryn Bushley, helped me in different capacities over the years. Mingjun and Gi-Ho were wonderful hosts to me when I visited their countries, and Barbara and Conrad were also gracious hosts to me on the other side of this country during a conference. Former graduate student Cedar Hesse really got me started in bioinformatics and is still a source of advice for dealing with fungal community data. Former graduate student/ former housemate/ climbing buddy, *etc*., Ryan Kepler has remained a close friend, a *Cordyceps* ally, field hand, and collaborator since leaving the lab. His willingness to share knowledge, databases, and literature with me at the beginning of my project gave me a headstart, and his feedback and edits on my NSF GRF proposal are not forgotten or unappreciated. Fellow graduate student Alija Mujic has always been there through long Illumina preps to beers at Bombs, and I am grateful for the time we spent together. Working with graduate students Derek Johnson, Dabao Lu, and Rheannon Arvidson has been a pleasure. In preparing for these dissertation acknowledgements over the past few weeks, it struck me how many visitors the lab has had during my time here. This experience of working with, lauging with, and learning from (often foreign) visiting scientists has been a valuable source of inspiration and information and an asset to my education. I credit these experiences to my advisor, Joey.

I could not give enough thanks and credit to Joey Spatafora for the ways that he has contributed to the development and execution of this dissertation and my education. Like only the best mentors, Joey has encouraged me and pushed me to accomplish things that I didn't know I was capable of doing. He has created so many opportunities for me in my career, and I have tried to take full advantage of these. It has been humbling and eye-opening working with him on a daily basis, and it has been a pleasure and an honor to work with him. Joey and his family have been so welcoming to me since I joined the lab, and as a young adult, thousands of miles aways from parents and other family, I really appreciated this.

Finally, my wonderful and extremely supportive husband, Dustin, deserves an enormous amount of appreciation. Every time I doubted myself along the way, he encouraged me to keep going, to look at the larger accomplishments instead of momentary failures, and to pursue my love of fungi travelling all over the world even when he wasn't able to travel along with me. He has seen me stressed about work more times than is fair and put up with me during the writing and analysis of this dissertation, and for that I am thankful. Mostly, though, I am so grateful for his love.

This dissertation could not have taken place without the financial support of many organizations. My field work collecting cordyceps-like fungi in China was supported by a National Science Foundation (NSF) East Asian and Pacific Summer Institute (EAPSI) Fellowship. An NSF Graduate Research Fellowship supported my education for three years during this dissertation, and also funded travel to the Asian Mycological Congress and field work in South Korea, and the Joint Genomes Institute (JGI) User Meeting in Walnut Creek, CA. A BPP Anita Summers Travel Grant and a Mycological Society of

America (MSA) Trappe Mentor Travel Award funded travel to the MSA annual meeting at Yale University in New Haven, NH. A BPP Anita Summers Travel Grant supported travel to the Fungal Genetics Conference in Asilomar, CA. A Visiting Scientist Award of the Lab of Excellence ARBRE supported my stay in Nancy, France, while working on the *Elaphomyces* metagenomics project. Funding as a part of an NSF Doctoral Dissertation Improvement Grant (DDIG) provided support for Chapters 4 and 5. A BPP Graduate Student Association travel grant and an MSA Alexopoulos Mentor Travel Award supported my travel to Austin, TX for the MSA annual meeting, where I received an MSA Graduate Fellowship to support my research.

CONTRIBUTION OF AUTHORS

Chapter 2: Ryan M. Kepler and Joseph W. Spatafora helped design, analyze, and edit the study. Walter Gams, João P. M. Araújo, Sayaka Ban, Harry C. Evans, David Hughes, Richard Humber, Nigel Hywel-Jones, Zengzhi Li, J. Jennifer Luangsa-ard, Stephen A Rehner, Tatiana Sanjuan, Hiroki Sato, Bhushan Shrestha, Gi-Ho Sung, Yi-Jian Yao, and Rasoul Zare had expertise for particular taxa and provided comments, suggestions, and edits to the manuscript.

Chapter 3: Joseph W. Spatafora helped design, analyze, and edit the study. Kathryn Bushley helped with computational guidance and early access to data collected for *T. inflatum*.

Chapter 4: Joseph W. Spatafora helped design, analyze, and edit the experiment and article. Yanming Di helped with statistical analysis and identification of differentially expressed genes. Alex Buchannan provided computational support. Pankaj Jaiswal and Justin Elser produced GO annotations of *T. ophioglossoides* genes.

Chapter 5: Joseph W. Spatafora helped design, analyze, and edit the study. Will Patterson helped with bioinformatic analyses. Pankaj Jaiswal and Justin Elser produced GO annotations of *T. capitatum* and *T. paradoxum*.

Chapter 6: Joseph W. Spatafora and Francis Martin helped design, analyze, and edit the study. Annegret Kohler helped with analysis and sequencing of *E. granulatus* RNA. Cedar Hesse sequenced and assembled the *C. orientalis* genome. Emmanuelle Morin provided computational support and analysis. Claude Murat helped with study design and analysis. Aurelie Deveau and Thomas Sharpton provided support with study design and analysis.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

Table Page

Chapter 1. Introduction

Introduction

One of the long-standing questions in fungal biology is how fungi, many of which are specialized to a specific host, are able to switch from parasitizing one host to another and therefore presumably switch from utilizing one source of nutrition to another. This answers to this question are intriguing ecologically, medicinally, and economically and also inherently difficult to ascertain. However, advances in fungal cellular biochemistry, genetics, and most recently genomics, transcriptomics, and metabolomics have broadened the understanding of the molecular toolkits from which fungi draw to utilize and consume their preferred substrate. Importantly, fungi live on or within their source of nutrition, and must excrete degradative enzymes into their substrate and subsequently absorb the simplified byproducts. Therefore, the knowledge of which enzymes and other metabolites are possessed and produced by fungi utilizing different nutritional resources is a key element in understanding the mechanisms by which fungi gain access to novel sources of nutrition.

This dissertation uses systematics, genomics, and transcriptomics to add to our growing understanding of how fungi utilize specific sources of nutrition, and what gene repertoires may tell us about an organism's life history. *Tolypocladium*, composed of both truffle parasites and insect pathogens, is used as a study system to address some of the following questions. Specifically, what gene families are specific to and expressed by truffle parasitic fungi? How do these genes compare to those genes hypothesized to play roles in other types of mycoparasitism? How do the gene repertoires of truffle parasites differ from closely related insect pathogens within the genus and outside of the genus? How does a truffle parasite recognize its host? And how are the truffle parasitic

Tolypocladium spp. related phylogenetically to the insect pathogenic species of the genus? The *Elaphomyces* hosts of most *Tolypocladium* spp. also provide the opportunity to ask questions about evolution of the ectomycorrhizal truffle niche from its saprobic relatives.

Systematics of Hypocreales

Hypocreales is a diverse and speciose fungal order characterized by a dynamic evolutionary history of inter-kingdom host jumping (Spatafora *et al.* 2007), with members that parasitize animals, plants, and other fungi (Sung *et al.* 2007). Hypocrealean species vary greatly in their host associations and ecological niches, and therefore represent an exceptional opportunity to study the evolution of host affiliation among other topics not covered here (such as the evolution of pathogenicity in general).

Most species in the earliest diverging lineages of the order derive their nutrition from plant material, either as pathogens or saprobes. These include Bionectriaceae, Nectriaceae, Niessliaceae, and the *Stachybotrys* clade (Castlebury *et al.* 2004, Sung *et al.* 2007). At the divergence of the four most derived families however, there was a major shift away from plant-based nutrition to either insect or fungal based nutrition (Sung *et al.* 2008). The largest described group of mycoparasites, Hypocreaceae, is sister to Cordycipitaceae, one of the three *Cordyceps*-like lineages. Traditionally, most insect pathogenic hypocrealean species were classified in the genus *Cordyceps*, but with the advent of molecular phylogenetics these insect pathogens were reclassified within three families and four genera: *Cordyceps* of Cordycipitaceae, *Elaphocordyceps*

(=*Tolypocladium*) and *Ophiocordyceps* of Ophiocordycipitaceae, and *Metacordyceps* (=*Metarhizium*) of Clavicipitaceae (Sung *et al.* 2007).

While generalizations can be made about the most frequent type of host for each clade (*e.g.*, insect pathogens in Ophiocordycipitaceae, plant pathogens/associates in Nectriaceae), there exist exceptions in every clade, and the order is characterized by frequent instances of host switching. Notably, mycoparasitic species exist in almost every lineage within the order (Kepler *et al.* 2012), and this provides an opportunity to compare the mycoparasites of different clades, to examine whether convergent evolution has led to some mycoparasitic adaptations, and if there exist mycoparasitic traits which were present in the ancestor to the four most derived families. The largest two clades of mycoparasites are Hypocreaceae (as mentioned above) and *Tolypocladium* species, a genus of truffle parasites in Ophiocordycipitaceae. *Trichoderma* species of Hypocreaceae have received much research attention due to their use in industry and potential as biocontrol agents against plant pathogenic fungi (see Schuster & Schmoll 2010). Excluding the type, *Tolypocladium inflatum* W. Gams, little is known about the genetics of *Tolypocladium* species, save for sporadic reports of metabolite isolation. Below an in depth review of what is known about gene content and expression in these groups is discussed.

The genus *Tolypocladium*

There are currently 32 described *Tolypocladium* species (Quandt *et al.* Ch. 2), most of which are parasites of the hypogeous, truffle-like fruiting bodies of fungi in the

ectomycorrhizal genus *Elaphomyces* (see section below). Several species are known to be insect pathogens however (Samson & Soares 1984, Weiser *et al.* 1991, Sung *et al.* 2007), including pathogens of nymphs in the family Cicadidae [Hemiptera] (cicadas) (Sung *et al.* 2007) and pathogens of coleopteran (beetle) larvae buried in wood, dung, or soil (Petch 1937). Several rotifer and nematode pathogens have also been described in the genus (Barron 1980, 1981, 1983), and some species have only been isolated from soil (Bissett 1983). *Tolypocladium* was erected by Gams (1971) to describe the distinctive inflated bases of the phialides of the anamorph of *T. inflatum* (Figure 1.1), which is the type species of the genus. This phialide morphology is not typical of most described species in the genus, but where known, most asexual stages produce phialides arranged in whorls including verticillium-like, chaunopycnis-like, and *Tolypocladium*. In the most extensive taxonomic revision of the order to date, *Elaphomyces* parasites (and related species) were renamed *Elaphocordyceps* by Sung *et al.* (2007), based on multi-gene phylogenetics. However, a revision of Article 59 in the International Code of Nomenclature for algae, fungi, and plants (ICN) abandoned the dual nomenclature which allowed for separate names for the anamorphic and teleomorphic states of fungi. Due to these changes in the ICN which took effect January 1, 2013, the name *Tolypocladium* was applied to the entire genus (including species of *Elaphocordyceps*, *Chaunopycnis*, and *Tolypocladium*) based on the criteria of monophyly and nomenclatural priority (Quandt *et al.*, Ch. 2).

The morphological similarities between some of the Cicadidae pathogens and the *Elaphomyces* parasites were noted by others (Lloyd 1918, Kobyasi 1939) before molecular phylogenetics could confirm this relationship. Nikoh and Fukatsu (2000) were

the first to collect molecular data supporting this relationship, and they proposed that the host habitat hypothesis (observed in other systems [*e.g.* Shaw 1988]) could best explain the evolution of host switching from cicada to *Elaphomyces*.

The premise behind the host habitat hypothesis is that cicadas and *Elaphomyces* truffles are found in the same subterranean habitat, attached to the roots of trees. Their hypothesis follows that this shared habitat allowed the cicada pathogen (proposed to be *T. paradoxum* or the cicada pathogenic ancestor to *T. paradoxum*) to transition or switch hosts to *Elaphomyces* (Nikoh and Fukatsu 2000). Published multi-gene phylogenies resolved the insect pathogenic species of *Tolypocladium*, including *T. paradoxum*, to be nested within the clade of *Elaphomyces* parasites (Sung *et al.* 2007). There has yet to be an extensive sampling of the diversity of the genus.

Currently, 18 species of *Tolypocladium* parasitic on *Elaphomyces* have been described, along with several described subspecific form taxa. Because no other *Elaphomyces* parasites exist outside of the genus *Tolypocladium* and *Elaphomyces* is such an atypical host for fungi in Hypocreales, it is hypothesized that there has been a single transition to *Elaphomyces* parasitism (Nikoh & Fukatsu 2000).

The life cycle of *Tolypocladium* species is not well known. It is presumed that asexually produced conidia or sexually produced ascospores land on soil or litter and germinate. The haploid hyphae then continue to grow saprobicly in the soil until encountering their host. When spores or hyphae encounter the truffle or insect host, they penetrate its outer layer (either the insect cuticle or *Elaphomyces* peridium), and then proliferate, consuming the host tissue and finally produce a stroma(ta). Perithecia then develop on the terminal portion of the stroma (Figure 1.1). It is presumed that each

perithecium is the product of an individual mating event initiated by the fertilization of an ascogonium by a spermatium or antheridium of an opposite mating type. Within perithecia mulitiple asci form and produce ascospores that disarticulate into small or medium sized partspores and are released from the ascus at maturity rendering the lifecycle complete. Many of these species are known to reproduce asexually as well, through the production of conidia (see below). Rhizomorphs (aggregated threads of hyphae) may connect the fruiting body to the *Elaphomyces* or insect host in some cases. Most parts of this life cycle are based on very few studies done in entomopathogenic *Cordyceps*-like organisms (see Clarkson & Charnley 1996), and therefore is speculative in the case of the *Elaphomyces* parasites.

Tolypocladium ophioglossoides

Tolypocladium ophioglossoides (Ehrh. ex J.F. Gmel.) Quandt, Kepler & Spatafora is the oldest described species in the genus and was one of the first species of *Cordyceps s.l.* ever described (Fries 1818). It is one of the most often collected species in the genus with often large (up to 14 cm), yellow-olive to brown spathulate sporocarps (hence the epithet "ophioglossoides" meaning "snake tongue like") that fade to black with age or desiccation. The anamorph or asexual conidia-producing state of this species is verticillium-like. This species is most likely a species complex as it has a wide distribution throughout many parts of the Northern Hemisphere, with collections reported from but not limited to Austria, Belgium, the Czech Republic, Finland, France, Germany, Great Britain, Hungary, Italy, Japan, Korea, the Netherlands, Russia, and Switzerland in

Eurasia, and Canada and the United States of America in North America (Mains 1957, Kobayasi & Shimizu 1960).

There has yet to be an examination of the genes that may be involved in *Elaphomyces* parasitism, and one aim of this research is to examine gene losses and expansions that may be unique to truffle parasitic *Tolypocladium* spp., including *T. ophioglossoides*.

To date, only a few secondary metabolites have been isolated from *T. ophioglossoides*. These include the compound ophiocordin (or balanol), a weak antifungal antibiotic (Kneifel *et al.* 1977, Boros *et al.* 1994), and ophiosetin and closely related equisetin, which both possess antibiotic properties (Putri *et al.* 2010) and the latter of which also displays inhibitory activity of HIV (human immunodeficiency virus) -1 integrase (Hazuda *et al.* 1999).

Tolypocladium ophioglossoides and some, but not all, species in the genus have the ability to easily produce asexual spores or conidia in the laboratory (Table 1.1). There are many reasons that it is advantageous to work with species that readily grow in culture, including (relevant to this study) the ability to produce large quantities of axenic tissue for uncomplicated production of genome and transcriptome libraries. It may be the case that species which are unable to grow in laboratory settings lack the same amount of saprobic potential as those species that do readily grow in culture on standard media, and this should be kept in mind when considering an organism's primary metabolic potential. Many of the insect pathogens readily grow in culture, but most of the parasites of *Elaphomyces* do not, with *T. ophioglossoides* being an obvious exception (Table 1.1).

Elaphomyces biology

Elaphomyces Nees is one of two genera in Elaphomycetaceae (Dictionary of the Fungi 10th Ed. Kirk *et al.* 2008). More than 90 species have been described and have been demonstrated to form ectomycorrhizae (Miller & Miller 1984) with various woody conifers and angiosperms (Castellano *et al.* 2012). The genus has a near global distribution, and new species continue to be described as new habitats are sampled (Castellano *et al.* 2011, Castellano *et al.* 2012). *Elaphomyces* produce subglobose, hypogeous "truffle" fruiting bodies, which possess a hard, minutely warted outer peridium and an internal fertile tissue, the gleba, where asci and ascospores are produced. Asci mature rapidly, and it is more common to find specimens that possess only mature highly ornamented, darkly pigmented, hydrophilic ascospores (Trappe 1979, Castellano *et al.* 2011). At maturity, the gleba tissue is a dry, powdery mass of spores. The common name is Hart's truffle (hart being an old alternative word for deer or stag) as fruiting bodies are often dug up and consumed by deer and other smaller woodland mammals (Maser *et al.* 1978, Trappe 1979).

Despite striking differences in ecology and morphology, Elaphomycetaceae is a member of Eurotiales, which is not known to include any other ectomycorrhizal taxa. This association was first hypothesized by Korf (1973), but it was the use of molecular data that confirmed the placement of *Elaphomyces* within Eurotiales (Landvik *et al.* 1996, LoBuglio *et al.* 1996). One uniting character for the order is the production of cleistothecia where sexual fruiting bodies are known. The phylogenetic placement of Elaphomycetaceae within class Eurotiomycetes means that this family represents an independent origin of the mycorrhizal symbiosis.

Initial characterization of *Tuber melanosporum* Vittad., the only ectomycorrhizal ascomycete with a published draft genome, has revealed several genome scale differences as compared to sequenced basidiomycete ectomycorrhizal taxa (Martin *et al.* 2010). The number of genes present in the *Tuber* genome is highly reduced compared to other ascomycetes, and in contrast to the basidiomycete *Laccaria bicolor* (Maire) P.D. Orton (Martin *et al.* 2008) it has an extremely expanded genome size (125 megabases [MB]) with a high number of transposable elements, especially those annotated as Gypsy long terminal repeats (LTRs). Despite the large genome size, very few protein models are predicted – just under 7,500 (Martin *et al*. 2010). The *T. melanosporum* genome has relatively few carbohydrate active enzymes (CAZymes) and lacks several specific genes that are also absent from *L. bicolor* (Martin *et al.* 2008, 2010).

One of the challenges a mycoparasite must overcome is gaining access to the contents of its fungal host by penetrating the host cell wall. Like all fungi, *Elaphomyces* hyphae possess cell walls. While the composition of the heterogenous cell wall matrix varies wildly between fungal lineages, all contain some amounts of chitin and β-1,3 and β-1,6 glucans (polymers of glucose) (Osherov & Yarden 2010). Chitin is a polymer of N-acetyl-glucosamine produced by a limited number of organisms, including fungi and arthropods (see below). The cell walls of very few fungi have been studied in detail, but one of the model systems is *Aspergillus fumigatus* Fresen. of Eurotiales (Latgé 2007). The cell wall of *Aspergillus* is mainly composed of branched β-1,3-glucan which is linked to chitin and short chain mannans (polymers of mannose) (Latgé 2010). Species of *Tolypocladium* parasitic on *Elaphomyces* would need to excrete enzymes capable of degrading these polysaccharides in order to gain access to cellular contents/nutrients.

Insect biology

Cordyceps sensu lato are pathogens of at least twelve orders of Arthropoda. Insect pathogenic species of *Tolypocladium* are only known to attack insects in the orders Coleoptera, rarely Diptera, and Hemiptera (in the family Cicadidae) (Table 1.1). Importantly, insect pathogenic *Tolypocladium* spp. mainly attack immature hosts (*e.g.* beetle larvae which are buried in substrates or cicada nymphs attached to tree roots in soil) (Sung *et al.* 2007). The amount of sclerotization of the insect cuticle is typically less in immature states (Hopkins & Kramer 1992). This means that less mechanical pressure or enzymatic degradation may be required to penetrate the host. Chitin, found in aggregated microfibers surrounded by proteins, is a major component of the insect endocuticle (Neville 1984). However, excluding initial penetration and final stroma formation, the cuticle typically remains intact and is not consumed by *Tolypocladium* or other *Cordyceps s.l.* Insect pathogenic *Cordyceps* are hypothesized to use a combination of mechanical (turgor) pressure and proteases to penetrate the exoskeleton (St. Leger *et al.* 1991).

New studies have suggested that some entomopathogens are rhizosphere competent (Hu & St. Leger 2002, Bruck 2005), and that they may be translocating nitrogen and other nutrients from insect hosts to plant associates (Behie *et al.* 2012). Older studies have also reported the antagonistic effect of entomopathogenic *Tolypocladium* spp. when grown with other fungi (Lundgren *et al.* 1978).

Mycoparasitsm in *Trichoderma*

Trichoderma species have been extensively studied for a variety of reasons (Samuels 1996) including their cellulase production for industrial purposes (Buchert *et al.* 1998, Galante *et al.* 1998a, 1998b), and their potential use in biocontrol of plant pathogenic fungi (Schuster & Schmoll 2010). Because of this, *Trichoderma reesei* E.G. Simmons was one of the first Sordariomycete genomes published (Martinez *et al.* 2008). The strain of *Tr. reesei* published was the prolific cellulase producer, the "industrial workhorse" strain, which was isolated degrading a tent canvas in the Pacific during World War II (Mandels & Reese 1957). This is a very atypical habitat for Hypocreaceae, as the vast majority of its species are mycoparasites.

Two other mycoparasitic *Trichoderma* spp. have now been sequenced (Kubicek *et al.* 2011), and several others are soon to be published (Grigoriev *et al.* 2014). One of the major conclusions from the comparative genomic analysis of three species (*Tr. reesei*, *Tr. atroviride* Bissett, and *Tr. virens* (J.H. Mill., Giddens, & A.A. Foster) Arx) was that these species (especially the mycoparasitic species) had more chitinases, chitosanases and some proteins of unknown function than other sequenced species at that time (Kubicek *et al.* 2011). Previous pre-genomic work on these taxa had provided evidence that one group of chitinases, Subgroup C, was essential to mycoparasitic interaction (Siedl *et al.* 2005, Seidl 2008). Subsequently, with access to genome scale data, several of these *Trichoderma* subgroup C chitinases were found to be significantly upregulated during growth antagonizing other fungi (Gruber *et al.* 2011). Interestingly, transcriptome data from *Tr. atroviride* grown in the presence of the plant pathogen, *Rhizoctonia solani* J.G.

Kühn, show that secondary metabolite cluster genes were downregulated (Reithner *et al.* 2011).

Genomics of insect pathogenesis

In 2011, the first draft genomes of entomopathogenic *Cordyceps s.l.*, *Metarhizium anisopliae* (Metschn.) Sorokīn and *M. acridum* (Driver & Milner) J.F. Bisch., were published (Gao *et al.* 2011). These two genomes provided the first clues into the genome structure and content of hypocrealean entomopathogens. Notably, the genomes contained numerous secondary metabolite genes (see below), numerous proteases, and a large repertoire of cell surface receptors called PTH11-related G-protein coupled receptors (GPCRs). PTH11-related GPCRs were first identified in the sordariomycetous plant pathogen, *Magnaporthe grisea* (Kulkarni *et al.* 2005), and they are proposed to play a role in pathogenicity. In *Metarhizium* spp. Gao *et al.* (2011) documented differential regulation of some of these PTH11-related GPCRs during growth on media containing their host insect and non-host insect, and therefore hypothesized that these genes may play an important role in host recognition.

Two species from the Cordycipitaceae, *Cordyceps militaris* and *Beauveria bassiana* (=*Cordyceps bassiana*), were subsequently sequenced (Zheng *et al.* 2011, Xiao *et al.* 2012). A reduced number of CAZymes in the class glycoside hydrolase (GH) were reported from these two genomes, as was seen in *Metarhizium* spp. Whether this represents independent and parallel or an ancestral reduction GHs in the entomopathogenic lineages is unknown.

Bushley *et al.* (2013) published the draft genome sequence of *T. inflatum*, the first species sequenced from the genus *Tolypocladium* and the family Ophiocordycipitaceae. *Tolypocladium inflatum*, a beetle pathogen, was the first source from which the widely used immunosuppressant drug cyclosporin was isolated (Borel 2002). Bushley *et al.* (2013) focused on the regulation and evolution of that secondary metabolite cluster, which had only partially been identified prior to genome sequencing (Weber *et al.* 1994). Many of the genes in the cyclosporin cluster were upregulated in media containing beetle (*Otiorhyncus sulcatus*) hemolymph, but only a few genes were upregulated on media containing beetle cuticle. Another interesting finding of that study was the large number of homologous gene clusters shared between *T. inflatum* and the mainly mycoparasitic *Trichoderma spp*. (1749), especially as compared to those shared between *T. inflatum* and the insect pathogenic *Metarhizium* spp. (closest sequenced relatives) and *C. militaris* (Cordycipitaceae) (194).

A very preliminary and incomplete draft of the *Ophiocordyceps sinensis* genome was published in 2013 (Hu *et al.* 2013), and it represents the only genome sequence in Ophiocordycipitaceae outside of *Tolypocladium* published to date. Due to difficulties in sequencing and assembly, the authors could only estimate the complete size of the genome. The genome is estimated to be more than twice as large as any published hypocrealean genome (~120 Mb), and has low GC content and a reduced number of protein coding genes (6,972).

The genomes of several plant-associated species of Clavicipitaceae (*e.g.* *Claviceps* spp., *Epichloë* spp.) were recently published with a focus on the production and diversity of ergot alkaloids and other secondary metabolites and their clusters (see

below) (Schardl *et al.* 2013). Most of these species were grass endophytes or grass seed parasites. This grass endophyte lineage of Clavicipitaceae is derived from an insect pathogenic ancestor (Spatafora *et al.* 2007), but the extent to which CAZyme cellulase families were expanded or contracted as compared to *Metarhizium* spp. and other insect pathogens was not discussed.

Fungal secondary metabolism

In addition to the variety of nutritional sources used by hypocrealean species, they are also prolific producers of secondary metabolites, some of which are widely used in modern medicine (*e.g.*, cyclosporin, ergot alkaloids) (Borel 2002, van Dongen & de Groot 1995), and others of which are potent mycotoxins (*e.g.*, fumonisins, trichothecenes) (Desjardins & Proctor 2007, Rocha *et al.* 2005). Here, secondary metabolism is defined as the synthesis of small, often bioactive molecules that are not essential to an organism's growth. Within fungal genomes genes involved in the synthesis, modification and transport of secondary metabolites are often clustered in close proximity within a genome and coregulated (Keller *et al.* 2005).

Many secondary metabolite clusters found in fungal genomes have no known products or the species were previously unknown to produce known compounds, but it has been documented that production of secondary metabolites may be environmentally dependent (Schroeckh *et al.* 2009). Futhermore, genome sequencing has revealed that standard chemical analyses (*e.g.* mass spectrometry) greatly underestimated the secondary metabolite potential of most fungi with known compounds linked to secondary

metabolite gene clusters. Some secondary metabolites are proposed to play an important role in pathogenesis. For example, destruxins are cyclic peptides produced by *Metarhizium spp.* which are known to have insecticidal activity (Roberts 1981, Wang *et al.* 2012), and peptaibols, produced by the members of Hypocreaceae (Whitmore & Wallace 2004), form ion channels and have antifungal and antibiotic activity by inserting into cell membranes and disrupting membrane potential (Chugh & Wallace 2001).

Many secondary metabolites are produced by Nonribosomal Pepitde Synthetases (NRPSs), which are multimodular, often large enzymes that produce short peptides without the aid of a ribosome. These NRPS proteins are encoded by large genes and transcribed and translated via normal cellular mechanisms. Each module in an NRPS protein incorporates a single residue, often non-proteinogenic amino acid into the growing peptide (Marahiel *et al.* 1997). A module can contain a few to several domains, but most commonly three core domains exist in a module including adenylation, thiolation, and condensation domains. Adenylation (A-) domains are responsible for the activation of residues; thiolation domains covalently link the residue to the NRPS enzyme; and the condensation domain covalently links residues from neighboring thiolation domains, forming a peptide that is typically co-linear with the modules of the NRPS enzyme. The A-domains are the most conserved domains within NRPS genes, and therefore have been used to construct phylogenies of these genes in both bacteria and fungi (*e.g.* Stachelhaus *et al.* 1999, Wei *et al.* 2005, Bushley & Turgeon 2010). Because each NRPS gene may have between one and 23 A domains, each gene may be represented by multiple leaves in a phylogeny, and sometimes these domains have different evolutionary origins, representing instances of non-canonical evolutionary

processes (*e.g.* gene fusion, gene conversion). By analyzing A-domains in a phylogenetic context information can be gained about what kinds of metabolites are produced by an NRPS, about the evolution of these secondary metabolite genes themselves, and their distribution throughout the fungal tree of life.

Another class of secondary metabolite producing enzymes is Polyketide Synthetases (PKSs). Fungal PKS products are some of the most potent mycotoxins known (*e.g.* Aflatoxin, Yu *et al.* 2004). PKS genes contain several domains as well, but at a minimum contain a ketoacyl CoA synthase (KS), an acyltransferase (AT), and an acylcarrier (AC) domain (Keller *et al.* 2005). Importantly, fungi can also possess what are called Hybrid NRPS-PKS genes. This is a gene consisting of domains typical of both PKS and NRPS genes, and there is wide variation of domain combinations. Notable products of hybrids are the mycotoxin equisetin (Sims *et al.* 2005) and the antifungal compound, Pseurotin A (Maiya *et al.* 2007). PKS products can also be incorporated into NRPS peptides, and are therefore often found clustered together within a genome (Keller *et al.* 2005). An example of this is the NRPS and PKS containing cluster that produces cyclosporin in *T. inflatum* (Bushley *et al.* 2013).

Fungal metabolism in light of comparative genomics and transcriptomics

Analysis of whole genome content has produced meaningful, and sometimes unexpected, insights into the genetic underpinnings of fungal metabolism. CAZymes, for instance, are classes of proteins responsible for degradation of carbohydrates, and the number of copies of CAZymes in a genome tends to be positively correlated with the

ability to break down certain organic compounds (*e.g.*, lignin peroxidases to degrade lignin in most white rot fungi [Floudas *et al.* 2012]).

Regulation of proteins can also be key to an organism's survival during certain growth conditions. For instance in several ectomycorrhizal species (*e.g. Laccaria bicolor*, a basidiomycete, and *T. melanosporum*, of Pezizales), genes that could be potentially degradative to their hosts (*e.g.*, cellulases) are not expressed in symbiotic conditions (Martin & Bonito 2012). Obligate symbionts are often reduced in secondary metabolic potential as well (Martin *et al.* 2010), and several have undergone genome expansion including *T. melanosporum* (Martin *et al.* 2010) and *B. graminis* (Spanu *et al.* 2010). *Tuber melanosporum* has relatively few genes associated with secondary metabolism (Martin *et al*. 2010).

Comparative genomics has drastically changed the perception of the secondary metabolic potential for many fungi, especially those in Ascomycota. Although known to produce many secondary metabolites, the genome sequencing of *Aspergillus* spp. revealed tens more secondary metabolite clusters than were expected based on direct chemical isolation of its metabolites (Galagan *et al.* 2005, Keller *et al.* 2005, Sanchez *et al.* 2012). Examination of genomes in other fungi has also revealed greater genomic capacity for metabolite production than was known to exist, including with the genomes of the hypocrealean species *M. robertsii*, *M. acridum*, and *T. inflatum* (Gao *et al.* 2011, Bushley *et al.* 2013). Comparative genomic studies have also provided evidence for horizontal gene transfer (HGT) of entire secondary metabolite clusters in fungi (Khaldi *et al.* 2008, Slot & Rokas 2011, Campbell *et al.* 2012), although a mechanism for HGT between fungi remains to be found. In addition to HGT, individual gene and whole gene

cluster loss has also occurred (Wang *et al.* 2012, Campbell *et al.* 2012). A complex pattern of gene loss and HGT are shaping the disparate pattern of secondary metabolic cluster distribution that genomic sequencing is revealing, but how these events take place on a genetic level and what ecological and evolutionary circumstances drive them is an emerging frontier of research.

Conclusion and thesis content

The ability to generate whole genomic and transcriptomic scale data is now enabling mycologists to analyze fundamental questions about how fungi occupy ecological niches and utilize available resources (Hibbett *et al.* 2013). Recent advances in whole genome sequencing have spurred this movement and are revolutionizing fungal biology. This thesis project has taken place in the context of this burgeoning revolution, and as a result is genome focused. *Tolypocladium* and their *Elaphomyces* hosts are the systems in which this thesis will use genome and transcriptome data to explore questions around host switching, host and niche specific gene sets, and convergent evolution of gene content. Both *Tolypocladium* and *Elaphomyces* are ideal systems to examine these questions for several reasons. They are well-supported , monophyletic genera. The hosts of *Tolypocladium* spp. are very different: *Elaphomyces* truffles for most species, and Coleoptera or Cicadidae juveniles for others, and the ecological niche (ectomycorrhizae) of *Elaphomyces* differs from its close relatives, which are mainly saprobes. Differences in content and expression of genes involved primary and secondary metabolism in these species will be focused on here.

Chapter 2 addresses the systematics and nomenclature of Ophiocordycipitaceae in the light of changes made to the ICN prohibiting the use of more than one name for a single fungal species. Especially pertinent to the remainder of the thesis, *Tolypocladium* is redefined and new combinations are created. In Chapter 3, the genome sequence and secondary metabolite gene and gene cluster content of *T. ophioglossoides* are explored, with a focus on the diversity and evolution of peptaibiotic genes. An RNA-Seq experiment that examined *T. ophioglossoides* differential expression on media containing host and non-host tissue is the subject of Chapter 4. Differential expression of genes related to host recognition, adhesion, degradation, and secondary metabolism are discussed. Chapter 5 is a comparative genomic study of *Tolypocladium*, including the previously sequenced beetle pathogen, *T. inflatum*, the *Elaphomyces* parasite *T. ophioglossoides* (sequenced as a part of Chapter 3), and two newly generated genomes, including the cicada pathogen *T. paradoxum* and another *Elaphomyces* parasite, *T. capitatum*. Evolution of the genus and gene tree/ species tree incongruence are examined in addition to comparisons of primary and secondary metabolic gene content in the truffle parasites compared to the insect pathogens. Finally, in Chapter 6 the genome structure and content of *Elaphomyces granulatus*, a common host for *T. ophioglossoides* and *T. captitatum*, is analyzed, along with a survey of the microbial community associated with the sporocarp as determined by metagenomic sequencing.

References

Barron GL (1980) Fungal parasites of rotifers: a new *Tolypocladium* with underwater conidiation. *Canadian Journal of Botany* **58**: 439–442.

Barron GL (1981) Two new fungal parasites of bdelloid rotifers. *Canadian Journal of Botany* **59**: 1449–1455.

Barron GL (1983) Structure and biology of a new *Tolypocladium* attacking bdelloid rotifers. *Canadian Journal of Botany* **61**: 2566–2569.

Behie SW, Zelisko PM, Bidochka MJ (2012) Endophytic insect-parasitic fungi translocate nitrogen directly from insects to plants. *Science* **336**: 1576-1577.

Bissett J (1983) Notes on *Tolypocladium* and related genera. *Canadian Journal of Botany* **61**: 1311–1329.

Borel JF (2002) History of the discovery of cyclosporin and of its early pharmacological development. *Wiener Klinische Wochenschrift* **114**: 433-437.

Boros C, Hamilton, SM Katz, B, Kulanthaivel P (1994) Comparison of balanol from *Verticillium balanoides* and ophiocordin from *Cordyceps ophioglossoides*. *The Journal of Antibiotics* **47**: 1010-1016.

Bruck DJ (2005) Ecology of *Metarhizium anisopliae* in soilless potting media and the rhizosphere: implications for pest management. *Biological Control* **32**: 155-163.

Buchert J, Oksanen J, Pere, J, Siika-Aho M, Suurnäkki A, Viikari L (1998) Applications of Trichoderma reesei enzymes in the pulp and paper industry. In: *Trichoderma and Gliocladium Volume 2*. (Harman GE, Kubicek CP, eds.): 343-364 London: Taylor & Francis.

Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, Owensby CA, Knaus BJ, Elser J, Miller D, Di Y, McPhail KL, Spatafora JW (2013) Draft genome sequence of the Cyclosporin producing fungus *Tolypocladium inflatum* reveals complex patterns of secondary metabolite evolution and expression. *PLoS Genetics* **9**: e1003496.

Bushley KE, Turgeon BG (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology* **10**: 26.

Campbell MA, Rokas A, Slot JC (2012) Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome biology and evolution* **4**: 289-293.

Castellano MA, Trappe JM, Vernes K (2011) Australian species of *Elaphomyces* (Elaphomycetaceae, Eurotiales, Ascomycota). *Australian Systematic Botany* **24**: 32-57.

Castellano MA, Beever RE, Trappe JM (2012) Sequestrate fungi of New Zealand: *Elaphomyces* and *Rupticutis* gen. nov. (Ascomycota, Eurotiales, Elaphomycetaceae). *New Zealand Journal of Botany* **50**: 423-433.

Castlebury LA, Rossman AY, Sung GH, Hyten AS, Spatafora JW (2004) Multigene phylogeny reveals new lineage for *Stachybotrys chartarum*, the indoor air fungus. *Mycological research* **108**: 864-872.

Chugh JK, Wallace BA (2001) Peptaibols: models for ion channels. *Biochemical Society Transactions* **29**: 565-570.

Clarkson JM, Charnley AK (1996) New insights into the mechanisms of fungal pathogenesis in insects. *Trends in microbiology* **4**: 197-203.

Desjardin AE, andProctor RH. 2007. Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology* **119**: 47-50.

van Dongen PWJ, de Groot ANJA (1995) History of ergot alkaloids from ergotism to ergometrine. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **60**: 109-116.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Gorecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kues U, Arun Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Duenas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St. John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev I, Hibbett DS (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715-1719.

Fries EM (1918) *Obersvationes Mycologicae Pars Secunda*. Hafniae.

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee S-I, Baş türkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MÁ, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105-1115.

Galante YM, De Conti A, Monteverdi R (1998) Applications of *Trichoderma reesei* enzymes in the food and feed industries. In: *Trichoderma and Gliocladium Volume 2*. (Harman GE, Kubicek CP, eds.): 338-342 London: Taylor & Francis.

Galante YM, De Conti A, Monteverdi R (1998) Applications of Trichoderma reesei enzymes in the textile industry. In: *Trichoderma and Gliocladium Volume 2*. (Harman GE, Kubicek CP, eds.): 311-326 London: Taylor & Francis.311-326.

Gams W (1971) *Tolypocladium*, eine Hyphomycetengattung mit geschwollenen Phialiden. *Persoonia* **6**: 185–191.

Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie X-Q, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma L-J, St. Leger RJ, Zhao G-P, Pei Y, Feng M-G, Xia Y, Wang C (2011) Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics* **7**: 1-18.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699-704.

Gruber S, Vaaje-Kolstad G, Matarese F, López-Mondéjar R, Kubicek CP, Seidl-Seiboth V (2011) Analysis of subgroup C of fungal chitinases containing chitin-binding

and LysM modules in the mycoparasite *Trichoderma atroviride*. *Glycobiology* **21**: 122-133.

Hibbett DS, Stajich JE, Spatafora JW (2013) Toward genome-enabled mycology. *Mycologia* **105**: 1339-1349.

Hopkins TL, Kramer KJ (1992) Insect cuticle sclerotization. *Annual Review of Entomology* **37**: 273-302.

Hu G, St. Leger RJ (2002) Field studies using a recombinant mycoinsecticide (*Metarhizium anisopliae*) reveal that it is rhizosphere competent. *Applied and Environmental Microbiology* **68**: 6383-6387.

Hu X, Zhang Y, Xiao G, Zheng P, Xia Y, Zhang X, St. Leger RJ, Liu X, Wang C (2013) Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chinese Science Bulletin* **58**: 2846–2854.

Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism - from biochemistry to genomics. *Nature Reviews Microbiology* **3**: 937-947.

Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E, Hosoya T, Bischoff JF, Spatafora JW (2012) Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilaginoidea* (Clavicipitaceae). *American Journal of Botany* **99**: 552-561.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736-741.

Kirk PM, Cannon PF, Minter DW, Stalpers JA (eds) (2008) *Ainsworth & Bisby's Dictionary of the Fungi.* 10th edn. Wallingford: CAB International.

Kneifel H, König WA, Loeffler W, Müller R (1977) Ophiocordin, an antifungal antibiotic of *Cordyceps ophioglossoides*. *Archives of Microbiology* **113**: 121-130.

Kobayasi, Y. (1939). On the genus *Cordyceps* and its allies on cicadae from Japan. *Bulletin of the Biogeographical Society of Japan* **9**: 145-176.

Kobayasi Y, Shimizu D (1960) Monographic studies of *Cordyceps* 1. Group parasitic on *Elaphomyces*. *Bulletin of the National Science Museum Tokyo* **5**: 69-85.

Korf RP (1973) Discomycetes and tuberales. *The fungi*: 249-319.

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Dohren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gomez-Rodriguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernandez-Onate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lubeck M, Lubeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, and Grigoriev IV (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology* **12**: R40.

Kulkarni RD, Thon MR, Pan H, Dean RA (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biology* **6**: R24.

Kurth F, Zeitler K, Feldhahn L, Neu TR, Weber T, Krištůfek V, Wubet T, Herrmann S, Buscot F, Tarkka MT (2013) Detection and quantification of a mycorrhization helper bacterium and a mycorrhizal fungus in plant-soil microcosms at different levels of complexity. *BMC Microbiology* **13**: 205.

Landvik S, Shaller NFJ, Eriksson OE (1996) SSU rDNA sequence support for a close relationship between the Elaphomycetales and the Eurotiales and Onygenales. *Mycoscience* **37**: 237-241.

Latgé (2007) The cell wall: carbohydrate armour for the fungal cell. *Molecular Microbiology* **66**: 279-290.

Latgé (2010) Tasting the fungal cell wall. *Cellular Microbiology* **12**: 863-872.

Lloyd CG (1918) Mycological Notes. *Mycological Notes* **5**: 809.

LoBuglio KF, Berbee ML, Taylor JW (1996) Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the Ascomycetes. *Molecular Phylogenetics and Evolution* **6**: 287-294.

Lundgren B, Bååth E, Söderström BE (1978) Antagonistic effects of *Tolypocladium* species. *Transactions of the British Mycological Society* **70**: 305-307.

Mains EB (1957) Species of *Cordyceps* parasitic on *Elaphomyces*. *Bulletin of the Torrey Botanical Club* **84**: 243-251.

Maiya S, Grundmann A, Li X, Li SM, Turner G (2007) Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen *Aspergillus fumigatus*. *ChemBioChem* **8**: 1736-1743.

Mandels M, Reese ET (1957) Induction of cellulas in *Trichoderma viride* as influenced by carbon sources and metals. *Journal of Bacteriology* **73**: 269-278.

Marahiel MA, Stachelhaus T, Mootz HD (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical Reviews* **97**: 2651-2673.

Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buee M, Brokstein P, Canback B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbe J, Lin YC, Legue V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculity-Hirzel H, Oudot-Le Secq MP, Quesneville H, Rajaskekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kues U, Lucas S, Van de Peer Y, Podila GK, Polles A, Pukkila PJ, Richardson PM, Rouze P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-93.

Martin F, Bonito GM (2012) Ten Years of Genomics for Ectomycorrhizal Fungi: What Have We Achieved and Where Are We Heading? In *Edible Ectomycorrhizal Mushrooms* (Zambonelli A, Bonito GM, eds.): 383-401. Berlin Heidelberg: Springer.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, Porcel B, Rubini A, Amicucci A, Amselem J, Anthouard V, Arcioni S, Artiguenave F, Aury J-M, Ballario P, Bolchi A, Brenna A, Brun A, Buée M, Cantarel B, Chevalier G, Couloux A, Da Silva C, Denoeud F, Duplessis S, Ghignone S, Hilselberger B, Iotti M, Marçais B, Mello A, Miranda M, Pacioni G, Quesneville H, Riccioni C, Ruotolo R, Splivallo R, Stocci V, Tisserant E, Roberto Viscomi A, Zambonelli A, Zampieri E, Henrissat B, Lebrun M-H, Paolocci F, Bonfante P, Ottonello S, Wincker P (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033-1038.

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, Lopez de Leon A, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature biotechnology* **26**: 553-560.

Maser C, Trappe JM, Nussbaum RA (1978) Fungal-small mammal interrelationships with emphasis on Oregon coniferous forests. *Ecology* **59**: 799-809.

Miller SL, Miller OK Jr. 1984. Synthesis of *Elaphomyces muricatus+ Pinus sylvestris* ectomycorrhizae. *Canadian journal of botany* **62**: 2363-2369.

Neville AC (1984) Cuticle: organization. In: *Biology of the Integument* (Bereiter-Hahn J, Matoltsy AG, Richards KS, eds.): 611-625 Berlin Heidelberg: Springer Verglag.

Nikoh N, Fukatsu T (2000) Interkingdom host jumping underground: phylogenetic analysis of entomoparasitic fungi of the genus *Cordyceps*. *Molecular Biology and Evolution* **17**: 629-638.

Osherov N, Yarden O (2010) The Cell Wall of Filamentous Fungi. In: *Cellular and Molecular Biology of Filamentous Fungi*. (Borkovich, Ebbole, eds.): 224-237 Washington DC: ASM Press.

Putri SP, Kinoshita H, Ihara F, Igarashi Y, Nihira T (2010) Ophiosetin, a new tetramic acid derivative from the mycopathogenic fungus *Elaphocordyceps ophioglossoides*. *The Journal of Antibiotics* **63**: 195-198.

Roberts (1981) Toxins of Entomopathogenic Fungi. In: *Microbial control of pests and plant diseases*. (Burges HD, ed.): 441-463 London, New York: Academic Press.

Rocha O, Ansari K, Doohan FM (2005) Effects of trichothecene mycotoxins on eukaryotic cells: a review. *Food additives and contaminants* **22**: 369-378.

Samson RA, Soares GG Jr (1984) Entomopathogenic species of the hyphomycete genus *Tolypocladium*. *Journal of Invertebrate Pathology* **43**: 133–139.

Samuels GJ (1996) *Trichoderma*: a review of biology and systematics of the genus. *Mycological Research* **100***:* 923-935.

Sanchez JF, Somoza AD, Keller NP, Wang CC (2012) Advances in *Aspergillus* secondary metabolite research in the post-genomic era. *Natural product reports* **29**: 351-371.

Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J,M Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach JE, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z (2013) Plant-symbiotic fungi as chemical engineers: Multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics* **9**: e1003323.

Schuster A, Schmoll M (2010) Biology and biotechnology of *Trichoderma*. *Applied Microbiology and Biotechnology* **87**: 787-799.

Seidl V (2008) Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. *Fungal Biology Reviews* **22**: 36-42.

Seidl V, Huemer B, Seiboth B, Kubicek CP (2005) A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS Journal* **272**: 5923-5939.

Shaw SR (1988) Euphorine phylogeny: the evolution of diversity in host-utilization by parasitoid wasps (Hymenoptera: Braconidae). *Ecological Entomology* **13**: 323-335.

Sims JW, Fillmore JP, Warner DD, Schmidt EW (2005) Equisetin biosynthesis in *Fusarium heterosporum*. *Chemical Communications* 186-188.

Slot JC, Rokas A (2011) Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Current Biology* **21:** 134-139.

Spatafora JW, Sung GH, Sung JM, Hywel-Jones NL, White JF (2007) Phylogenetic evidence for an animal pathogen origin of ergot and the grass enndophytes. *Molecular Ecology* **16**: 1701-1711.

St. Leger RJ, Goettel M, Roberts DW, Staples RC (1991) Prepenetration events during infection of host cuticle by *Metarhizium anisopliae*. *Journal of Invertebrate Pathology* **58**: 168-179.

Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology* **6**: 493-505.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, Lebrun M-H, Ridout CJ, Schulze-Lefert P, Talbot NJ, Ahmadinejad N, Ametz C, Barton GR, Benjdia M, Bidzinski P, Bindschedler LV, Both M, Brewer MT, Cadle-Davidson L, Cadle-Davidson MM, Collermare J, Cramer R, Frenkel O, Godfrey D, Harriman J, Hoede C, King BC, Klages S, Kleemann J, Knoll D, Koti PS, Kreplak J, López-Ruiz FJ, Lu X, Maekawa T, Mahanil S, Micali C, Milgroom MG, Montana G, Noir S, O'Connell RJ, Oberhaensli S, Parlange F, Pedersen C, Quesneville H, Reinhardt R, Rott M, Sacristán S, Schmidt SM, Schön M, Skamnioti P, Sommer H, Stephans A, Takahara H, Thordal-Christensen H, Vigouroux M, Weßling R, Wicker T, Panstruga R (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543-1546.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5–59.

Sung G-H, Poinar GO, Spatafora JW (2008) The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal–arthropod symbioses. *Molecular Phylogenetics and Evolution* **49**: 495-502.

Trappe JM (1979) The orders, families, and genera of hypogeous Ascomycotina (truffles and their relatives). *Mycotaxon* **9**: 297-340.

Wang B, Kang Q, Lu Y, Bai L, Wang C (2012) Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proceedings of the National Academy of Sciences* **109**: 1287-1292.

Wei X, Yang F, Straney DC (2005) Multiple non-ribosomal peptide synthetases genes determine peptaibol synthesis in *Trichoderma virens*. *Canadian Journal of Microbiology* **51**: 423-429.

Weiser J, Matha V, Jegorov A (1991) *Tolypocladium terricola* sp. n., a new mosquito-killing species of the genus *Tolypocladium* Gams (hyphomycetes). *Folia Parasitologica* **38**: 363–369.

Whitmore L, Wallace BA (2004) The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Research* **32**: D593-D594.

Xiao G, Ying S-H, Zheng P, Wang Z-L, Zhang S, Xie X-Q, Shang Y, St. Leger RJ, Zhao G-P, Wang C, Feng M-G. 2012. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Scientific Reports* **2**: 483.

Yu J, Chang PK., Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, Payne GA, Linz JE, Woloshuk CP, Bennett JW (2004) Clustered pathway genes in aflatoxin biosynthesis. *Applied and Environmental Microbiology* **70**: 1253-1262.

Zheng P, Xia L, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S, Zhao G-P, Liu X, St. Leger RJ, Wang C (2011) Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biology* **12**: R116.

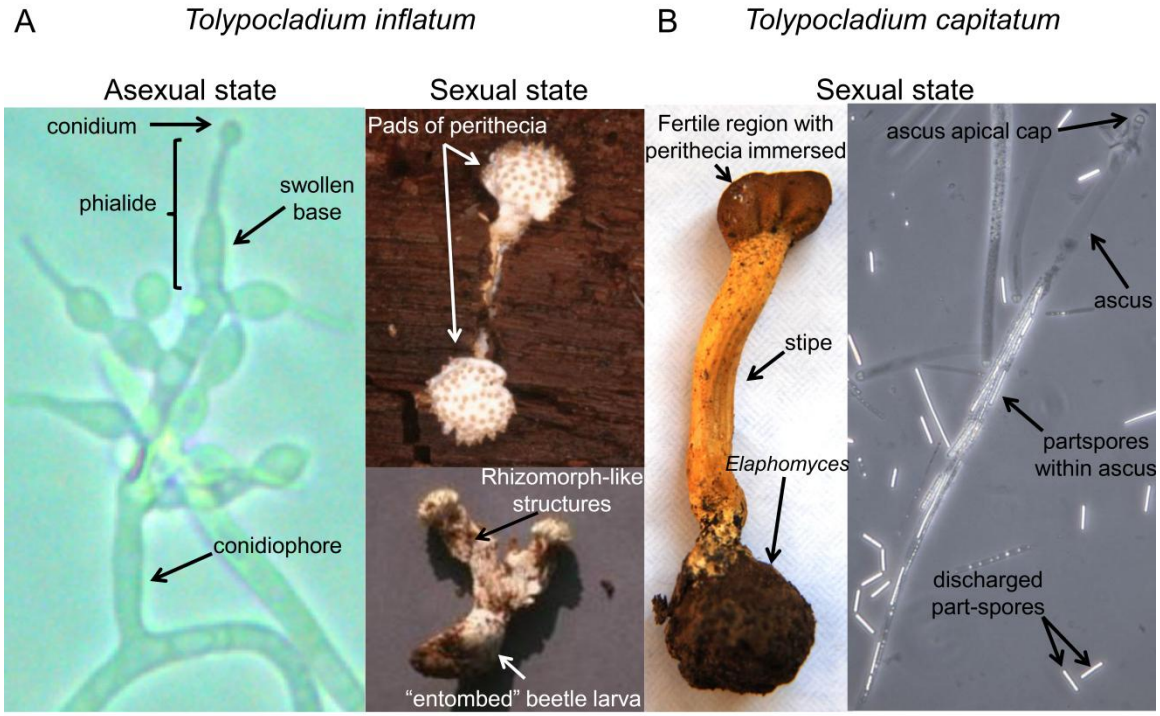


Figure 1.1. Fruiting structures of *Tolypocladium*. A. Asexual and sexual reproductive structures in *T. inflatum*. B. Sexual reproductive structures in *T. capitatum*, for which no asexual reproductive structures have been described. Photo credits: *T. inflatum* in panel A, Joseph Spatafora, and *T. capitatum* in panel B, C. Alisha Quandt.

Table 1.1. *Tolypocladium* species, their hosts, and ability to grow in culture. Y = yes, N = no, R = rarely, and U = unknown. Host/Substrate colors: Brown = soil, Peach = animal, no color = *Elaphomyces*. Growth in culture color: green = yes, no color = no/rarely. Notice correlation between taxa growing in culture (where known) and the host/substrate.

Species	Host/ Substrate	Growth in culture?
T. album	Soil	Y
T. capitatum	*Elaphomyces*	R
T. cylindrosporum	Soil	Y
T. delicatastipitatum	*Elaphomyces*	U
T. extinguens	Diptera	Y
T. fractum	*Elaphomyces*	U
T. geodes	Soil	Y
T. inegoense	Cicadidae	Y
T.inflatum	Coleoptera/ Soil	Y
T. intermedium	*Elaphomyces*	U
T. japonicum	*Elaphomyces*	N
T. jezoense	*Elaphomyces*	N
T. lignicola	Rotifera	Y
T. longisegmentum	*Elaphomyces*	N
T. minazukiense	*Elaphomyces*	U
T. miometeana	*Elaphomyces*	U
T. nubicola	Soil	Y
T. oligosporum	Soil	Y
T. ophioglossoides	*Elaphomyces*	Y
T. paradoxum	Cicadidae	Y
T. pustulatum	Soil	Y
T. ramosum	*Elaphomyces*	U
T. rouxii	*Elaphomyces*	N
T. szemaoense	*Elaphomyces*	U
T. sinense	*Ophiocordyceps sinensis*	Y
T. tenuisporum	*Elaphomyces*	U
T. terricola	Soil/ Diptera	Y
T. trigonosporum	Rotifera	Y
T. tundrense	Soil	Y
T. valliformum	*Elaphomyces*	U
T. valvatostipitatum	*Elaphomyces*	U
T. virens	*Elaphomyces*	U

Chapter 2. Phylogenetic-based nomenclatural proposals for *Ophiocordycipitaceae* (*Hypocreales*) with new combinations in *Tolypocladium*

C. Alisha Quandt, Ryan M. Kepler, Walter Gams, João P. M. Araújo, Sayaka Ban, Harry C. Evans, David Hughes, Richard Humber, Nigel Hywel-Jones, Zengzhi Li, J. Jennifer Luangsa-ard, Stephen A Rehner, Tatiana Sanjuan, Hiroki Sato, Bhushan Shrestha, Gi-Ho Sung, Yi-Jian Yao, Rasoul Zare, and Joseph W. Spatafora

Journal: *IMA Fungus*
Milford House, The Mead, Ashtead, Surrey KT21, 2LZ, UK.
Article published June 19, 2014, vol. 5(1): 121-134. doi:10.5598

Background

The revision of Art. 59 in the *International Code of Nomenclature for algae, fungi, and plants* (ICN; McNeill *et al.* 2012) has created a major task for mycologists, who must now reconcile under one name various possible names existing for different morphs of the same species of fungus (Hibbett & Taylor 2013). Groups have already begun to propose names which should be protected or suppressed within *Hypocreales* in accordance with the 'one fungus one name' policy (Geiser *et al.* 2013, Rossman *et al.* 2013, Leuchtmann *et al.* 2014, Johnston *et al.* 2014, Kepler *et al.* 2014) and others are in progress. Here, we seek to retain names in *Ophiocordycipitaceae* with the goal of harmonizing priority, monophyly, simplicity of taxonomic revisions, and minimization of disruption to the research community.

The family *Ophiocordycipitaceae* was described by Sung *et al.* (2007) to accommodate species that were determined to be phylogenetically distinct from *Cordycipitaceae* and *Clavicipitaceae s.s.* Asexual morphologies in *Ophiocordycipitaceae* show a tremendous range of variation, some of which are restricted in their phylogenetic distribution while others are often found in disparate lineages. For example, *Verticillium* is a common asexual morph of many species in several hypocrealean families, including *Ophiocordycipitaceae*, *Cordycipitaceae* and *Clavicipitaceae* (see Zare *et al.* 2000, Sung *et al.* 2001, 2007, and Gams & Zare 2001).

Ophiocordyceps is the most speciose genus of the family, and was described originally by Petch (1931a) for species of *Cordyceps* that have septate ascospores that do not disarticulate into part-spores at maturity and asci with inconspicuous apical caps (Petch 1931a, 1933). Kobayasi (1941) later used *Ophiocordyceps* as a subgeneric

classification of the genus *Cordyceps*, but Sung *et al.* (2007) restored *Ophiocordyceps* to the rank of genus to include those *Cordyceps* species within *Ophiocordycipitaceae* forming a sister clade with the genus *Elaphocordyceps* (see below). The type of the genus is *O. blattae*, a rarely collected cockroach pathogen for which no culture or molecular data are available.

Asexual generic names associated with *Ophiocordyceps* include *Sorosporella*, the oldest name still in use for species in the clade, *Hirsutella*, *Hymenostilbe*, *Stilbella*, *Syngliocladium*, and *Paraisaria*. *Hirsutella* species typically produce one to several conidia in a limited mucus droplet borne on basally subulate phialides that taper into slender necks (Gams & Zare 2003). *Hymenostilbe* was proposed by Petch (1931b), and there is some evidence to support restricting its use within the genus *Ophiocordyceps* to the '*O. sphecocephala* clade', most species of which sporulate from adult insects (Sung *et al.* 2007, Luangsa-ard *et al.* 2011a). These taxa produce conidia singly from multiple denticles on conidiogenous cells forming a palisade-like layer along the entire outer surface of synnemata (Mains 1950). The *Stilbella* morphology has been applied broadly among species associated with *Ophiocordyceps*, as well as to fungi later reclassified in other genera (Seifert 1985, Gräfenhan *et al.* 2011). *Stilbella* species often produce aggregate synnemata with a fertile, terminal head of conidia. *Syngliocladium* spp. often have laterally arising conidiophores similar in morphology to the hypocreaceous asexual morph *Gliocladium*, and they may be either synnematous or mononematous on their arthropod hosts (Petch 1932, Hodge *et al.* 1998). *Sorosporella*, a chlamydospore producing spore state, has been linked as a synasexual morph of *Syngliocladium* (Speare 1917, 1920), but the two morphologies are not always produced by all species (Hodge *et*

*al.*1998, Evans & Shah 2002). Species of *Paraisaria* possess feathery synnemata which fruit from arthropod hosts, and several species have been linked via cultural and molecular data to the *O. gracilis* clade (Samson & Brady 1983, Sung *et al.* 2007, Evans *et al.* 2010). Names of genera associated with *Ophiocordyceps* whose types are located outside of *Hypocreales* include *Tilachlidiopsis* and *Podonectria*, members of the *Agaricomycetes* and *Dothideomycetes*, respectively (Rossman 1978, Stalpers *et al.*1991, Hughes *et al.* 2001, Boonmee *et al.* 2011). Despite the large number of taxa associated with *Ophiocordyceps*, a lack of support for internal nodes resulting in equivocal topologies has limited inferences about relationships within the genus in previous studies (Sung *et al.*2007).

The most notable species in the *Ophiocordyceps* clade is *O. sinensis*, which is nearly double the price of gold by weight (Stone 2008, Shrestha & Bawa 2013) and the subject of intense research, especially in China (Shrestha *et al.* 2010, Hu *et al.* 2013, Ren & Yao 2013, Bushley *et al.* 2013a, *etc.*). Almost exclusively found parasitizing the larvae of ghost moths (*Hepialidae*: *Thitarodes*) in the alpine and sub-alpine pastures of the Tibetan plateau and the Himalayas, this species is undergoing heavy, possibly unsustainable, and destructive harvesting (Cannon *et al.* 2009, Shrestha & Bawa 2013).

The recently described genus *Elaphocordyceps* is typified by *E. ophioglossoides*, one of the first *Cordyceps* species to be described. Species in *Elaphocordyceps* are mostly parasites of the ectomycorrhizal truffle genus *Elaphomyces* (*Ascomycota*, *Eurotiales*). The majority of *Elaphocordyceps* species have no known asexual morph, but where known they produce ones which are verticillium-like or *Tolypocladium* (Sung *et al.* 2007). There are a few *Elaphocordyceps* species known to be entomopathogens,

including three cicada pathogens (*E. inegoensis*, *E. paradoxa*, and *E. toriharamontana*), and one beetle pathogen, *E. subsessilis* (syn.*Tolypocladium inflatum*) (Hodge *et al.* 1996, Sung *et al.* 2007). *Tolypocladium inflatum* (a name conserved by the rejection of *Pachybasium niveum*; Dreyfuss & Gams 1994), is a medicinally important fungus and the subject of much research due to its production of the immunosuppressant drug, cyclosporin A (Survase *et al.* 2011, Bushley *et al.* 2013b). The other species of *Tolypocladium* have no known sexual morphs and have mainly been isolated from soil (Gams 1971, Bissett 1983) or observed parasitizing rotifers or insects (Barron 1980, 1981, 1983, Samson & Soares 1984, Weiser *et al.* 1991). The asexually typified genus *Chaunopycnis* is also related to this clade (Bills *et al.* 2002) and has been isolated mainly from soil samples (Gams 1980, Bills *et al.* 2002), although one species was isolated from epilithic Antarctic lichens (Möller & Gams 1993). The similarity of conidiogenesis between *Chaunopycnis* and *Tolypocladium* was noted in the original description of *Chaunopycnis* (Gams 1980), and its phialides often taper in a manner similar to those of *Tolypocladium*. Interestingly, these two genera have also been linked by their shared production of cyclosporin A (Traber & Dreyfuss 1996). Two of the described *Chaunopycnis* species produce loosely enclosed conidiomata, a morphology not seen in other members of the clade or within *Ophiocordycipitaceae* as a whole.

The relationships among the species of the *Purpureocillium* clade were recently reviewed by Luangsa-ard *et al.* (2011b). The genus was proposed to encompass taxa closely related to *Purpureocillium lilacinum* (syn. *Paecilomyces lilacinus*) and consists of species with purple-hued conidia, including *Nomuraea atypicola* and *Isaria takamizusanensis*. The type of *Nomuraea* is *N. rileyi* (syn. *N. prasina*), which has

recently been synonymized with *Metarhizium* (Kepler *et al.* 2014). The type of *Isaria* is a member of *Cordycipitaceae* (Gams *et al.* 2005, Hodge *et al.* 2005, Luangsa-ard *et al.* 2011b). While *N. atypicola* and *I. takamizusanensis* have not been addressed taxonomically, other studies found close relationships between these taxa and *Purpureocillium* (Sung *et al.*2007, Perdomo *et al.* 2013). *Nomuraea atypicola* is the asexual morph of *C. cylindrica* (Hywel-Jones & Sivichai 1995), the only sexual morph described for this clade and one of the "residual" *Cordyceps s. lat*. left without reassignment to any phylogenetically redefined genus by Sung *et al.* (2007).

Nematode pathogens have been described in many genera throughout *Hypocreales*. The largest and oldest of these is the asexually typified genus *Harposporium*. Most *Harposporium* species produce crescent-shaped or helicoid conidia that are ingested by their hosts and become lodged in the upper portions of the digestive tract (Barron 1977). Conidia are produced on spherical conidiogenous cells, and several species are known to produce hirsutella-like synasexual morphs (Hodge *et al.* 1997, Chaverri *et al.* 2005, Li *et al.* 2005). While the majority of *Harposporium* species are known from nematodes, these fungi are common in the soil and several studies have reported an entomopathogenic ecology as well (*e.g.*, Shimazu & Glockling 1997, Evans & Whitehead 2005). In 2005, Chaverri *et al.* reported the asexual-sexual morph connection between *Harposporium* and *Podocrella*, an arthropod-pathogenic genus. Several researchers initially described nematophagous taxa in the originally plant-pathogenic genus *Meria* (Vuillemin 1896, Drechsler 1941), but this genus was found to be polyphyletic (Gams & Jansson 1985), and for this reason *Drechmeria* was erected for the nematophagous meria-like taxa in *Hypocreales*. The type of *Drechmeria*, *D.*

coniospora, has cone-shaped conidia whose conidiogenous cells are not basally swollen as in *Harposporium*. One protozoan-infecting species of *Drechmeria, D. harposporioides*, produces crescent-shaped conidia similar to those of *Harposporium* (Barron & Szijarto 1982). *Haptocillium* was erected for asexual nematode pathogens bearing verticillate phialides and whose conidia are not ingested but adhere to the surface of their hosts (Zare & Gams 2001).

Polycephalomyces represents a diverse clade that is currently *incertae sedis* within *Hypocreales*, as its placement has lacked support in previous molecular studies (Kepler *et al.* 2013). Of particular uncertainty was whether *Polycephalomyces* and its closest related taxon, *C. pleuricapitata*, formed a sister clade to *Ophiocordycipitaceae*, or if it was more closely related to *Clavicipitaceae*. Many morphological characters are shared between *Ophiocordycipitaceae* and *Polycephalomyces*. For example, numerous species in both clades produce hirsutella-like anamorphs with conidia often borne in a slimy mass (Seifert 1985). In addition, sexual fruiting structures of *Polycephalomyces* often possess a wiry, tough, carbonaceous stipe which is a common morphology of *Ophiocordyceps* (Kepler *et al.* 2013). Many species within this genus are known mycoparasites of other hypocrealean entomopathogens and myxomycetes, but there are also several species of entomopathogens. *Cordyceps pleuricapitata* was deemed a residual species of *Cordyceps* of uncertain placement by Kepler *et al.* (2013), due to a lack of statistical support joining that species and *Polycephalomyces*.

In this paper we expand the taxon sampling presented in Sung *et al.* (2007) by 222 hypocrealean isolates. This includes sexual and asexual states which provide the

framework for addressing the nomenclatural issues demanded by changes to the most recent ICN.

Materials and Methods

Sequences from five nuclear loci, including the small and large subunits of the rDNA (SSU and LSU), the transcription elongation factor-1α (TEF), and the first and second largest subunits of RNA polymerase II (RPB1 and RPB2) were used for phylogenetic analyses. DNA extraction and PCR amplification were carried out as previously described (Kepler *et al.* 2013). Sequencing reactions were performed at the University of Washington High-Throughput Genomics Center (Seattle, WA) with the primers used for the initial amplifications. All other sequences were collected from GenBank. Efforts were made for all specimens to have data for at least three of the five genes to be considered in our analyses. However, certain taxa for which only one or two genes were available were included due to the importance in addressing the taxonomic issues at hand (Table 2.1).

Raw sequences were processed, aligned, and gaps excluded as in Kepler *et al.* (2013), using the programs MAFFT v. 6 (Katoh *et al.* 2002, Katoh & Toh 2008), Geneious v. 7.0.6 (Biomatters, available http://www.geneious.com), and Gblocks (Talavera & Castresana 2007). The final alignment length was 4570 nucleotides - 1023 for SSU, 879 for LSU, 987 for TEF, 646 for RPB1, and 1035 for RPB2. RAxML v. 7.6.6 (Stamatakis 2006) was used to perform Maximum likelihood (ML) estimation of the phylogeny with 500 bootstrap replicates on the concatenated dataset using eleven data

partitions. These included one each for SSU and LSU, and three for each of the three codon positions of the protein coding genes, TEF, RPB1, and RPB2. The GTR-GAMMA model of nucleotide substitution was used.

Results and Discussion

Our results are in agreement with the overall phylogenetic structure of the order *Hypocreales* put forth by Sung *et al.* (2007). Nomenclatural issues for taxa in the other two families of cordyceps-like organisms, *Cordycipitaceae* and *Clavicipitaceae*, will be presented elsewhere or have already been published (Leuchtmann *et al.* 2014, Kepler *et al.* 2014). Based on this exhaustive phylogenetic reconstruction (Figure 2.1), we recognize six genera within *Ophiocordycipitaceae* *Ophiocordyceps*, *Tolypocladium*, *Purpureocillium*, *Harposporium*, *Drechmeria*, and *Polycephalomyces* (Table 2.2). This framework will provide clarity for researchers, ease of communication for instructors, and phylogenetic taxonomy around which to investigate the evolution of life histories (*e.g.* morphology, ecology).

Taxonomy

Ophiocordyceps Petch 1931

Ophiocordyceps sensu Sung *et al.* (2007) is resolved as a well-supported (MLBP=77) clade (Fig. 1, Node 3). This clade is speciose, diverse, and almost exclusively comprises insect pathogens. In spite of increased taxon sampling, current reconstructions fail to find strong statistical support at the internal nodes, and therefore we refrain from defining

infrageneric groupings (Fig. 1). While *Sorosporella* is the oldest name for any members in this clade, there are only two described species, and Evans & Shah (2002) argued *Sorosporella* should be synonymized with *Syngliocladium* instead of being recognized as a an asexual morph, as *Synnematium* was previously treated with respect to *Hirsutella* (Evans & Samson 1982). We propose, therefore, to suppress the use of *Sorosporella* for this clade. *Hirsutella* is the next oldest name, but the type, *H. entomophila*, which was described growing from adult Coleoptera, has not been sampled and no culture of this species is available. Sung *et al.* (2007) argued that the *Hirsutella* morphology was phylogenetically informative for the '*O. unilateralis* group' which they resolved as paraphyletic, a topology recovered in the current analyses as well (Fig 1, Nodes 4 and 5). However, the *Hirsutella* morphology is observed in other clades (*e.g. Harposporium*, *Polycephalomyces*, *Clavicipitaceae*), and while it is difficult to place the type species based on morphology alone, it appears from its original description to be morphologically and ecologically (as a parasite of adults) similar to species of *Hymenostilbe* found in the '*O. sphecocephala*' clade and not *Hirsutella* of the '*O. unilateralis* group' (Patouillard 1892). Another reason for suppressing the use of *Hirsutella* for this clade is the larger number of new combinations that would have to be made – 178 for *Ophiocordyceps vs.* 77 for *Hirsutella* – as the vast majority of species encompassed here are currently described as *Ophiocordyceps*. Also, preservation of the name "cordyceps" within the name of *O. sinensis* is considered paramount given its economic, medicinal, and cultural importance in addition to being the most widely known and researched species in the clade (Shrestha *et al.* 2010).

At this time, we also propose to suppress the use of the other names proposed for taxa in this clade, including *Hymenostilbe*, *Synglíocladium*, and *Paraisaria*, because these names are younger, and they contain fewer associated taxa than either *Ophiocordyceps* or *Hirsutella*. Our results suggest the restriction of *Hymenostilbe* to the '*O. sphecocephala* clade' (Fig. 1, Node 6) which occupies a long branch and has strong support (MLBP=100), however, because the other internal nodes of the clade do not receive support, we refrain from making this distinction now as it would result in a paraphyletic *Ophiocordyceps*. These analyses place one species of *Stilbella*, *S. buquetii,* in this clade, while other studies (Seifert 1985, Gräfenhan *et al.* 2011) have placed other *Stilbella* species in *Nectriaceae*, *Bionectriaceae*, or *Polycephalomyces*, and the current placement of *Stilbella* remains *Hypocreales incertae sedis* (Kirk *et al.* 2008). The type of *Stilbella*, a coprophile, has yet to be considered in a phylogenetic context, and for these reasons we do not address that name here, but reject the use of that name for this clade. Therefore, we propose to protect *Ophiocordyceps* as the genus name for the entire clade, while acknowledging that future studies including more data and taxonomic sampling may provide better resolution of the relationships within the genus and a narrower concept of *Ophiocordyceps*.

***Tolypocladium* W. Gams 1971**

Tolypocladium is proposed for protection over the other two generic names in the clade, *Elaphocordyceps* and *Chaunopycnis*. The clade itself is well supported (MLBP=97) in this and other published analyses (Sung *et al.* 2007, Kepler *et al.* 2013). However, relationships between species in this clade are very sensitive to taxon sampling, and there

is little bootstrap support for internal branches from the current data to justify more than one name for this clade. The asexual-sexual morph connection between *Tolypocladium* and some *Elaphocordyceps* species has been known for several years (Hodge *et al.* 1996), although where known most *Elaphocordyceps* spp. do not possess the morphology associated with *Tolypocladium* (Sung *et al.* 2007). While this may cause some short-term confusion, the alternative would be to name the clade *Elaphocordyceps* (which would cause the fewest name changes, 12 *vs*. 26 for *Tolypocladium*) and suppress *Tolypocladium*, a much more widely known, medicinally important, and older name, and therefore we find this a poor option. In this analysis the *Chaunopycnis* species sampled form a monophyletic clade which is the most divergent group within the clade. However, this may be the result of limited taxon and genetic sampling; only small subunit rDNA data for the sampled *Chaunopycnis* species were available for these analyses.

Here, we present a list of 26 new combinations within the genus *Tolypocladium*, which we emend to include species whose anamorphic forms do not possess inflated phialide bases, but that do form a single monophyletic clade encompassing a large number of truffle parasites, several insect pathogens, rotifer pathogens, and several fungi isolated to date only from soil.

Tolypocladium W. Gams, *Persoonia* **6**: 185 (1971)

Synonyms: *Chaunopycnis* W. Gams, *Persoonia* **11**: 75 (1980).

Elaphocordyceps G.H. Sung & Spatafora, *Stud. Mycol.* **57**: 36 (2007).

Circumscription: The genus *Tolypocladium* is emended here to apply to all descendants of the node defined in the reference phylogeny (Fig. 1) as the terminal *Tolypocladium* clade. It is the least inclusive clade containing *T. album, T. capitatum, T. cylindrosporum, T. fractum, T. inflatum, T. japonicum, T. longisegmentum, T. ophioglossoides,* and *T. pustulatum.* No definitive synapomorphies are known for the clade. Morphologies associated with sexual reproductive states include robust stipitate stroma with clavate to capitate clava (*e.g.*,*T. capitatum*) to highly reduced stroma comprising rhizomorphs and aggregated perithecia (*e.g.*, *T. inflatum*); perithecia may be immersed and ordinal to the long axis of the stroma or superficial and produced on a highly reduced stromatic pad; asci are single-walled, long and cylindrical with a pronounced apical cap; ascospores are filiform, approximately as long as asci, septate and typically disarticulate into part-spores. Where known, asexual states include morphologies described as *Tolypocladium sensu* Gams (1970), *Chaunopycnis sensu* Gams (1979), or verticillium-like. Ecologies include parasites and pathogens of insects, rotifers and fungi, as well as, soil-inhabiting.

Type: *Tolypocladium inflatum* W. Gams 1971.

Tolypocladium inflatum W. Gams, *Persoonia* **6**: 185 (1971), *nom. cons*.

Synonyms: *Cordyceps subsessilis* Petch, *Trans. Brit. Mycol. Soc*. **21**: 39 (1937).

Elaphocordyceps subsessilis (Petch) G.H. Sung *et al., Stud. Mycol*. **57**: 37 (2007).

Cordyceps facis Kobayasi & Shimizu, *Trans. Mycol. Soc. Japan* **23**: 361 (1982); as '*Codyceps*'.

Tolypocladium album (W. Gams) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808698

Basionym: *Chaunopycnis alba* W. Gams, *Persoonia* **11**: 75 (1979).

Tolypocladium capitatum (Holmsk. : Fr.) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB 808699.

Basionym: Clavaria capitata Holmsk., *Beata Ruris Otia Fung. Dan.* **1**: 38 (1790).

Synonyms: *Sphaeria capitata* (Holmsk. : Fr.) Pers., *Comm. Fung. Clav*.: **13** (1797): Fr., *Syst. Mycol*. **2**: 324 (1822).

Cordyceps capitata (Holmsk.: Fr.) Link, *Handb. Erk. Gew.* **3**: 347 (1833).

Torrubia capitata (Holmsk. : Fr.) Tul. & C. Tul., *Sel. Fung. Carpol*. **3**: 22 (1865).

Elaphocordyceps capitata (Holmsk. : Fr.) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Sphaeria agariciformis Bolton, *Hist. Fung. Halifax*: 130 (1789).

Cordyceps agariciformis (Bolton) Seaver, *N. Amer. Fl*. **3**: 53 (1910).

Cordyceps canadensis Ellis & Everh., *Bull. Torrey Bot. Club* **25**: 501 (1898).

Cordyceps capitata var. *canadensis* (Ellis & Everh.) Lloyd, *Mycol. Writ*. **5**: 609 (1916).

Cordyceps nigriceps Peck, *Bull. Torrey Bot. Club* **27**: 21 (1900).

Tolypocladium delicatistipitatum (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank:MB808700

Basionym: *Cordyceps delicatistipitata* Kobayasi ,*Bull. Natn. Sci. Mus., Tokyo* **5** (2, no. 47): 79 (1960); as '*delicatostipitata'*.

Synonym: Elaphocordyceps delicatistipitata (Kobayasi) G.H. Sung *et al., Stud. Mycol*. **57**: 37 (2007).

Tolypocladium fractum (Mains) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808701

Basionym: *Cordyceps fracta* Mains, *Bull. Torrey. Bot. Club* 84: 250 (1957).

Synonym: *Elaphocordyceps fracta* (Mains) G.H. Sung, *et al.*, *Stud. Mycol*. 57: 37 (2007).

Tolypocladium inegoense (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808702

Basionym: *Cordyceps inegoensis* Kobayasi, *Bull. Natn. Sci. Mus*., *Tokyo* **6**: 292 (1963

Synonyms: *Elaphocordyceps inegoensis* (Kobayasi) G.H. Sung *et al., Stud. Mycol*. **57**: 37 (2007); as '*inegoënsis'*.

Tolypocladium intermedium (S. Imai) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808703

Basionym: *Cordyceps intermedia* S. Imai, *Proc. Imp. Acad. Japan* **10**: 677 (1934).

Synonyms: *Elaphocordyceps intermedia* (S. Imai) G.H. Sung *et al., Stud. Mycol*. **57**: 37 (2007).

Tolypocladium intermedium* f. *michinokuense (Kobayasi & Shimizu) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808704

Basionym: *Cordyceps intermedia* f. *michinokuensis* Kobayasi & Shimizu, *Bull. Natn. Sci. Mus*., *Tokyo*, B **8**: 116 (1982).

Synonym: *Elaphocordyceps intermedia* f. *michinokuensis* (Kobayasi & Shimizu) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007); as '*michinokuënsis'*.

Tolypocladium japonicum (Lloyd) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808705

Basionym: *Cordyceps japonica* Lloyd, *Mycol. Writ*. **6** (Letter 62): 913 (1920).

Synonyms: *Elaphocordyceps japonica* (Lloyd) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Cordyceps umemurae S. Imai, *Trans. Sapporo Nat. Hist. Soc*. **11**: 32 (1930) [1929]; as '*umemurai'*.

Tolypocladium jezoense (S. Imai) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808706

Basionym:*Cordyceps jezoensis* S. Imai, *Trans. Sapporo Nat. Hist. Soc*. **11**: 33 (1930) [1929]. *Synonym*: *Elaphocordyceps jezoensis* (S. Imai) G.H. Sung *et al.*, *Stud. Mycol.* **57**: 37 (2007);

as '*jezoënsis'*.

Tolypocladium longisegmentum (Ginns) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808856

Basionym: *Cordyceps longisegmentis* Ginns, *Mycologia* **80**: 219 (1988).

Synonym:Elaphocordyceps longisegmentis (Ginns) G.H. Sung*et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium minazukiense (Kobayasi & Shimizu) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808857

Basionym: *Cordyceps minazukiensis* Kobayasi & Shimizu, *Bull. Natn. Sci. Mus*., *Tokyo*, B **8**: 117 (1982).

Synonym: *Elaphocordyceps minazukiensis* (Kobayasi & Shimizu) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (1982).

Tolypocladium miomoteanum (Kobayasi & Shimizu) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808858

Basionym: *Cordyceps miomoteana* Kobayasi & Shimizu, *Bull. Natn. Sci. Mus*., *Tokyo*, B **8**: 118 (1982).

Synonym: *Elaphocordyceps miomoteana* (Kobayasi & Shimizu) G.H. Sung*et al.*, *Stud. Mycol*. **57**: 37 (1982).

Tolypocladium ophioglossoides (Ehrh. ex J.F. Gmel.) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808859

Basionym: *Sphaeria ophioglossoides* Ehrh. ex J.F. Gmel., *Syst. Na.*, 13th edn **2**: 1474 (1792).

Synonyms: *Sphaeria ophioglossoides* Ehrh., *Pl. Crypt. Exs.* fasc. 16 no. 160 (1789); nom. inval. (Art. 38.1).

Cordyceps ophioglossoides (Ehrh. ex G.F. Gmel.) Link, *Handb. Erk. Gew.* **3**: 347 (1833) : Fr., *Syst. Mycol.* **2**: 324 (1822).

Torrubia ophioglossoides (Ehrh. ex G.F. Gmel.) Tul. & C. Tul., *Sel.. Fung. Carp.* **3**: 20 (1865). *Elaphocordyceps ophioglossoides* (Ehrh. ex G.F. Gmel.) G.H. Sung *et al.*, *Stud. Mycol.* **57**: 37 (2007).

Clavaria parasitica Willd., *Fl. Berol. Prodr.*: 405 (1787). *Cordyceps parasitica* (Willd.) Henn., *Nerthus* **6**: 4 (1904).

Tolypocladium ophioglossoides* f. *album (Kobayasi & Shimizu ex Y.J. Yao) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808860

Basionym: *Cordyceps ophioglossoides* f. *alba* Kobayasi & Shimizu ex Y.J. Yao, *Acta Mycol. Sin*. **14**: 257 (1995).

Synonym: *Elaphocordyceps ophioglossoides* f. *alba* (Kobayasi & Shimizu ex Y.J. Yao) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium ophioglossoides* f. *cuboides (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808861

Basionym: *Cordyceps ophioglossoides* f. *cuboides* Kobayasi, *Bull. Natn. Sci. Mus*., *Tokyo* **5** (2, no. 47): 77 (1960).

Synonym: *Elaphocordyceps ophioglossoides* f. *cuboides* (Kobayasi) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium ovalisporum (C. Möller & W. Gams) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808862

Basionym: *Chaunopycnis ovalispora* C. Möller & W. Gams, *Mycotaxon* **48**: 442 (1993).

Tolypocladium paradoxum (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808863

Basionym:*Cordyceps paradoxa* Kobayasi, *Bulletin of the Biogeogr. Soc. Jap*. **9**: 156 (1939). *Synonym*: *Elaphocordyceps paradoxa* (Kobayasi) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium pustulatum (Bills *et al.*) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808864

Basionym:*Chaunopycnis pustulata* Bills *et al.*, *Mycol. Progr*. **1**: 8 (2002).

Tolypocladium ramosum (Teng) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808865

Basionym: *Cordyceps ramosa* Teng, *Sinensia* **7**: 810 (1936).

Synonym: *Elaphocordyceps ramosa* (Teng) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium rouxii (Cand.) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808866

Basionym: *Cordyceps rouxii* Cand., *Mycotaxon* **4**: 544 (1976).

Synonym: *Elaphocordyceps rouxii* (Cand.) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 37 (2007).

Tolypocladium szemaoense (M. Zang) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808867

Basionym:*Cordyceps szemaoensis* M. Zang, *Acta Bot. Yunn*. **23**: 295 (2001).

Synonym: *Elaphocordyceps szemaoensis* (M. Zang) G.H. Sung*et al., Stud. Mycol*. **57**: 38 (2007); as '*szemaoënsis'*.

Tolypocladium tenuisporum (Mains) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808868

Basionym: *Cordyceps tenuispora* Mains, *Bull. Torrey Bot. Club* **84**: 247 (1957).

Synonym: *Elaphocordyceps tenuispora* (Mains) G.H. Sung *et al., Stud. Mycol*. **57**: 38 (2007).

Tolypocladium toriharamontanum (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808869

Basionym: *Cordyceps toriharamontana* Kobayasi, *Bull. Natn. Sci. Mus., Tokyo* **6**: 305 (1963).

Synonym: *Elaphocordyceps toriharamontana* (Kobayasi) G.H. Sung*et al., Stud. Mycol*. **57**: 38 (2007).

Tolypocladium valliforme (Mains) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808870

Basionym: *Cordyceps valliformis* Mains, *Bull. Torrey Bot. Club* **84**: 250 (1957).

Synonym: *Elaphocordyceps valliformis* (Mains) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 38 (2007).

Tolypocladium valvatistipitatum (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov**.

MycoBank: MB808871

Basionym: *Cordyceps valvatistipitata* Kobayasi, *Bull. Natn. Sci. Mus*., *Tokyo* **5(**2, no. 47): 81 (1960); as '*valvatostipitata'*].

Synonym: *Elaphocordyceps valvatistipitata* (Kobayasi) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 38 (2007).

Tolypocladium virens (Kobayasi) Quandt, Kepler & Spatafora, **comb. nov.**

MycoBank: MB808872

Basionym: *Cordyceps virens* Kobayasi, *J. Jap. Bot*. **58**: 222 (1983).

Synonym: *Elaphocordyceps virens* (Kobayasi) G.H. Sung *et al.*, *Stud. Mycol*. **57**: 38 (2007).

Purpureocillium **Luangsa-ard *et al.* 2011**

Our findings support those reported by Luangsa-ard *et al.* (2011b) for the *Purpureocillium* clade, and the change in Art. 59 allows for the inclusion of *N. atypicola*

(syn. *Cordyceps cylindrica*) and *Isaria takamizusanensis* within this genus. Shared characters for this clade include purple-hued conidia and pathogenesis of arthropods, although *P. lilacinum* and *P. lavendulum* have been cultured from various substrates (Perdomo *et al* .2013), and *P. lilacinum* can cause keratitis and other mycoses in humans and other vertebrates (Pastor & Guarro 2006, Rodríguez *et al.* 2010). Because this genus is well supported (MLBP=76) as sister to the nematode pathogen clade (Fig. 1), it is important to mention that *P. lilacinum* is frequently collected from nematodes (Luangsa-ard *et al.* 2011b), and has been used in the biocontrol of plant pathogenic nematodes (Kalele *et al.* 2006, Castillo *et al.* 2013).

***Harposporium* Lohde 1874 and *Drechmeria* W. Gams & H.-B. Jansson 1985**

Our analyses reconstruct a well-supported (MLBP=76) monophyletic origin of the mostly nematophagous clade of *Ophiocordycipitaceae* (Fig. 1 Node 2). Within this clade, there is strong phylogenetic support for two clades: one containing *Harposporium* and *Podocrella*, and the other consisting of *Drechmeria*, *Haptocillium*, and *Cordyceps gunnii*. The relationship between *Harposporium* and *Podocrella* has already been described (Chaverri *et al.* 2005), but the revision of Art. 59 requires that one name be chosen for this genus. *Harposporium* is an older name, and the morphology of at least somewhat crescent-shaped conidia is a shared character for this clade. Suppression of *Podocrella* also requires the fewest taxonomic revisions (three *vs* 30). For these reasons, we propose to protect *Harposporium* over *Podocrella* (Table 2.2).

Within the other nematophagous subclade, *Drechmeria* is an older name than *Haptocillium*, and the isolate included in these analyses is nested within the *Haptocillium*

isolates sampled. For this reason, we propose to protect *Drechmeria* over *Haptocillium*. The inclusion of *C. gunnii* in this clade also provides a name for this residual taxon of *Cordyceps*. Most species however, are nematophagous (*C. gunnii* being the exception), and conidia may be cone-shaped, formed on conidiogenous cells in rosettes or verticils, or in the case of *C. gunnii*, paecilomyces-like. We did not have access to molecular data from *D. harposporioides*, but given our finding that the two nematophagous clades in *Ophiocordycipitaceae* are monophyletic in origin, it will be interesting to see if this species, a protozoan pathogen with helical conidia, is truly a member of the *Drechmeria* clade or in fact a species within *Harposporium* that simply lacks the basally swollen conidiogenous cells.

***Polycephalomyces* Kobayasi 1941**

This study is the first to have definitive ML support (MLBP=82) for the sister relationship between the *Polycephalomyces* clade and *Ophiocordycipitaceae* (Fig. 1 Node 1). Support for this relationship remains even with the exclusion of *C. pleuricapitata*, which is on an early-diverging, long branch within the clade. Two options remain to deal with this finding. Either a new family must be erected to account for this clade, or *Polycephalomyce*s and related taxa must be moved into *Ophiocordycipitaceae*. We propose to accept *Polycephalomyce*s and *C. pleuricapitata* in *Ophiocordycipitaceae*, where it will be the earliest diverging lineage of the family. The taxonomy of *C. pleuricapitata* will be addressed elsewhere.

Conclusions

We present a concise, thorough, phylogenetically relevant, and taxonomically accurate revision of the family *Ophiocordycipitaceae* with the aim of complying with the changes to Art. 59 of the ICN. With the criteria of naming monophyletic taxa, and where possible, of adhering to priority while avoiding changes that would be disruptive to the wider community of researchers, we have proposed to protect six genera within *Ophiocordycipitaceae*, including incorporation of the genus *Polycephalomyces* within the family. We have also formally revised the genus *Tolypocladium*, to reflect the nomenclature suggested by our results.

Acknowledgements

We would like to acknowledge the following individuals for their contribution to this manuscript: Priscila Chaverri, Mingjun Chen, Tsuyoshi Hosoya, Jae-Mo Sung, and James White. An NSF Graduate Research Fellowship supported CAQ.

References

Barron GL (1977) *The Nematode-destroying Fungi*. [Topics in Mycobiology no. 1.] Guelph, ON: Canadian Biological Publications.

Barron GL (1980) Fungal parasites of rotifers: a new *Tolypocladium* with underwater conidiation. *Canadian Journal of Botany* **58**: 439–442.

Barron GL (1981) Two new fungal parasites of bdelloid rotifers. *Canadian Journal of Botany* **59**: 1449–1455.

Barron GL (1983) Structure and biology of a new *Tolypocladium* attacking bdelloid rotifers. *Canadian Journal of Botany* **61**: 2566–2569.

Barron GL, Szijarto E (1982) A new hyphomycete parasitic on the ciliated protozoans *Vorticella* and *Opercularia*. *Canadian Journal of Botany* **60**: 1031–1034.

Bills GF, Polishook JD, Goetz MA, Sullivan RF, White Jr. JF (2002) *Chaunopycnis pustulata* sp. nov., a new clavicipitalean anamorph producing metabolites that modulate potassium ion channels. *Mycological Progress* **1**: 3–17.

Bissett J (1983) Notes on *Tolypocladium* and related genera. *Canadian Journal of Botany* **61**: 1311–1329.

Boonmee S, Zhang Y, Chomnunti P, Chukeatirote E, Tsui CKM, Bahkali AH, Hyde KD (2011) Revision of lignicolous *Tubeufiaceae* based on morphological reexamination and phylogenetic analysis. *Fungal Diversity* **51**: 63–102.

Bushley KE, Li Y, Wang WJ, Wang XL, Jiao L, Spatafora JW, Yao YJ (2013a). Isolation of the MAT1-1 mating type idiomorph and evidence for selfing in the Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biology* **117**: 599–610.

Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, Owensby CA, Knaus BJ, Elser J, Miller D, Di Y, McPhail KL, Spatafora JW (2013b) Draft genome sequence of the Cyclosporin producing fungus *Tolypocladium inflatum* reveals complex patterns of secondary metabolite evolution and expression. *PLoS Genetics* **9**: e1003496.

Cannon PF, Hywel-Jones NL, Maczey N, Norbu L, Samdup T, Lhendup P (2009) Steps towards sustainable harvest of *Ophiocordyceps sinensis* in Bhutan. *Biodiversity and Conservation* **18**: 2263–2281.

Castillo JD, Lawrence KS, Kloepper JW (2013) Biocontrol of the reniform nematode by *Bacillus firmus* GB-126 and *Paecilomyces lilacinus* 251 on cotton. *Plant Disease* **97**: 967–976.

Chaverri P, Samuels GJ, Hodge KT (2005) The genus *Podocrella* and its nematode-killing anamorph *Harposporium. Mycologia* **97**: 433–443.

Drechsler C (1941) Some hyphomycetes parasitic on free-living terricolous nematodes. *Phytopathology* **31**: 773–802.

Dreyfuss M, Gams W (1994) Proposal to reject *Pachybasium niveum* Rostr. in order to retain the name *Tolypocladium inflatum* W. Gams for the fungus that produces cyclosporin. *Taxon* **43**: 660–661.

Evans HC, Samson RA (1982) *Cordyceps* species and their anamorphs pathogenic on ants (*Formicidae*) in tropical forest ecosystems 1.The *Cephalotes* (*Myrmicinae*) complex. *Transactions of the British Mycological Society* **79**: 431–453.

Evans HC, Shah PA (2002) Taxonomic status of the genera *Sorosporella* and *Syngliocladium* associated with grasshoppers and locusts (*Orthoptera*: *Acridoidea*) in Africa. *Mycological Research* **106**: 737–744.

Evans HC, Whitehead PF (2005) Entomogenous fungi of arboreal *Coleoptera* from Worcestershire, England, including the new species *Harposporium bredonense*. *Mycological Progress* **4**: 91–99.

Evans HC, Groden E, Bischoff JF (2010) New fungal pathogens of the red ant *Myrmica rubra* from the UK and implications for ant invasions in the USA. *Fungal Biology* **114**: 451–466.

Gams W (1971) *Tolypocladium*, eine Hyphomycetengattung mit geschwollenen Phialiden. *Persoonia* **6**: 185–191.

Gams W (1980) *Chaunopycnis alba*, gen. et sp. nov., a soil fungus intermediate between *Moniliales* and *Sphaeropsidales*. *Persoonia* **11**: 75–79.

Gams W, Jansson H-B (1985) The nematode parasite *Meria coniospora* Drechsler in pure culture and its classification. *Mycotaxon* **22**: 33–38.

Gams W, Hodge KT, Samson RA, Korf RP, Seifert KA (2005) (1684) Proposal to conserve the name *Isaria* (anamorphic fungi) with a conserved type. *Taxon* **54**: 537.

Gams W, Zare R (2001) A revision of *Verticillium* section *Prostrata*. III. Generic classification. *Nova Hedwigia* 72: 329–337.

Gams W, Zare R (2003) A taxonomic review of the clavicipitaceous anamorphs parasitizing nematodes and other microinvertebrates. In: *Clavicipitalean Fungi: evolutionary biology, chemistry, biocontrol and cultural Impacts* (White JF jr., Bacon CW, Hywel-Jones NL, Spatafora JW, eds): 17-73. New York: Marcel Dekker.

Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MB, *et al.*(2013) One Fungus, One Name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400–408.

Gräfenhan T, Schroers H-J, Nirenberg HI, Seifert KA (2011) An overview of the taxonomy, phylogeny, and typification of nectriaceous fungi in *Cosmospora*, *Acremonium*, *Fusarium*, *Stilbella*, and *Volutella*. *Studies in Mycology* **68**: 79–113.

Hibbett DS, Taylor JW (2013) Fungal systematics: is a new age of enlightenment at hand? *Nature Reviews Microbiology* **11**: 129–133.

Hodge KT, Gams W, Samson RA, Korf RP, Seifert KA (2005) Lectotypification and status of *Isaria* Pers. : Fr. *Taxon* **54**: 485–489.

Hodge KT, Humber RA, Wozniak CA (1998) *Cordyceps variabilis* and the genus *Synglíocladium*. *Mycologia* **90**: 743–753.

Hodge KT, Krasnoff SB, Humber RA (1996) *Tolypocladium inflatum* is the anamorph of *Cordyceps subsessilis*. *Mycologia* **88**: 715–719.

Hodge KT, Viaene NM, Gams W (1997) Two *Harposporium* species with *Hirsutella* synanamorphs. *Mycological Research* **101**: 1377–1382.

Hu X, Zhang Y, Xiao G, Zheng P, Xia Y, Zhang X, St. Leger RJ, Liu X, Wang C (2013) Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chinese Science Bulletin* **58**: 2846–2854.

Hughes KW, Petersen RH, Johnson JE, Moncalvo JM, Vilgalys R, Redhead SA, Thomas T, McGhee LL (2001) Infragenic phylogeny of *Collybia s. str.* based on sequences of ribosomal ITS and LSU regions. *Mycological Research* **105**: 164–172.

Hywel-Jones NL, Sivichai S (1995) *Cordyceps cylindrica* and its association with *Nomuraea atypicola* in Thailand. *Mycological Research* **99**: 809–812.

Kalele DN, Affokpon A, Coosemans J (2006) Efficacy of *Paecilomyces lilacinus* strain 251 against root knot nematodes in tomato under greenhouse conditions. *Communications in Agricultural and Applied Biological Sciences* **72**: 209–213.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**: 286–298.

Kepler RM, Ban S, Nakagiri A, Bischoff J, Hywel-Jones N, Owensby CA, Spatafora JW (2013) The phylogenetic placement of hypocrealean insect pathogens in the genus *Polycephalomyces*: an application of One Fungus One Name. *Fungal Biology* **117**: 611–622.

Kepler RM, Humber RA, Bischoff JF, Rehner SA (2014) Clarification of generic and species boundaries for *Metarhizium* and related fungi through multigene phylogenetics. *Mycologia* doi:10.3852/13-319

Kirk PM, Cannon PF, Minter DW, Stalpers JA (eds) (2008) *Ainsworth & Bisby's Dictionary of the Fungi.* 10th edn. Wallingford: CAB International.

Kobayasi Y (1941) The genus *Cordyceps* and its allies. *Science Reports of the Tokyo Bunrika Daigaku*, sect. B, **84** (5**)**: 53–260.

Leuchtmann A, Bacon CW, Schardl CL, White JF jr, Tadych M (2014) Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia* doi:10.3852/13-251.

Li X, Luo H, Zhang K (2005) A new species of *Harposporium* parasitic on nematodes. *Canadian Journal of Botany* **83**: 558–562.

Luangsa-ard JJ, Houbraken J, van Doorn T, Hong S-B, Borman AM, Hywel-Jones NL, Samson RA (2011b) *Purpureocillium*, a new genus for the medically important *Paecilomyces lilacinus*. *FEMS Microbiology Letters* **321**: 141–149.

Luangsa-ard JJ, Ridkaew R, Tasanathai K, Thanakitpipattana D, Hywel-Jones NL (2011a) *Ophiocordyceps halabalaensis*: a new species of *Ophiocordyceps* pathogenic to *Camponotus gigas* in Hala Bala Wildlife Sanctuary, Southern Thailand. *Fungal Biology* **115**: 608–614.

Mains EB (1950) Entomogenous species of *Akanthomyces*, *Hymenostilbe* and *Insecticola* in North America. *Mycologia* **42**: 566–588.

McNeill J (2012) Guidelines for requests for binding decisions on application of the Code. *Taxon* **61**: 477-478.

Möller C, Gams W (1993) Two new hyphomycetes isolated from Antarctic lichens. *Mycotaxon* **48**: 441–450.

Pastor FJ, Guarro J (2006) Clinical manifestations, treatment and outcome of *Paecilomyces lilacinus* infections. *Clinical Microbiology and Infection* **12**: 948-960.

Patouillard NT (1892) Une Clavariée entomogène. *Revue Mycologique* **14**: 67–70.

Perdomo H, Cano J, Gené J, García D, Hernández M, Guarro J (2013) Polyphasic analysis of *Purpureocillium lilacinum* isolates from different origins and proposal of the new species *Purpureocillium lavendulum*. *Mycologia* **105**: 151–161.

Petch T (1931a) Notes on entomogenous fungi. *Transactions of the British Mycological Society* **16**: 55–75.

Petch T (1931b) New species of *Cordyceps*, collected during the Whitby foray. *The Naturalist, Hull* **1931**: 101–103.

Petch T (1932) A list of the entomogenous fungi of Great Britain. *Transactions of the British Mycological Society* **17**: 170–178.

Petch T (1933) Notes on entomogenous fungi. *Transactions of the British Mycological Society* **18**: 48–75.

Ren SY, Yao YJ (2013) Evaluation of nutritional and physical stress conditions during vegetative growth on conidial production and germination in *Ophiocordyceps sinensis*. *FEMS Microbiology Letters* **346**: 29–35.

Rodríguez MM, Pastor FJ, Serena C, Guarro J (2010) Efficacy of voriconazole in a murine model of invasive paecilomycosis. *International Journal of Antimicrobial Agents* **35**: 362–365.

Rossman AY (1978) *Podonectria*, a genus in the *Pleosporales* on scale insects. *Mycotaxon* **7**: 163–182.

Rossman AY, Seifert KA, Samuels GJ, Minnis AM, Schroers H-J, Lombard L, Crous PW, Põldmaa K, Cannon PF, Summerbell RC, Geiser DM, Zhuang W, Hirooka Y, Herrera C, Salgado-Salazar C, Chaverri P (2013) Genera in *Bionectriaceae*, *Hypocreaceae*, and *Nectriaceae* (*Hypocreales*) proposed for acceptance or rejection. *IMA Fungus* **4**: 41–51.

Samson RA, Brady BL (1983) *Paraisaria*, a new genus for *Isaria dubia*, the anamorph of *Cordyceps gracilis*. *Transactions of the British Mycological Society* **81**: 285–290.

Samson RA, Soares GG jr (1984) Entomopathogenic species of the hyphomycete genus *Tolypocladium*. *Journal of Invertebrate Pathology* **43**: 133–139.

Seifert KA (1985) A monograph of *Stilbella* and some allied hyphomycetes. *Studies in Mycology* **27**: 1–234.

Shimazu M, Glockling SL (1997) A new species of *Harposporium* with two spore types isolated from the larva of a cerambycid beetle. *Mycological Research* **101**: 1371–1376.

Shrestha B, Bawa KS (2013) Trade, harvest, and conservation of caterpillar fungus (*Ophiocordyceps sinensis*) in the Himalayas. *Biological Conservation* **159**: 514–520.

Shrestha B, Zhang WM, Zhang YJ, Liu XZ (2010) What is the Chinese caterpillar fungus *Ophiocordyceps sinensis* (*Ophiocordycipitaceae*)? *Mycology* **1**: 228–236.

Speare AT (1917) *Sorosporella uvella* and its occurrence in cutworms in America. *Journal of Agricultural Research* **8**: 189–194.

Speare AT (1920) Further studies of *Sorosporella uvella*, a fungous parasite of noctuid larvae. *Journal of Agricultural Research* **18**: 399–439.

Stalpers JA, Seifert KA, Samson RA (1991) A revision of the genera *Antromycopsis*, *Sclerostilbum*, and *Tilachlidiopsis* (hyphomycetes). *Canadian Journal of Botany* **69**: 6–15.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stone R (2008) Last stand for the body snatcher of the Himalayas? *Science* 322: 1182.

Sung G-H, Spatafora JW, Zare R, Hodge KT, Gams W (2001) A revision of *Verticillium* sect. *Prostrata*. II. Phylogenetic analyses of SSU and LSU nuclear rDNA sequences from anamorphs and teleomorphs of the *Clavicipitaceae*. *Nova Hedwigia* **72**: 311–328.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5–59.

Survase SA, Kagliwal LD, Annapure US, Singhal RS (2011) Cyclosporin A - A review on fermentative production, downstream processing and pharmacological applications. *Biotechnology Advances* **29**: 418–435.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564–577.

Traber R, Dreyfuss MM (1996) Occurrence of cyclosporins and cyclosporin-like peptolides in fungi. *Journal of Industrial Microbiology* **17**: 397–401.

Vuillemin P (1896) Les Hypostomacées, nouvelle famille de champignons parasites. *Bulletin de la Société des Sciences de Nancy*, sér. 2 **14**: 15–67.

Weiser J, Matha V, Jegorov A (1991) *Tolypocladium terricola* sp. n., a new mosquito-killing species of the genus *Tolypocladium* Gams (hyphomycetes). *Folia Parasitologica* **38**: 363–369.

Zare R, Gams W, Culham A (2000) A revision of *Verticillium* sect. *Prostrata*. I. Phylogenetic studies using ITS sequences. *Nova Hedwigia* **71**: 465–480.

Zare R, Gams W (2001) A revision of *Verticillium* section *Prostrata*. VI. The genus *Haptocillium*. *Nova Hedwigia* **73**: 271–292.

Figure 2.1 ML phylogeny of *Ophiocordycipitaceae* obtained using RAxML to analyze the concatenated five gene dataset (SSU, LSU, TEF, RPB1, and RPB2). Proposed genus level names to protect are delimited, but names of individual species have not been changed on the leaves of the tree, to demonstrate the diversity of taxa sampled. Values above branches represent MLBP proportions greater than or equal to 70 % from 500 replicates.

Figure 2.1

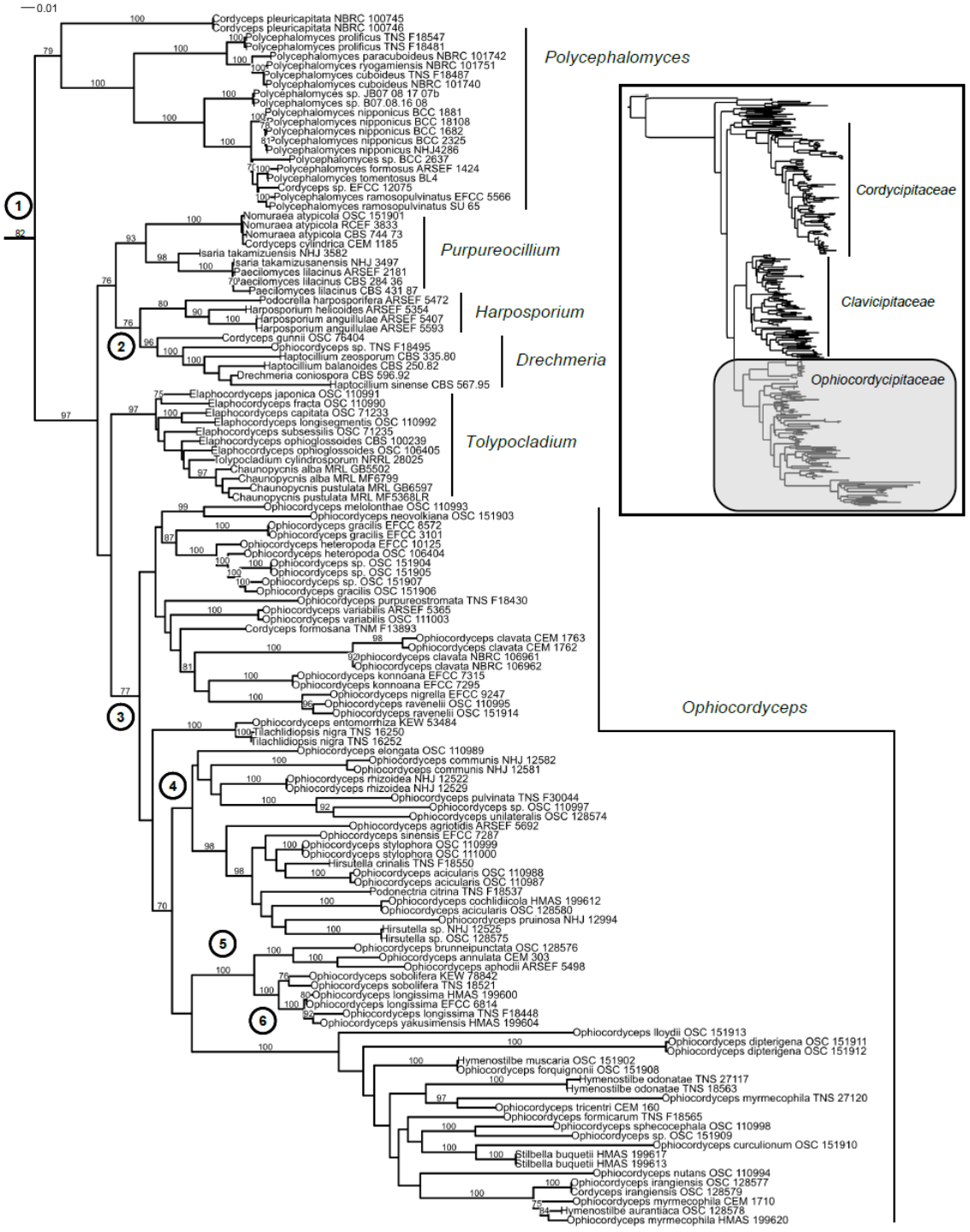

0.01
Cordyceps pleuricapitata NBRC 100745
Cordyceps pleuricapitata NBRC 100746
Polycephalomyces prolificus TNS F18547
Polycephalomyces prolificus TNS F18481
Polycephalomyces paracuboideus NBRC 101742
Polycephalomyces ryogamiensis NBRC 101751
Polycephalomyces cuboideus TNS F18487
Polycephalomyces cuboideus NBRC 101740
Polycephalomyces sp. JB07 08 17 07b
Polycephalomyces sp. B07.08.16 08
Polycephalomyces nipponicus BCC 1881
Polycephalomyces nipponicus BCC 18108
Polycephalomyces nipponicus BCC 1682
Polycephalomyces nipponicus BCC 2325
Polycephalomyces nipponicus NHJ4286
Polycephalomyces sp. BCC 2637
Polycephalomyces formosus ARSEF 1424
Polycephalomyces tomentosus BL4
Cordyceps sp. EFCC 12075
Polycephalomyces ramosopulvinatus EFCC 5566
Polycephalomyces ramosopulvinatus SU 65
Polycephalomyces
Nomuraea atypicola OSC 151901
Nomuraea atypicola RCEF 3833
Nomuraea atypicola CBS 744 73
Cordyceps cylindrica CEM 1185
Isaria takamizuensis NHJ 3582
Isaria takamizusanensis NHJ 3497
Paecilomyces lilacinus ARSEF 2181
Paecilomyces lilacinus CBS 284 36
Paecilomyces lilacinus CBS 431 87
Purpureocillium
Podocrella harposporifera ARSEF 5472
Harposporium helicoides ARSEF 5354
Harposporium anguillulae ARSEF 5407
Harposporium anguillulae ARSEF 5593
Harposporium
Cordyceps gunnii OSC 76404
Ophiocordyceps sp. TNS F18495
Haptocillium zeosporum CBS 335.80
Haptocillium balanoides CBS 250.82
Drechmeria coniospora CBS 596.92
Haptocillium sinense CBS 567.95
Drechmeria
Elaphocordyceps japonica OSC 110991
Elaphocordyceps fracta OSC 110990
Elaphocordyceps capitata OSC 71233
Elaphocordyceps longisegmentis OSC 110992
Elaphocordyceps subsessilis OSC 71235
Elaphocordyceps ophioglossoides CBS 100239
Elaphocordyceps ophioglossoides OSC 106405
Tolypocladium cylindrosporum NRRL 28025
Chaunopycnis alba MRL GB5502
Chaunopycnis alba MRL MF6799
Chaunopycnis pustulata MRL GB6597
Chaunopycnis pustulata MRL MF5368LR
Tolypocladium
Ophiocordyceps melolonthae OSC 110993
Ophiocordyceps neovolkiana OSC 151903
Ophiocordyceps gracilis EFCC 8572
Ophiocordyceps gracilis EFCC 3101
Ophiocordyceps heteropoda EFCC 10125
Ophiocordyceps heteropoda OSC 106404
Ophiocordyceps sp. OSC 151904
Ophiocordyceps sp. OSC 151905
Ophiocordyceps sp. OSC 151907
Ophiocordyceps gracilis OSC 151906
Ophiocordyceps purpureostromata TNS F18430
Ophiocordyceps variabilis ARSEF 5365
Ophiocordyceps variabilis OSC 111003
Cordyceps formosana TNM F13893
Ophiocordyceps clavata CEM 1763
Ophiocordyceps clavata CEM 1762
Ophiocordyceps clavata NBRC 106961
Ophiocordyceps clavata NBRC 106962
Ophiocordyceps konnoana EFCC 7315
Ophiocordyceps konnoana EFCC 7295
Ophiocordyceps nigrella EFCC 9247
Ophiocordyceps ravenelii OSC 110995
Ophiocordyceps ravenelii OSC 151914
Ophiocordyceps entomorrhiza KEW 53484
Tilachlidiopsis nigra TNS 16250
Tilachlidiopsis nigra TNS 16252
Ophiocordyceps elongata OSC 110989
Ophiocordyceps communis NHJ 12582
Ophiocordyceps communis NHJ 12581
Ophiocordyceps rhizoidea NHJ 12522
Ophiocordyceps rhizoidea NHJ 12529
Ophiocordyceps pulvinata TNS F30044
Ophiocordyceps sp. OSC 110997
Ophiocordyceps unilateralis OSC 128574
Ophiocordyceps agriotidis ARSEF 5692
Ophiocordyceps sinensis EFCC 7287
Ophiocordyceps stylophora OSC 110999
Ophiocordyceps stylophora OSC 111000
Hirsutella crinalis TNS F18550
Ophiocordyceps acicularis OSC 110988
Ophiocordyceps acicularis OSC 110987
Podonectria citrina TNS F18537
Ophiocordyceps cochlidiicola HMAS 199612
Ophiocordyceps acicularis OSC 128580
Ophiocordyceps pruinosa NHJ 12994
Hirsutella sp. NHJ 12525
Hirsutella sp. OSC 128575
Ophiocordyceps brunneipunctata OSC 128576
Ophiocordyceps annulata CEM 303
Ophiocordyceps aphodii ARSEF 5498
Ophiocordyceps sobolifera KEW 78942
Ophiocordyceps sobolifera TNS 18521
Ophiocordyceps longissima HMAS 199600
Ophiocordyceps longissima EFCC 6814
Ophiocordyceps longissima TNS F18448
Ophiocordyceps yakusimensis HMAS 199604
Ophiocordyceps lloydii OSC 151913
Ophiocordyceps dipterigena OSC 151911
Ophiocordyceps dipterigena OSC 151912
Hymenostilbe muscaria OSC 151902
Ophiocordyceps forquignonii OSC 151908
Hymenostilbe odonatae TNS 27117
Hymenostilbe odonatae TNS 18563
Ophiocordyceps myrmecophila TNS 27120
Ophiocordyceps tricentri CEM 160
Ophiocordyceps formicarum TNS F18565
Ophiocordyceps sphecocephala OSC 110998
Ophiocordyceps sp. OSC 151909
Ophiocordyceps curculionum OSC 151910
Stilbella buquetii HMAS 199617
Stilbella buquetii HMAS 199613
Ophiocordyceps nutans OSC 110994
Ophiocordyceps irangiensis OSC 128577
Cordyceps irangiensis OSC 128579
Ophiocordyceps myrmecophila CEM 1710
Hymenostilbe aurantiaca OSC 128578
Ophiocordyceps myrmecophila HMAS 199620
Ophiocordyceps
Cordycipitaceae
Clavicipitaceae
Ophiocordycipitaceae
1
2
3
4
5
6

Table 2.1 Specimen information and Genbank accession numbers for sequences used in this study.

Species	Voucher Information	SSU	LSU	TEF	RPB1	RPB2
Chaunopycnis alba	MRL GB5502		AF245297			
	MRL MF6799		AF373284			
Chaunopycnis pustulata	MRL GB6597		AF389190			
	MRL MF5368LR		AF373282			
Cordyceps cylindrica	CEM 1185	KJ878907	KJ878872	KJ878955		
Cordyceps formosana	TNM F13893	KJ878908		KJ878956	KJ878988	KJ878943
Cordyceps gunnii	OSC 76404	AF339572	AF339522	AY489616	AY489650	DQ522426
Cordyceps irangiensis	OSC 128579	EF469123	EF469076	EF469060	EF469089	EF469107
Cordyceps nipponica	BCC 18108	KF049608	KF049626	KF049681	KF049644	
Cordyceps pleuricapitata	NBRC 100745	KF049606	KF049624	KF049679	KF049642	KF049667
Cordyceps pleuricapitata	NBRC 100746	KF049607	KF049625	KF049680	KF049643	KF049668
Cordyceps sp.	EFCC 12075	KJ878909	KJ878873	KJ878957	KJ878989	
Drechmeria coniospora	CBS 596.92	AF106012				
Elaphocordyceps capitata	OSC 71233	AY489689	AY489721	AY489615	AY489649	DQ522421
Elaphocordyceps fracta	OSC 110990	DQ522545	DQ518759	DQ522328	DQ522373	DQ522425
Elaphocordyceps japonica	OSC 110991	DQ522547	DQ518761	DQ522330	DQ522375	DQ522428
Elaphocordyceps longisegmentis	OSC 110992		EF468816		EF468864	EF468919
Elaphocordyceps ophioglossoides	CBS 100239	KJ878910	KJ878874	KJ878958	KJ878990	KJ878944
	OSC 106405	AY489691	AY489723	AY489618	AY489652	DQ522429
Elaphocordyceps subsessilis	OSC 71235	EF469124	EF469077	EF469061	EF469090	EF469108
Haptocillium balanoides	CBS 250.82	AF339588	AF339539	DQ522342	DQ522388	DQ522442
Haptocillium sinense	CBS 567.95	AF339594	AF339545	DQ522343	DQ522389	DQ522443
Haptocillium zeosporum	CBS 335.8	AF339589	AF339540	EF469062	EF469091	EF469109

Table 2.1 (Continued)

Harposporium anguillulae	ARSEF 5407		AY636080			
	ARSEF 5593		AY636081			
Harposporium helicoides	ARSEF 5354	AF339577	AF339527			
Hirsutella crinalis	TNS F18550	KJ878911	KJ878875	KJ878959		
Hirsutella sp.	OSC 128575	EF469126	EF469079	EF469064	EF469093	EF469110
Hirsutella sp.	NHJ 12525	EF469125	EF469078	EF469063	EF469092	EF469111
Hymenostilbe aurantiaca	OSC 128578	DQ522556	DQ518770	DQ522345	DQ522391	DQ522445
Hymenostilbe muscaria	OSC 151902	KJ878912	KJ878876		KJ878991	KJ878945
Hymenostilbe odonatae	TNS F18563		KJ878877		KJ878992	
	TNS F27117		KJ878878			
Isaria takamizuensis	NHJ 3582	EU369097	EU369034	EU369015		
Isaria takamizusanensis	NHJ 3497	EU369096	EU369033	EU369014	EU369053	EU369074
Nomuraea atypicola	RCEF 3833	KJ878913	KJ878879	KJ878960	KJ878993	
	OSC 151901	KJ878914	KJ878880	KJ878961	KJ878994	
	CBS 744.73	EF468987	EF468841	EF468786	EF468892	
Ophiocordyceps acicularis	OSC 110987	EF468950	EF468805	EF468744	EF468852	
	OSC 110988	EF468951	EF468804	EF468745	EF468853	
	OSC 128580	DQ522543	DQ518757	DQ522326	DQ522371	DQ522423
Ophiocordyceps agriotidis	ARSEF 5692	DQ522540	DQ518754	DQ522322	DQ522368	DQ522418
Ophiocordyceps annulata	CEM 303	KJ878915	KJ878881	KJ878962	KJ878995	
Ophiocordyceps aphodii	ARSEF 5498	DQ522541	DQ518755	DQ522323		DQ522419
Ophiocordyceps brunneipunctata	OSC 128576	DQ522542	DQ518756	DQ522324	DQ522369	DQ522420
Ophiocordyceps clavata	CEM 1762	KJ878916	KJ878882	KJ878963	KJ878996	

Table 2.1 (Continued)

	CEM 1763		KJ878883	KJ878964	KJ878997	
	NBRC 106961	JN941727	JN941414		JN992461	
	NBRC 106962	JN941726	JN941415		JN992460	
Ophiocordyceps cochlidiicola	HMAS_199612	KJ878917	KJ878884	KJ878965	KJ878998	
Ophiocordyceps communis	NHJ 12581	EF468973	EF468831	EF468775		EF468930
	NHJ 12582	EF468975	EF468830	EF468771		EF468926
Ophiocordyceps curculionum	OSC 151910	KJ878918	KJ878885		KJ878999	
Ophiocordyceps dipterigena	OSC 151911	KJ878919	KJ878886	KJ878966	KJ879000	
	OSC 151912	KJ878920	KJ878887	KJ878967	KJ879001	
Ophiocordyceps elongata	OSC 110989		EF468808	EF468748	EF468856	
Ophiocordyceps entomorrhiza	KEW 53484	EF468954	EF468809	EF468749	EF468857	EF468911
Ophiocordyceps formicarum	TNS F18565	KJ878921	KJ878888	KJ878968	KJ879002	KJ878946
Ophiocordyceps forquignonii	OSC 151908	KJ878922	KJ878889		KJ879003	KJ878947
Ophiocordyceps gracilis	EFCC 3101	EF468955	EF468810	EF468750	EF468858	EF468913
	EFCC 8572	EF468956	EF468811	EF468751	EF468859	EF468912
	OSC 151906	KJ878923	KJ878890	KJ878969		
Ophiocordyceps heteropoda	EFCC 10125	EF468957	EF468812	EF468752	EF468860	EF468914
	OSC 106404	AY489690	AY489722	AY489617	AY489651	
Ophiocordyceps irangiensis	OSC 128577	DQ522546	DQ518760	DQ522329	DQ522374	DQ522427
Ophiocordyceps konnoana	EFCC 7295	EF468958			EF468862	EF468915
Ophiocordyceps konnoana	EFCC 7315	EF468959		EF468753	EF468861	EF468916
Ophiocordyceps lloydii	OSC 151913	KJ878924	KJ878891	KJ878970	KJ879004	KJ878948
Ophiocordyceps longissima	EFCC 6814		EF468817	EF468757	EF468865	

Table 2.1 (Contiued)

	TNS F18448	KJ878925	KJ878892	KJ878971	KJ879005	
Ophiocordyceps longissima	HMAS_199600	KJ878926		KJ878972	KJ879006	KJ878949
Ophiocordyceps melolonthae	OSC 110993	DQ522548	DQ518762	DQ522331	DQ522376	
Ophiocordyceps myrmecophila	HMAS_199620	KJ878929	KJ878895	KJ878975	KJ879009	
	CEM 1710	KJ878927	KJ878893	KJ878973	KJ879007	
	TNS 27120	KJ878928	KJ878894	KJ878974	KJ879008	
Ophiocordyceps neovolkiana	OSC 151903	KJ878930	KJ878896	KJ878976	KJ879010	
Ophiocordyceps nigrella	EFCC 9247	EF468963	EF468818	EF468758	EF468866	EF468920
Ophiocordyceps nutans	OSC 110994	DQ522549	DQ518763	DQ522333	DQ522378	
Ophiocordyceps pruinosa	NHJ 12994	EU369106	EU369041	EU369024	EU369063	EU369084
Ophiocordyceps pulvinata	TNS-F 30044	GU904208		GU904209	GU904210	
Ophiocordyceps purpureostromata	TNS F18430	KJ878931	KJ878897	KJ878977	KJ879011	
Ophiocordyceps ravenelii	OSC 110995	DQ522550	DQ518764	DQ522334	DQ522379	DQ522430
	OSC 151914	KJ878932		KJ878978	KJ879012	KJ878950
Ophiocordyceps rhizoidea	NHJ 12522	EF468970	EF468825	EF468764	EF468873	EF468923
	NHJ 12529	EF468969	EF468824	EF468765	EF468872	EF468922
Ophiocordyceps ryogamiensis	NBRC 101751	KF049614	KF049633	KF049688	KF049650	
Ophiocordyceps sinensis	EFCC 7287	EF468971	EF468827	EF468767	EF468874	EF468924
Ophiocordyceps sobolifera	KEW 78842	EF468972	EF468828		EF468875	EF468925
	TNS F18521	KJ878933	KJ878898	KJ878979	KJ879013	
Ophiocordyceps sp.	TNS F18495	KJ878937	KJ878901		KJ879017	
Ophiocordyceps sp.	OSC 110997	EF468976		EF468774	EF468879	EF468929
Ophiocordyceps sp.	OSC 151904	KJ878934	KJ878899	KJ878980	KJ879014	
Ophiocordyceps sp.	OSC 151905	KJ878935		KJ878981	KJ879015	KJ878951

Table 2.1 (Continued)

Ophiocordyceps sp.	OSC 151909	KJ878936	KJ878900	KJ878982	KJ879016	KJ878952
Ophiocordyceps sphecocephala	OSC 110998	DQ522551	DQ518765	DQ522336	DQ522381	DQ522432
Ophiocordyceps stylophora	OSC 110999	EF468982	EF468837	EF468777	EF468882	EF468931
	OSC 111000	DQ522552	DQ518766	DQ522337	DQ522382	DQ522433
Ophiocordyceps tricentri	CEM 160	AB027330	AB027376			
Ophiocordyceps unilateralis	OSC 128574	DQ522554	DQ518768	DQ522339	DQ522385	DQ522436
Ophiocordyceps variabilis	OSC 111003	EF468985	EF468839	EF468779	EF468885	EF468933
	ARSEF 5365	DQ522555	DQ518769	DQ522340	DQ522386	DQ522437
Ophiocordyceps yakusimensis	HMAS_199604	KJ878938	KJ878902		KJ879018	KJ878953
Paecilomyces lilacinus	ARSEF 2181	AF339583	AF339534	EF468790	EF468896	
	CBS 431.87	AY624188	EF468844	EF468791	EF468897	EF468940
	CBS 284.36	AY624189	AY624227	EF468792	EF468898	EF468941
Podocrella harposporifera	ARSEF 5472	AF339569	AF339519	DQ118747	DQ127238	
Podonectria citrine	TNS F18537		KJ878903	KJ878983		KJ878954
Polycephalomyces cuboideus	TNS F18487	KF049609	KF049628	KF049683		
Polycephalomyces cuboideus	NBRC 101740	KF049610	KF049629	KF049684	KF049646	
Polycephalomyces formosus	ARSEF 1424	KF049615	AY259544	DQ118754	DQ127245	KF049671
Polycephalomyces nipponicus	BCC 1881	KF049618	KF049636	KF049692		KF049674
	BCC 1682	KF049620	KF049638	KF049694		
	NHJ4286	KF049621	KF049639	KF049695	KF049654	KF049676
	BCC2325	KF049622	KF049640	KF049696	KF049655	KF049677
Polycephalomyces paracuboideus	NBRC 101742	KF049611	KF049630	KF049685	KF049647	KF049669
Polycephalomyces prolificus	TNS F18481	KF049612	KF049631	KF049686	KF049648	
	TNS F18547	KF0496613	KF049632	KF049687	KF049649	KF049670

Table 2.1 (Continued)

Polycephalomyces ramosopulvinatus	SU-65		DQ118742	DQ118753	DQ127244	
	EFCC 5566		KF049627	KF049682	KF049645	
Polycephalomyces sp.	JB07.08.16_08	KF049616	KF049635	KF049690	KF049652	KF049672
Polycephalomyces sp.	JB07.08.17_07b	KF049617		KF049691	KF049653	KF049673
Polycephalomyces sp.	BBC 2637	KF049619	KF049637	KF049693		KF049675
Polycephalomyces tomentosus	BL4	KF049623	AY259545	KF049697	KF049656	KF049678
Stilbella buquetii	HMAS_199613	KJ878939	KJ878904	KJ878984	KJ879019	
	HMAS_199617	KJ878940	KJ878905	KJ878985	KJ879020	
Tilachlidiopsis nigra	TNS 16252	KJ878941	KJ878906	KJ878986		
	TNS 16250	KJ878942		KJ878987	KJ879021	
Tolypocladium cylindrosporum	NRRL 28025	AF049153	AF049173			

Table 2.2 Proposed list of generic names in *Ophiocordycipitaceae* to be protected and their competing synonyms. Names previously synonymized are in gray.

Proposed to protect	**Proposed to suppress**
Ophiocordyceps Petch, *Trans. Br. Mycol. Soc.* **16**: 74 (1931). *Type*: *O. blattae* Petch 1931.	*Sorosporella* Sorokin *Zentbl. Bakt. ParasitKde.*, Abt. II **4**: 644 (1888). *Type*: *S. agrotidis* Sorokin 1888. *Hirsutella* Pat., *Revue Mycol.* **14**: 67 (1892). *Type*: *H. entomophila* Pat. 1892. *Didymobotryopsis* Henn., *Hedwigia* **41**: 149 (1902). *Type*: *D. parasitica* Henn. 1902. *Mahevia* Lagarde, *Archs Zool. Exp. Gen*. **56**: 292 (1917). *Type*: *M. guignardii* (Maheu) Lagarde 1917. *Synnematium* Speare, *Mycologia* **12**: 74 (1920). *Type*: *S. jonesii* Speare 1920. *Trichosterigma* Petch, *Trans. Br. Mycol. Soc*. **8**: 215 (1923). *Type*: *T. clavisporum* Petch 1923. *Didymobotrys* Clem. & Shear, *Gen. Fungi*: 228 (1931). *Type*: *D. parasitica* (Henn.) Clem. & Shear 1931. *Troglobiomyces* Pacioni, *Trans. Br. Mycol. Soc*. **74**: 244 (1980). *Type*: *T. guignardii* (Maheu) Pacioni 1980. *Hymenostilbe* Petch, *Naturalist* (*Hull*), ser. 3, **1931**: 101 (1931). *Type*: *H. muscaria* Petch 1931. *Syngliocladium* Petch, *Trans. Br. Mycol. Soc.* **17**: 177 (1932). *Type*: *S. aranearum* Petch 1932. *Cordycepioideus* Stifler, *Mycologia* **33**: 83 (1941). *Type*: *C. bisporus* Stifler 1941. *Paraisaria* Samson & B.L. Brady, *Trans. Br. Mycol. Soc.* **81**: 285 (1983). *Type*: *P. dubia* (Delacr.) Samson & B.L. Brady 1983.

Table 2.2 (Continued)

Purpureocillium Luangsa-ard *et al.*, *FEMS Microbiol Lett* **321**: 144 (2011). *Type*: *P. lilacinum* (Thom) Luangsa-ard *et al.* 2011 (syn. *Penicillium lilacinum* Thom 1920).	
Tolypocladium W. Gams, *Persoonia* **6**: 185 (1971). *Type*: *T. inflatum* W. Gams 1971.	*Chaunopycnis* W. Gams, *Persoonia* **11**: 75 (1980). *Type*: *C. alba* W. Gams 1980. *Elaphocordyceps* G.H. Sung & Spatafora, *Stud. Mycol.* **57**: 36 (2007). *Type*: *E. ophioglossoides* (Ehrh. ex J.F. Gmel. : Fr.) G.H. Sung *et al.* 2007.
Harposporium Lohde, *Tagbl. Versamml. Ges. Deutsch. Naturf.* **47**: 206 (1874). *Type*: *H. anguillulae* Lohde 1874.	*Polyrhina* Sorokin, *Annls Sci. Nat., Bot.*, sér 6, **4**: 65 (1876). *Type*: *P. multiformis* Sorokin 1876. *Podocrella* Seaver, *Mycologia* 20: 57 (1928). *Type*: *P. poronioides* Seaver 1928. *Atricordyceps* Samuels, *N.Z. Jl. Bot.* **21**: 174 (1983). Type *A. harposporifera* Samuels 1983.
Drechmeria W. Gams & H.-B. Jansson, *Mycotaxon* **22**: 36 (1985). *Type*: *D. coniospora* (Drechsler) W. Gams & H.-B. Jansson 1985 (syn. *Meria coniospora* Drechsler 1941).	*Haptocillium* W. Gams & Zare, *Nova Hedwigia* **73**: 334 (2001). *Type*: *H. balanoides* (Drechsler) Zare & W. Gams 2001.
Polycephalomyces Kobayasi, *Sci. Rep. Tokyo Bunrika Daig.*, sect. B **5**: 245 (1941). *Type*: *P. formosus* Kobayasi 1941.	*Blistum* B. Sutton, *Mycol. Pap.* **132**: 16 (1973). *Type*: *B. tomentosum* (Schrad.) B. Sutton 1973.

Chapter 3. The genome of the truffle-parasite *Tolypocladium ophioglossoides* and the evolution of antifungal Peptaibiotics.

C. Alisha Quandt, Kathryn E. Bushley, Joseph W. Spatafora

Target Journal: *BMC Genomics*

Background

Hypocreales is home to a wide array of ecologically diverse fungi. Some are devastating plant pathogens (*e.g.*, *Fusarium* spp.), while others form numerous lineages of both insect pathogens (*e.g.*, *Cordyceps*) and mycoparasites (*e.g.*, *Trichoderma*; Sung *et al.* 2007). At the divergence of the four most derived families of Hypocreales, there was a major shift away from plant-based nutrition to either insect pathogenesis or fungal parasitism, *i.e.*, mycoparasites (Sung *et al.* 2008). Two major lineages of mycoparasites are found within the order, although other mycoparasites exist (*e.g.*, *Polycephalomyces*; Kepler *et al.* 2012, 2013). The first and larger of these two lineages is the family Hypocreaceae, most notable for mycoparasitic *Trichoderma* spp. used in biological control of plant pathogenic fungi, and *Trichoderma reesei* (Papavizas 1985), the industrial workhorse for cellulase production (Durand *et al.* 1988, Sukumaran *et al.* 2005). The second major lineage of mycoparasites, the genus *Tolypocladium*, is nested within the insect pathogenic family, Ophiocordycipitaceae (Sung *et al.* 2007, Quandt *et al.* Ch. 2). Most species of *Tolypocladium* parasitize the truffles of *Elaphomyces* [Eurotiales, Ascomycota], ectomycorrhizal fungi closely related to *Aspergillus* and *Penicillium* (Landvik *et al.* 1996, LoBuglio *et al.* 1996, Quandt *et al.* Ch. 6). *Tolypocladium ophioglossoides* (Ehrh. ex J.F. Gmel.) Quandt, Kepler & Spatafora is a commonly collected truffle parasite with a broad geographic distribution throughout many parts of the Northern Hemisphere (Mains 1957, Kobayasi & Shimizu 1960). Historically it was used in religious ceremonies along with other neurotropic fungi in parts of Mexico (Guzmán 2008). There are, however, a few *Tolypocladium* species that attack insects and rotifers, and based on current multigene phylogenies some of these are

inferred to be reversals to insect pathogenesis (Sung *et al.* 2007, 2008, Quandt *et al.*, Ch. 2). One of these is a beetle pathogen, *T. inflatum*, which was the first source of the immunosuppressant drug, cyclosporin A (Borel 2002). Evidence from multigene studies has also shown a close phylogenetic relationship between *T. ophioglossoides* and *T. inflatum* (Sung *et al.* 2007).

Secondary metabolism is defined as synthesis of often bioactive, small molecules that are not essential to the growth of an organism, and genes related to production of secondary metabolites are often clustered together in close proximity within a genome and coregulated (Keller *et al.* 2005). A wide variety of secondary metabolites including the ergot alkaloids, fumonisins, and destruxins, is produced by species of Hypocreales (Isaka *et al.* 2003, Desjardin & Proctor 2007, Molnar *et al.* 2010). Many of these metabolites are produced by nonribosomal peptide synthetases (NRPSs), which are often large, multi-modular proteins that produce short peptides frequently incorporating non-standard amino acids. NRPS modules are further broken down into the functional domains, which include adenylation (A), thiolation (T), and condensation (C) domains (Marahiel *et al.* 1997). Due to their high level of amino acid and nucleotide conservation, the A-domains are frequently used to reconstruct the evolutionary histories of these genes (Bushley & Turgeon 2010, Wei *et al.* 2005, *etc.*). Polyketide synthases (PKSs) are another class of secondary metabolite producing enzymes that are common in fungi and are also modular in nature. They are related to fatty acid synthases (Jenke-Kodama *et al.* 2005), and assemble small bioactive molecules based on acetyl-CoA subunits (Fischbach & Walsh 2006). The other major classes of secondary metabolite-producing enzymes are terpene synthases and dimethylallyltryptophan synthases, both of which have been

reported from hypocrealean taxa. Fungal secondary metabolites clusters often include genes required for regulation of expression of the gene cluster and decoration, epimerization, and transport of the mature secondary metabolite (Zhang *et al.* 2005, Hoffmeister & Keller 2007).

Peptaibols, or peptaibiotics, are antibiotic secondary metabolites products produced by very large NRPS proteins (up to 21,000 amino acids in length). Their name is a derivative of their structure as they are **Pept**ides containing the uncommon non-proteinogenic amino acid, α-amino isobutryic acid (**AIB**), and a C-terminal amino ethan**ol** (Chugh & Wallace 2001). The presence of AIB residues promotes helix formation, and several of these helices form multimeric units that in turn form voltage gated ion channels capable of inserting into cell membranes where they disrupt membrane potential causing leakiness or apoptosis (Chugh *et al.* 2002, Fox and Richards 1982). Peptaibols are produced by *Trichoderma* spp. and other members of Hypocreaceae (Peptabiol Database, Whitmore and Wallace 2004), leading to the proposition that they may play a role in mycoparasitism. There is at least one empirical study to support this in *Trichoderma* (Röhrich *et al.* 2012). Further studies have found that peptaibols function, along with cell wall degrading enzymes, to synergistically inhibit new cell wall synthesis in fungal prey of *Tr. harzianum* (Schrimböck *et al.* 1994, Lorito *et al.* 1994, 1996). Wiest *et al.* (2002) identified and characterized the first peptaibol synthetase NRPS (Tex1) from *Tr. virens* along with the 18 residue peptaibol product. Since that time, several other peptaibol NRPS genes have been identified in *Tr. virens* and other species (Wei *et al.* 2005, Mukherjee *et al.* 2011, Degenkolb *et al.* 2012).

Efrapeptins are another class of peptaibiotics described almost exclusively from *Tolypocladium* spp. that have antifungal and insecticidal properties (Krishna *et al.* 1990, Krasnoff & Gupta 1992, Bandani *et al.* 2000). They differ from orthodox peptaibols by the presence of a mitochondrial ATPase inhibiting C-terminal "blocking group" N-peptido-1-isobutyl-2[1-pyrrole-(1-2-α)-pyrimidinium,2,3,4,6,7,8-hexahydro]-ethylamine (Gupta *et al.* 1991). Six of these, named efrapeptins C-I, have been isolated from *T. inflatum* (Bullough *et al.* 1982, Krasnoff & Gupta 1992, Nagaraj *et al.* 2001).

Numerous genomes from species of the mycoparasitic genus *Trichoderma* (Hypocreaceae) have been sequenced (Martinez *et al.* 2008, Kubicek *et al.* 2011), and more recently the genomes of several insect pathogens in Hypocreales have been completed (*e.g.*, *Cordyceps militaris, Beauveria bassiana, Metarhizium* spp., and *Ophiocordyceps sinensis*) (Zheng *et al.* 2011, Xiao *et al.* 2012, Gao *et al.* 2011, Hu *et al.* 2013) including T. inflatum, the beetle pathogenic congener of *T. ophioglossoides* (Bushley *et al.* 2013). The genomes of all of these species are rich in secondary metabolite genes and clusters, ranging from 23 to 51 secondary metabolite gene clusters per genome. Comparisons of gene content and expression are beginning to shed light on mechanisms underlying host specificity and the evolution of primary and secondary metabolism. In this study the draft genome of the truffle parasite *T. ophioglossoides* was generated to compare the gene content and secondary metabolite content of this truffle parasite to those of closely related insect pathogens and more distantly related mycoparasites. The secondary metabolite potential of *T. ophioglossoides* is characterized with a focus on understanding the evolution of gene clusters encoding for peptaibiotics.

Results and discussion

Genome assembly and structure

The draft genome assembly of the *T. ophioglossoides* CBS 100239 is approximately 31.3 megabases (Mb) and is assembled on 173 scaffolds. With an n50 of 668,222 base pairs (bp), there is sufficient approximation of gene space and order [Table 3.1], in spite of the exclusive use of short read technology. In addition, the Core Eukaryotic Genes Mapping Approach (CEGMA) identified 239 complete and 242 complete and partial core eukaryotic genes, estimating the assembly completeness at 96.4% or 97.6%. We predict 10,135 protein-coding genes resulting in 10,308 protein models, of which 9,476 have support from RNA, making the size and number of protein models similar to that of the beetle pathogen, *T. inflatum*, which is 30.3 Mb and has 9,998 protein models. They both also share high GC contents, 57.3% and 58%, for *T. ophioglossoides* and *T. inflatum*, respectively. These species are very closely related, which is reflected in the genome scale phylogeny (Figure 3.1) and the large regions of shared synteny (Figure 3.2). There are, however, two large scale inversions and some small rearrangements with disagreement, suggesting these species have diverged despite the relatively short branch lengths separating them in the species phylogeny (Figure 3.1). The large of number of small points in the mummerplot alignment could be due to the reliance on short read technology and some of the very short contigs in both assemblies.

An overview of secondary metabolites in *T. ophioglossoides*

The genomes of *T. ophioglossoides* and *T. inflatum* both harbor a large number of core secondary metabolite genes – NRPS, PKS, terpene synthase, and DMATs – and

gene clusters with 45 and 55 genes in 38 and 38 secondary metabolite gene clusters, respectively. Similar to the contrast observed between closely related *Metarhizium* species (Gao *et al.* 2011), *T. ophioglossoides* and *T. inflatum* differ in the types of core secondary metabolite genes they possess with 21 shared between the two species, 34 unique to *T. inflatum*, and 24 unique to *T. ophioglossoides* (Table 3.S1). Notably, *T. ophioglossoides* does not contain the NRPS gene, or any of the other genes in the *simA* cluster responsible for the production of cyclosporin A in *T. inflatum* (Bushley *et al.* 2013). However, *T. ophioglossoides* does share a Hypocreales-conserved core set of genes that flank the *simA* region with *T. inflatum*. Cyclosporin A production has been reported from several species of *Tolypocladium* (Sedmera *et al.* 1995, Traber & Dreyfuss 1996), however, none of the truffle parasites, including *T. ophioglossoides*, have been demonstrated to produce the compound. Whether possession of the *simA* gene is a derived character state, or whether there has been a single or multiple losses within the genus, remains unknown and requires further sampling.

Secondary metabolites predicted from the *T. ophioglossoides* genome include 15 PKSs, two PKS-like genes, 15 NRPSs, six NRPS-like genes, three hybrid NRPS-PKS genes, and four terpenes (Table 3.S1). No DMAT synthases were identified in the *T. ophioglossoides* genome. Based on A-domain homology, two putative siderophore synthetases (one intracellular [TOPH_02853] and one extracellular [TOPH_02629]) were among the predicted NRPSs (Figure 3.S1, Table 3.S1). This is in contrast to *T. inflatum*, which possesses three putative siderophore synthetases (two extracellular and one intracellular) (Bushley *et al.* 2013). The entire Pseurotin-A precursor synthetase hybrid NRPS-PKS cluster (TOPH_07102) was identified in the *T. ophioglossoides* genome.

Pseurotin-A , an antifungal compound described from several *Aspergillus* spp. [Eurotiales, Ascomycota], was also recently identified in the genome of *M. robertsii* (Wiemann *et al.* 2013). The disjunct distribution of this secondary metabolite cluster raises several questions about the evolutionary mechanisms (*e.g.*, horizontal gene transfer vs. complex patterns of gene loss) that may have led to this distribution.

Tolypocladium ophioglossoides possesses the destruxins synthetase NRPS gene. The Maker-generated annotation of this gene results in two protein models (TOPH_08872 and TOPH_08873), while AUGUSTUS annotation of the gene, produces a single protein model spanning the region of the genome containing TOPH_08872 and TOPH_08873. Destruxins are known for their insecticidal properties in *Metarhizium* spp. (Kershaw *et al.* 1999), and recently, the entire destruxins synthetase cluster was characterized by Wang *et al.* (2012) in *M. robertsii*. Homologs of the other genes in the destruxins cluster are present in *T. ophioglossoides*, except for DtxS4 (Figure 3.S2), an aspartic decarboxylase responsible for producing β-alanine, one of the amino acids incorporated into destruxins. There are inversions in this cluster between *M. robertsii* and *T. ophioglossoides* as well, including the DtxS2 aldo-keto reductase homolog (TOPH_08871) and an ABC transporter (TOPH_08869), which was not found to be essential in destruxins production in *M. robertsii* (Wang *et al.* 2012). For these reasons, it is not likely that *T. ophioglossoides* produces destruxins, but possibly produces another group of related compounds. The sequenced strain of *Tr. virens* shares a homolog of the destruxins NRPS gene (Tv62540), but destruxins have not been reported to be produced by that species either (Molnar *et al.* 2010).

To date, only two secondary metabolites have been reported to be produced by *T. ophioglossoides*: ophiocordin (also reported as balanol) and ophiosetin (Kneifel *et al.* 1977, Boros *et al.* 1994, Putri *et al.* 2010). Ophiosetin is structurally similar to equisetin, an antibiotic with inhibitory activity of HIV-1 integrase, and both are produced by NRPS-PKS hybrid genes. Based on phylogenetic analysis of A-domains, cluster synteny with the equisetin cluster (Sims *et al.* 2005), and sequence homology, this study identifies the putative ophiosetin synthetase cluster around the hybrid NRPS-PKS, TOPH_07403 (Table 3.S1). Further studies involving transformations and chemical verification and characterization of this cluster will be necessary to confirm this genotype-chemotype linkage. Ophiocordin is a polyketide and no putative gene or gene cluster thought to be related to its production was identified here. Except for the two peptaibiotic clusters discussed below, the remaining 33 secondary metabolite gene clusters are not yet associated with a specific gene product.

Peptaibiotics of *Tolypocladium*

Among the wide assortment of secondary metabolite genes and gene clusters, the draft genome of *T. ophioglossoides* possesses four peptaibiotic NRPS genes located within three gene clusters. Phylogenetic analyses of fungal A-domains from a variety of NRPS genes known to produce specific products and from whole genome mining of the hypocrealean species in Figure 3.1, revealed that *Trichoderma* peptaibol A-domains group into three clades, one of which is well supported in all analyses (Figure 3.3, Figure 3.S1), the other two clades have strong support with the exception of a few A-domains from one of the *T. ophioglossoides* peptaibiotics (TOPH_08469) (see below). All A-

domains from the four peptaibiotic genes in *T. ophioglossoides* fall within these three clades (Figure 3.3), and representation in the three peptaibol clades is exclusive to A-domains from *Trichoderma* and *Tolypocladium*. *T. inflatum* also possesses three peptaibiotic NRPS genes. Strikingly, this limits the presence of peptaibiotic A-domains and genes to the two sampled mycoparasitic lineages in Hypocreales.

There are differences between species and gene membership within A-domain clades, however. For instance, Clade 1 is enriched in A-domains from the *Trichoderma* peptaibols (62 *Trichoderma* A-domains vs. 27 *Tolypocladium* A-domains), while Clade 2 contains more A-domains from *Tolypocladium* spp. (17 *Trichoderma* A-domains vs. 33 *Tolypocladium* A-domains). Importantly Clade 1 contains A-domains that encode for incorporation of AIB, as well as other A-domains that encode for incorporation of isovaline, leucine, isoleucine, alanine, glycine, valine, and serine (Wei *et al.* 2005, Mukherjee *et al.* 2011). Clade 2 is known to include A-domains that encode for valine, glutamine, asparagines, leucine, and isoleucine (Wei *et al.* 2005, Mukherjee *et al.* 2011). Clade 3, which is known only to incorporate a single amino acid, proline (Wei *et al.* 2005, Mukherjee *et al.* 2011), has a relatively equal distribution of both *Trichoderma* (10) and *Tolypocladium* (8) A-domains. Prolines are proposed to play an important structural role in peptaibols by creating a kink in the peptaibol chain, and Clade 3 occupies a long branch within the tree, suggesting it is highly diverged from the other A-domains (Figure 3.S1).

Peptaibols were reported from *T. geodes* (Tsantrizos *et al.* 1996) based on chemical isolation, but all previous reports of the compounds were identified from fungi in Hypocreaceae (Whitmore & Wallace 2002) (or Boletaceae [Basidiomycota] which are

likely produced by hypocreaceous parasites of Boletaceae fruiting bodies [Lee *et al.* 1999a, b]). A gene cluster responsible for the production of these peptaibols has not been identified, and to date no genomic sequence data have been produced for *T. geodes* from which to predict which genes or clusters may be responsible for its production. Efrapeptins, which are Peptaibiotics originally described from *T. inflatum*, have been reported from several species of *Tolypocladium* (Krasnoff & Gupta 1992), and it remains unknown whether the products produced by these clusters in *T. ophioglossoides* are of the efrapeptin class, or more traditional class of peptaibols. Regardless, the phylogenetic diversity of peptaibiotic NRPSs, as revealed by phylogenomic analyses of A-domains, supports a greater chemical diversity of peptaibiotics than currently known from chemical analyses.

To interpret the evolutionary history of the peptaibol A-domains within Hypocreales, the peptaibol A-domain clade tree was reconciled with the species tree (Figure 3.4). Because Clade 3 is inferred to have an independent origin from Clades 1 and 2 within the A-domain phylogeny, it was reconciled with the species tree separately (Figure 3.S3). We have included closely related A-domains that are not a part of the three peptaibiotic A-domain clades in order to root the peptaibol A-domain clades. The deep coalescence of peptaibol A-domains from all three clades at the common ancestor of *Trichoderma* and *Tolypocladium* (Figure 3.4) suggests that the presence of peptaibiotics in *Tolypocladium* is an ancient attribute of *Tolypocladium* genomes and not a product of a more recent horizontal transfer. Analyzing Clades 1 and 2, only 31 of the A-domains in *T. ophioglossoides* and *T. inflatum* coalesce at their most recent common ancestor,

whereas 20 coalesce more deeply at the divergence of the four most derived families (Figure 3.4).

There exist at least two possible explanations to these findings in *Tolypocladium* spp. First, the divergence of theA-domains could be significant enough so as to distort the evolutionary history as represented in the phylogenetic tree, especially since many of the branches leading to *Tolypocladium* A-domains do not have support in the tree. This explanation is not supported by the data, because based on the species phylogeny nucleotide divergence levels between *Tolypocladium* species is less than that of *Trichoderma* species (Figure 3.1). Second, incomplete lineage sorting could lead to this pattern of coalescence, suggesting that the ancestor to the genus possessed a multitude of these A-domains within one or several peptaibiotic genes that have undergone a complex history of ancient gains (duplications) and losses.

In contrast, five of the A-domains inferred to be present in the common ancestor of *Trichoderma* spp. deeply coalesce, and 59 are shared in the common ancestor. This pattern indicates a higher degree of domain tree – species tree congruence, which could be explained by two different mechanisms including descent with expansion and maintenance or horizontal gene transfer from *Tolypocladium* to common ancestor of *Trichoderma*. To test for signatures of horizontal gene transfer, the A-domain tree was reconciled, with a modified species tree in which *Trichoderma* was sister to *Tolypocladium*, and outgroup A-domains were included to root the domain tree. This produced a smaller deep coalescent cost (212 v. 275), but this is due to the fewer number of extinctions (in Cordycipitaceae, Clavicipitaceae, and *O. sinensis*) required (118 v. 181), because the number of duplications remains the same (91 v. 91). In this simulated

reconciliation, 14 A-domains were inferred to deeply coalesce, and of these, seven were inherited in each lineage. Taken together, this suggests that the diversity of *Trichoderma* A-domains cannot solely be characterized as the product of *Tolypocladium* A-domains and that the common ancestor of *Trichoderma* possessed a small number of peptaibiotic NRPSs and A domains that largely diversified in a manner consistent with speciation of the genus.

T. ophioglossoides peptaibiotic gene clusters

The four peptaibiotic genes (TOPH_03025, TOPH_03035, TOPH_08469 and TOPH_08528) are located in three gene clusters on three different scaffolds (Figure 3.5, Table 3.S1). Genes TOPH_03035 (10 modules) and TOPH_03025 (16 modules) are located within the same gene cluster and separated by only nine genes, including one PKS. In total, the cluster contains18 genes, many of which are typically found within secondary metabolite clusters including two multidrug transporters (TOPH_03022 and TOPH_3034), a decarboxylase (TOPH_03027), an esterase(TOPH_03026), an epimerase (TOPH_03033), and a leucine zipper transcription factor (TOPH_03024) among others (Figure 3.5a). At more than 17,000 amino acids in length, TOPH_03025 encodes for the largest peptaibiotic NRPS producted by *T. ophioglossoides*, and is the largest gene (52.5 kb) in the genome. Gene TOPH_08528 (2 modules), located on a separate scaffold, is only 2,587 amino acids in length and shares a high degree of amino acid identity with modules 8 and 9 of TOPH_03025 (Figure 3.6). No short Peptaibiotics have been described so far, but the two A-domains of this gene both fall within peptaibiotic Clade 2 (Figure 3.3), and it is clearly orthologous to part of TOPH_03025. So while it may not be

producing a peptaibiotic, it is analyzed here within the evolutionary context of these genes.

The remaining *T. ophioglossoides* peptaibiotic NRPS gene, TOPH_08469 (10 modules), is located within a cluster containing two PKS genes (TOPH_08457 and TOPH_08462), several ABC transporters (TOPH_08453, TOPH_08459, and TOPH_08470), an esterase (TOPH_08466), an epimerase (TOPH_08468) , a hydrolase (TOPH_08465) , two cytochrome p450s (TOPH_08458 and TOPH_08455), and a RadR transcription regulator (TOPH_08461) (Figure 3.5b). It remains to be seen if the products of the PKSs are incorporated into the peptide created by TOPH_08469. Some of the A-domains within TOPH_08469 are divergent (Figure 3.3), especially those that group (without support) as the earliest diverging lineages of Clade 2, and the presence of these A-domains in the tree, causes the support for this clade to weaken substantially.

Cluster synteny between *Tolypocladium* spp.

Despite also possessing three large peptaibiotic genes in two clusters, *T. inflatum* peptaibiotic genes are highly divergent from *T. ophioglossoides* and the gene clusters are not located in the same regions of the genome (Figure 3.7c). Alignment of *T. ophioglossoides* scaffold containing two peptaibol NRPS genes (TOPH_03025 and TOPH_03035) with the *T. inflatum* scaffold containing the single peptaibol NRPS gene (Tinf05969), revealed high synteny in the *T. ophioglossoides* two-peptaibol gene cluster region, except for the absence of the two peptaibol genes, themselves (Figure 3.7a). No other genes have been inserted into the *T. inflatum* scaffolds, where the peptaibiotic NRPS genes are located in the *T. ophioglossoides* genome. This alignment did reveal,

however, the presence of a truncated N terminal portion of an NRPS that aligns with the N terminal region of TOPH_03025. When included in the entire hypocrealean A-domain phylogeny from Figure 3.S1, this *T. inflatum* NRPS "relic" is most closely related to the first A-domain of TOPH_03025 (Figure 3.7b). While retrotransposon relics are commonly reported from secondary metabolite gene clusters (Scott *et al.* 2009, Fleetwood *et al.* 2007) and secondary metabolite gene modules may be truncated (Hoffmeister & Keller 2007), this is the first report of a relic NRPS domain that remains within an otherwise intact secondary metabolite cluster including the PKS located within the cluster. Due to the truncated nature of this protein model (only 68 amino acids in length) and lack of other functional domains, it is unlikely that this relic produces a secondary metabolite peptide. There is no evidence of transposable elements in or near this cluster in either of the *Tolypocladium* genomes, although RepeatMasker identified a partial sequence for a TAD1-like Long Interspersed Element (LINE) at the N-terminal end of one of the other *T. inflatum* peptaibol NRPS genes (TINF08927). As mentioned above, this *T. inflatum* scaffold which is syntenous with the gene cluster in *T. ophioglossoides* containing two peptaibol NRPS genes, also contains the NRPS gene Tinf05969, but this gene is located approximately 380 Kb downstream of the relic cluster (Figure 3.7a).

The region of the scaffold containing the *T. ophioglossoides* third large peptaibiotic gene, TOPH_08469, does not align well with any portion of the *T. inflatum* genome. Similarly, the final peptaibol cluster in *T. inflatum*, containing two peptaibol NRPS genes (TINF07827 and TINF07876), does not align well to any portion of the *T. ophioglossoides* genome. The lack of synteny between these clusters in *Tolypocladium*

spp. highlights the significant amount of genomic rearrangements between these closely related taxa. Campbell *et al.* (2012) observed patterns of differential gene loss in *Botrytis* spp. within an ancient, horizontally-transferred, secondary metabolite gene cluster, leading to a patchy distribution of the genes within the clusters. This is not the pattern seen in the peptaibiotic clusters in *Tolypocladium* spp., in which the protein models are not reciprocal best BLAST hits (except for the protein models in the "relic" cluster in *T. inflatum*). Thus, despite the fact that their products may have similar functions, they are highly divergent NRPS genes located within nonhomologous gene clusters.

Mixed homology of Peptaibiotic A-domains in Hypocreales

Using the moderate to strongly supported nodes (≥50 MLBP) in the A-domain phylogeny as a guide for module homology, the peptaibol NRPS genes are more conserved among the *Trichoderma* species examined as compared to species of *Tolypocladium* (Figure 3.6), a finding reflected in the domain tree – species tree reconciliation analyses (Figure 3.4). Using whole genome data of the sampled species of *Trichoderma*, A-domains from the *Tr. virens* three peptaibol NRPS genes (Tex1, Tex2, and Tex3) (Mukherjee *et al.* 2011) were identified for the phylogenetic analyses; *Tr. virens* peptaibols contain 18, 14, and 7 modules respectively. *Trichoderma reesei* has two peptaibol NRPSs (Tr_23171 and Tr_123786) which possess 18 and 14 modules. In the annotation of *Tr. atroviride* IMI 206040, the A-domain HMM identified one 19 module peptaibol NRPS (Ta_317938), and several single A-domain protein models that group within the three peptabioitic clades and are all located on scaffold 29. Further examination of this *Tr. atroviride* gene region using the JGI genome browser (Grigoriev

et al. 2014) revealed that all of the *ab initio* gene predictions of that region predict a single protein model that is the approximate length of the 14 module peptaibol genes in *Tr. virens* and *Tr. reesei*. Degenkolb *et al.* (2012) identified a homolog of the 14 modular peptaibol gene from a different strain of *Tr. atroviride*, and thus this is likely a mis-annotation of *Tr. atroviride* scaffold 29. Alignment of this scaffold in *Tr. atroviride* (scaffold 29) to those of *Tr. virens* and *Tr. reesei* (Figure 3.S4) revealed high nucleotide homology. However, the flanking regions did not align well, and a BLAST search revealed that the 14 modular peptaibol gene in *Tr. atroviride* is located within a different portion of the genome than in the other two *Trichoderma* spp. (Figure 3.S4).

Comparing the *Trichoderma* A-domains, each species possesses one ortholog of the large 18 or 19 module NRPS gene (Figure 3.6), and at the nucleotide level these regions of their genomes also align and are syntenous (Figure 3.S4). The A-domains are syntenous in their arrangement within the large peptaibol NRPS genes across the species, except for: (a) the insertion of a Clade 3 domain at the third module, and (b) a duplication of either the seventeenth or eighteenth A-domains, which are most closely related. Within the two 14 module NRPS genes in *Tr. virens* and *Tr. reesei*, there is complete synteny of the A-domains. In *Tr. virens*, it has been demonstrated that this 14 module NRPS, Tex2, is responsible for two different sizes of peptaibols (11 and 14 residues in length) (Mukherjee *et al.* 2011). Due to differences in annotation (see above), the 14 modular peptaibol gene in *Tr. atroviride* is not compared in this analysis. The short 7 module peptaibol synthetase from *Trichoderma* spp. is found only in *Tr. virens*. Between these three groups of peptaibol NRPSs, the terminating residues are all orthologous, as well as the initiating residues in the larger classes of peptaibol NRPSs.

In *Tolypocladium*, there is a very different pattern of homology and synteny between the peptaibiotic NRPSs of the two species. Only a few of the A-domain relationships within *Tolypocladium* are statistically supported (Figure 3.3, Figure 3.6). The first, third, and last A-domains of the largest NRPS in both species (TOPH_03025 and TINF07827) are orthologous, but not the other domains within those two genes. There are several instances of intragenic module duplications which are known to occur within NRPS genes and have been proposed to play a role in the evolution of novel metabolites (Fischbach *et al.* 2008). Within TOPH_03035, for example, there is strong support for a shared ancestry between modules 2, 3 and 6 (Figure 3.6), indicating that these modules are the product of lineage specific duplications (Figure 3.4). This indicates a more complicated evolutionary history of these genes in *Tolypocladium*.

The lack of module synteny and orthology between *Tolypocladium* peptaibiotic gene modules is comparable to the lack of genomic synteny observed between their clusters. Part of this is due to the deep coalescence of the *Tolypocladium* A-domains, as discussed earlier. This evidence indicates that *Tolypocladium* peptaibiotic genes are not highly similar but are the products of more ancient divergences. This is notable, because in contrast to *Trichoderma* spp., all of which exhibit some degree of mycoparasitism (Druzhinina *et al.* 2011), *T. ophioglossoides* and *T. inflatum* have different ecologies, which are characterized by mycoparasitism and insect pathogenicity, respectively. Thus, if peptaibols are important in successful mycoparasitism (as the case has been made in *Trichoderma* spp. [Schrimböck *et al.* 1994, Röhrich *et al.* 2012]), then there may be less selective pressure to maintain a specific mycoparasitic function of these extremely large (>10,000 amino acid) NRPS genes in more ecologically diverse lineages.

Conclusions

The genome of *T. ophioglossoides* is rich in secondary metabolite gene clusters, and 31 out of 38 of these clusters have no known putative product. Given this potential and its life history as a mycoparasite, this species should be targeted for future studies to discover novel natural compounds with potential antibiosis, including antifungal, activity. The *simA* NRPS gene cluster, responsible for the production of the immunosuppressant cyclosporin, is not present in the *T. ophioglossoides* genome, but three large peptaibiotic genes are present within two clusters. These are the first data to suggest peptaibiotic production from a mycoparasitic species of *Tolypocladium*. This study confirms the presence of three phylogenetic clades of peptaibiotic NRPS A-domains from *Tolypocladium* and *Trichoderma* spp., and that peptaibiotics in general are limited to the mycoparasitic lineages of Hypocreales based on current sampling. Reconciliation of the A-domain tree with the organismal phylogeny reveals that the peptaibiotic NRPSs of *Trichoderma* and *Tolypocladium* are likely the product of different mechanisms of diversification. *Trichoderma* is characterized by A-domain diversification that is largely consistent with speciation whereas *Tolypocladium* is characterized by A-domain diversity that results in deep coalescence. This pattern of deep coalescence is inconsistent with peptaibiotic NRPS diversity being the product of HGT to *Tolypocladium*, rather it is the product of complex patterns of gains and losses for A-domains from hypocrealean ancestors. While the diversity of peptaibiotic NRPSs in *Trichoderma* could possibly be explained by HGT in the common ancestor of the three species, none of the

Tolypocladium peptaibiotic NRPSs analyzed here are candidates for HGT. Further research is required to identify the structures of specific metabolites of the *Tolypocladium* gene clusters are and if these peptaibiotics are produced during mycoparasitism by *T. ophioglossoides* or if these genes are present in other mycoparasitic lineages of Hypocreales.

Materials and methods

Genome sequencing

T. ophioglossoides strain CBS 100239 was grown for 7 days in a shaking incubator in potato dextrose broth (PDB) inoculated with plugs of tissue growing on potato dextrose agar for collection of tissue for DNA extraction. Tissue was harvested via filtration, frozen at -80° C in 1.5 mL tubes, and then lyophilized for 24 hours. Lyophilized tissue was ground using a mortar and pestle, and DNA was extracted using a Qiagen DNeasy Plant Mini kit following the standard protocol starting at the step with the addition of lysis buffer AP1 and eluted in 50 μL water. Tissue for RNA extraction was grown in Yeast Malt (YM) broth, minimal media (MM) containing autoclaved insect cuticle with proteins removed using the protocol in Andersen (1980) (Bushley *et al.* 2013), and MM containing lyophilized *Elaphomyces muricatus* peridium for 24 hours and harvested into liquid nitrogen and stored at -80° until extraction. RNA was extracted using the Qiagen RNeasy Plant kit following the manufacturer's protocol. The small insert DNA library was prepared using New England Biomedicals NEBNext reagents, and size selection (350 bp) was performed using gel extraction. Nextera Mate Pair

Sample Preparation of a large insert (6800bp) library and sequencing was conducted by the Core Labs at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University. The Illumina TruSeq RNA Sample Preparation Kit v2 was using for RNA library construction, using the manufacturer's suggested protocols including Agencourt AMPure magnetic beads for cleaning steps. All libraries were sequenced on the Illumina HiSeq2000 at the Core Labs of the CGRB with paired-end 101 cycles for DNA libraries and single-end 51 cycles for RNA libraries.

Assembly, annotation, and bioinformatic analyses

Using scripts in the fastx toolkit (Gordon 2011), raw reads were trimmed (to 50 bp in length) and filtered based on quality score (all bases ≥q20). Initial de novo assembly of the short insert reads was conducted in Velvet v. 1.19 (Zebrino & Birney 2008) with over 156 million reads where the assembly had a median coverage depth of 74.45. The final trim length (50 bp) used in the assembly was chosen after trimming to different lengths (40-80 bp) followed by quality filtering and then the assembly with highest n50 and fewest number of contigs was selected as the "best" assembly. From that assembly, 50 million overlapping 150 bp paired reads were simulated with a 250 bp insert size using the program wgsim v. 0.3.1-r13 in "haploid" mode (Li 2011). Final assembly using the simulated overlapping short insert library reads and the mate pair reads from the 6 kb library was conducted in AllPaths-LG with default settings (Gnerre *et al.* 2011). The Core Eukaryotic Mapping Genes Approach (CEGMA) was used to estimate the completeness of the *T. ophioglossoides* genome (Parra *et al.* 2007, 2009). Nucmer v. 3.07 and other scripts in that package were used to create a mummerplot between *T. ophioglossoides* and *T. inflatum*.

Gene model predictions were created using the Maker annotation pipeline (Cantarel *et al.* 2008) incorporating RNA data assembled in Trinity (Grabherr *et al.* 2011) using the Jellyfish v. 2.0 method of kmer counting (Marçais & Kingford 2011). Other information given to Maker included a custom HMM for *T. ophioglossoides* built by Genemark-ES v 2.0 (Ter-Hovhannisyan *et al.* 2008), a SNAP HMM (Korf 2004) trained on *Fusarium graminearum,* which was also set as the species model for AUGUSTUS (Stanke *et al.* 2006), and protein and/or EST data from the following hypocrealean taxa: *F. graminearum, N. haematococca, Tr. reesei, Tr. virens, M. robertsii, T. inflatum, C. militaris, B. bassiana*. Annotation of transposable elements was performed in RepeatMasker v 3.2.8 with organism set to "fungi" (Smit & Green 1996), and custom repeat content was estimated using RepeatScout v 1.0.3 and scripts associated with that package (Price *et al.* 2005). Non-overlapping *ab initio* protein models were aligned using BLAST (Altschul *et al*. 1990) against a custom database of all the protein models of all the hypocrealean taxa used in this study. Any of these protein models with a significant hit ($\leq 1e^{-5}$) were included in the final protein set and used for downstream analyses.

Using a set of NRPS A-domains from a wide array of published fungal genomes (Bushley *et al.* 2010) a hidden markov model (HMM) was created for this study using the program Hmmer 3.0 (Eddy 2011). This hidden markov model was then used to mine the 18 hypocrealean genomes used for this study for the identification of A-domains. Putative A-domains identified were filtered for short sequences (less than 100 bp), and where applicable cross referenced with published reports of NRPS from those species (e.g *Tr.virens* Tex1) (Wiest *et al.* 2002). Additional annotation of secondary metabolite clusters was completed using the antiSMASH (Blin *et al.* 2013) and SMURF (Khaldi *et*

al. 2010) pipelines. A-domain trees were reconciled with species trees in Mesquite v. 2.75 with the contained tree treated as unrooted (Maddison & Maddison 2011).

Whole scaffold alignments were performed in the program Mauve (Darling *et al.* 2010) with default progressivemauve alignment settings.

Phylogenetic analyses

Predicted A-domain amino acids sequences were aligned using MUSCLE v 3.8.31 (Edgar 2004) under default settings. Gaps were removed manually, and all alignments were analyzed using RAxML v 7.2.6 (Stamatakis 2006) using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates.

Whole genome phylogenomic analyses were executed in the HAL pipeline (Robbertse *et al.* 2011). Orthologous clusters of proteins were identified in MCL (Enright *et al.* 2002) across inflation parameters 1.2, 3 and 5. Briefly, orthologous clusters were filtered for retention of clusters with one sequence per genome and removal of any redundant clusters. The resulting unique, single-copy orthologous clusters of proteins were aligned in MUSCLE (Edgar 2004) with default settings; poorly aligned regions were identified using Gblocks (Talavera & Castresana 2007; gap removal setting = l, for liberal) and excluded from subsequent analyses. The aligned clusters were concatenated into a superalignment and maximum likelihood analysis was performed using RAxML v 7.2.6 with the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates.

Acknowledgments

The authors would like to thank Mark Desanko, Cedar Hesse, Chris Sullivan, for providing technical and computational support for this project. This project was supported with funds from the NSF grants DEB-0732993 to JWS, and CAQ was supported by an NSF Graduate Research Fellowship.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Bandani AR, Khambay BPS, Faull JL, Newton R, Deadman M, Butt TM (2000) Production of efrapeptins by *Tolypocladium* species and evaluation of their insecticidal and antimicrobial properties. *Mycological Research* **104**: 537-544.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic acids research* **41**: W204-W212.

Borel JF (2002) History of the discovery of cyclosporin and of its early pharmacological development. *Wiener Klinische Wochenschrift* **114**: 433-437.

Boros C, Hamilton, SM Katz, B, Kulanthaivel P (1994) Comparison of balanol from *Verticillium balanoides* and ophiocordin from *Cordyceps ophioglossoides*. *The Journal of Antibiotics* **47**: 1010-1016.

Bullough DA, Jackson CG, Henderson PJF, Cottee FH, Beechey RB, Linnett PE (1982) The amino acid sequence of Efrapeptin-D. *Biochemistry International* **4**: 543-549.

Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, Owensby CA, Knaus BJ, Elser J, Miller D, Di Y, McPhail KL, Spatafora JW (2013) Draft genome sequence of the Cyclosporin producing fungus *Tolypocladium inflatum* reveals complex patterns of secondary metabolite evolution and expression. *PLoS Genetics* **9**: e1003496.

Bushley KE, Turgeon BG (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology* **10**: 26.

Campbell MA, Rokas A, Slot JC (2012) Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome biology and evolution* **4**: 289-293.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188-196.

Chugh JK, Brückner H, Wallace BA (2002) Model for a helical bundle channel based on the high-resolution crystal structure of trichotoxin_A50E. *Biochemistry* **41**: 12934-12941.

Chugh JK, Wallace BA (2001) Peptaibols: models for ion channels. *Biochemical Society Transactions* **29**: 565-570.

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Degenkolb T, Berg A, Gams W, Schlegel B, Gräfe U (2003) The occurrence of peptaibols and structurally related peptaibiotics in fungi and their mass spectrometric identification via diagnostic fragment ions. *Journal of Peptide Science* **9**: 666-678.

Degenkolb T, Von Doehren H, Fog Nielsen K, Samuels GJ, Brueckner H (2008) Recent advances and future prospects in peptaibiotics, hydrophobin, and mycotoxin research, and their importance for chemotaxonomy of *Trichoderma* and *Hypocrea*. *Chemistry & Biodiversity* **5**: 671-680.

Degenkolb T, Karimi Aghcheh R, Dieckmann R, Neuhof T, Baker SE, Druzhinina IS, Kubicek CP, Brückner H, von Döhren, H (2012) The Production of Multiple Small Peptaibol Families by Single 14 Module Peptide Synthetases in Trichoderma/ Hypocrea. Chemistry & biodiversity 9: 499-535.

Desjardin AE, andProctor RH (2007) Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology* **119**: 47-50.

Druzhinina IS, Seidl-Seiboth V, Herrera-Estrella A, Horwitz BA, Kenerley CM, Monte E, Mukherjee PK, Zeilinger S, Grigoriev IV, Kubicek CP (2011) *Trichoderma*: the genomics of opportunistic success. *Nature Reviews Microbiology* **9**: 749-759.

Durand H, Clanet M, Tiraby G (1988) Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme and microbial technology* **10**: 341-346.

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology **7***: e1002195.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575-1584.

Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews* **106**: 3468-3496.

Fischbach MA, Walsh CT, Clardy J (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proceedings of the National Academy of Sciences* **105**: 4601-4608.

Fox EM, Howlett BJ (2008) Secondary metabolism: regulation and role in fungal biology. *Current Opinion in Microbiology* **11**: 481-487.

Fox RO, Richards FM (1982) A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. *Nature* **300**: 325-330.

Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie X-Q, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma L-J, St. Leger RJ, Zhao G-P, Pei Y, Feng M-G, Xia Y, Wang C (2011) Genome

sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics* **7**: 1-18.

Gnerre S, MacCallum I, Prxybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Wiliams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* **108**: 1513-1518.

Gordon A (2011) "FASTX-Toolkit" FASTQ/A short-reads pre-processing tools. Available at http://hannonlab.cshl.edu/fastx_toolkit/

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644-652.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699-704.

Guzmán G (2008) Hallucinogenic mushrooms in Mexico: an overview. *Economic Botany* **62**: 404-412.

Hoffmeister D, Keller NP (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports* **24**: 393-416.

Hu X, Zhang Y, Xiao G, Zheng P, Xia Y, Zhang X, St. Leger RJ, Liu X, Wang C (2013) Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chinese Science Bulletin* **58**: 2846–2854.

Isaka M, Kittakoop P, Thebtaranonth Y (2003) Secondary metabolites of Clavicipitalean fungi. *Mycology Series* **19**: 355-398.

Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution* **22**: 2027-2039.

Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism - from biochemistry to genomics. *Nature Reviews Microbiology* **3**: 937-947.

Kepler RM, Ban S, Nakagiri A, Bischoff JF, Hywel-Jones N, Owensby CA, Spatafora JW (2013) The phylogenetic placement of hypocrealean insect pathogens in the genus *Polycephalomyces*: An application of One Fungus One Name. *Fungal Biology* **117**: 611-622.

Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E, Hosoya T, Bischoff JF, Spatafora JW (2012) Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilaginoidea* (Clavicipitaceae). *American Journal of Botany* **99**: 552-561.

Kershaw MJ, Moorhouse ER, Bateman R, Reynolds SE, Charnley AK (1999) The role of destruxins in the pathogenicity of *Metarhizium anisopliae* for three species of insects. *Journal of Invertebrate Pathology* **74**: 213-223.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736-741.

Kneifel H, König WA, Loeffler W, Müller R (1977) Ophiocordin, an antifungal antibiotic of *Cordyceps ophioglossoides*. *Archives of Microbiology* **113**: 121-130.

Kobayasi Y, Shimizu D (1960) Monographic studies of *Cordyceps* 1. Group parasitic on *Elaphomyces*. *Bulletin of the National Science Museum Tokyo* **5**: 69-85.

Korf I (2004) Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.

Krasnoff SB, Gupta S (1992) Efrapeptin production by *Tolypocladium* fungi (Deuteromycotina: Hyphomycetes): intra- and interspecific variation. *Journal of Chemical Ecology* **18**: 1727-1741.

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Dohren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gomez-Rodriguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernandez-Onate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lubeck M, Lubeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, and Grigoriev IV (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology* **12**: R40.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

Landvik S, Shaller NFJ, Eriksson OE (1996) SSU rDNA sequence support for a close relationship between the Elaphomycetales and the Eurotiales and Onygenales. *Mycoscience* **37**: 237-241.

Lee S-J, Yeo W-H, Yun B-S, Yoo I-D (1999) Isolation and sequence analysis of new peptaibol, boletusin, from *Boletus* ssp. *Journal of Peptide Science* **5**: 374–378.

Lee S-J, Yun B-S, Cho D-H, Yoo I-D (1999) Tylopeptins A and B, new antibiotic peptides from *Tylopilus neofelleus*. *Journal of Antibiotics* **52**: 998–1006.

Li H (2011) Wgsim – read simulator for next generation sequencing. Available at http://github.com/lh3/wgsim.

LoBuglio KF, Berbee ML, Taylor JW (1996) Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the Ascomycetes. *Molecular Phylogenetics and Evolution* **6**: 287-294.

Lorito M, Peterbauer C, Hayes CK, Harman GE (1994) Synergistic interaction between fungal cell wall degrading enzymes and different antifungal compounds enhances inhibition of spore germination. *Microbiology* **140**: 623-629.

Lorito M, Woo SL, D'ambrosio M, Harman GE, Hayes CK, Kubicek CP, Scala F (1996) Synergistic interaction between cell wall degrading enzymes and membrane affecting compounds. *MPMI-Molecular Plant Microbe Interactions* **9**: 206-213.

Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org.

Mains EB (1957) Species of *Cordyceps* parasitic on *Elaphomyces*. *Bulletin of the Torrey Botanical Club* **84**: 243-251.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764-770.

Marahiel MA, Stachelhaus T, Mootz HD (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical Reviews* **97**: 2651-2673.

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, Lopez de Leon A, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature biotechnology* **26**: 553-560.

Molnar I, Gibson DM, Krasnoff SB (2010) Secondary metabolites from entomopathogenic hypocrealean fungi. *Natural Product Reports* **27**: 1241-1275.

Mukherjee PK, Wiest A, Ruiz N, Keightley A, Moran-Diez ME, McCluskey K, Pouchus YF, Kenerley CM (2011) Two classes of new peptaibols are synthesized by a single non-ribosomal peptide synthetase of *Trichoderma virens*. *Journal of Biological Chemistry* **286**: 4544-4554.

Nagaraj G, Uma MV, Shivayogi MS, Balaram H (2001) Antimalarial activities of peptide antibiotics isolated from fungi. *Antimicrobial agents and chemotherapy* **45**: 145-149.

Papavizas GC (1985) *Trichoderma* and *Gliocladium*: biology, ecology, and potential for biocontrol. *Annual Review of Phytopathology* **23**: 23-54.

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

Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 298-297.

Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351-i358.

Putri SP, Kinoshita H, Ihara F, Igarashi Y, Nihira T (2010) Ophiosetin, a new tetramic acid derivative from the mycopathogenic fungus *Elaphocordyceps ophioglossoides*. *The Journal of Antibiotics* **63**: 195-198

Robbertse B, Yoder RJ, Boyd A, Reeves J, Spatafora JW (2011) Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLoS Currents* **3**: RRN1213.

Röhrich CR, Iversen A, Jaklitsch WM, Voglmayr H, Berg A, Dörfelt H, Thrane U, Vilcinskas A, Nielsen KF, von Döhren, Brückner H, Degenkolb T (2012) Hypopulvins, novel peptaibiotics from the polyporicolous fungus *Hypocrea pulvinata*, are produced during infection of its natural hosts. *Fungal biology* **116**: 1219-1231.

Schirmböck M, Lorito M, Wang YL, Hayes CK, Arisan-Atac I, Scala F, Harman GE, Kubicek CP (1994) Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. *Applied and Environmental Microbiology* **60**: 4364-4370.

Sedmera P, Havliček V, Jegorov A, Segre AL (1995) Cyclosporin d hydroperoxide a new metabolite of *Tolypocladium terricola*. *Tetrahedron letters* **36**: 6953-6956.

Sims JW, Fillmore JP, Warner DD, Schmidt EW (2005) Equisetin biosynthesis in *Fusarium heterosporum*. *Chemical Communications* **2**: 186-188.

Smit AF, Green P (1996) RepeatMasker. *Published on the web at http://www. Repeatmasker.org*.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.

Sukumaran RK, Singhania RR, Pandey A (2005) Microbial cellulases-production, applications and challenges. *Journal of Scientific and Industrial Research* **64**: 832-844.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5–59.

Sung G-H, Poinar GO, Spatafora JW (2008) The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal–arthropod symbioses. *Molecular Phylogenetics and Evolution* **49**: 495-502.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564-577.

Traber R, Dreyfuss MM (1996) Occurrence of cyclosporins and cyclosporin-like peptolides in fungi. *Journal of Industrial Microbiology* **17**: 397–401.

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

Tsantrizos YS, Pischos S, Sauriol F, Widden P (1996) Peptaibol metabolites of *Tolypocladium geodes*. *Canadian journal of chemistry* **74**: 165-172.

Wang B, Kang Q, Lu Y, Bai L, Wang C (2012) Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proceedings of the National Academy of Sciences* **109**: 1287-1292.

Wei X, Yang F, Straney DC (2005) Multiple non-ribosomal peptide synthetases genes determine peptaibol synthesis in *Trichoderma virens*. *Canadian Journal of Microbiology* **51**: 423-429.

Wiemann P, Guo CJ, Palmer JM, Sekonyela R, Wang CC, Keller NP (2013) Prototype of an intertwined secondary-metabolite supercluster. *Proceedings of the National Academy of Sciences* **110**: 17065-17070.

Whitmore L, Wallace BA (2004) The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Research* **32**: D593-D594.

Wiest A, Grzegorski D, Xu B-W, Goulard C, Rebuffat S, Ebbole DJ, Bodo B, Kenerley C (2002) Identification of peptaibols from *Trichoderma virens* and cloning of peptaibols synthetases. *The Journal of Biological Chemistry* **277**: 20862-20868.

Xiao G, Ying S-H, Zheng P, Wang Z-L, Zhang S, Xie X-Q, Shang Y, St. Leger RJ, Zhao G-P, Wang C, Feng M-G (2012) Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Scientific Reports* **2**: 483.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821-829.

Zheng P, Xia L, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S, Zhao G-P, Liu X, St. Leger RJ, Wang C (2011) Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biology* **12**: R116.

Zhang YQ, Wilkinson H, Keller NP, Tsitsigiannis D, An Z (2005) Secondary metabolite gene clusters. *Handbook of Industrial Mycology* New York: Marcel Dekker 355-386.

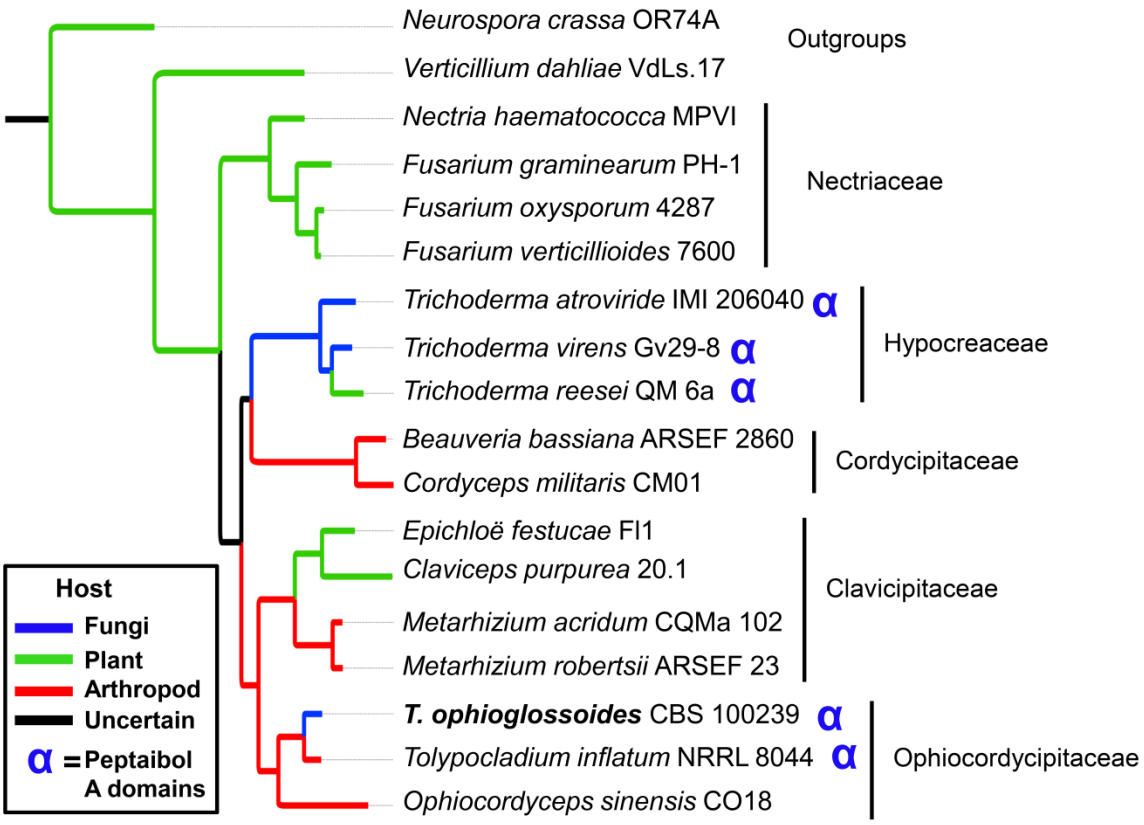


Figure 3.1 ML phylogeny of hypocrealean taxa (with strain information) analyzed in this study. Branches are colored based on host/nutritional association with plants and saprobes of plant material in green. Species that possess peptaibiotic A-domains are denoted by a blue 'α'.

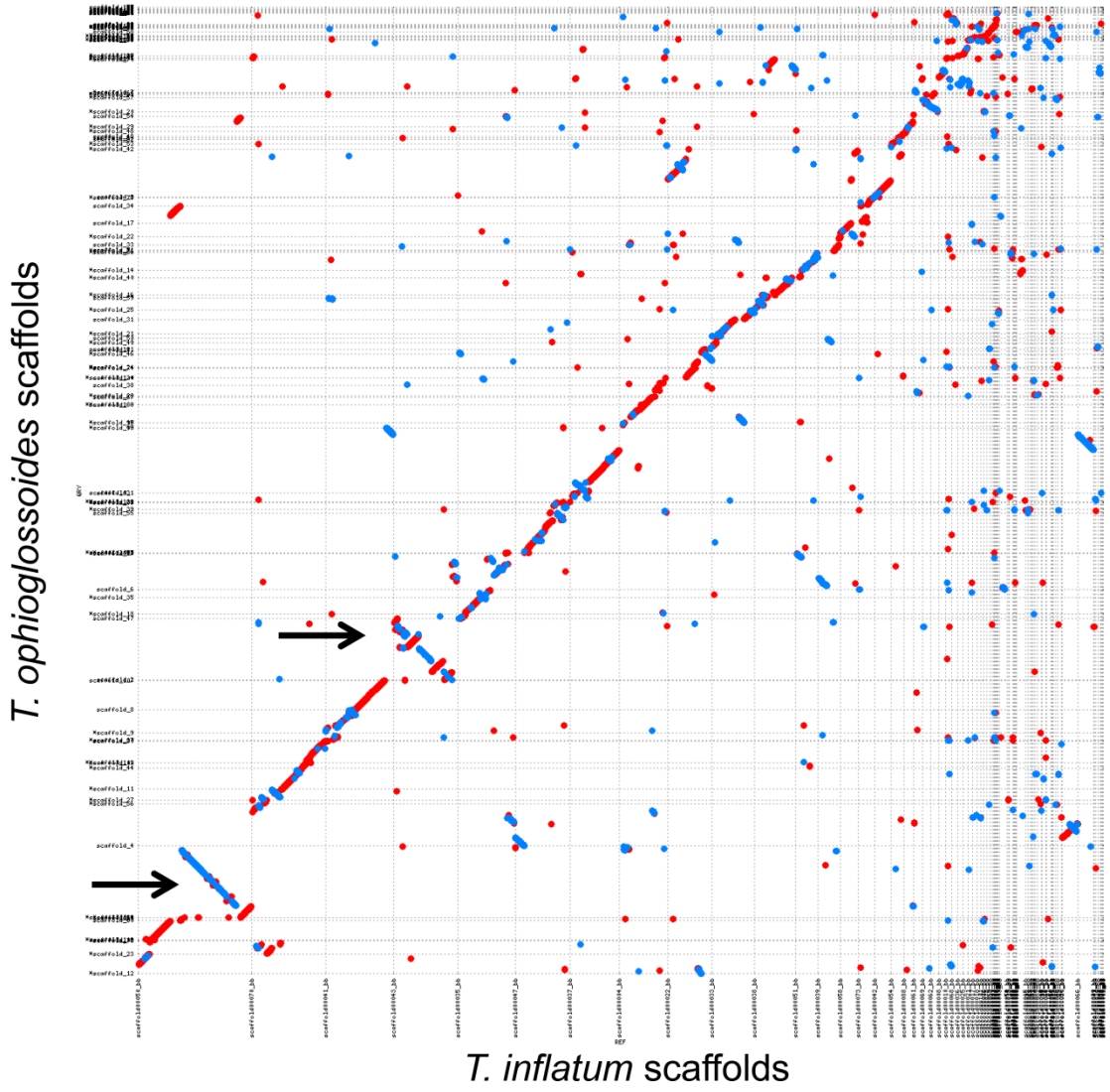


Figure 3.2 Mummerplot visualization of nucmer alignment of all 173 *T. ophioglossoides* scaffolds against the reference 101 *T. inflatum* scaffolds. Red lines represent sequence aligning in the same direction while blue lines represent inversions. The black arrows point out the major inversions between the two genomes.

Figure 3.3 Peptaibiotic focused A-domain phylogeny created using RAxML, showing the 3 major clades. Green branches represent those supported by ≥70% MLBP support. *Tolypocladium* spp. A-domains are colored in blue and teal, and *Trichoderma* spp. A-domains are colored in yellow, red, or orange.

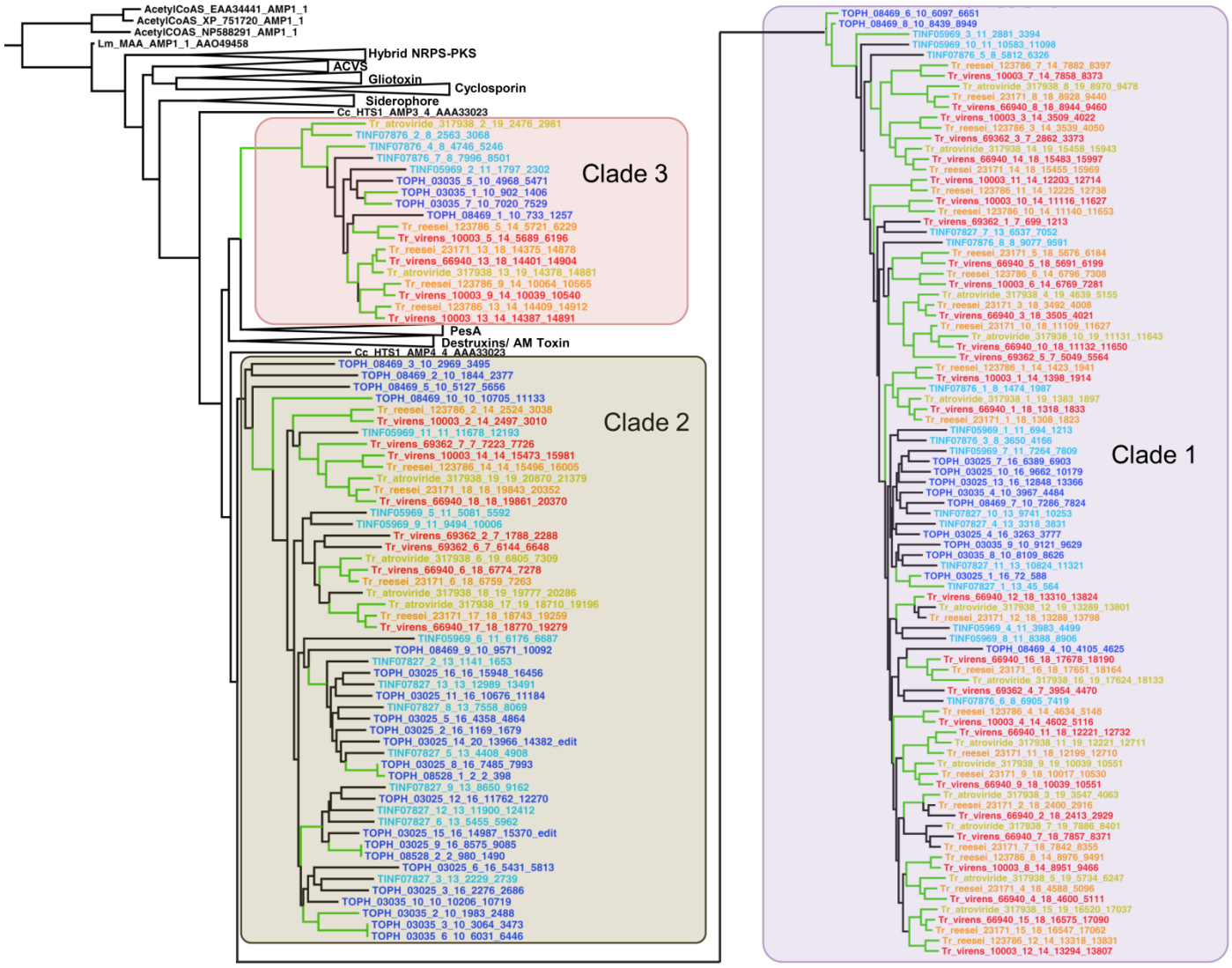

AcetylCoAS_EAA34441_AMP1_1
AcetylCoAS_XP_751720_AMP1_1
AcetylCOAS_NP588291_AMP1_1
Lm_MAA_AMP1_1_AAO49458
Hybrid NRPS-PKS
ACVS
Gliotoxin
Cyclosporin
Siderophore
Cc_HTS1_AMP3_4_AAA33023
Clade 3
Tr_atroviride_317938_2_19_2476_2981
TINF07876_2_8_2563_3068
TINF07876_4_8_4746_5246
TINF07876_7_8_7996_8501
TINF05969_2_11_1797_2302
TOPH_03035_5_10_4968_5471
TOPH_03035_1_10_902_1406
TOPH_03035_7_10_7020_7529
TOPH_08469_1_10_733_1257
Tr_reesei_123786_5_14_5721_6229
Tr_virens_10003_5_14_5689_6196
Tr_reesei_23171_13_18_14375_14878
Tr_virens_66940_13_18_14401_14904
Tr_atroviride_317938_13_19_14378_14881
Tr_reesei_123786_9_14_10064_10565
Tr_virens_10003_9_14_10039_10540
Tr_reesei_123786_13_14_14409_14912
Tr_virens_10003_13_14_14387_14891
PesA
Destruxins/ AM Toxin
Cc_HTS1_AMP4_4_AAA33023
Clade 2
TOPH_08469_3_10_2969_3495
TOPH_08469_2_10_1844_2377
TOPH_08469_5_10_5127_5656
TOPH_08469_10_10_10705_11133
Tr_reesei_123786_2_14_2524_3038
Tr_virens_10003_2_14_2497_3010
TINF05969_11_11_11678_12193
Tr_virens_69362_7_7_7223_7726
Tr_virens_10003_14_14_15473_15981
Tr_reesei_123786_14_14_15496_16005
Tr_atroviride_317938_19_19_20870_21379
Tr_reesei_23171_18_18_19843_20352
Tr_virens_66940_18_18_19861_20370
TINF05969_5_11_5081_5592
TINF05969_9_11_9494_10006
Tr_virens_69362_2_7_1788_2288
Tr_virens_69362_6_7_6144_6648
Tr_atroviride_317938_6_19_6805_7309
Tr_virens_66940_6_18_6774_7278
Tr_reesei_23171_6_18_6759_7263
Tr_atroviride_317938_18_19_19777_20286
Tr_atroviride_317938_17_19_18710_19196
Tr_reesei_23171_17_18_18743_19259
Tr_virens_66940_17_18_18770_19279
TINF05969_6_11_6176_6687
TOPH_08469_9_10_9571_10092
TINF07827_2_13_1141_1653
TOPH_03025_16_16_15948_16456
TINF07827_13_13_12989_13491
TOPH_03025_11_16_10676_11184
TINF07827_8_13_7558_8069
TOPH_03025_5_16_4358_4864
TOPH_03025_2_16_1169_1679
TOPH_03025_14_20_13966_14382_edit
TINF07827_5_13_4408_4908
TOPH_03025_8_16_7485_7993
TOPH_08528_1_2_2_398
TINF07827_9_13_8650_9162
TOPH_03025_12_16_11762_12270
TINF07827_12_13_11900_12412
TINF07827_6_13_5455_5962
TOPH_03025_15_16_14987_15370_edit
TOPH_03025_9_16_8575_9085
TOPH_08528_2_2_980_1490
TOPH_03025_6_16_5431_5813
TINF07827_3_13_2229_2739
TOPH_03025_3_16_2276_2686
TOPH_03035_10_10_10206_10719
TOPH_03035_2_10_1983_2488
TOPH_03035_3_10_3064_3473
TOPH_03035_6_10_6031_6446
Clade 1
TOPH_08469_6_10_6097_6651
TOPH_08469_8_10_8439_8949
TINF05969_3_11_2881_3394
TINF05969_10_11_10583_11098
TINF07876_5_8_5812_6326
Tr_reesei_123786_7_14_7882_8397
Tr_virens_10003_7_14_7858_8373
Tr_atroviride_317938_8_19_8970_9478
Tr_reesei_23171_8_18_8928_9440
Tr_virens_66940_8_18_8944_9460
Tr_virens_10003_3_14_3509_4022
Tr_reesei_123786_3_14_3539_4050
Tr_virens_69362_3_7_2862_3373
Tr_atroviride_317938_14_19_15458_15943
Tr_virens_66940_14_18_15483_15997
Tr_reesei_23171_14_18_15455_15969
Tr_virens_10003_11_14_12203_12714
Tr_reesei_123786_11_14_12225_12738
Tr_virens_10003_10_14_11116_11627
Tr_reesei_123786_10_14_11140_11653
Tr_virens_69362_1_7_699_1213
TINF07827_7_13_6537_7052
TINF07876_8_8_9077_9591
Tr_reesei_23171_5_18_5676_6184
Tr_virens_66940_5_18_5691_6199
Tr_reesei_123786_6_14_6796_7308
Tr_virens_10003_6_14_6769_7281
Tr_atroviride_317938_4_19_4639_5155
Tr_reesei_23171_3_18_3492_4008
Tr_virens_66940_3_18_3505_4021
Tr_reesei_23171_10_18_11109_11627
Tr_atroviride_317938_10_19_11131_11643
Tr_virens_66940_10_18_11132_11650
Tr_virens_69362_5_7_5049_5564
Tr_reesei_123786_1_14_1423_1941
Tr_virens_10003_1_14_1398_1914
TINF07876_1_8_1474_1987
Tr_atroviride_317938_1_19_1383_1897
Tr_virens_66940_1_18_1318_1833
Tr_reesei_23171_1_18_1308_1823
TINF05969_1_11_694_1213
TINF07876_3_8_3650_4166
TINF05969_7_11_7264_7809
TOPH_03025_7_16_6389_6903
TOPH_03025_10_16_9662_10179
TOPH_03025_13_16_12848_13366
TOPH_03035_4_10_3967_4484
TOPH_08469_7_10_7286_7824
TINF07827_10_13_9741_10253
TINF07827_4_13_3318_3831
TOPH_03025_4_16_3263_3777
TOPH_03035_9_10_9121_9629
TOPH_03035_8_10_8109_8626
TINF07827_11_13_10824_11321
TOPH_03025_1_16_72_588
TINF07827_1_13_45_564
Tr_virens_66940_12_18_13310_13824
Tr_atroviride_317938_12_19_13289_13801
Tr_reesei_23171_12_18_13288_13798
TINF05969_4_11_3983_4499
TINF05969_8_11_8388_8906
TOPH_08469_4_10_4105_4625
Tr_virens_66940_16_18_17678_18190
Tr_reesei_23171_16_18_17651_18164
Tr_atroviride_317938_16_19_17624_18133
Tr_virens_69362_4_7_3954_4470
TINF07876_6_8_6905_7419
Tr_reesei_123786_4_14_4634_5148
Tr_virens_10003_4_14_4602_5116
Tr_virens_66940_11_18_12221_12732
Tr_atroviride_317938_11_19_12221_12711
Tr_reesei_23171_11_18_12199_12710
Tr_atroviride_317938_9_19_10039_10551
Tr_reesei_23171_9_18_10017_10530
Tr_virens_66940_9_18_10039_10551
Tr_atroviride_317938_3_19_3547_4063
Tr_reesei_23171_2_18_2400_2916
Tr_virens_66940_2_18_2413_2929
Tr_atroviride_317938_7_19_7886_8401
Tr_virens_66940_7_18_7857_8371
Tr_reesei_23171_7_18_7842_8355
Tr_reesei_123786_8_14_8976_9491
Tr_virens_10003_8_14_8951_9466
Tr_atroviride_317938_5_19_5734_6247
Tr_reesei_23171_4_18_4588_5096
Tr_virens_66940_4_18_4600_5111
Tr_atroviride_317938_15_19_16520_17037
Tr_virens_66940_15_18_16575_17090
Tr_reesei_23171_15_18_16547_17062
Tr_reesei_123786_12_14_13318_13831
Tr_virens_10003_12_14_13294_13807

Figure 3.3

Figure 3.4 Reconciliation of the peptaibiotic A-domain (clades one and two) tree with the species tree. To root the tree, closely related outgroup A-domains in from *Cochliobolus carbonum* (Ccarb), *Fusarium graminearum* (Fgram), *Cordyceps militaris* (Cmil), and *Epichloë festucae* (Efest) were included. Other abbreviations include: Neurospora crassa (Ncrass), *Verticillium dahliae* (Vdah), *F. verticillioides* (Fvert), *F. oxysporum* (Foxy), *Nectria haematococcum* (Nhaem), *Tr. atroviride* (Tatro), *T. reesei* (Trees), *Tr. virens* (Tvir), *Beauveria bassiana* (Bbass), *Claviceps purpurea* (Cpurp), *Metarhizium robertsii* (Mrob), *M. acridum* (Mac), *T. ophioglossoides* (Toph), *T. inflatum* (Tinf), *Ophiocordyceps sinensis* (Osin).

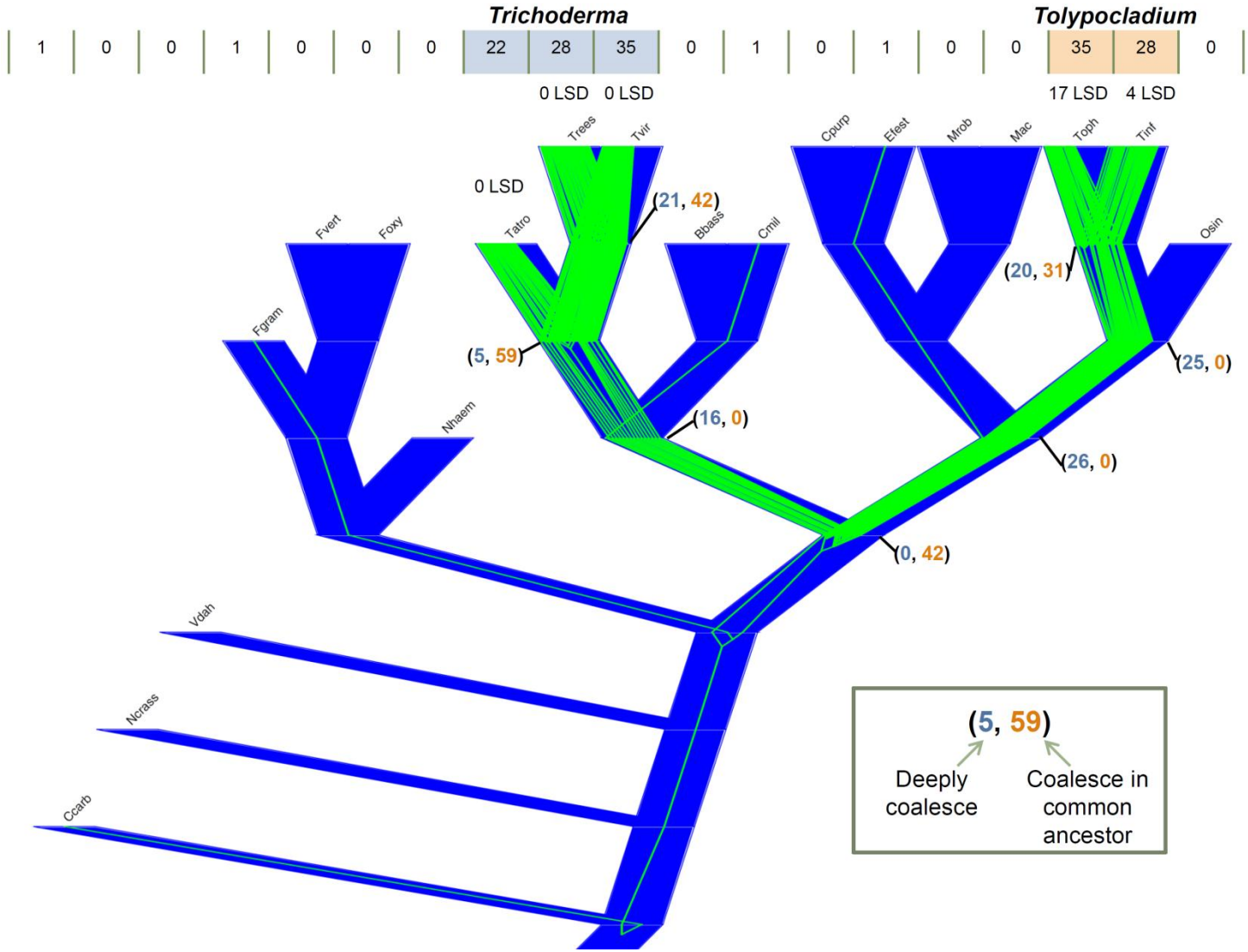

Trichoderma
Tolypocladium
1 0 0 1 0 0 0 22 28 35 0 1 0 1 0 0 35 28 0
0 LSD 0 LSD
17 LSD 4 LSD
Trees
Tvir
Cpurp
Efest
Mrob
Mac
Toph
Tinf
0 LSD
(21, 42)
Fvert
Foxy
Tatro
Bbass
Cmil
Osin
(20, 31)
Fgram
(5, 59)
(25, 0)
Nhaem
(16, 0)
(26, 0)
(0, 42)
Vdah
Ncrass
Ccarb
(5, 59)
Deeply coalesce
Coalesce in common ancestor

Figure 3.4

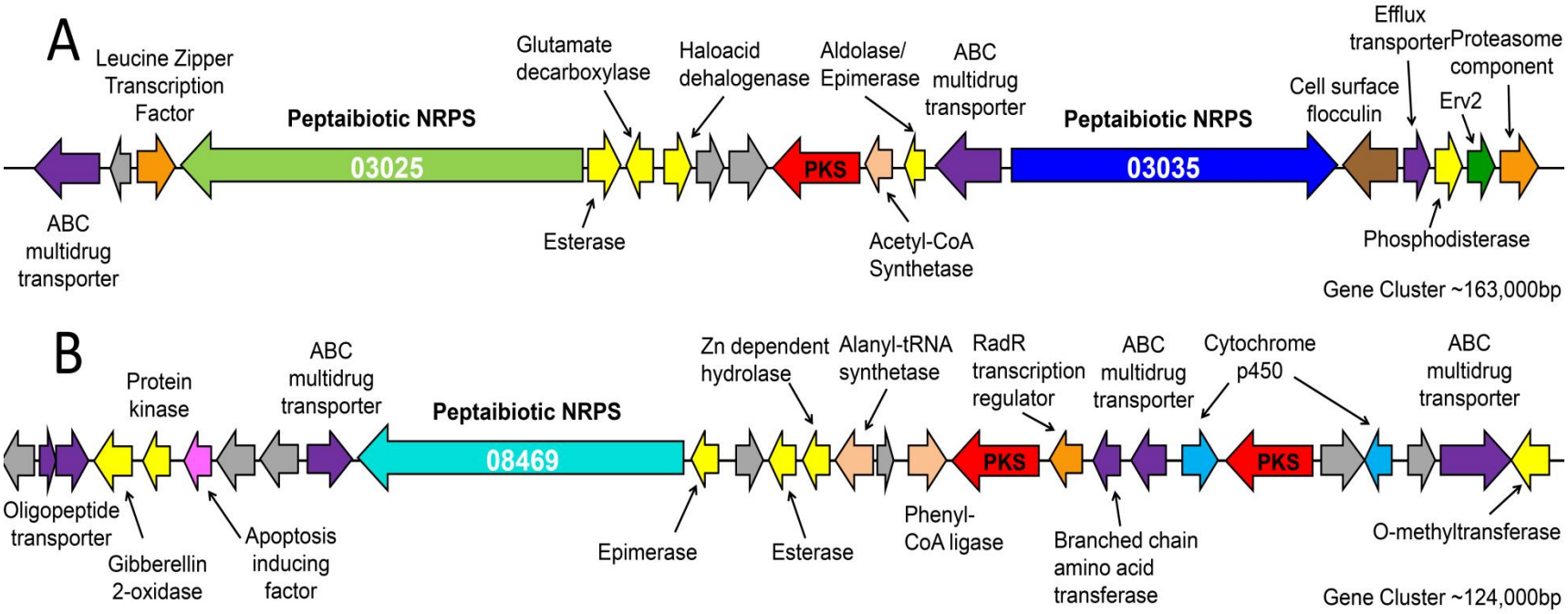


Figure 3.5 Large peptaibiotic clusters *in T. ophioglossoides* as predicted by the antiSMASH pipeline, and putative annotation of genes within these clusters. Proteins without putative functions are colored gray, and size of genes and spacing approximate in scale.

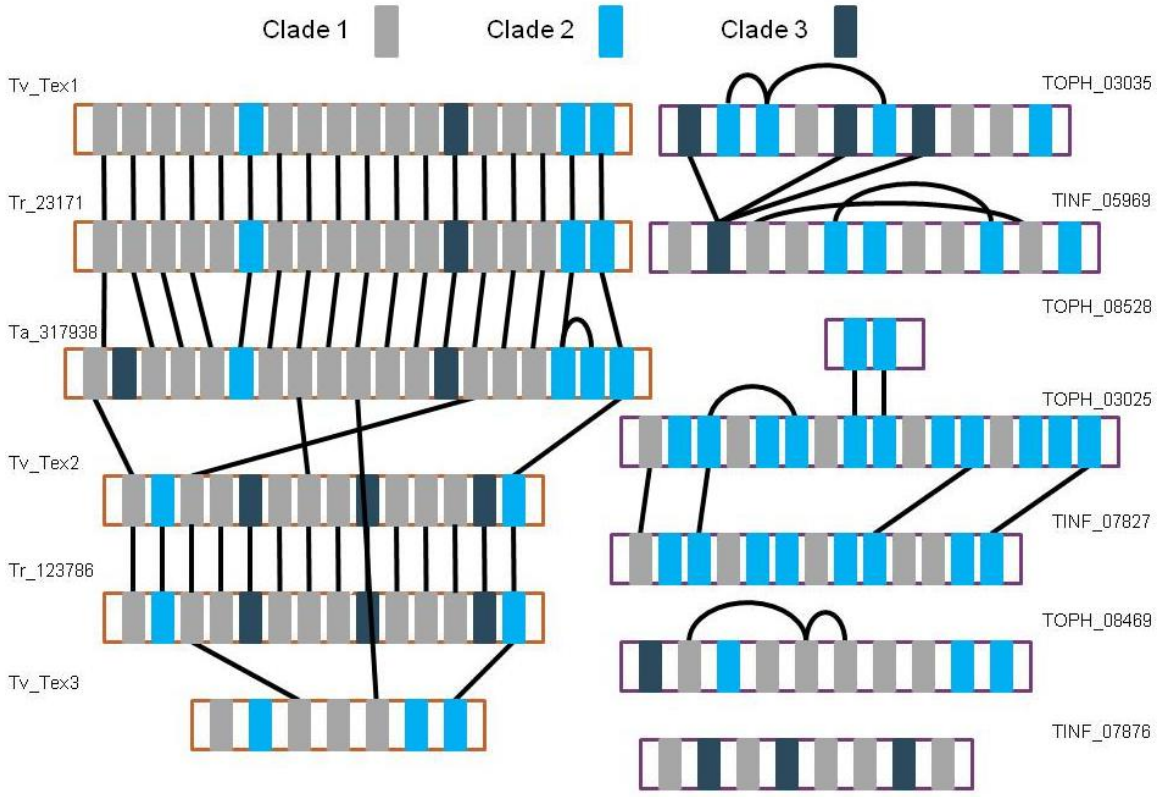


Figure 3.6 A-domain synteny map of for peptaibiotic genes in *Tolypocladium* and *Trichoderma*. Black lines connect orthologous A-domains as determined by MLBP ≥50% in the small peptaibol tree (Figure 3.2).

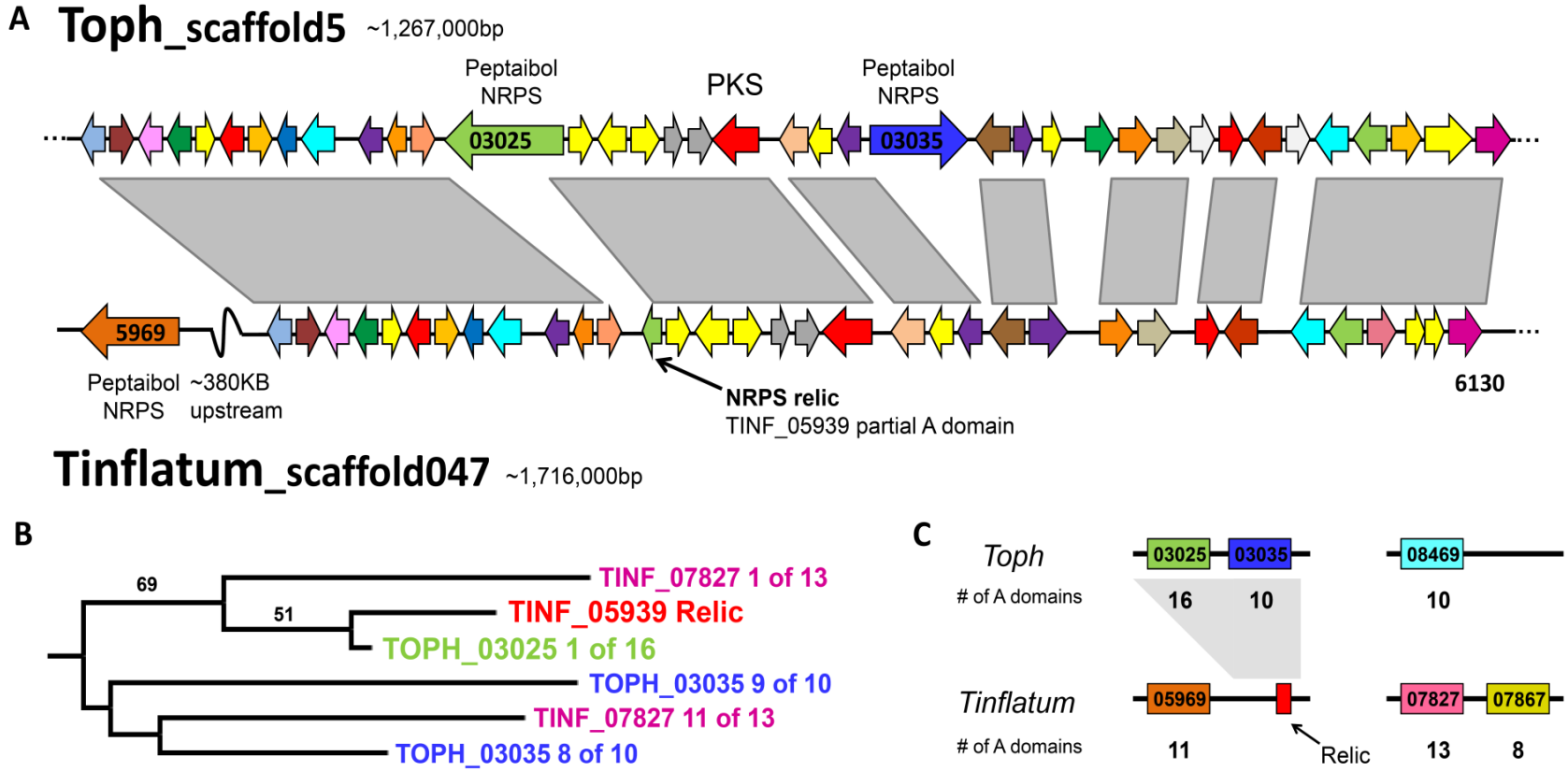


Figure 3.7 Analysis of cluster containing two peptaibiotic genes in *T. ophioglossoides* and corresponding region of the *T. inflatum* genome. A. Alignment showing high degree of synteny between *T. ophioglossoides* scaffold containing two peptaibiotic genes,TOPH_03025 and TOPH_03035, and scaffold 047 in *T. inflatum*, which also contains the peptaibiotic gene TINF05969, that is over 380 kb upstream of the syntenic region. B. Subsection of the large A-domain tree showing the relationship between the TINF05939 relic NRPS and the first A domain of TOPH_03025. C. Map of major peptaibiotic genes within the *Tolypocladium* genomes with the number of modules per gene shown below the scaffold. Gray area represents the syntenic regions between these clusters.

Table 3.1 Genome statistics for *T. ophioglossoides* compared to *T. inflatum* (Bushley *et al.* 2013).

	T. ophioglossoides	*T. inflatum*
Size (Mb)	31.3	30.3
# Scaffolds	173	101
N50	668,222	1,509,745
Longest Scaff.	2,309,933	3,562,345
%GC	57.3	58
Protein-coding genes	10,134	9,998
# SM clusters	38	38

Figure 3.S1 RAxML phylogeny of A-domains mined from the hypocrealean genomes sampled (see Figure 1). A-domains from *T. ophioglossoides* are highlighted in blue.

Figure 3.S1


AcetylCoAS_XP_751720_AMP1_1
AcetylCoAS_EAA34441_AMP1_1
AcetylCOAS_NP588291_AMP1_1
Efest_076270_592_18461_21389_1_3_109_325
FGSG_07659_1_3_80_235
CCM_08018_1_1_2722_3236
MAC_01111_1_1_2803_3322
MAA_05622_1_1_2824_3342
Tr_virens_28610_1_1_2774_3269
Tr_reesei_59315_1_1_2778_3274
FGSG_07798_1_1_2970_3483
FVEG_11086_1_1_2910_3418
Gf_FUSS_AMP1_1_AAT28740
Nhaem_70660_1_1_2791_3303
TINF07675_1_1_3006_3519
MAA_08518_1_1_2971_3484
TINF07495_1_1_3043_3542
TOPH_08247_1_1_3077_3574
Tr_reesei_58285_1_1_2842_3341
MAC_05687_1_1_1286_1759
Bbass_95906_gb_EJP63694.1_1_1_3197_3714
CCM_08261_1_1_3170_3677
MAA_01626_1_1_3014_3525
FVEG_13421_1_1_31_245
FOXG_14587_1_1_2821_3316
Bbass_93204_gb_EJP61198.1_1_1_3058_3564
TOPH_07102_1_1_3078_3581
MAA_08352_2_2_2978_3480
Tr_atro_212102_1_1_2829_3332
VDAG_07928T0_1_1_3112_3583
Efest_008600_52_61077_73766_1_1_2994_3509
MAA_10114_1_1_2982_3481
Tr_virens_47905_1_1_3020_3525
Mgris_Ace1_MGG15097_AMP1_1
TINF04403_1_2_3039_3278
Tr_virens_152250_1_1_675_1164
TINF04403_2_2_3303_3547
CCM_04722_1_1_3083_3584
Bbass_99045_gb_EJP65765.1_1_1_2994_3494
FOXG_03945_1_1_2854_3370
FVEG_11932_1_1_3005_3521
Fh_Equisetin_AMP1_1_AAV66106
TOPH_07403_1_1_2366_2870
FOXG_15296_1_1_3010_3508
FVEG_12610_1_1_2774_3267
Cpurp70120_emb_CCE28893.1_1_1_32_486
Efest_081170_665_25373_29650_1_1_26_486
TINF01692_2_2_139_492
Bbass_95643_gb_EJP63435.1_1_1_53_518
CCM_08771_1_1_58_530
MAA_08622_1_1_31_509
MAC_09549_1_1_31_509
Lmac_MAA_AMP1_1_AAO49458
VDAG_09763T0_1_1_13_519
Nhaem_68854_1_1_13_517
FGSG_06507_1_1_13_517
FOXG_02458_1_1_13_517
FVEG_05643_1_1_13_517
TOPH_08366_1_1_13_517
TINF09755_1_1_13_517
Tr_atro_228361_1_1_13_518
Tr_virens_39004_1_1_13_518
Tr_reesei_81014_1_1_13_522
CCM_02691_1_1_13_518
Bbass_99031_gb_EJP66735.1_1_1_13_518
MAC_06316_1_1_13_517
MAA_07148_1_1_13_517
Cpurp71213_emb_CCE27843.1_1_1_13_518
Efest_062170_439_40786_45444_1_1_14_519
ACVS_AMP3_3_Anid_XP660225
Ac_PCBAB_AMP3_3_P25464
ACVS_AMP1_3_Anid_XP660225
Ac_PCBAB_AMP1_3_P25464
Ac_PCBAB_AMP2_3_P25464
ACVS_AMP2_3_Anid_XP660225
AlphaA_UM01697_AMP1_1
AlphaA_RO3G12433_AMP1_1
VDAG_07112T0_1_1_86_591
TOPH_00033_1_1_23_613
Tr_atro_131513_1_1_33_566
Tr_virens_41206_1_1_22_612
Nhaem_78435_1_1_29_617
MAC_09675_1_1_24_612
AlphaA_Afu4g11240_AMP1_1
Cpurp72176_emb_CCE26861.1_1_1_23_606
Osin531867413_gb_EQL04172.1_1_1_24_606
TINF08448_1_1_24_606
FGSG_06041_1_1_23_602
Ncrass_2289_NCU03010T0_1_1_25_608
AlphaA_MGG_02611_AMP1_1
Bbass_94908_gb_EJP62735.1_1_1_21_607
Nhaem_81329_1_1_507_736
MAA_03758_1_2_20_149
CCM_03390_1_1_20_592
MAA_03758_2_2_199_549
FOXG_01867_1_2_72_260
FVEG_08245_1_1_20_590
Tr_atro_290106_1_2_20_610
Tr_virens_55364_1_1_20_601
Tr_reesei_4117_1_1_20_586
FOXG_11115_1_1_19_587
MAA_01054_2_2_108_455
Osin531862745_gb_EQL00224.1_1_1_26_501
FVEG_03415_1_1_2_387
FOXG_04709_1_1_12_489
CCM_07302_2_2_1107_1329
Efest_109990_1163_2238_4402_1_2_1_262
TINF00159_1_11_490_1008
TINF00159_5_11_6036_6498
TINF00159_2_11_1576_2032
TINF00159_10_11_12620_13062
TINF00159_3_11_3073_3508
TINF00159_8_11_10068_10527
TINF00159_6_11_7513_8006
TINF00159_9_11_11540_12053
Bbass_93382_gb_EJP61332.1_2_3_1525_1981
Bbass_00960_gb_EJP68628.1_2_3_1585_2017
FVEG_09993_3_4_936_1223
FOXG_11647_2_3_1572_2021
Fe_Esyn1_AMP2_2_CAA79245
Tr_atro_30015_2_3_1559_2016
Tr_virens_46961_2_3_1569_2019
TINF00159_11_11_14118_14623
Efest_106340_1047_5531_9418_1_1_29_363
TINF00159_4_11_4543_5004
TINF00159_7_11_8574_9022
FGSG_11989_1_2_1_433
CCM_07302_1_2_929_1097
Nhaem_106280_5_5_4817_5341
Tr_virens_68467_1_8_314_766
Cpurp66504_emb_CCE32633.1_1_5_308_760
Tr_virens_68467_6_8_4404_4859
Tr_virens_68467_3_8_1819_2297
Tr_virens_62540_7_8_6541_6990
TOPH_08872_and8873_6_6_5803_5842_1_549
MAA_10043_6_6_6464_6911
Tr_virens_62540_5_8_5052_5507
TOPH_08872_5_6_4905_5368
MAA_10043_5_6_4989_5444
VDAG_02911T0_1_1_38_518
FVEG_11841_1_1_38_496
FGSG_11294_1_1_37_495
FOXG_13405_1_1_38_496
Nhaem_106273_1_1_38_518
VDAG_07711T0_2_2_303_484
VDAG_07711T0_1_2_32_235
Nhaem_54080_1_1_58_499
FGSG_03245_1_1_58_492
FVEG_08516_2_2_103_355
FOXG_09998_3_3_166_355
Nhaem_106355_1_1_36_503
FGSG_13153_1_1_36_500
CCM_01648_2_2_204_518
Tr_atro_40037_1_1_43_485
Tr_virens_50383_1_1_18_487
Tr_reesei_68204_1_1_18_485
TOPH_02978_1_1_72_536
TINF06128_1_1_3_245
Efest_080450_428_27317_31002_1_1_61_508
Cpurp72481_emb_CCE28828.1_1_1_34_504
MAC_07837_2_2_190_435
MAA_04824_1_1_34_507

Figure 3.S1 (Continued)

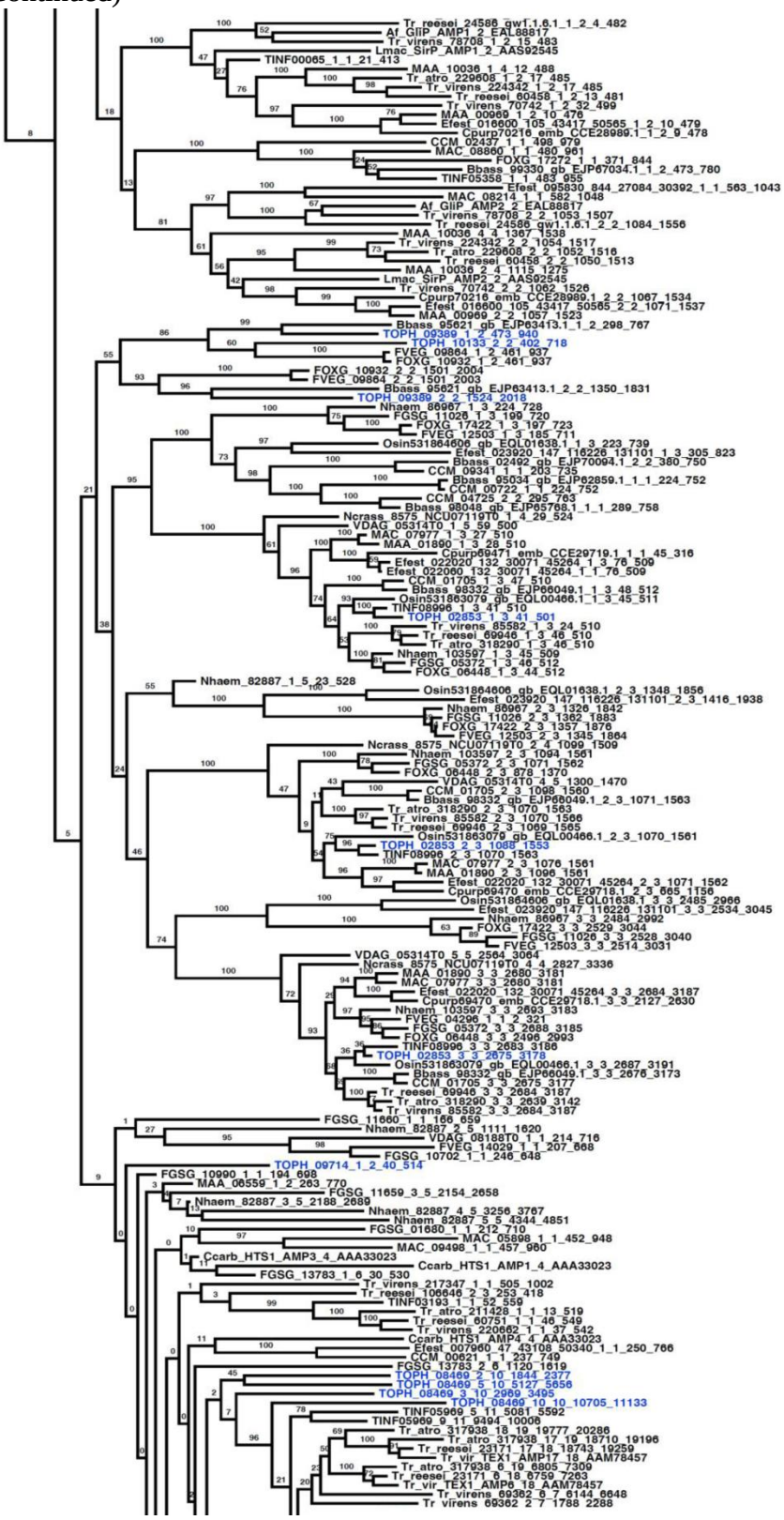


Figure 3.S1 (Continued)

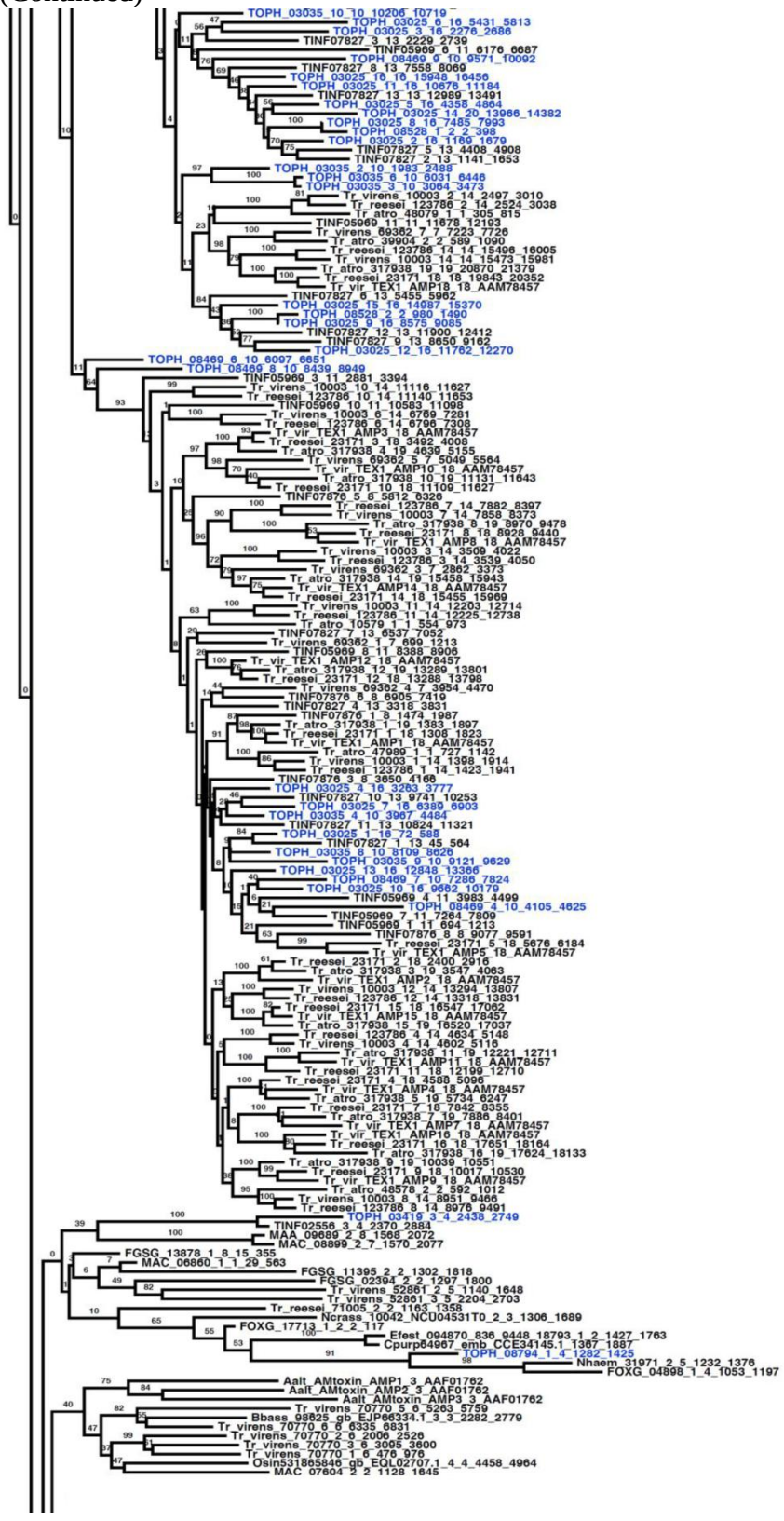


Figure 3.S1 (Continued)

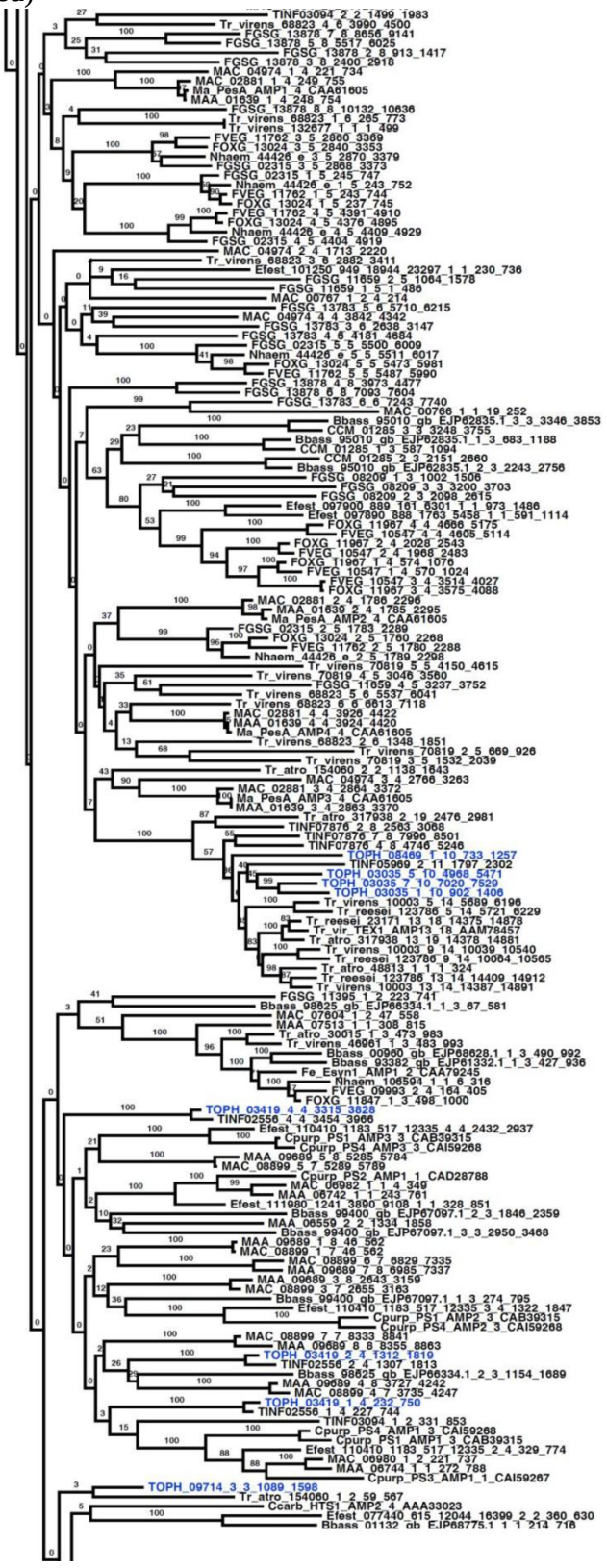
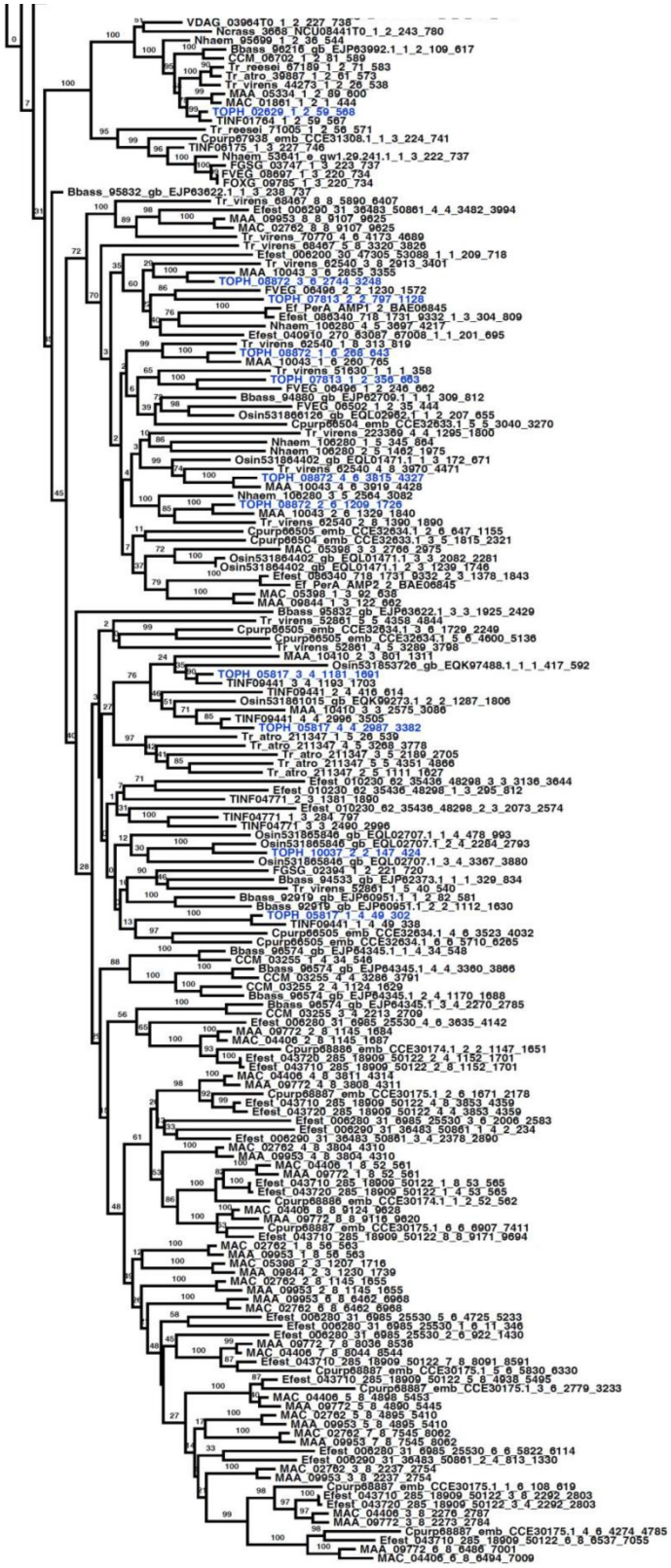


Figure 3.S1 (Continued)



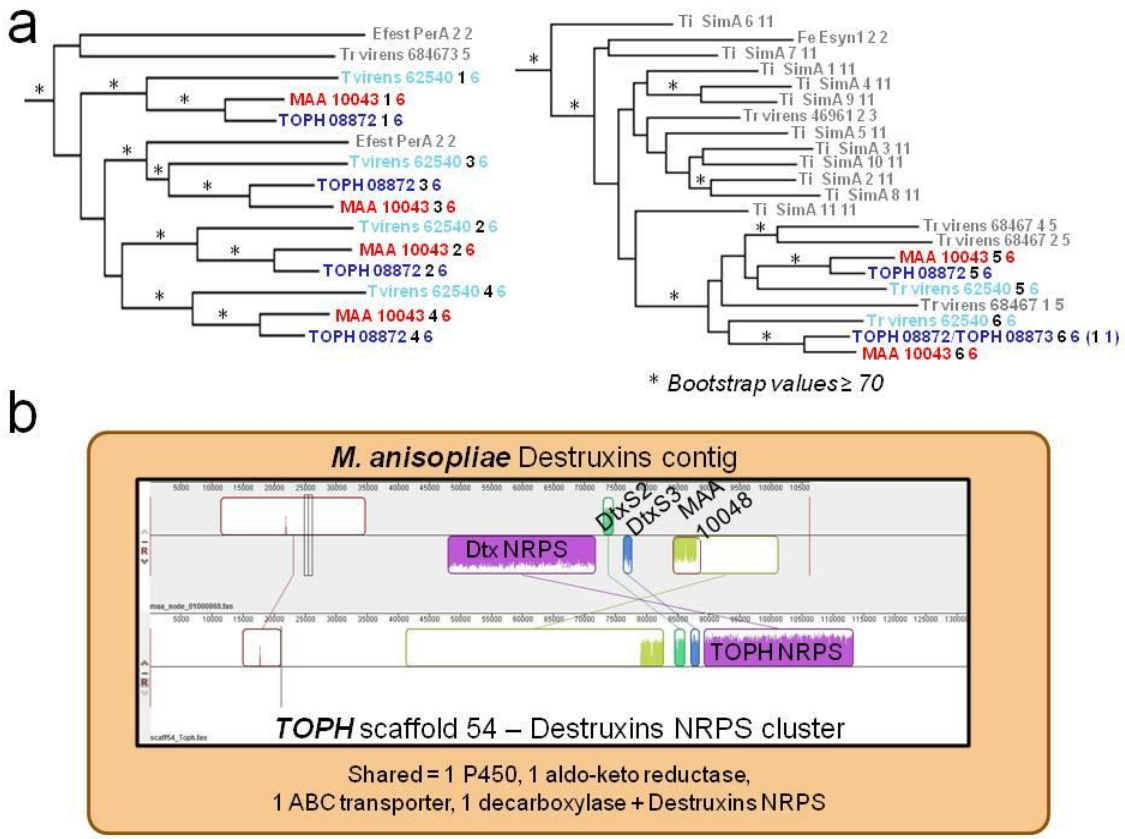


Figure 3.S2 Destruxins NRPS and cluster in *T. ophioglossoides* and other taxa. A. Excerpts from Figure 3.S1 showing the phylogenetic relationships of destruxins A-domains (which group into two areas within the A-domain phylogeny; the PerA clade, and then *T. inflatum simA* clade) in *M. robertsii*, *T. ophioglossoides*, and *Tr. virens*. B. Nucleotide alignment of *M. robertsii* destruxins cluster with homologous cluster in *T. ophioglossoides*.

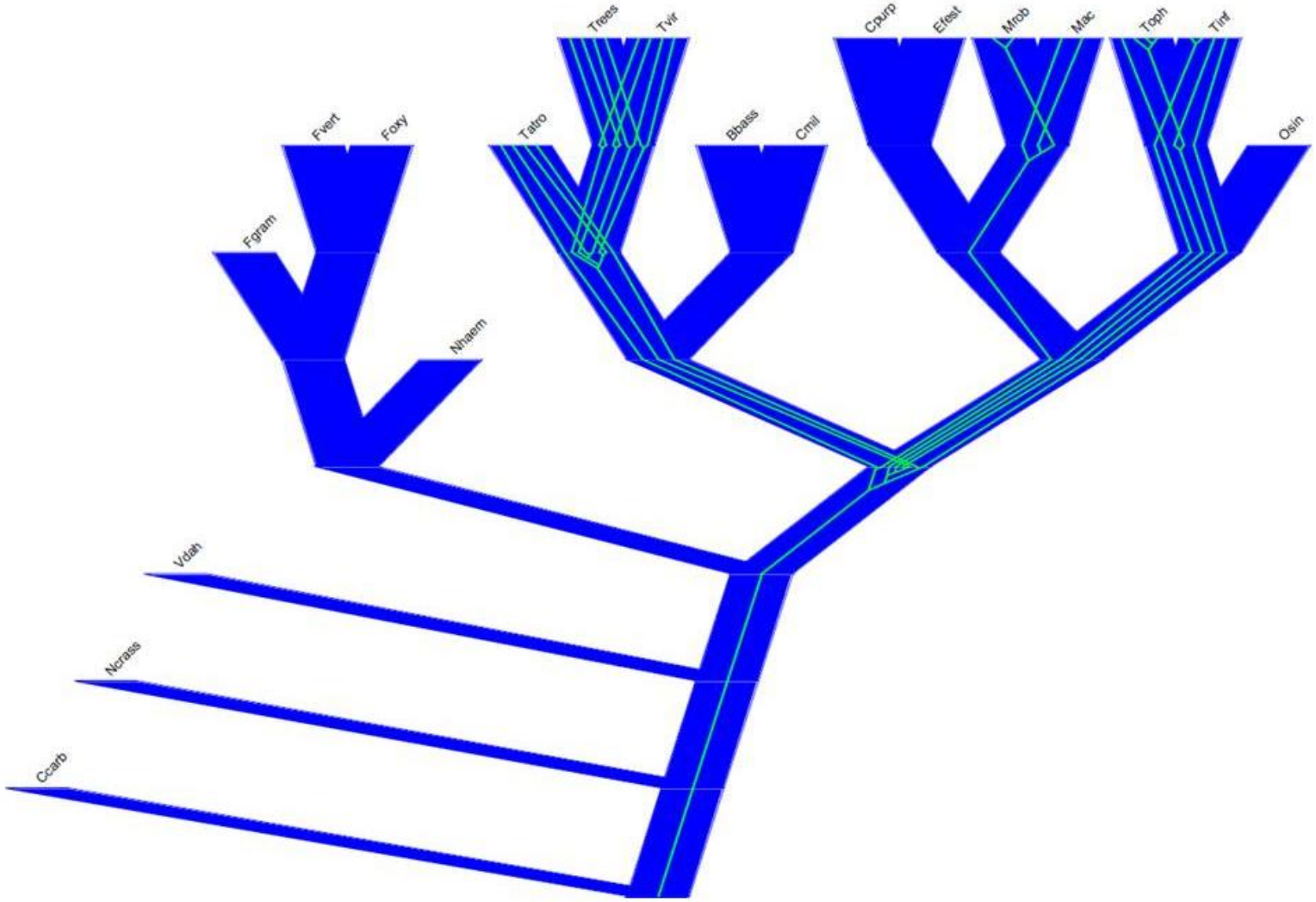


Figure 3.S3 A-domain Clade 3 tree/ species tree reconciliation. Abbreviations as in Figure 3.4.

Figure 3.S4 Alignment of *Trichoderma* peptaibol scaffolds. Top panel depicts alignments of the regions of *Trichoderma* genomes that contain the large (18-19 module) peptaibol genes (which are boxed in blue). Bottom panel shows alignments of scaffold containing the *Trichoderma* 14 module peptaibol genes. Alignment one demonstrates that the region around the 14 module peptaibols in *Tr. reesei* and *Tr. virens* are conserved and syntenic with *Tr. atroviride* contig 25, which does not contain an NRPS. Alignment two shows the same two scaffolds in *Tr. reesei* and *Tr. virens*, but aligned with *Tr. atroviride* contig 29, which does not share synteny with the other two species except for the peptaibol NRPS

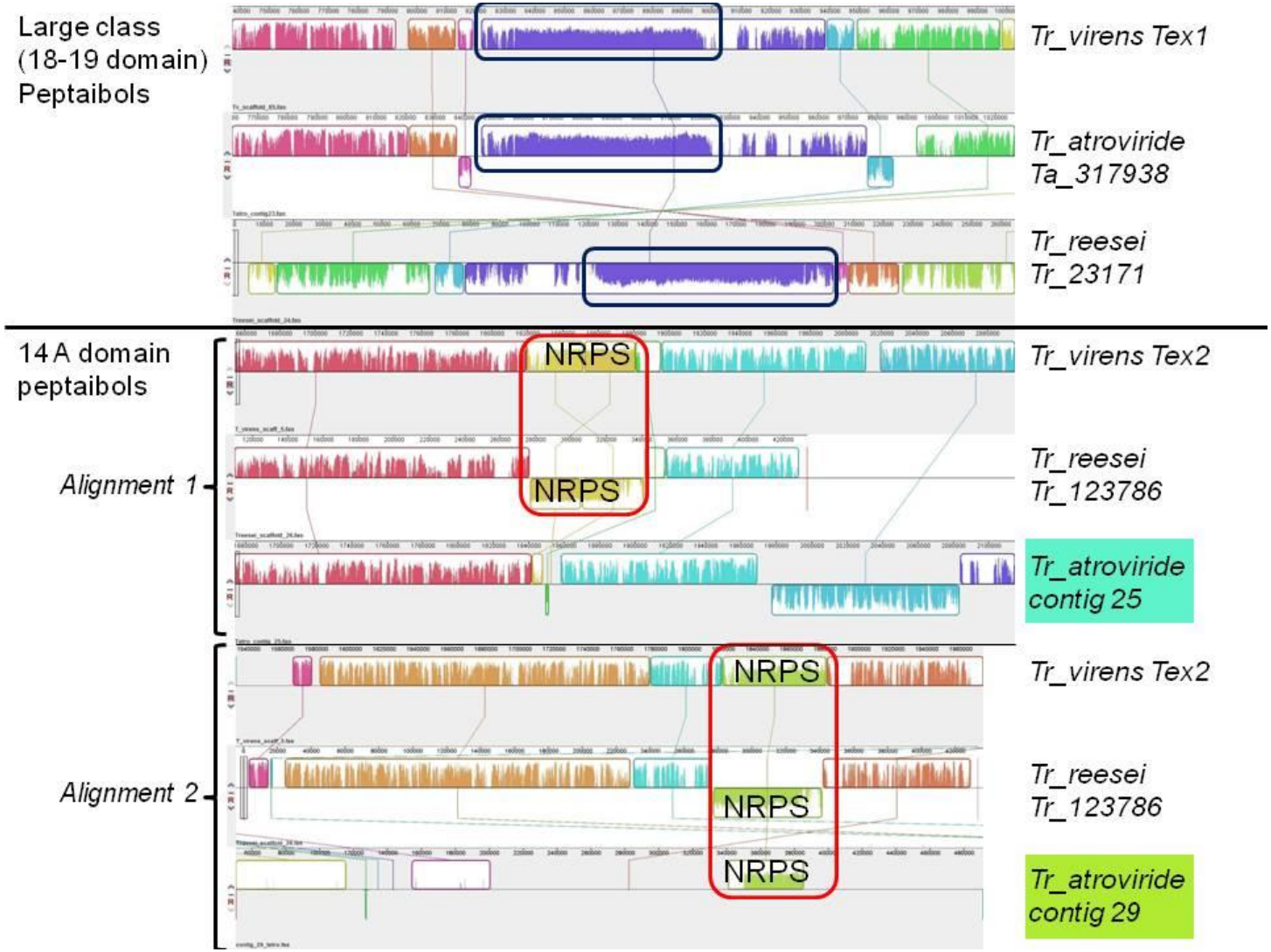

Large class
(18-19 domain)
Peptaibols
Tr_virens Tex1
Tr_atroviride
Ta_317938
Tr_reesei
Tr_23171
14 A domain
peptaibols
Alignment 1
NRPS
NRPS
Tr_virens Tex2
Tr_reesei
Tr_123786
Tr_atroviride
contig 25
Alignment 2
NRPS
NRPS
NRPS
Tr_virens Tex2
Tr_reesei
Tr_123786
Tr_atroviride
contig 29

Figure 3.S4

Table 3.S1. Secondary metabolite genes and clusters in *T. ophioglossoides*. Genes belonging to the same clusters are highlighted in the same color. Cluster prediction based on antiSMASH. Siderophores are marked as either intracellular (i) or extracellular (e). Orthologs of *T. inflatum* (*Ti*) were annotated as reciprocal best BLAST hits and further based on amino acid alignment and A-domain phylogeny.

Protein Model	Description	anti-SMASH	SMURF	HMM A domain?	# A doms predicted	Putative product	*Ti* ortholog?
TOPH_00033	NRPS-like	X		yes	1		
TOPH_00044	PKS	X	x				x
TOPH_00667	PKS-like		x				x
TOPH_01006	PKS	X	x				x
TOPH_01395	PKS	X	x				x
TOPH_01399	PKS	X	x				x
TOPH_01953	Terpene	X					
TOPH_02629	NRPS	X	x	yes	1	Siderophore (e)	x
TOPH_02853	NRPS	X	x	yes	3	Siderophore (i)	x
TOPH_02894	PKS	X	x				x
TOPH_03025	NRPS	X	x	yes	16	Peptaibiotic	
TOPH_03031	PKS	X	x				x
TOPH_03035	NRPS	X	x	yes	10	Peptaibiotic	
TOPH_03217	NRPS-like		x				
TOPH_03419	NRPS	X	x	yes	4		x
TOPH_03459	NRPS-like	X	x				x
TOPH_03628	Terpene	X					
TOPH_04325	Terpene	X					
TOPH_04431	PKS	X	x				
TOPH_05458	PKS	X	x				x
TOPH_05817	NRPS	X	x	yes	3		x

Table 3.S1 (Continued)

TOPH_06344	NRPS-Like	X	x				x
TOPH_06345	PKS		x				x
TOPH_07065	PKS	X	x				
TOPH_07087	PKS	X	x				x
TOPH_07102	Hybrid	X	x	yes	1	Pseurotin-A	
TOPH_07403	Hybrid	X	x	yes	1	Ophiosetin	
TOPH_07700	PKS	X	x				
TOPH_07813	NRPS	X	x	yes	2		
TOPH_08016	PKS	X	x				
TOPH_08020	PKS	X	x				
TOPH_08068	Terpene	X					
TOPH_08168	PKS-like		x				
TOPH_08247	Hybrid	X	x	yes	1		x
TOPH_08386	NRPS-like	X	x	yes	1		x
TOPH_08411	NRPS	X		yes	1		
TOPH_08457	PKS	X	x				x
TOPH_08462	PKS	X	x				
TOPH_08469	NRPS	X	x	yes	10	Peptaibiotic	
TOPH_08528	NRPS	X	x	yes	2		
TOPH_08794	NRPS	X	x				x
TOPH_08872	NRPS	X	x	yes	4	Destruxin	
TOPH_08873	NRPS	X	x	yes	2	Destruxin	
TOPH_08946	NRPS-like	X	x				x
TOPH_09389	NRPS	X	x	yes	2		
TOPH_09714	NRPS	X	x	yes	2		

Chapter 4. Differential expression of genes involved in host recognition, attachment, and degradation in the mycoparasite *Tolypocladium ophioglossoides*.

C. Alisha Quandt, Yanming Di, Justin Elser, Pankaj Jaiswal, Alex Buchanan, Joseph W. Spatafora

Target: *Applied and Environmental Microbiology*

Introduction

Hypocrealean fungi are ecologically diverse and include endophytes, saprobes, and pathogens of plants, insects, and other fungi (Sung *et al.* 2007, Spatafora *et al.* 2007). *Tolypocladium* is a genus in Ophiocordycipitaceae, which is a large family of fungi that are primarily pathogens of insects (Sung *et al.* 2007, Quandt *et al.* Ch. 2). While there are a few species of insect pathogenic *Tolypocladium*, most species in the genus are parasites of truffle fruiting bodies (Quandt *et al.* Ch. 2). The truffle parasitizing members of *Tolypocladium* have a narrow host range restricted to the ectomycorrhizal genus *Elaphomyces* [Eurotiales, Ascomycota]. Despite having very different ecologies, the insect pathogenic and mycoparasitic species of *Tolypocladium* shared a recent common ancestor from approximately 50-70 MYA and appear to be the product of a rapid radiation event (Sung *et al.* 2007, 2008, Spatafora *et al.* 2007, Quandt *et al.* Ch. 2, Quandt *et al.* Ch. 5). *Tolypocladium*, therefore, represents an ideal system for investigating the mechanisms associated with host-jumping between distantly related hosts.

Another larger group of mycoparasites within Hypocreales is the family Hypocreaceae, which includes industrially important species of *Trichoderma* (Samuels 1996, Rossman *et al.* 1999). Several *Trichoderma* spp. are used in biocontrol of plant-pathogenic fungi (Schuster & Schmoll 2010) and studies have examined which genes are expressed during growth on plant-pathogenic fungi. Hartl *et al.* (2009), for example, used expressed sequence tags (ESTs) from a custom *Trichoderma atroviride* library and found several classes of proteins to be upregulated during mycoparasitism, including those involved in post-translational processing, amino acid metabolism, and catabolism of lipids and aminosugars. Reithner *et al.* (2001) found differential expression of several

proteases when cultures of *Tr. atroviride* were grown antagonizing the plant pathogen *Rhizoctonia solani*. Another study used RT-PCR and chitinase-specific primers to determine which chitinases in *Tr. virens* were upregulated during mycoparasitism (see below; Gruber *et al.* 2011). A logical follow up to these studies would be a high-throughput direct RNA sequencing (RNA-Seq) study examining differential expression of *Trichoderma* spp. when grown on fungal hosts, but to date no such study has been published.

A crucial step in parasitizing a specific host involves the ability to recognize its specific host and differentiate it from other objects it encounters. G-protein-coupled receptors (GPCRs) are seven transmembrane domain proteins embedded in the cell membrane of eukaryotes which play critical roles in sensing the environment and signal transduction of downstream regulatory cascades (Rosenbaum *et al.* 2009). In fungi, one kind of GPCR, referred to as PTH11-related, is abundant in pathogenic species of the class Sordariomycetes (Kulkarni *et al.* 2005, Zheng *et al.* 2010). In the insect pathogen specialist *M. acridum*, a close relative of *Tolypocladium*, different PTH11-related receptors were found to be upregulated under host and non-host conditions, while the generalist insect pathogen, *M. roberstii,* used a single PTH11-related GPCR to recognize both hosts (Gao *et al.* 2011). It is currently unknown if GPCRs function in a similar way in mycoparasites and the recognition of their hosts.

Attachment to host tissue is also an important process in pathogenicity. In EST studies of *M. robertsii* grown on insect hemolymph and root exudate (Wang *et al.* 2005), two adhesins, MAD1 and MAD2, were identified and highly expressed. A subsequent study demonstrated that, (1) mutant conidia with *Mad1* deleted were unable to adhere to

locust wings, (2) similar *Mad2* mutants were unable to attach to onion epidermis, and (3) that *Mad1* deletion also interfered with conidial germination and hyphal differentiation (Wang & St. Leger 2007). These findings suggest that MAD1 is important for insect attachment in *M. robertsii*, while MAD2 is necessary for interactions with plants and the rhizosphere. Studies in other fungal pathogens including *Candida* spp. and *Coccidioides immitis* have demonstrated the importance of adhesins in pathogenicity (Cormack *et al.* 1999, Hung *et al.* 2002, Argimón *et al.* 2007).

The correlation between the quantity of certain carbohydrate active enzymes (CAZymes) (Henrissat & Davies 1997) and life history has been well demonstrated in fungi (Floudas *et al.* 2012, Sharpton *et al.* 2009, Martin *et al.* 2008, *etc.*). Understanding how these quantitative differences arise (*e.g.*, lineage specific losses/expansions, incomplete lineage sorting) can provide more information about the utilization of novel host resources. Chitinases, specifically those in CAZyme family GH18 are abundant in members of the mycoparasitic lineage Hypocreaceae that have been sequenced (Kubicek *et al.* 2011). Within GH18 proteins, there are three subgroups that vary in their possession of carbohydrate binding modules (CBMs) (Boraston *et al.* 2004). CBMs enhance chitin binding and greatly improve the efficiency of chitinase activity (Seidl *et al.* 2005). In particular, chitinases in subgroup C contain CBMs 18 and 50 (Hartl *et al.* 2012), which some studies suggest may be especially important for mycoparasitism (Gruber *et al.* 2011). In *Tr. virens*, differential expression of chitinases was specifically examined during various environmental and nutritional conditions as well as in confrontation assays with fungal hosts (Gruber *et al.* 2011). A subset of the

fungal/bacterial Class V chitinases in the subgroup C were found to be upregulated during mycoparasitic conditions and in media containing cell walls of fungal hosts.

In addition to using a variety of nutritional sources, hypocrealean species are also prolific producers of secondary metabolites. Some of these metabolites are thought to play an important role in pathogenesis. For example, destruxins are cyclic peptides produced by *Metarhizium spp.*, which are known to have insecticidal activity. Also, peptaibiotics, produced by the members of the Hypocreaceae (Whitmore & Wallace 2004), form ion channels and have antifungal and antibiotic activity by inserting into cell membranes and disrupting membrane potential (Chugh & Wallace 2001). One prevailing hypothesis proposes that the numerous secondary metabolites produced by *Trichoderma* spp. play a role in the mycoparasitism process (Schrimböck *et al.* 1994, Röhrich *et al.* 2012). However, studies examining the initial stages of mycoparasitism in *Tr. atroviride* and *Tr. harzianum* revealed that genes involved in secondary metabolism were down-regulated (Vinale *et al.* 2009, Reithner *et al.* 2011).

To understand the evolution of mycoparasitism in *Tolypocladium*, it is key to understand the suite of genes that are differentially regulated when parasites first encounter their hosts. To this aim, the sequenced strain of *T. ophioglossoides* was grown under different growth conditions, including media containing its *Elaphomyces* host, and media containing insect tissue (the ancestral host affiliation for the genus), as well as a standard "reference" medium. Analysis of differential expression was used to determine which genes are important in *Elaphomyces* recognition with relevant comparisons to related insect pathogens and mycoparasites.

Materials and Methods

Media preparation

Yeast Malt (YM) broth was prepared by dissolving 4 grams (g) of yeast extract and 20 g of malt extract in 1 Liter of deionized water. Minimal media was made following the recipe in St Leger *et al.* (1999). Truffles of *Elaphomyces muricatus* were cleaned of debris and surface sterilized using 85% ethanol for 30 seconds. A spatula was used to physically remove the warty outer peridium; the truffle was then cut in half and as much of the gleba as possible was removed from the peridium using a sterilized spatula. Peridium and gleba tissues were lyophilized separately in 1.5 mL Eppendorf tubes, and the peridium was ground using a sterilized mortar and pestle. Using a sterilized stir bar and magnetic stirrer, 0.25 g of gleba and peridium were separately mixed into 25 mL of MM to produce 1% concentration solutions in separate 50 mL flasks and autoclaved for 15 minutes, resulting in two media conditions, *E. muricatus* peridium (EMP) and *E. muricatus* gleba (EMG). For the cuticle media (CUT), fifth instar larvae of *Otiorhyncus sulcatus* (Black Vine Weevil) were dissected and decapitated using razors and sterilized spatulas were used to remove the organs and scrape as much of the hemolymph off of the cuticle as possible. Then following the protocol from Andersen (1980), cuticle was soaked in 1% potassium tetraborate solution for 24 hours to remove any hemolymph or epidermal cells and dried completely at room temperature. After grinding with mortar and pestle in liquid nitrogen, the insect cuticle was added to 25 mL of MM in 50 mL flasks at the same concentration as the *Elaphomyces* media, and autoclaved.

Growth experiment

Strain CBS 100239 of *T. ophioglossoides* was grown on Corn Meal Agar for 1 week to produce ample conidia. Plates were washed with molecular grade water, and filtered with Miracloth (EMD Millipore, Germany) to remove hyphae. A hemocytometer was used to estimate spore concentration, and 20uL of spore solution containing $3.6*10^4$ conidia was added to 2 mL of YM media in 3 mL wells in 12 multi-well (24 well) covered, individually-packaged disposable sterile plates (BD Falcon). After 7 days of growth, the plates were centrifuged for one minute to pull hyphae and spore tissue to the bottoms of the wells, and then a micropipette with sterile filter tips was used to remove YM media from all the wells. Samples used for the YM growth condition were harvested at this time point. The tissue in the remaining wells was rinsed with molecular grade water, and then 2 mL of the alternative media was added to sets of replicate wells. Alternative media (EMP, EMG, and CUT) were then added in 2 mL aliquots to 3 replicate wells per condition (Figure 1), resulting in 12 biological samples (4 treatments X 3 replicates). After 24 hours, tissue was harvested. To harvest tissue, plates were centrifuged for 30 seconds and a micropipette was used to remove as much media as possible. *T. ophioglossoides* tissue was transferred via pipette to 1.5 mL Eppendorf tubes and frozen in liquid nitrogen and then immediately transferred to -80°C where samples were held until RNA extraction.

Nucleic acid preparation and sequencing

RNA was extracted using the Qiagen RNeasy Plant kit following the manufacturer's protocol. The Illumina TruSeq RNA Sample Preparation Kit v2 was used for RNA library construction, using manufacturer suggested protocols, including Agencourt AMPure magnetic beads for cleaning steps. The twelve RNA libraries

described above were barcoded with the TruSeq kit adapters, multiplexed twelve deep per lane, and sequenced across three technical replicates (Auer & Doerge 2010) resulting in a total of 36 individual RNA-Seq datasets. Sequencing reactions were conducted for 51 single-end cycles on the Illumina HiSeq 2000 at the Center for Genome Research and Biocomputing at Oregon State University.

Bioinformatic analyses

Illumina-generated reads were filtered based on Illumina quality flagging using the casava filtering programs from the Short Read Toolbox (Knaus 2012; http://brianknaus.com), and the first and last five base pairs (bp) were removed. This produced 45 bp trimmed and filtered RNA reads. The reads were aligned to the reference genome sequence of *T. ophioglossoides* using Bowtie 2 v. 2.0.6 with default settings (Langmead & Salzberg 2012). Unique reads were counted and sorted using a custom set of perl scripts. For each of the three biological replicates, there are three technical replicates. Verification of count variation among technical replicates was near Poisson distribution, suggesting it was safe to combine the counts from the technical replicates. For assessing differential expressed genes (DEG), the software package NBPSeq (Di *et al.* 2013) in R (R Core Team 2014) was used to fit negative binomial regression models to RNA read counts where one of the regression coefficient corresponds to the log fold change between two treatments. DEGs were assessed by testing whether that regression coefficient was 0 using a likelihood ratio test with higher-order asymptotic adjustment (Di 2014). The DEG test was performed on each gene separately and false discovery rates (q-values) were estimated according to Storey and Tibshirani (2003). The DEG analysis was repeated for each pair of the *T. ophioglossoides* replicates for EMP, EMG,

YM, and CUT. A cutoff for DEGs was set at a false discovery rate q-value ≤ 0.10. Reads per thoused base pairs per million reads (RPKM) was calculated using the following formula: ((total reads across all replicates) / (the length of a gene (in bp) / 1000)) / (total reads for that treatment across all loci / 1,000,000).

To identify Interpro domains of *T. ophioglossoides* peptides, InterProScan v 5.44 (http://code.google.com/p/interproscan/wiki/Interproscan5_44_ReleaseNotes) (Jones *et al.* 2014) analysis was performed on the Discovery Environment of the iPlantCollaborative (https://de.iplantcollaborative.org/de/). The returned results were processed through customized scripts (provided by Cathy Gresham from Mississippi State University) to generate the Gene Ontology (GO) assignments. InParanoid 7 (http://inparanoid.sbc.su.se/cgi-bin/index.cgi) was run to identify homologs (orthologs and paralogs) between *T. ophioglossoides* and the reference *Saccharomyces cerevisae*, *Neurospora crassa* and *Schizosaccharomyces pombe*. The resulting species pair homologs against the reference species (3 of them, which have the most number of experimentally validated gene function assignments) were used to project similarity based GO projections for *T. ophioglossoides*. The BiNGO 2.44 plugin (Maere *et al.* 2005) of Cytoscape 2.8 (Smoot *et al.* 2011) was used to identify statistically overrepresented GO categories of DEGs. Heatmap clustering in R was based on default Euclidean distances as a part of the heatmap.2 program.

Hidden Markov models (HMMs) from dbCAN (Yin *et al.* 2012) were used to annotate CAZymes including GH18 family chitinases in the *T. ophioglossoides* and other hypocrealean genomes. PTH11-related GPCRs were identified using a custom hidden Markov model based on those PTH11-related GPCRs identified by Gao *et al.* (2011) in

Metarhizium anisopliae and *M. acridum* created for this study using the program Hmmer 3.0 (Eddy 2011). Prediction of the number of transmembrane helices associated with the PTH11-related proteins was performed by the TMHMM Server v. 2.0 (Krogh *et al.* 2001). The SignalP 4.1 webserver identified eukaryotic signal peptides with default cutoffs (Petersen *et al.* 2011). Additional annotations of DEGs were made using BLAST searches against the NCBI non-redundant protein (nr) database with a minimum e-value of 1e-10. Secondary metabolite genes and clusters in *T. ophioglossoides* were based on those identified in Quandt *et al.* (Ch. 3).

Phylogenetic analysis of protein families were performed in the following manner. First, sequences were aligned using MUSCLE v 3.8.31 (Edgar 2004) under default settings; gaps were removed manually. Alignments were then analyzed using RAxML v 7.2.6 (Stamatakis 2006) using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates.

Results and Discussion

The growth conditions EMP, EMG, YM, and CUT resulted in 199 million, 164 million, 182 million, and 214 million raw reads, respectively, after filtering for quality (Table S1). Using bowtie at default setting, 48.9% of these reads aligned to the genome of *T. ophioglossoides*, and 55.8% were unique alignments. Thus, there were 55.5 million, 66.6 million, 44.9 million, and 113.9 million uniquely aligned reads for EMP, EMG, YM, and CUT, respectively. A large number of the reads aligned ambiguously (i.e. they aligned at two or more locations within the genome), a result that is not surprising given

the short read length of 45 bp. Still, despite this, there were more than a sufficient number of unambiguously aligned reads for downstream analyses. A larger percentage of uniquely aligned reads were obtained from fungi grown in cuticle media; possibly due to differences in library quality. The differences in read counts between the conditions and replicates were accounted for in the statistical modeling used to identify DEGs.

Of the 10,135 *T. ophioglossoides* single copy genes, 9,659, or 95.3% were expressed under one or more of the growth conditions in this experiment. The most commonly expressed genes in all conditions were several heat shock proteins, a circadian clock controlled protein, a DNA topoisomerase, translation elongation factor 1-α, several genes involved in the citric acid cycle, and two proteins with no known function (Table 1). A great number of these were differentially expressed in the growth conditions examined.

This experiment identified many differentially expressed genes between the growth conditions examined (Table S2). The largest difference in gene expression was observed between EMG and CUT (47.6% of genes were differentially expressed), meaning that almost half of the transcriptome is differentially expressed between these two growth conditions. Similarly, 37% of genes are differentially expressed in EMP compared to CUT. These large differences in expression between the *Elaphomyces* containing media and the insect cuticle media highlight the difference between the natural fungal host of *T. ophioglossoides* and the ancestral insect host of the genus *Tolypocladium*. The fewest number of DEGs (284, 3.1% of genes) were found between EMG and EMP, the two media containing *Elaphomyces* tissue. Expression changes between EMP, EMG, and CUT compared to YM, the rich media condition, were 8.6%,

16.3%, and 14.3% of genes, respectively. Of the 133 genes upregulated on EMP compared to both YM and CUT, only 3 were identified as orphans, having no homolog in the NCBI nr database or closely related taxa. One of these orphans (TOPH_00374) has a signal peptide and is only 99 amino acids in length indicating it may be a small secreted protein possibly involved in host-pathogen interactions, and it has moderate expression levels in EMP (reads per kilobase per million reads mapped [RPKM] = 12.3).

Differential expression of PTH11-related G-protein coupled receptors

Several PTH11-related GPCRs were identified in *T. ophioglossoides* and other related insect pathogens including *T. inflatum* and *O. sinensis* of Ophiocordycipitaceae, and *M. robertsii* and *M. anisopliae* of the sister family, Clavicipitaceae (Figure 2). Growth on EMP and CUT elicited different patterns of expression of these putative GPCRs in *T. ophioglossoides*. Specifically, TOPH_07673 and TOPH_08741 were upregulated on EMP compared to cuticle. Of these, TOPH_08741 is part of a statistically well supported clade (Figure 2) containing only two other proteins; one from the moth pathogen, *O. sinensis*, and the other, a *T. ophioglossoides* paralog that is not differentially expressed between growth conditions. TOPH_07673 has one ortholog in *T. inflatum* and one in *M. acridum*, but not in *M. robertsii* or *O. sinensis*.

Several PTH11-related GPCRs were upregulated on CUT as compared to EMP, including TOPH_00772, TOPH_01767, TOPH_06678, TOPH_07233, and TOPH_08210. One of these, TOPH_01767 is orthologous to the PTH11-related GPCR, MAC_00494, of *M. acridum* that was upregulated on cuticle of locust, its natural host (Figure 3). Encountering insect cuticle elicits the expression of the same receptor in both

M. acridum and *T. ophioglossoides*, despite the loss of insect pathogenesis in the truffle parasite.

The combination of the large number of PTH11-related GPCRs and their differential expression supports the hypothesis that they function in detection of different hosts of *Tolypocladium*. The upregulation of a *T. ophioglossoides* ortholog of a GPCR identified as highly expressed in a distantly related insect pathogen encountering insect tissue provides additional evidence for the PTH11-related GPCR host-recognition hypothesis.

Chitinase expression

Tolypocladium ophioglossoides possesses numerous GH18 chitinases in both Class III and Class V. One chitinase, TOPH_09828, was statistically upregulated during growth on both EMP and EMG compared to YM and CUT (Figure 3). This chitinase, while having undergone divergence at the sequence level and despite lacking carbohydrate binding modules (CBMs), is phylogenetically related to subgroup C of Class V chitinases. It possesses a signal peptide, which suggests that this chitinase is excreted into the extracellular environment, where it presumably degrades exogenous chitin. There is no homolog of this chitinase in the beetle pathogen, *T. inflatum*, but TOPH_09828 is an ortholog of one of the four subgroup C chitinases identified (Trive 112097) by Gruber *et al.* (2011) as upregulated in *Tr. virens* when grown during mycoparasitism and on fungal cell walls. It is not surprising that a *T. ophioglossoides* chitinase is upregulated on *Elaphomyces* tissue, as chitinases play a crucial role in breaking down fungal cell walls (Siedl 2008). The fact that this chitinase is closely

related to those found to be upregulated during mycoparasitism by *Trichoderma* chitinases (Figure 3; Gruber *et al.* 2008) suggests two things. First, this chitinase may also play role in mycoparasitism (in Hypocreales mycoparasites), especially given the log (base2) fold change, 1.8x and 2.0x, between conditions and the moderately high RPKM, 36 and 37, on the media containing *Elaphomyces*, EMP and EMG, respectively, compared to YM. Second, the shared ancestry of this particular gene, suggests that the common ancestor of these two genera (and the four most divergent families to which they belong) may have had some mycoparasitic ability. This is intriguing because mycoparasites are found in all of the most divergent families (Kepler *et al.* 2012), and previous studies using a large phylogenetic sampling found equivocal support for the ancestral character state of the ancestor of these families as either pathogens of insect, fungi, or plants (Spatafora *et al.* 2007, Sung *et al.* 2008)

The most highly expressed chitinase under all conditions was TOPH_05538, a Class III chitinase that does not possess any CBMs or a signal peptide (Figure 3). It was the 33rd most common transcript in EMP, 30th in EMG, 35th in YM, and 56th in CUT. Statistically, TOPH_05538 was upregulated in EMP and EMG compared to CUT, but not to YM. The lack of a signal peptide suggests this highly expressed chitinase may play an important role in hyphal growth by restructuring the endogenous chitin of *T. ophioglossoides*' cell wall.

Other CAZymes upregulated in EMP and EMG compared to YM and CUT include two β-1,3-glucanases, which break bonds in a common component of cell walls of filamentous fungi, β-1,3-glucans (polymers of glucose). Both of these glucanases have signal peptides indicating they are secreted proteins, but one (TOPH_03534) is annotated

as part of the CAZy family GH64 which are endo-acting and implicated in fungal cell wall degradation, whereas the other (TOPH_03859) is exo-acting and posseses an extra chitinase-like domain. β-1,3-glucans are one of the most abundant carbohydrates in the cell walls of *Aspergillus fumigatus* (Latgé *et al.* 2005, Latgé 2007, 2010), a close relative of *Elaphomyces*, the host of *T. ophioglossoides*, and expression of β-1,3-glucanases is hypothesized here to facilitate *T. ophioglossoides* parasitism of its host fungus.

Expression of Adhesins

One of the most highly expressed *T. ophioglossoides* genes in EMP was the ortholog (TOPH_02818) of the *M. robertsii* adhesin gene, *Mad1*. Statistically, *Mad1* is upregulated on both EMP and EMG compared to the other growth conditions (Figure 4). MAD1 was identified in *M. robertsii* because of high expression during insect pathogenesis (Wang *et al.* 2005), and later it was found to be involved in both cytoskeletal orientation and blastospore formation (Wang & St. Leger 2007). Blastospores are a yeast-like growth morphology frequently taken on by fungi when growing in liquids, where filamentous growth can be restricted, and it is unlikely *T. ophioglossoides* produces blastospores during its infection of *Elaphomyces* truffles in nature. While this experiment was conducted exclusively using liquid media (where blastospore growth could have been induced but was not monitored during the experiment), the *Mad1* homolog was still statistically upregulated on *Elaphomyces* media (EMP and EMG) compared to the other conditions. This indicates that MAD1 may play a role in host infection that is not solely involved in changes in growth morphology. The upregulation of *Mad1* on EMP and EMG could be related to adhesion to *Elaphomyces*

cells. It is also possible that the *T. ophioglossoides* MAD1 homolog plays a role in increased overall growth on its natural host and progression of the cell cycle. This is supported by the fact that all studies examining *Mad1* expression including this one, have found it to be highly expressed regardless of condition (Wang *et al.* 2005a, Wang & St. Leger 2007). Furthermore, if MAD1 is necessary for normal cellular functioning this could explain why its deletion in *M. robertsii* caused a reduction in insecticidal ability.

In contrast, the *T. ophioglossoides* ortholog of *Mad2* (TOPH_07267) had low expression under all of the growth conditions tested and not differentially expressed (Figure 4). This protein has been proposed to promote attachment to plant surface (Wang & St. Leger 2007), and this was a condition not tested in our experiment, but this gene is not upregulated in the presence of *Elaphomyces*.

Overrepresented gene ontologies

BiNGO analysis of the GO categories assigned to *T. ophioglossoides* DEGs identified three major categories of GOs that were statistically overrepresented in EMP compared to YM (Figure 5). Two GO IDs, 55114 and 16491, corresponding to the ontologies oxidation reduction and oxidoreductase, respectively, and are part of a large group of redox-related GOs overrepresented in *T. ophioglossoides* genes upregulated in EMP (Figure 5, Table S2). This result suggests that certain types of oxidative stress may be employed by *T. ophioglossoides* when it encounters its host and could play a role in its mycoparasitic ability.

The second most overrepresented GO during *T. ophioglossoides* growth on EMP is FMN binding (10181) (Table S2). Flavin mononucleotide (FMN) is the coenzyme of

flavoprotein oxidoreductase enzymes, and this GO category is a child of many parent GOs, including cofactor binding, signal transduction activity, receptor activity, and ion binding. It is possible this category of proteins is involved in redox reactions as well, and one of the proteins in this category is annotated as cytochrome p450 (TOPH_05009). Upregulation of enzymes involved in oxidation and reduction has been observed in studies examining fungi growing on, or near, their hosts (Digustini *et al.* 2007). Some studies have examined oxidoreductases used by plant-pathogenic fungi for detoxification of the host environment (Idnurm & Howlett 2001). Also a high number of both cytochrome p450s and monoxygenases were expressed in *Metarhizium* spp. grown on host tissue (Gao *et al.* 2011).

The other major group of statistically overrepresented GO categories that were upregulated in EMP are related to transmembrane transport (Figure 5). Specifically, transmembrane transporters of amino acids, organic acids, and carboxylic acids were overrepresented. The transport of amino acids and their derivatives is crucial to cellular processes including energy generation, cell wall synthesis, and intercellular signaling (Saier 2000), indicating an increase in *T. ophioglossoides* growth and proliferation. Overall growth of tissue was not quantified at harvesting, but based on coarse visual inspection of the plates, the wells containing EMP had the most tissue at the end of the experiment.

Expression patterns of secondary metabolite genes

Based on DEGs and the RPKMs for each of the core secondary metabolite (SM) genes, secondary metabolism appears to be reduced when *T. ophioglossoides* is grown on

tissue containing its host (Figure 6, Figure S1, Table S3). The most highly expressed SM core genes were an NRPS-like gene (TOPH_03459) and two terpene synthases (TOPH_03628 and TOPH_04325). The NRPS-like SM gene is statistically upregulated in YM compared to the other conditions (Table S3). There are no known products for these gene clusters.

Some of the secondary metabolite core genes are upregulated exclusively during growth on insect cuticle, while others are equally expressed in YM and CUT. Notable clusters that are upregulated on cuticle include almost the entire peptaibiotic cluster around the NRPS gene (TOPH_08469), including the two PKS genes in that cluster (Figure S1). Other SM core genes that are upregulated in CUT compared to EMP and EMG include a terpene synthase (TOPH_08068), an NRPS (TOPH_05817), a PKS (TOPH_01395), and a putative extracellular siderophore synthetase NRPS (TOPH_02629) (Table S3). The upregulation of SM genes when *T. ophioglossoides* is grown on insect cuticle could be a response to a stressful, or potentially foreign environment, or could represent an ancestral response to the insect environment. Secondary metabolism has not been a topic of discussion in studies examining genes expressed on insect tissue in *M. anisopliae* (Wang *et al.* 2005a, 2005b), but several studies have reported secondary metabolite production during infection of insect hosts in that species (Roberts 1981, Wang *et al.* 2012). The low expression of *T. ophioglossoides* secondary metabolite core genes during growth on *Elaphomyces* media mirrors reports from mycoparasitism assays in *Tr. atroviride*. The cause of this differential expression remains to be determined, but *T. ophioglossoides* must survive in soil where it interacts with a large number of microorganisms including bacteria, and secondary metabolism

could be involved in defense against these. There are also many bacteria living within the fruiting body of *Elaphomyces* (Quandt *et al.* Ch. 6), and it could be that during growth on its host in nature, *T. ophioglossoides* may express higher amounts of secondary metabolites, but this was not addressed within the context of this study.

Conclusions

Simulating growth under different host conditions provides insights into how pathogenic fungi perceive and begin to antagonize host tissue. Almost half of the genes in the *T. ophioglossoides* genome were differentially expressed depending on whether the growth media contained its host, the truffle *Elaphomyces*, or an unnatural host, the beetle, *Otiorhyncus*. This experiment provides the first glimpse into mycoparasitic interactions in a non-*Trichoderma* system. Some patterns of *T. ophioglossoides* expression mirror those seen in *Trichoderma* mycoparasitism assays including an upregulation of similar types of chitinases in subgroup C of Class V and a downregulation of secondary metabolism. PTH11-related GPCRs were found to be differentially expressed under the different growth conditions, and two GPCRs were identified as candidate receptors involved in *Elaphomyces* recognition. Analysis of GO categories that were overrepresented when *T. ophioglossoides* was grown on its host, revealed that redox reactions and transmembrane transport of amino acids and their derivates were dominant. The adhesin gene, *Mad1*, was highly expressed during all growth conditions, but was statistically upregulated in media containing *Elaphomyces* as compared to the other media. *Mad2* showed low expression under all conditions. These findings suggest that

T. ophioglossoides differentially uses a number of genes and gene categories when faced with variable growth media. Some of these responses may represent expression of pliesiomorphic responses to other hosts (e.g., insects) but many appear to be specific to current hosts: *Elaphomyces*.

Acknowledgements

The authors thank Mark Desanko for providing technical support for sequencing in this project. We also acknowledge the contributions of the members of the North American Truffling Society (Corvallis, OR) for their assistance in gathering *Elaphomyces* sporocarps and Amanda Lake for *O. sulcatus* larvae used in this study. This project was supported with funds from the NSF grants DEB-0732993 to JWS, and DEB-1311572 to JWS and CAQ, and an NSF Graduate Research Fellowship funded CAQ.

References

Argimón S, Wishart JA, Leng R, Macaskill S, Mavor A, Alexandris T, Nicholls S, Knight AW, Enjalbert B, Walmsley R, Odds FC, Gow NAR, Brown AJ (2007) Developmental regulation of an adhesin gene during cellular morphogenesis in the fungal pathogen *Candida albicans*. *Eukaryotic cell* **6**: 682-692.

Auer PL Doerge RW (2010) Statistical design and analysis of RNA sequencing data. *Genetics* **185**: 405-416.

Boraston AB, Bolam D, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* **382**: 769-781.

Chugh JK, Wallace BA (2001) Peptaibols: models for ion channels. *Biochemical Society Transactions* **29**: 565-570.

Cormack BP, Ghori N, Falkow S (1999) An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* **285**: 578-582.

Di Y (2014) Single-gene negative binomial regression models for RNA-Seq data with higher-order asymptotic inference" *Statistics and Its Interface*. Accepted.

Di Y, Emerson SC, Schafer DW, Kimbrel JA, Chang JH (2013) Higher order asymptotics for negative binomial regression inferences from RNA-sequencing data. *Statistical Applications in Genetics and Molecular Biology* **12**: 49-70.

DiGuistini S, Ralph SG, Lim YW, Holt R, Jones S, Bohlmann J, Breuil C (2007) Generation and annotation of lodgepole pine and oleoresin-induced expressed sequences from the blue-stain fungus *Ophiostoma clavigerum*, a Mountain Pine Beetle-associated pathogen. *FEMS Microbiology Letters* **267**: 151-158.

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology **7***: e1002195.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Gorecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kues U, Arun Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Duenas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St. John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev I, Hibbett DS (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715-1719.

Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie X-Q, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma L-J, St. Leger RJ, Zhao G-P, Pei Y, Feng M-G, Xia Y, Wang C (2011) Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics* **7**: 1-18.

Gruber S, Vaaje-Kolstad G, Matarese F, López-Mondéjar R, Kubicek CP, Seidl-Seiboth V (2011) Analysis of subgroup C of fungal chitinases containing chitin-binding and LysM modules in the mycoparasite *Trichoderma atroviride*. *Glycobiology* **21**: 122-133.

Hartl L, Zach S, Seidl-Seiboth V (2012) Fungal chitinases: diversity, mechanistic properties, and biotechnological potential. *Applied Microbiology and Biotechnology* **93**: 533-543.

Henrissat B, Davies GJ (1997) Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7**: 637-644.

Hung CY, Yu JJ, Seshan KR, Reichard U, Cole GT (2002) A parasitic phase-specific adhesin of *Coccidioides immitis* contributes to the virulence of this respiratory fungal pathogen. *Infection and Immunity* **70**: 3443-3456.

Idnurm A, Howlett BJ (2001) Pathogenicity genes of phytopathogenic fungi. *Molecular Plant Pathology* **2**: 241–255.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M,

Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification.*Bioinformatics* **30**: 1236-1240.

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Dohren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gomez-Rodriguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernandez-Onate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lubeck M, Lubeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, and Grigoriev IV (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology* **12**: R40.

Kulkarni RD, Thon MR, Pan H, Dean RA (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biology* **6**: R24.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Latgé (2007) The cell wall: carbohydrate armour for the fungal cell. *Molecular Microbiology* **66**: 279-290.

Latgé (2010) Tasting the fungal cell wall. *Cellular Microbiology* **12**: 863-872.

Latgé JP, Mouyna I, Tekaia F, Beauvais A, Debeaupuis JP, Nierman W (2005) Specific molecular features in the organization and biosynthesis of the cell wall of *Aspergillus fumigatus*. *Medical Mycology* **43**: S15-S22.

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**: 3448-3449.

Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buee M, Brokstein P, Canback B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbe J, Lin YC, Legue V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculity-Hirzel H, Oudot-Le Secq MP, Quesneville H, Rajaskekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kues U, Lucas S, Van de Peer Y, Podila GK, Polles A, Pukkila PJ, Richardson PM, Rouze P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-93.

Östlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer EL (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Research* **38**: D196-D203.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reithner B, Ibarra-Laclette E, Mach RL, Herrera-Estrella A (2011) Identification of mycoparasitism-related genes in *Trichoderma atroviride*. *Applied and Environmental Microbiology* **77**: 4361-4370.

Roberts (1981) Toxins of Entomopathogenic Fungi. In: *Microbial control of pests and plant diseases*. (Burges HD, ed.): 441-463 London, New York: Academic Press.

Röhrich CR, Iversen A, Jaklitsch WM, Voglmayr H, Berg A, Dörfelt H, Thrane U, Vilcinskas A, Nielsen KF, Von Döhren H, Brückner H, Degenkolb T (2012) Hypopulvins, novel peptaibiotics from the polyporicolous fungus *Hypocrea pulvinata*, are produced during infection of its natural hosts. *Fungal Biology* **116**: 1219-1231.

Rosenbaum DM, Rasmussen SG, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* **459**: 356-363.

Saier Jr MH (2000) Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology* **146**: 1775-1795.

Samuels GJ (1996) *Trichoderma*: a review of biology and systematics of the genus. *Mycological Research* **100***:* 923-935.

Schirmböck M, Lorito M, Wang YL, Hayes CK, Arisan-Atac I, Scala F, Harmon GE, Kubicek CP (1994) Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. *Applied and Environmental Microbiology* **60**: 4364-4370.

Schuster A, Schmoll M (2010). Biology and biotechnology of *Trichoderma*. *Applied Microbiology and Biotechnology* **87**: 787-799.

Seidl V (2008) Chitinasess of filamentous fungi: a large group of diverse proteins with multiple physiological functions. *Fungal Biology Reviews* **22**: 36-42.

Seidl V, Huemer, B, Seiboth, B, Kubicek CP (2005) A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS Journal* **272**: 5923-5939.

Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, Maiti R, Kodira CD, Neafsey DE, Zeng Q, Hung C-Y, McMahan C, Muszewska A, Grynberg M, Mandel MA, Kellner EM, Barker BM, Galgiani JN, Orbach MJ, Kirkland TN, Cole GT, Henn MR, Birren BW, Taylor JW (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Research* **19**: 1722-1731.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**: 431-432.

Spatafora JW, Sung G-H, Sung J-M, Hywel-Jones NL, White JF (2007) Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. *Molecular Ecology* **16**: 1701-1711.

St. Leger RJ, Nelson JO, Screen SE (1999) The entomopathogenic fungus *Metarhizium anisopliae* alters ambient pH, allowing extracellular protease production and activity. *Microbiology* **145**: 2691-2699.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies," *Proceedings of the National Academy of Sciences* **100**: 9440–9445.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5–59.

Sung G-H, Poinar GO, Spatafora JW (2008) The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal–arthropod symbioses. *Molecular Phylogenetics and Evolution* **49**: 495-502.

Vinale F, Ghisalberti EL, Sivasithamparam K, Marra R, Ritieni A, Ferracane R, Woo S, Lorito M (2009) Factors affecting the production of *Trichoderma harzianum* secondary metabolites during the interaction with different plant pathogens. *Letters in applied microbiology* **48**: 705-711.

Wang C, Hu G, St Leger RJ (2005) Differential gene expression by *Metarhizium anisopliae* growing in root exudate and host (*Manduca sexta*) cuticle or hemolymph reveals mechanisms of physiological adaptation. *Fungal Genetics and Biology* **42**: 704-718.

Wang C, St Leger RJ (2005) Developmental and transcriptional responses to host and nonhost cuticles by the specific locust pathogen *Metarhizium anisopliae* var. *acridum. Eukaryotic Cell* **4**: 937-947.

Wang C, St Leger RJ (2007) The MAD1 adhesin of *Metarhizium anisopliae* links adhesion with blastospore production and virulence to insects, and the MAD2 adhesin enables attachment to plants. *Eukaryotic cell* **6**: 808-816.

Whitmore L, Wallace BA (2004) The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Research* **32**: D593-D594.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: W445-W451.

Zheng H, Zhou L, Dou T, Han X, Cai Y, Zhan X, Tang C, Huang J, Wu Q (2010) Genome-wide prediction of G protein-coupled receptors in *Verticillium* spp. *Fungal Biology* **114**: 359-368.

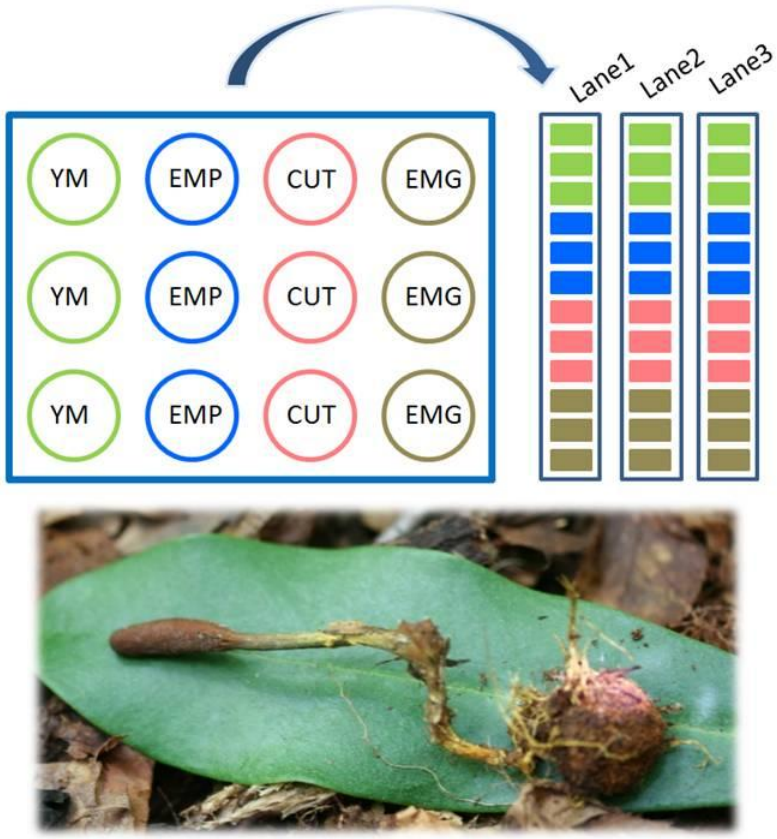
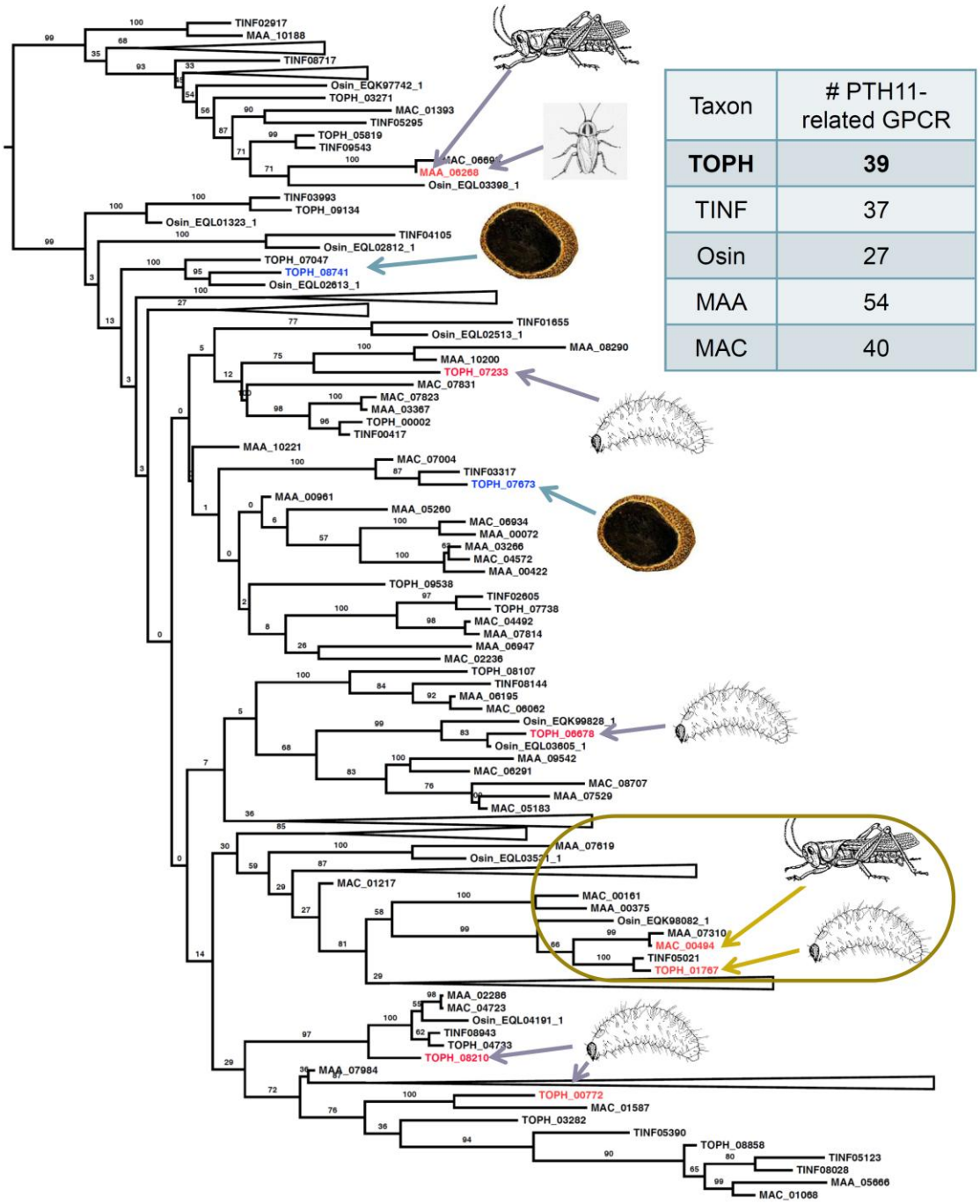


Figure 4.1. Experimental design for the study, showing biological and technical replication, and an image of *T. ophioglossoides* infecting and fruiting from its host, *Elaphomyces* sp., in nature. Yeast Malt (YM), minimal media containing *Elaphomyces muricatus* peridium (EMP), minimal media containing insect cuticle (CUT), and minimal media containing *E. muricatus* gleba (EMG).

Figure 4.2. PTH11-related GPCR phylogeny and numbers per taxon sampled. Genes in *T. ophioglossoides* that are upregulated on EMP are colored blue and marked with a truffle, and those that are upregulated on CUT are colored red and marked with a beele larva. Genes in *Metarhizum* spp. that are upregulated on grasshopper or roach hosts are colored red and marked with the host(s) eliciting expression of those genes. Abbreviations: *T. ophioglossoides* (TOPH), *T. inflatum* (TINF), *Ophiocordyceps sinensis* (Osin), *M. acidrum* (MAC), *M. robertsii* (MAA). Illustration references: *Elaphomyces* (Latvian Nature, Askolds Klavins, Ltd. Gandrs and authors, http://www.latvijasdaba.lv/), Grasshopper (Pearson Scott Foresman), Cockroach (Snodgrass, http://en.wikipedia.org/wiki/Cockroach), *Otiorhyncus sulcatus* larva (http://ipm.ncsu.edu/AG189/html/Black_Vine_Weevil.HTML).

Figure 4.2



Taxon	# PTH11-related GPCR
TOPH	**39**
TINF	37
Osin	27
MAA	54
MAC	40

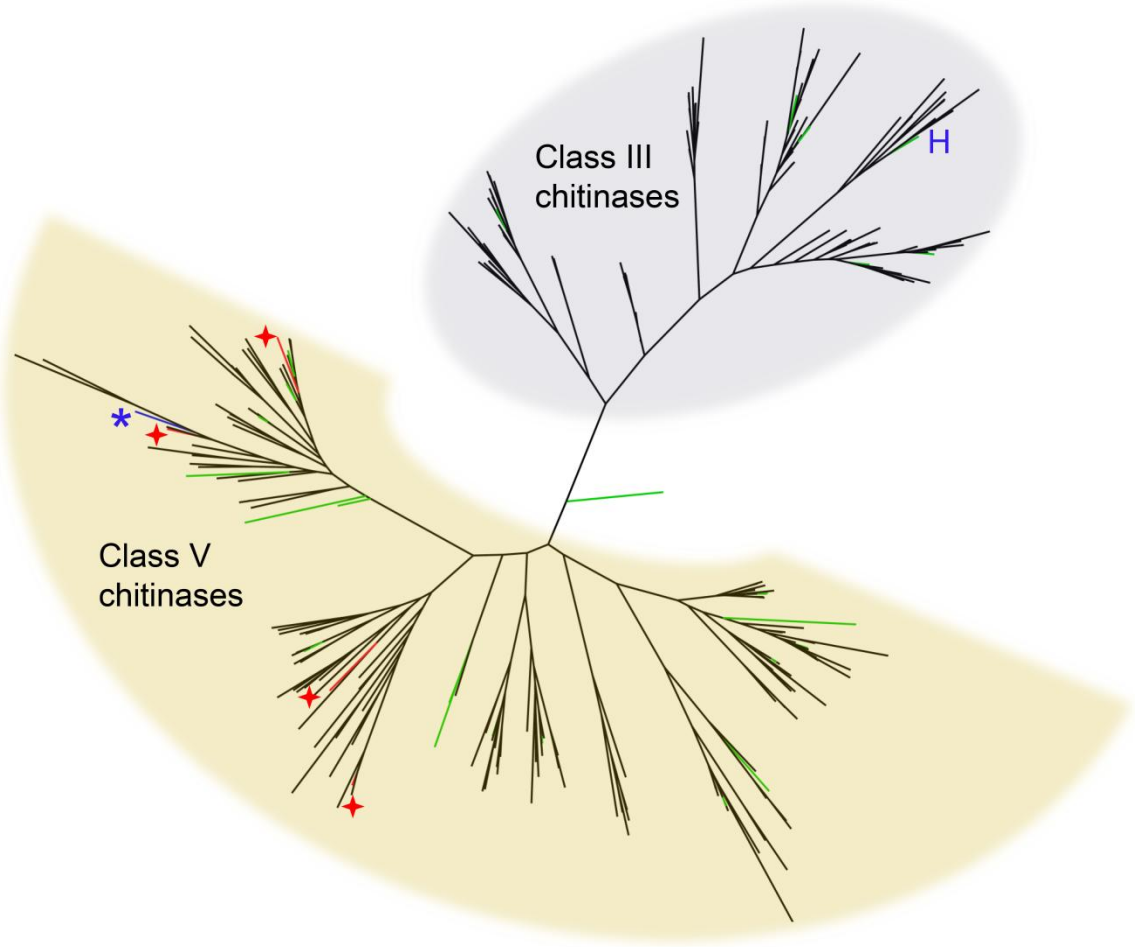
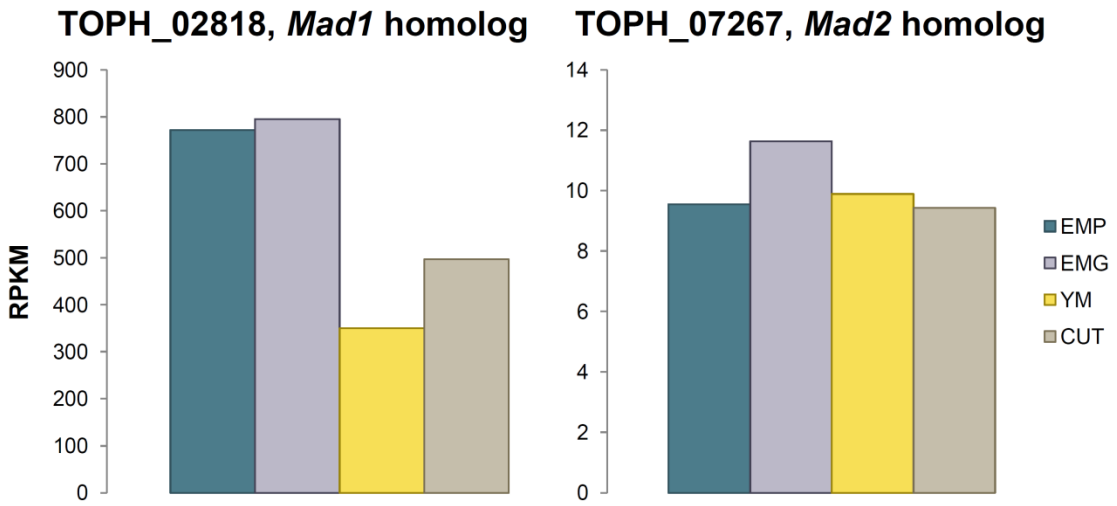


Figure 4.3. Phylogeny of chitinases showing two major fungal classes. Green branches highlight those chitinases specific to *T. ophioglossoides*. The *T. ophioglossoides* chitinase upregulated on EMP (represented by an asterisk) which is closely related to one of the four *Tr. virens* chitinases upregulated during mycoparasitism (represented by stars). The most highly expressed *T. ophioglossoides* chitinase in all conditions (represented by an 'H') is part of the Class III chitinases, a phylogenetically distinct group. Taxa sampled: *Tr. virens*, *Tr. atroviride*, *Tr. reesei*, *Beauveria bassiana*, *Cordyceps militaris*, *Epichloë festucae*, *Claviceps purpurea*, *M. acridum*, *M. robertsii*, *O. sinensis*, *T. inflatum*, and *T. ophioglossoides*.

Figure 4.4 Adhesin expression a) Upregulation of *Mad1* homolog, TOPH_02818, in media containing *Elaphomyces* (EMP and EMG) as seen in the difference in RPKM. b) Upregulation of *Mad1* and *Mad2* homologs in three hypocrealean spp. from data collected in this study combined with data from *M. robertsii* and *M. acridum* in other studies (Wang *et al.* 2005, Wang & St. Leger 2005, Gao *et al.* 2011) Illustration references: *Elaphomyces* (Latvian Nature, Askolds Klavins, Ltd. Gandrs and authors, http://www.latvijasdaba.lv/), Grasshopper (Pearson Scott Foresman), and *Manduca sexta* (Dave Pape), and tree roots by Nathan Eady (https://openclipart.org) .



Species	*Mad1* upregulated	*Mad2* upregulated
T.oph.	 Host	—
M. robertsii	 Host	
M. acridum	 Host	—

Figure 4.5. BiNGO network of statistically overrepresented GO categories in the 360 *T. ophioglossoides* genes upregulated in EMP compared to YM. Colors are based on p-values and follow the scale provided.

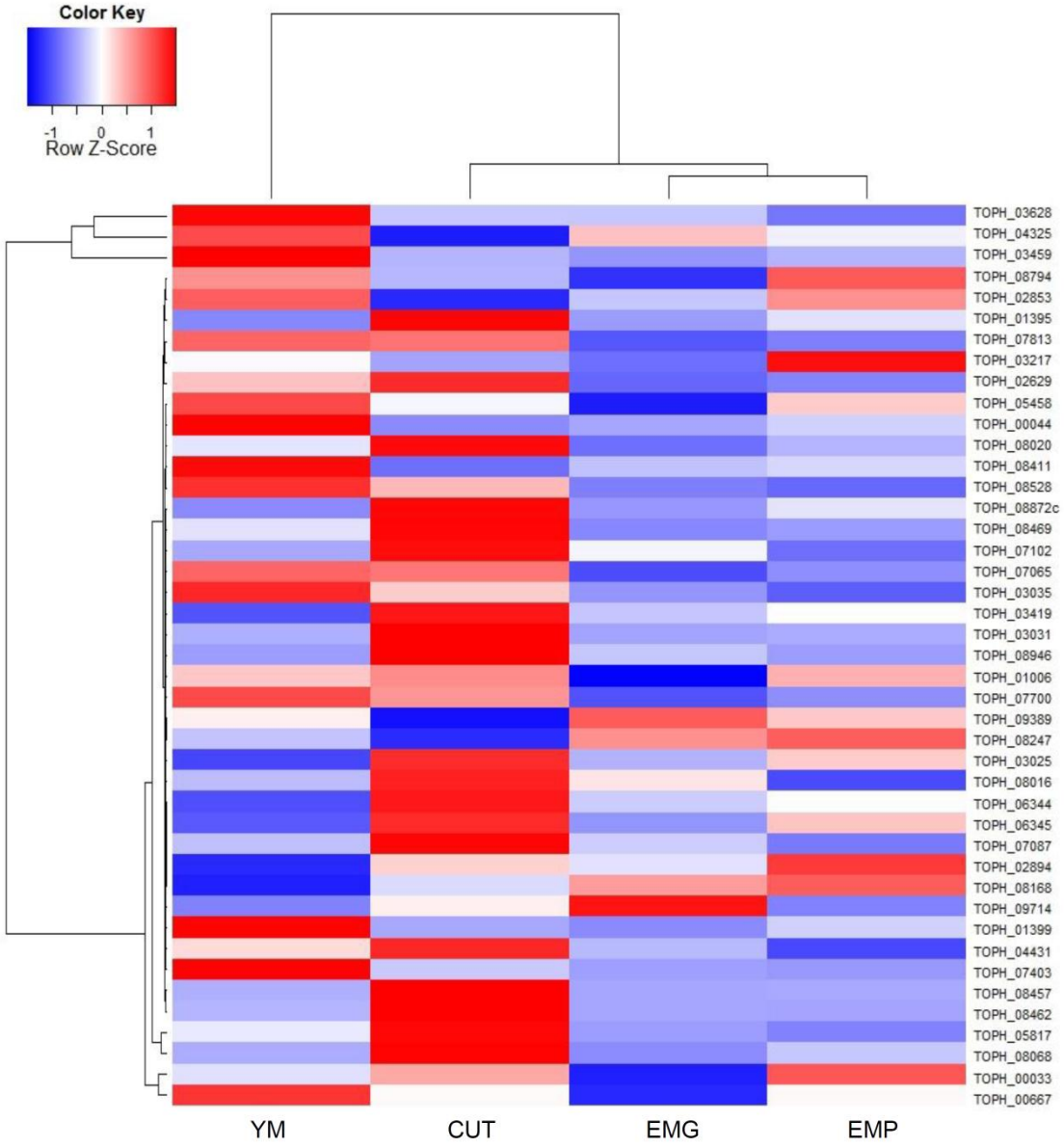


Figure 4.6. Heatmap of RPKM of secondary metabolite core genes in *T. ophioglossoides* under the experimental conditions. See text for explantation of treatments.

Table 4.1. Top twenty most highly expressed *T. ophioglossoides* genes when grown in *Elaphomyces muricatus* peridium (EMP), and their respective placement in the other treatments, showing that these genes were highly expressed during all conditions. *E. muricatus* gleba (EMG). Yeast Malt (YM). Black vine weevil cuticle (CUT). Those genes that are statistically differentially expressed in one or more conditions are denoted by an (*).

Protein Model	EMP	EMG	YM	CUT	Annotation?
TOPH_05155	1	1	1	1	HSP 30
TOPH_08949*	2	2	2	4	no putative
TOPH_03606	3	4	3	7	CCG-6 clock-controlled protein 6 (ccg-6)
TOPH_00884*	4	5	8	2	HSP 70-2
TOPH_09013	5	3	4	6	SNF2-related protein
TOPH_08965	6	6	6	3	HSP 30
TOPH_03227	7	7	13	5	HSP 101
TOPH_03499	8	11	11	9	globin-like-protein
TOPH_02789	9	10	10	10	DNA topoisomerase 2
TOPH_09312*	10	8	5	12	no putative
TOPH_02300*	11	9	9	11	C-4 sterol methyl oxidase
TOPH_01093	12	12	7	13	glyceraldehyde-3-phosphate dehydrogenase
TOPH_01292	13	14	14	15	alcohol dehydrogenase I
TOPH_08416	14	17	35	14	NO HIT
TOPH_08031	15	15	15	16	Endoplasmic oxidoreductin-1
TOPH_03470*	16	16	24	8	HSP 90
TOPH_02137*	17	19	12	17	pyruvate decarboxylase
TOPH_07279*	18	21	19	26	Ef-1 alpha
TOPH_02581	19	18	17	22	Histone H3
TOPH_02697	20	20	18	20	no putative

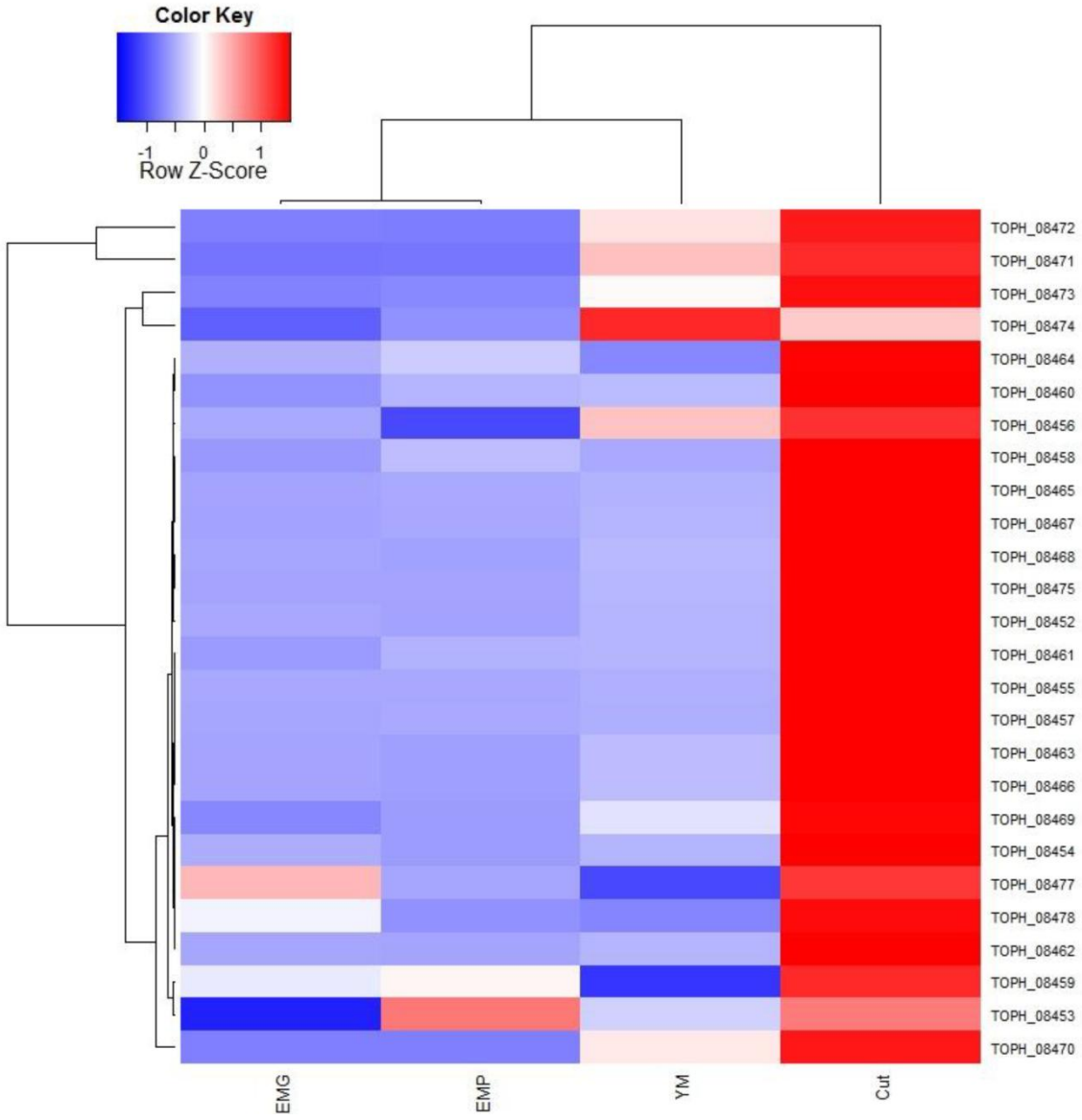


Figure 4.S1Heatmap of peptaibiotic TOPH_08469 cluster expression in the different growth conditions.

Table 4.S1. Number of reads obtained per biological and technical replicate after passing the Illumina quality filter. Biological replicates of the same treatment are in the same shade.

YM	Read Count	EMP	Read Count	CUT	Read Count	EMG	Read Count
YM1 lane5	16984502	EMP5 lane5	19455818	CUT4 lane5	21557554	EMG3 lane 1	17774680
YM1 lane6	16884615	EMP5 lane6	19493806	CUT4 lane6	21607741	EMG3 lane6	17510987
YM1 lane7	16746146	EMP5 lane7	19391547	CUT4 lane7	21475196	EMG3 lane7	17505495
YM3 lane5	23053582	EMP6 lane5	19876712	CUT6 lane5	25456069	EMG6 lane1	18222718
YM3 lane6	23099212	EMP6 lane6	19925044	CUT6 lane6	25496300	EMG6 lane6	17990853
YM3 lane7	22977797	EMP6 lane7	19803261	CUT6 lane7	25379915	EMG6 lane7	17937229
YM4 lane5	20631881	EMP8 lane5	20342464	CUT8 lane5	24398390	EMG9 lane1	18984453
YM4 lane6	20643393	EMP8 lane6	20274380	CUT8 lane6	24471158	EMG9 lane6	18824350
YM4 lane7	20500989	EMP8 lane7	20238837	CUT8 lane7	24357937	EMG9 lane7	18824350
Total	181522117	Total	199040706	Total	214200260	Total	163575115

Table 4.S2. Numbers of statistically significant differentially expressed genes between treatment comparisons (q-value < 0.1).

Comparison	# of DEGs
EMG and CUT	4332
EMP and CUT	3363
EMG and YM	1486
YM and CUT	1303
EMP and YM	786
EMP and EMG	284

Table 4.S3. Overrepresented GO terms identified using BiNGO from *T. ophioglossoides* gene upregulated on EMP compared to YM.

GO-ID	Description	p-val	correct p-val	cluster freq	total freq	Protein models
55114	**oxidation reduction**	4.91E-06	3.70E-03	42/205 20.4%	524/5167 10.1%	TOPH_05009 TOPH_08668 TOPH_05655 TOPH_06301 TOPH_01520 TOPH_06771 TOPH_01519 TOPH_02834 TOPH_02373 TOPH_05757 TOPH_01110 TOPH_02507 TOPH_08259 TOPH_06457 TOPH_01967 TOPH_02434 TOPH_03747 TOPH_01965 TOPH_00010 TOPH_09943 TOPH_07165 TOPH_01108 TOPH_04762 TOPH_00730 TOPH_07628 TOPH_07487 TOPH_03670 TOPH_05310 TOPH_03281 TOPH_06636 TOPH_04789 TOPH_06021 TOPH_06186 TOPH_09004 TOPH_05056 TOPH_06565 TOPH_08404 TOPH_04264 TOPH_04245 TOPH_05576 TOPH_05267 TOPH_07739
10181	**FMN binding**	2.72E-05	1.02E-02	7/205 3.4%	24/5167 0.4%	TOPH_07165 TOPH_05009 TOPH_03747 TOPH_00730 TOPH_06771 TOPH_02834 TOPH_02373
16491	**oxidoreductase**	4.44E-05	1.12E-02	47/205 22.9%	670/5167 12.9%	TOPH_07679 TOPH_07377 TOPH_05009 TOPH_08668 TOPH_05655 TOPH_06301 TOPH_01520 TOPH_06771 TOPH_01519 TOPH_02834 TOPH_02373 TOPH_05757 TOPH_01110 TOPH_02507 TOPH_07609 TOPH_08259 TOPH_06457TOPH_01967 TOPH_03747 TOPH_02434 TOPH_01965 TOPH_00010 TOPH_06725 TOPH_07165 TOPH_09943 TOPH_01108 TOPH_04762 TOPH_00730 TOPH_07628 TOPH_07487 TOPH_03670 TOPH_05310 TOPH_03281 TOPH_06636 TOPH_04789 TOPH_06021 TOPH_03498 TOPH_06186 TOPH_09004 TOPH_05056 TOPH_06565 TOPH_08404 TOPH_04264 TOPH_04245 TOPH_05576 TOPH_05267 TOPH_07739
46943	**carboxylic acid transmembrane transporter activity**	3.25E-04	4.22E-02	8/205 3.9%	45/5167 0.8%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_01447 TOPH_07883 TOPH_01331 TOPH_05834
5342	**organic acid transmembrane transporter activity**	3.25E-04	4.22E-02	8/205 3.9%	45/5167 0.8%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_01447 TOPH_07883 TOPH_01331 TOPH_05834
4601	**peroxidase activity**	4.19E-04	4.22E-02	4/205 1.9%	10/5167 0.1%	TOPH_06021 TOPH_05056 TOPH_08668 TOPH_01520

Table 4.S3. (Continued)

16684	**oxidoreductase activity, acting on peroxide as acceptor**	4.19E-04	4.22E-02	4/205 1.9%	10/5167 0.1%	TOPH_06021 TOPH_05056 TOPH_08668 TOPH_01520
22804	**active transmembrane transporter activity**	5.21E-04	4.22E-02	15/205 7.3%	144/5167 2.7%	TOPH_07826 TOPH_07915 TOPH_07839 TOPH_02728 TOPH_04759 TOPH_01447 TOPH_08787 TOPH_05834 TOPH_05431 TOPH_01966 TOPH_02276 TOPH_02729 TOPH_07883 TOPH_01331 TOPH_06819
15171	**amino acid transmembrane transporter activity**	7.28E-04	4.22E-02	7/205 3.4%	39/5167 0.7%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_07883 TOPH_01331 TOPH_05834
3333	**amino acid transmembrane transport**	7.28E-04	4.22E-02	7/205 3.4%	39/5167 0.7%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_07883 TOPH_01331 TOPH_05834
5275	**amine transmembrane transporter activity**	7.28E-04	4.22E-02	7/205 3.4%	39/5167 0.7%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_07883 TOPH_01331 TOPH_05834
15837	**amine transport**	7.28E-04	4.22E-02	7/205 3.4%	39/5167 0.7%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_07883 TOPH_01331 TOPH_05834
6865	**amino acid transport**	7.28E-04	4.22E-02	7/205 3.4%	39/5167 0.7%	TOPH_07839 TOPH_07915 TOPH_01966 TOPH_02728 TOPH_07883 TOPH_01331 TOPH_05834
42221	**response to chemical stimulus**	8.01E-04	4.31E-02	6/205 2.9%	29/5167 0.5%	TOPH_06021 TOPH_05056 TOPH_08668 TOPH_02276 TOPH_08912 TOPH_01520

Table 4.S4. RPKM values of secondary metabolite core genes for the four treatments with more highly expressed genes in EMP highlighted in blue, in YM highlighted in yellow, and in CUT highlighted in gray, and those highly expressed in both YM and CUT highlighted in olive.

Protein Model	EMP RPKM	EMG RPKM	YM RPKM	CUT RPKM
TOPH_00033	29.5003266	23.3802201	26.401874	28.2389646
TOPH_00044	0.88166891	0.81384159	1.370452	0.76912002
TOPH_00667	17.5564046	15.1430029	19.670592	17.5370193
TOPH_01006	0.16689762	0.08255756	0.1612829	0.17670243
TOPH_01395	4.00461263	3.58929388	3.4769083	5.62541097
TOPH_01399	0.06599953	0.0529808	0.1222434	0.05854681
TOPH_01953	0	0	0	0
TOPH_02629	1.83717046	1.39616827	4.6358535	6.89002987
TOPH_02853	4.13719587	3.2036872	4.4136113	2.32989344
TOPH_02894	0.03361966	0.01999113	0.0088957	0.02456037
TOPH_03025	0.0493168	0.04365715	0.0390249	0.05615583
TOPH_03031	0.06865292	0.06503496	0.0701561	0.2236727
TOPH_03035	0.17700778	0.18833512	0.2554266	0.22138367
TOPH_03217	3.20674963	1.31067832	1.9869802	1.57303978
TOPH_03419	0.1090707	0.08631489	0.039905	0.2065779
TOPH_03459	57.3002808	28.7034526	350.9919	57.5848655
TOPH_03628	156.376393	181.264498	274.61887	181.625503
TOPH_04325	111.810774	122.113515	137.57464	84.0270712
TOPH_04431	0.26907797	0.38790427	0.5024345	0.69519041
TOPH_05458	1.10188755	0.59271393	1.3426588	0.98955071
TOPH_05817	6.28450019	7.79324116	11.944484	26.7120259
TOPH_06344	0.01638631	0.01364122	0.0067446	0.02926192
TOPH_06345	0.01651392	0.00687372	0.0033985	0.02546846
TOPH_07065	0.16407194	0.12292741	0.336307	0.32602033
TOPH_07087	0.01582921	0.01694242	0.0167535	0.02092497
TOPH_07102	0.03239777	0.1532408	0.0836457	0.38585575
TOPH_07403	0.18943525	0.20679577	1.0812957	0.31333409
TOPH_07700	0.07628872	0.05388605	0.1884012	0.15987661
TOPH_07813	1.88722692	1.63920507	3.5787223	3.49830038
TOPH_08016	0.01774608	0.04220911	0.0313039	0.06543806
TOPH_08020	0.93879911	0.8982865	0.9664828	1.12266187
TOPH_08068	7.14733121	2.10975544	4.9577287	33.8661684
TOPH_08168	0.02996666	0.02494653	0	0.01459448
TOPH_08247	0.04238412	0.04015051	0.0324841	0.02562465
TOPH_08386	0	0	0	0
TOPH_08411	0.64462668	0.56940818	1.7074058	0.26362093
TOPH_08457	0.15318001	0.13893833	0.1976145	2.53981909
TOPH_08462	0.02724859	0.03591603	0.0869196	1.35471865
TOPH_08469	0.08578981	0.0651164	0.1576525	0.44730976
TOPH_08528	0.50925659	0.62546011	2.2276776	1.57709641
TOPH_08794	4.21538525	3.29624879	4.0816929	3.62388492
TOPH_08872c	0.19097883	0.11391865	0.1011979	0.4679857
TOPH_08946	0.02781627	0.04631278	0.0274778	0.18966061
TOPH_09389	0.04145253	0.04601101	0.0398107	0.02916098
TOPH_09714	0	0.00686429	0	0.00267721

Chapter 5. Comparative genomics of the fungal genus *Tolypocladium* reveals a complex pattern of lineage sorting associated with interkingdom host-jumps and species radiation

C. Alisha Quandt, Will Patterson, Justin Elser, Pankaj Jaiswal, Joseph W. Spatafora

Introduction

Tolypocladium is a diverse genus of fungi including pathogens of insects, nematodes, and rotifers, as well as parasites of other fungi and saprobic soil inhabitants. Despite this diversity in ecologies, multigene phylogenetic analyses support the recognition of a monophyletic genus of closely related species (Sung *et al.* 2007, Kepler *et al.* 2013, Quandt *et al.* Ch. 2). *Tolypocladium* is a member of the family Ophiocordycipitaceae, one of three, mostly insect-pathogenic, lineages of the order Hypocreales. Previous studies have found support for an insect pathogenic origin of Ophiocordycipitaceae (Spatafora *et al.* 2007, Sung *et al.* 2008), and moderate support for a mycoparasitic origin of *Tolypocladium* (Sung *et al.* 2008). Other well-documented examples of host-jumping in Hypocreales include transitions from insect-pathogenesis to plant-endophytism (Spatafora *et al.* 2007) and plant-endophytism to mycoparasitism (Kepler *et al.* 2012) within the Clavicipitaceae.

The majority of *Tolypocladium* species are mycoparasites that parasitize truffles of the ectomycorrhizal genus *Elaphomyces* [Eurotiales, Ascomycota] (Quandt *et al.* Ch. 2). These mycoparasites infect and consume internal and external truffle tissue to varying degrees. They produce relatively large, stipitate, clavate to capitate sexual fruiting bodies that arise out of the ground while staying directly connected to their hosts or through strands of aggregated hyphae called rhizomorphs (Mains 1957, Kobayasi & Shimizu 1960). Unlike the insect pathogens and saprobic species of *Tolypocladium*, most of these *Elaphomyces*-parasites lack the ability to grow in culture. *Tolypocladium ophioglossoides*, however, readily germinates and grows in culture, and some species are rarely isolated into pure culture but do not produce conidia, including *T. capitatum*.

Like approximately half of the described *Elaphomyces*-parasites (Mains 1957, Kobayasi & Shimizu 1960), the fruiting body of *T. capitatum* arises directly from its host and does not produce rhizomorphs.

Of the insect pathogenic *Tolypocladium* spp., the beetle pathogen, *T. inflatum*, was the first source of the immunosuppressant cyclosporin (Borel 2002). Though *T. inflatum* produces its sexual fruiting bodies after parasitizing beetle larvae buried in wood or other substrates, it is commonly isolated from soil as well (Hodge *et al.* 1996, Gams 1971). Three cicada pathogens have been described in the genus, including *T. paradoxum*, which was so named due to the perplexing morphological similarities between its fruiting body and that of the mycoparasite, *T. ophioglossoides* (Lloyd 1918, Kobayasi 1939). They both produce singularly arising stipitate stromata terminating in a spathulate to clavate fertile area with olive-yellow tissue at the base of the stipe and attached to the host via rhizomorphs.

Nikoh & Fukatsu (2000) were the first to collect molecular data supporting the relationship between the insect pathogenic and mycoparasitic species of *Tolypocladium*, and they proposed that the host habitat hypothesis (observed in other systems [*e.g.* Shaw 1988]) could best explain the evolution of host switching from cicada to *Elaphomyces*. The premise behind their hypothesis is that cicadas and *Elaphomyces* truffles are found in the same subterranean habitat, attached to the roots of trees, and thus this shared habitat facilitated the transition or host-jump from cicada pathogenesis to *Elaphomyces*-parasitism (Nikoh & Fukatsu 2000). However, published phylogenetic sampling shows the insect pathogenic species of *Tolypocladium*, including *T. paradoxum* and *T. inflatum*, to be nested within the clade of *Elaphomyces* parasites, and therefore were

hypothesized to be reversions to insect pathogenesis from a mycoparasitic ancestor (Sung *et al.* 2007, 2008). There has yet to be an extensive molecular sampling of the diversity of the genus with more than just two genes for all species, and phylogenies have typically resulted in low support for the internal branching order within the genus and intrageneric relationships are poorly understood (Sung *et al.* 2007, Quandt *et al.* Ch. 2).

Genome scale data provide the opportunity to examine several aspects of organismal biology and evolution. Specifically genome scale datasets can provide the maximum amount of discrete genetic information to resolve the phylogenetic relationships among species that have undergone rapid diversification (Rokas *et al.* 2003, Salichos & Rokas 2013). In particular short internodes can be especially difficult to resolve with support even using genome scale data, because there is a greater likelihood that some of the genes will result in incomplete lineage sorting (Maddison 1997). Lineage sorting occurs when multiple alleles of a gene exist within a population, and subsequent to divergence some alleles persist and some are lost which can distort the ability to discern evolutionary histories (Doyle 1992). If the species populations were large during speciation events, lineage sorting is even more likely to occur.

In addition to phylogenetic information, multiple studies have shown that differences in nutritional or ecological niches are correlated with genome scale or proteome scale differences between related groups of fungi, especially in possession or absence of carbohydrate active enzymes (CAZymes) (*e.g.* white rot and brown rot Basidiomycota [Floudas *et al.* 2012], *Elaphomyces granulatus* [Quandt *et al.* Ch. 6]). Hypocrealean taxa also possess numerous and varied secondary metabolite (SM) genes

and gene clusters (Desjardins *et al.* 1993, Molnar *et al.* 2010), which are non-essential to growth and function of the organisms producing the metabolites, but may provide some type of selective advantage (Keller *et al.* 2005). Some of these metabolites have insecticidal properties (*e.g.* destruxins [Roberts 1981, Wang *et al.* 2012]). Others appear to be found only in lineages having similar ecologies (*e.g.* peptaibiotic genes have only been identified from mycoparasitic lineages [Quandt *et al.* Ch. 3]).

To better understand the phylogenetic relationships of *Tolypocladium* species, the directionality of host-jumping and the associated changes in genome structure and content seen in *Tolypocladium* species parasitizing different hosts, the genomes of the mycoparasite *T. capitatum* and the cicada pathogen *T. paradoxum* were sequenced and analyzed. These two genomes were combined with the previously sequenced genomes of *T. inflatum* and *T. ophioglossoides* (Bushley *et al.* 2013, Quandt *et al.* Ch. 3) and numerous hypocrealean outgroup taxa. Using phylogenomics, ancestral character state reconstruction, and comparative genomics, a new picture of *Tolypocladium* evolution emerges that supports mycoparasitism as a derived ecology for the genus and brings to light complex patterns of lineage sorting that are the product of a relatively rapid phylogenetic radiation.

Material and Methods

For collection of tissue for DNA extraction, *T. capitatum* strain CBS 113982 was grown on potato dextrose agar (PDA) from Difco (Detroit, MI), and then plugs were used to inoculate 150 mL flasks of Difco Sabouraud dextrose broth. Due to a slow

growth rate, liquid cultures of *T. capitatum* were grown on a shaking incubator for 14 days. Plugs of *T. paradoxum* strain NRBC 100945 tissue growing on PDA were used to inoculate 150 mL flasks of yeast malt broth (YM) using the previously described recipe (Quandt *et al.* Ch. 4) and shaken for 6 days prior to harvesting. Tissue of both species was harvested via filtration, frozen at -80° C in 1.5 mL tubes, and then lyophilized for 24 hours. Lyophilized tissue was ground using a mortar and pestle, and DNA was extracted using a Qiagen DNeasy Plant kit following the standard protocol starting with the addition of lysis buffer AP1 and eluted in 50 μL of water. Because the tissue of *T. capitatum* grown for a longer amount of time, the DNeasy protocol was modified by adding an extra 30 μL of the AP1 lysis buffer. Tissue for RNA extraction was grown in the same way in either Sabouraud dextrose broth (*T. capitatum*) from Difco (Detroit, MI) or YM (*T. paradoxum*), and harvested after 14 or four days, respectively, into liquid nitrogen and stored at -80°C where samples were held until RNA extraction, and RNA was extracted using the Qiagen RNeasy Plant kit.

DNA Illumina library preparation for *T. capitatum* was created using New England Biomedicals NEBNext reagents, and size selection (325 bp) was performed using gel extraction. For *T. paradoxum,* the Apollo 324 NGS Library Prep System created the DNA library using the IntegenX PrepX ILM DNA Library Kit with a 370 bp library size. Both libraries were sequenced on the Illumina HiSeq2000 for 101 paired-end cycles. For both species the Illumina TruSeq RNA Sample Preparation Kit v2 was used for RNA library construction, using the manufacturer's suggested protocols including Agencourt AMPure magnetic beads for cleaning steps. RNA libraries were

barcoded with the TruSeq kit adapters and sequenced for 51 single-end cycles on the Illumina HiSeq2000.

Using scripts in the fastx toolkit (Gordon 2011), raw reads were trimmed (to 50 bp in length) and filtered based on quality score (all bases ≥q20). This resulted in 121.1 million and 157.5 million remaining reads for *T. captitatum* and *T. paradoxum*, respectively. *De novo* assembly of reads was conducted in Velvet v. 1.19 (Zebrino & Birney 2008). Multiple trim lengths were initially employed, ranging from 40-80 bp. After subjecting the variously trimmed data to quality filtering and assembly, the resulting assemblies were compared based on their n50 score and number of contigs. A trim length of 50 bp was ultimately chosen, as it yielded the assembly with the highest n50 score and the fewer contigs. To estimate the completeness of these genomes, the Core Eukaryotic Mapping Genes Approach (CEGMA) was employed (Parra *et al.* 2007, 2009).

Gene model predictions were created using the Maker annotation pipeline (Cantarel *et al.* 2008) incorporating RNA data assembled in Trinity (Grabherr *et al.* 2011) using the Jellyfish v. 2.0 method of kmer counting (Marçais & Kingford 2011). Other information given to Maker included custom hidden Markov models (HMMs) for *T. capitatum and T. paradoxum* built by Genemark-ES v 2.0 (Ter-Hovhannisyan *et al.* 2008), a SNAP HMM (Korf 2004) trained on *Fusarium graminearum,* which was also set as the species model for AUGUSTUS (Stanke *et al.* 2006), and protein and/or EST data from the following hypocrealean taxa: *F. graminearum*, *N. haematococca*, *Tr. reesei*, *Tr. virens*, *M. robertsii*, *T. inflatum*, *C. militaris*, *B. bassiana*. Annotation of transposable elements for each species was performed in RepeatMasker v 3.2.8 with

organism set to "fungi" (Smit & Green 1996), and custom repeat contents were estimated using RepeatScout v 1.0.3 and scripts associated with that package (Price *et al.* 2005). Non-overlapping *ab initio* protein models were BLAST against a custom database of all the protein models of all the hypocrealean taxa used in this study (see Figure 1). Any of these protein models with significant E-value ($E \leq 1e^{-5}$) were included in the final protein set and used for downstream analyses.

An HMM created for a previous study (Quandt *et al.* Ch. 3) was used to mine *T. capitatum* and *T. paradoxum* genomes for NRPS adenylation domains (A-domains). Predicted A-domain amino acids sequences were aligned using MUSCLE v 3.8.31 (Edgar 2004) under default settings. Gaps were removed manually, and all alignments were analyzed using RAxML v 7.2.6 (Stamatakis 2006) using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates. Predictions of polyketide synthases (PKSs) and other types of secondary metabolite genes and gene clusters were performed using antiSMASH (Blin *et al.* 2013) and SMURF (Khaldi *et al.* 2010). HMMs from dbCAN were used to annotate CAZymes in the *T. capitatum* and *T. paradoxum* genomes (Yin *et al.* 2012).

Whole genome phylogenomic analyses of the selected hypocrealean genomes from Quandt *et al.* Ch. 3 were executed in the HAL pipeline (Robbertse *et al.* 2011). Orthologous clusters of proteins were identified in MCL (Enright *et al.* 2002) across inflation parameters 1.2, 3 and 5. Briefly, orthologous clusters were filtered for retention of clusters with one sequence per genome, removal of any redundant clusters, and retention of proteins whose best reciprocal BLAST hits were to intracluster proteins. The resulting unique, single-copy orthologous clusters of proteins were aligned in

MUSCLE (Edgar 2004) with default settings; poorly aligned regions were identified using Gblocks (Talavera & Castresana 2007; gap removal setting = liberal) and excluded from subsequent analyses. ProtTest v 1.2.7 (Abascal *et al.* 2005) was performed on the individual protein clusters and JTT was identified as the most common substitution matrix. The aligned clusters were concatenated into a superalignment and maximum likelihood analysis was performed using RAxML v 7.2.6 with the Gamma model of rate heterogeneity and the JTT substitution matrix with 100 bootstrap replicates.

All shared, single-copy, orthologous cluster alignments with liberal gap removal that were a part of the concatenated superalignment were analyzed individually in RAxML v 7.2.6 with *Neurospora crassa* set as the outgroup and with the Gamma model of rate heterogeneity and the JTT substitution matrix with 100 bootstrap replicates to generate individual gene trees. Custom python scripts were used to identify tree topologies with bootstrap support values ≥70 MLBP for all three nodes in a monophyletic *Tolypocladium*.

Of the gene trees that resolved the concatenated *Tolypocladium* as a monophyletic clade with support (MLBP ≥70), ten were randomly chosen and those orthologous cluster alignments were loaded into BEAUti to create an xml file with the following site parameters: substitution rate estimated, a gamma category count of 4, shape set to 1.0 and the JTT substitution model (Drummond *et al*. 2012). The relaxed log normal clock was chosen with a clock rate of 1.0. The Birth Death Model of evolution was chosen and the following fossil calibrations were used: *Beauveria bassiana* and *Cordyceps militaris* have a minimum age of 25 million years (Poinar & Thomas 1984); *Epichloë festucae* and *Claviceps purpurea* were calibrated based on

grass fossils (at 66 to 100 million years) (see Sung *et al.* 2008); and *Tolypocladium* spp. and *Ophiocordyceps sinensis* are calibrated based on the *Paleoophiocordyceps* amber fossil which is estimated to be between 99 to 105 years. Trees were sampled every 100 trees and 10,000,000 states were run. The xml file was run in BEAST v 2.1.3 (Bouckaert *et al.* 2014).

Orthologous clusters were further assigned exclusively analyzing the four *Tolypocladium* genomes using the fastortho (Wattam *et al.* 2014) reimplementation of OrthoMCL (Li *et al.* 2003).

Results and Discussion

The *T. capitatum* genome was assembled into 22.9 Megabases (Mb) across 1,345 scaffolds (Table 5.1) with a median coverage depth of 92x. Despite a relatively low N50 of 34,296 bp, CEGMA estimates of for core eukaryotic genes present in the *T. capitatum* assembly were as follows: 96.8% completely present, and 98.4 % partially present. The assembly of *T. paradoxum* genome encompasses 27.6 Mb on 1139 scaffolds, has an N50 of 79,523, and a median coverage depth of 80x. Of the core eukaryotic genes, 93.6% or 95.57% were completely or partially, respectively, identified in the *T. paradoxum* assembly. Both genomes are smaller than those reported for previously sequenced *Tolypocladium* spp. or any Hypocreales genome sequenced to date (the next smallest is *Claviceps paspali* with 28.9 Mb [Schardl *et al.* 2013]), with *T. capitatum* at least 4.7 Mb smaller than all other species. Both *T. capitatum* and *T.*

paradoxum have high GC contents, 59.7% and 58.0%, respectively, similar to the value recorded in two other sequenced *Tolypocladium* species (Table 5.1).

The Maker pipeline identified 13,375 non-overlapping *ab initio* protein model predictions for *T. capitatum*. Maker identifies only 7,603 of these which were supported by RNA evidence. Using a concatenated hypocrealean protein model database, an additional 503 protein coding genes were predicted based on BLAST matches to the database (e-value $<1e^{-5}$) resulting in a total of 8,106 predicted protein-coding genes in *T. capitatum*. Using the same approach, *T. paradoxum* yields 9,773 protein-coding genes that are supported either by RNA evidence (8,962) or by BLAST matches to hypocrealean proteins (811), out of the total 15,417 non-overlapping *ab initio* predictions. These results are similar to those found for *T. inflatum* (9,889) and *T. ophioglossoides* (10,135) (Quandt *et al*. Ch.3). The greatest difference is gene number (2,029) is seen between the two truffle parasites, *T. capitatum* and *T. ophioglossoides*.

Analysis of the orthologous gene clusters in *Tolypocladium* revealed that the majority, 5,600 of the 8,814 clusters identified, were shared by all four species (Figure 5.1), although the numbers of species-specific genes, both singletons and paralogs of gene families, was relatively high (743, 1,614, 1,628, and 1,370 for *T. capitatum*, *T. inflatum*, *T. ophioglossoides*, and *T. paradoxum* , respectively). The numbers of orthologous clusters shared by two or three, but not all species are lower than these species-specific genes and clusters for each species (Figure 5.1). The differences in orthologous protein cluster membership between species could either be due to species specific expansions or to differential retention and expansion of ancestral protein families.

Phylogenetic results

The concatenated dataset of 1350 single copy orthologous clusters with 100% membership from the 20 species including *Tolypocladium* spp., other selected hypocrealean taxa, and two outgroup taxa contained 610,821 amino acids. Maximum likelihood analysis of the concatenated dataset resulted in a phylogenetic hypothesis that was well supported at all nodes (MLBP=100), and a topology similar to other published phylogenies of Hypocreales except for relationships within *Tolypocladium* (Figure 5.2) (Sung *et al.* 2007, Bushley *et al.* 2013, Quandt *et al.* Ch. 3). *Tolypocladium* is resolved as a monophyletic genus with *T. paradoxum* reconstructed its earliest diverging species, followed by the divergence of *T. inflatum*, which is sister to both of the truffle parasites. These analyses resolve the insect pathogens as the earliest diverging lineages of the genus and based on this sampling there has been a single transition to *Elaphomyces* parasitism. This result is contrary to the previous studies (Sung *et al.* 2007, Sung *et al.* 2008, Quandt *et al.* Ch. 2) which found moderate to low support for the ancestral ecology being truffle parasitism with multiple reversals to insect pathogenicity.

Gene tree conflict and lineage sorting

The internal nodes separating *Tolypocladium* species are the shortest in the entire phylogeny – shorter than the internodes between the *Trichoderma* spp. and *Fusarium* spp. sampled. For this reason clusters were re-analyzed individually in RAxML. Of the 1,350 original clusters, 1,208 resulted in the ML tree topologies in which the genus *Tolypocladium* was monophyletic, with or without bootstrap support. Of these, all 15

possible four-taxon topologies of the genus were recovered (Figure 5.3), but not at equal frequencies (Figure 5.S1, Figure 5.3). The topology of the concatenated dataset was the most common topology recovered, but only accounted for 17% of the 1,208 trees. The other most common topologies included one similar to the concatenated tree, but with *T. inflatum* as the earliest diverging lineage instead of *T. paradoxum* (13%); and finally a topology with *T. capitatum* as the earliest diverging species, followed by *T. ophioglossoides* which is sister to *T. inflatum* and *T. paradoxum* (9%) (Figure 5.3).

Of the 1,208 individual cluster trees that resolve *Tolypocladium* as a monophyletic genus, a majority (1,099) did not receive bootstrap support (MLBP≥70) at all three nodes within and joining *Tolypocladium*, resulting in only 109 trees supporting a fully resolved, monophyletic *Tolypocladium*. Of the 109 trees with support, the concatenated tree topology was recovered almost three times more often than any of the other topologies (36 trees; 33%). The other two most commonly recovered topologies with bootstrap support both resolve *T. ophioglossoides* and *T. capitatum* as sister to one another but differ in their placement of the insect pathogens with either *T. inflatum* and *T. paradoxum* as sister to each other or with *T. paradoxum* as more closely related to the truffle parasites than *T. inflatum* (as mentioned before, this was the second most common topology overall). Combined, these three topologies with *T. capitatum* and *T. ophioglossoides* as sister taxa account for 57% of all supported trees.

The concatenated topology of *Tolypocladium* was the most commonly recovered tree topology and represented disproportionately more of the trees with bootstrap support. However, there is a significant amount of gene tree discordance, even within just the supported tree topologies (Figure 5.4), with 13 different topologies receiving

support in at least one of the cluster trees. This discordance is not likely the product of error associated with taxon sampling, because there are no long branches in this part of the tree (nor in any of the multi-gene phylogenies) suggesting that the addition of more taxa would not decrease gene tree discordance. The combination of closely related taxa, separated by short internodes and a high number of alternative gene tree topologies that receive bootstrap support is indicative of species that have undergone relatively rapid speciation events in which a large portion of the genes have undergone incomplete lineage sorting. A high degree of incomplete lineage sorting could explain the differing phylogenetic results obtained in published five-gene phylogenies where many of the *Tolypocladium* spp. sampled had only one or two ribosomal genes available (Sung *et al.* 2007, Quandt *et al.* Ch. 2). Based on these analyses it is likely that the diversification of *Tolypocladium* spp. happened relatively quickly and that truffle parasites form a monophyletic, derived lineage within the genus that is the result of a single ecological transition or host-jump from insects to fungi.

Divergence Dating

To test the rapid diversification hypothesis, two methods, strict and relaxed molecular clock models, were used for dating the divergence times within *Tolypocladium*. Ten cluster alignments were randomly chosen from the set of 36 inidividual trees with bootstrap support for the nodes joining *Tolypocladium* with the concatenated topology, resulting in an alignment with 7,280 amino acids. Fossil calibrations were used in both analyses for minimum ages of Cordycipitaceae (Poinar &

Thomas 1984), the grass associated Clavicipitaceae taxa with grass ages (Gandolfo *et al.* 2002), and Ophiocordypitaceae (Sung *et al.* 2008).

In the strict clock analysis, the mean time for occurrence of the node joining *Tolypocladium* is estimated at 57 million years ago (MYA) [40 - 81 MYA, 95% credibility interval (CI)] (Figure 5.5). The divergence of the most recent common ancestor (MRCA) to *T. inflatum* and the *Elaphomyces* parasites is estimated to have occurred 48 ± 18.5 MYA, and the divergence of *T. ophioglossoides* and *T. inflatum* is estimated to have occurred 41 ± 14 MYA.

The relaxed clock analysis resolved the mean estimate for the MCRA of *Tolypocladium* at 43 MYA [27 - 93, 95% CI] (Figure 5.5). The divergence of *T. inflatum* and the *Elaphomyces* parasites is estimated to have occurred 30 MYA [0.2 - 93 MYA, 95%CI], and finally *T. ophioglossoides* and *T. capitatum* are joined at a node estimated at 20 MYA [0.1 - 60.7 MYA, 95% CI]. The credibility intervals around these nodes are larger than in the strict clock analysis, but this is likely due to the nature of the relaxed clock model (as almost all of the 95% CIs are larger in this analysis). The relaxed clock estimation of the root of the tree (which represents MRCA to Sordariomycetes) is similar to other published reports using different fossil calibrations and a wider breadth of taxon sampling (Lücking *et al.* 2009, Berbee & Taylor 2010), although both analyses predict ages that are much older than estimates from other analyses (Wedin & Prieto 2013).

Despite topological differences in some of the other lineages (*i.e.*, placement of Cordycipitaceae and Hypocreaceae), the two analyses largely yield similar results with respect to the short amount of time (based on molecular evolution and fossil calibration)

separating the speciation events within *Tolypocladium*. Compared to the other genera in which more than two species are sampled (*e.g., Fusarium* and *Trichoderma*), the internodes separating the species of *Tolypocladium* are five to nine times shorter. These analyses support the hypothesis that *Tolypocladium* underwent a more rapid species diversification, and is consistent with a large percentage of gene trees being the product of incomplete lineage sorting within the ancestral populations. This rapid radiation occurred at the same time as a major shift in host association among insect hosts and between insects and fungi, with host-mediated isolation functioning as a strong speciation mechanism.

Based on these analyses, the estimation of the host-jump onto the ectomycorrhizal genus *Elaphomyces* is hypothesized to have taken place between 20 to 40 MYA, spanning from the middle Eocene to early Miocene. The first fossil of ectomycorrhizal root tips is described from the Eocene and estimated to be at least 50 million year old (LePage *et al.* 1997), and therefore the estimates for the emergence of *Elaphomyces*-parasitic *Tolypocladium* species is congruent with the presence of ectomycorrhizae in the fossil record.

Orthologous cluster analysis

To identify genes that may be related to the differences in host affiliation of *Tolypocladium* species, orthologous gene clusters, specific to either the insect pathogens (*T. inflatum* and *T. paradoxum*) or the mycoparasites (*T. capitatum* and *T. ophioglossoides*), were analyzed. There are 348 such clusters specific to the insect pathogens and 174 clusters specific to the mycoparasites. Large percentages (41% for

insect pathogens and 30% for the mycoparasites) of these host-specific clusters have no putative functions. Furthermore, orphans (gene clusters with no homologs in Genbank) account for 15.6% of the insect pathogen-specific and 6.9% of the mycoparasite-specifc clusters.

Of the clusters with putative annotations, one of the mycoparasitic-specific clusters includes two dextranses in *T. ophioglossoides* and one in *T. capitatum*. Most of the dextranses described in fungi are from Eurotiomycetes, close relatives of the *Elaphomyces* hosts of these mycoparasites (Quandt *et al.* Ch. 6), including *Penicillium*, and yeasts such as *Lipomyces* (Jimenéz 2009). While *E. granulatus* does possess a homolog of this dextranase, maximum likelihood analysis of these proteins, does not indicate that these are the product of horizontal gene transfer (Figure 5.S2), although very few hypocrealean taxa (and other Sordariomyctes) possess homologs to these proteins.

Of the orthologous clusters specific to insect pathogens, 16.7% of these proteins are secreted (Table 5.2), which is more than double the percentage of secreted proteins seen in the mycoparasite specific clusters (6.9%). These secreted protein clusters specific to insect pathogens may play a role in host-pathogen interactions specifically related to evading or subduing host immunity. Fungi do not possess an immune system comparable to those of insects, and little is known about how fungi or fungal fruiting bodies respond to parasitism. This pattern, however, is consistent with mycoparasitic species having lost genes for secreted proteins functioning in insect immune response evasion as a consequence of an interkingdom host-jump from insects to fungi. Many of these secreted proteins have no known function, but five of the *T. inflatum* and *T.*

paradoxum-specific secreted proteins (grouping into four orthologous clusters) are annotated as proteases, including subtilisins and metalloproteases, which other studies have identified as major pathogenicity factors in insect pathogens such as *M. robertsii* (Gao *et al.* 2011). Of the mycoparasite-specific clusters that are secreted, one is a S10 carboxypeptidase (Table 5.3), which was also upregulated in *T. ophioglossoides* when grown on media containing *Elaphomyces* (Quandt *et al.* Ch. 4). Both of these proteins are truncated in *T. ophioglossoides* and *T. capitatum*, which probably explains why they do not cluster with the other S10 carboxypeptidases in all four species. Whether this is an artifact of protein annotation or biologically relevant is unknown, but this gene was found to be upregulated in *T. ophioglossoides* when grown on media containing *Elaphomyces* (Quandt *et al.* Ch. 4).

For both insect pathogenic and mycoparasitic-species, four and six of the clusters, respectively, were transcription factors. Given that insect pathogenic lineages are relatively reduced in CAZymes compared to the earliest diverging Hypocreales (Bushley *et al.* 2013), differences in how these fungi regulate the gene expression could play a crucial role in the ability to parasitize a new host (see below).

Analysis of CAZymes

Overall, when comparing the abundances of CAZymes in *Tolypocladium* spp. to the other hypocrealean taxa sampled, broader patterns based on ecology begin to emerge. First, *Fusarium* spp., *Nectria haematococca*, *V. dahliae*, and the two mycoparasitic *Trichoderma* spp. have greater CAZyme quantities. All the other hypocrealean species analyzed, which include the insect pathogens and mycoparasites,

fall into two groups, one of which has slightly more CAZymes than the other. This warmer (or more CAZyme rich) group includes most of the insect pathogens (*B. bassiana, C. militaris, Metarhizium* spp., and *T. inflatum*) along with *Tr. reesei* and *Neurospora crassa* (one of the outgroups). The remaining *Tolypocladium* species group with the two grass-associates (*E. festucae* and *Cl. purpurea*) and the insect pathogen *O. sinensis*. Overall, sampled species in the Ophiocordycipitaceae and the grass associated Clavicipitaceae are the most CAZyme reduced in the order, with the exception of *T. inflatum.*

Within Tolypocladium, the difference in the quantity of CAZymes does correlate with ecology to some extent (Figure 5.S3), with *T. capitatum* and *T. ophioglossoides* grouping together based on Euclidean distances. Other than this, there are no obvious pattern of CAZyme distribution between insect pathogens and mycoparasites (Figure 5.S4). Specifically, the quantity of chitinases (CAZyme family GH18) does not correlate with ecology, as *T. ophioglossoides* has the most, followed by *T. paradoxum*, then *T. capitatum* , and finally *T. inflatum* has the fewest (Figure 5.S5). A reduced number of CAZymes in *T. inflatum* may not be that surprising given that it grows best on standard culture media. T. inflatum was also originally described growing in soil, suggesting saprobic growth may be common for that species; and if so, the chitinases possessed in the other species may not be as essential for its survival.

Secondary metabolism in *Tolypocladium*

To examine the secondary metabolic potential of the newly sequenced *Tolypocladium* species and compare the intrageneric diversity of these genes, SM core

genes including nonribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs), Hybrid NRPS-PKSs, terpene synthases, and DMAT synthases, and their corresponding SM clusters were identified. All sequenced species of *Tolypocladium* possess numerous secondary metabolite SM gene clusters including *T. capitatum* and *T. paradoxum*, which possess 39 and 34 core SM genes in 35 and 31 gene clusters, respectively (Table S1, Table S2). Each *Tolypocladium* spp. has a unique profile of core SM genes (Figure 5.6), and there are a few notable patterns. *Elaphomyces* parasites do not possess DMATs, whereas *T. inflatum* has one (Bushley *et al.* 2013) and *T. paradoxum* has two (Figure 5.6). In contrast, *T. ophioglossoides* and *T. capitatum* are expanded in their possession of terpene synthases (Figure 5.6). They each possess four terpene synthases, while *T. inflatum* possesses none and *T. paradoxum* has two of these genes. Terpenes are a part of a diverse class of compounds (Gershenzon & Dudareva 2007), but most of the terpenes described in hypocrealean taxa have toxic or immunosuppressive activities in animals (*e.g.*, trichothecenes in *Fusarium* and *Stachybotrys*, lolitrems in *Neotyphodium* [Sorenson *et al.* 1987, Desjardins *et al.* 1993, Young *et al.* 2006]). It is therefore interesting that the mycoparasites within *Tolypocladium* are enriched in terpene synthases.

Possession of the gene responsible for production of cyclosporin, *simA*, is limited to *T. inflatum* (Figure 5.S5). Peptaibiotic genes are found in all species except *T. capitatum*, which is interesting given its ecology and the distribution of peptaibiotic genes within Hypocreales, which is limited to the mycoparasitc lineages of *Trichoderma* and *Tolypocladium* (Quandt *et al.* Ch. 3). Peptaibiotics are proposed to play a role in mycoparasitism, however, expression of genes in the peptaibiotic clusters was not found

to be upregulated in *T. ophioglossoides* when grown on media containing *Elaphomyces*. Moreover, one of the peptaibiotic clusters (not shared with other species of *Tolypocladium*) was significantly upregulated on media containing insect cuticle (Quandt et al. Ch. 4).

Conclusion

Through the sequencing of two additional species of *Tolypocladium, T. capitatum* and *T. paradoxum*, a robust dataset was created with which to examine the evolution of host associations in this genus of mycoparasites and insect pathogens. In contrast to previous studies, phylogenomic analysis supported the hypothesis that insect pathogens form the early diverging lineages of the genus, while parasites of *Elaphomyces* are members the most derived lineage. Further analysis of individual gene trees identified a large number of gene tree - species tree incongruences, indicating that many genes may have undergone incomplete lineage sorting or a complicated pattern of gene duplication and subsequent loss. The concatenated dataset produced a phylogenetic hypothesis for the genus in which the internodes were short, and divergence dating using fossil calibration revealed that the estimated divergence times within *Tolypocladium* were several times shorter and overlapping than those observed in other sampled lineages, consistent with a rapid radiation event in concert with the major host shift from insect pathogenesis to mycoparasitism. The insect pathogens, *T. inflatum* and *T. paradoxum*, have a larger percentage of unique orthologous clusters that produce secreted proteins (some of which have been identified to function in insect

pathogenicity) than their mycoparasitic counterparts. Quantitative differences in CAZymes between insect pathogens and mycoparasites were minimal, but the mycoparasites, *T. capitatum* and *T. ophioglossoides*, are enriched in terpene synthases, while lacking DMATs, which the insect pathogens possess. This study exemplifies the power of genome scale data in addressing questions around host jumping and the evolutionary context in which that takes place.

Acknowledgements

The authors thank Mark Desanko for providing technical and robotics support for sequencing in this project. This project was supported with funds from the NSF grants DEB-0732993, and DEB-1311572 to JWS and CAQ, and an NSF Graduate Research Fellowship funded CAQ.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104-2105.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* **41**: W204-W212.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond, AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.

Borel JF (2002) History of the discovery of cyclosporin and of its early pharmacological development. *Wiener Klinische Wochenschrift* **114**: 433-437.

Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, Owensby CA, Knaus BJ, Elser J, Miller D, Di Y, McPhail KL, Spatafora JW (2013) Draft genome sequence of the Cyclosporin producing fungus *Tolypocladium inflatum* reveals complex patterns of secondary metabolite evolution and expression. *PLoS Genetics* **9**: e1003496.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188-196.

Desjardins AE, Hohn TM, McCormick SP (1993) Trichothecene biosynthesis in *Fusarium* species: chemistry, genetics, and significance. *Microbiological Reviews* **57**: 595-604.

Doyle JD (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* **17**: 144-163.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969-1973.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575-1584.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Gorecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kues U, Arun Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Duenas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St. John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev I, Hibbett DS (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715-1719.

Gams W (1971) *Tolypocladium*, eine Hyphomycetengattung mit geschwollenen Phialiden. *Persoonia* **6**: 185–191.

Gandolfo MA, Nixon KC, Crepet WL (2002) Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. *American Journal of Botany* **89**: 1940-1957.

Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie X-Q, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma L-J, St. Leger RJ, Zhao G-P, Pei Y, Feng M-G, Xia Y, Wang C (2011) Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics* **7**: 1-18.

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology* **3**: 408-414.

Gordon A (2011) "FASTX-Toolkit" FASTQ/A short-reads pre-processing tools. Available at http://hannonlab.cshl.edu/fastx_toolkit/

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644-652.

Hodge KT, Krasnoff SB, Humber RA (1996) *Tolypocladium inflatum* is the anamorph of *Cordyceps subsessilis*. *Mycologia* **88**: 715–719.
Jiménez ER (2009) Dextranase in sugar industry: A review. *Sugar Tech* **11**: 124-134.
Kepler RM, Ban S, Nakagiri A, Bischoff J, Hywel-Jones N, Owensby CA, Spatafora JW (2013) The phylogenetic placement of hypocrealean insect pathogens in the genus *Polycephalomyces*: an application of One Fungus One Name. *Fungal Biology* **117**: 611–622.
Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E, Hosoya T, Bischoff JF, Spatafora JW (2012) Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilaginoidea* (Clavicipitaceae). *American Journal of Botany* **99**: 552-561.
Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736-741.
Kobayasi Y (1939) On the genus *Cordyceps* and its allies on cicadae from Japan. *Bulletin of the Biogeographical Society of Japan* **9**: 145-176.
Kobayasi Y, Shimizu D (1960) Monographic studies of *Cordyceps* 1. Group parasitic on *Elaphomyces*. *Bulletin of the National Science Museum Tokyo* **5**: 69-85.
Korf I (2004) Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.
LePage B, Currah R, Stockey R, Rothwell G (1997) Fossil ectomycorrhizae from the Middle Eocene. *American Journal of Botany* **84**: 410-410.
Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* **13**: 2178-2189.
Lücking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT (2009) Fungi evolved right on track. *Mycologia* **101**: 810-822.
Maddison WP (1997) Gene trees in species trees. *Systematic biology* **46**: 523-536.
Mains EB (1957) Species of *Cordyceps* parasitic on *Elaphomyces*. *Bulletin of the Torrey Botanical Club* **84**: 243-251.
Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764-770.
Molnar I, Gibson DM, Krasnoff SB (2010) Secondary metabolites from entomopathogenic hypocrealean fungi. *Natural Product Reports* **27**: 1241-1275.
Nikoh N, Fukatsu T (2000) Interkingdom host jumping underground: phylogenetic analysis of entomoparasitic fungi of the genus *Cordyceps*. *Molecular Biology and Evolution* **17**: 629-638.
Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061-1067.
Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 298-297.
Poinar Jr GO, Thomas GM (1984) A fossil entomogenous fungus from Dominican amber. *Experientia* **40**: 578-579.
Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351-i358.
Prieto M, Wedin M (2013) Dating the diversification of the major lineages of Ascomycota (Fungi). *PloS One* **8**: e65576.

Roberts (1981) Toxins of Entomopathogenic Fungi. In: *Microbial control of pests and plant diseases*. (Burges HD, ed.): 441-463 London, New York: Academic Press.

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798-804.

Salichos L, Rokas A (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**: 327-331.

Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J,M Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach JE, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z (2013) Plant-symbiotic fungi as chemical engineers: Multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics* **9**: e1003323.

Shaw SR (1988) Euphorine phylogeny: the evolution of diversity in host?utilization by parasitoid wasps (Hymenoptera: Braconidae). Ecological Entomology 13: 323-335.

Smit AF, Green P (1996) RepeatMasker. *Published on the web at http://www. Repeatmasker.org*.

Sorenson WG, Frazer DG, Jarvis BB, Simpson J, Robinson VA (1987) Trichothecene mycotoxins in aerosolized conidia of *Stachybotrys atra*. *Applied and Environmental Microbiology* **53**: 1370-1375.

Spatafora JW, Sung GH, Sung JM, Hywel?Jones NL, White JF (2007) Phylogenetic evidence for an animal pathogen origin of ergot and the grass enndophytes. Molecular Ecology 16: 1701-1711.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5–59.

Sung, G-H, Poinar, GO, and JW Spatafora (2008) The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal–arthropod symbioses. *Molecular Phylogenetics and Evolution* **49**: 495-502.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564-577.

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

Wang B, Kang Q, Lu Y, Bai L, Wang C (2012) Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proceedings of the National Academy of Sciences* **109**: 1287-1292.

Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**: D581-D591.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: W445-W451.

Young, C. A., Felitti, S., Shields, K., Spangenberg, G., Johnson, R. D., Bryan, G. T., Saikia S, Scott B (2006) A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. *Fungal Genetics and Biology* **43**: 679-693.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821-829.

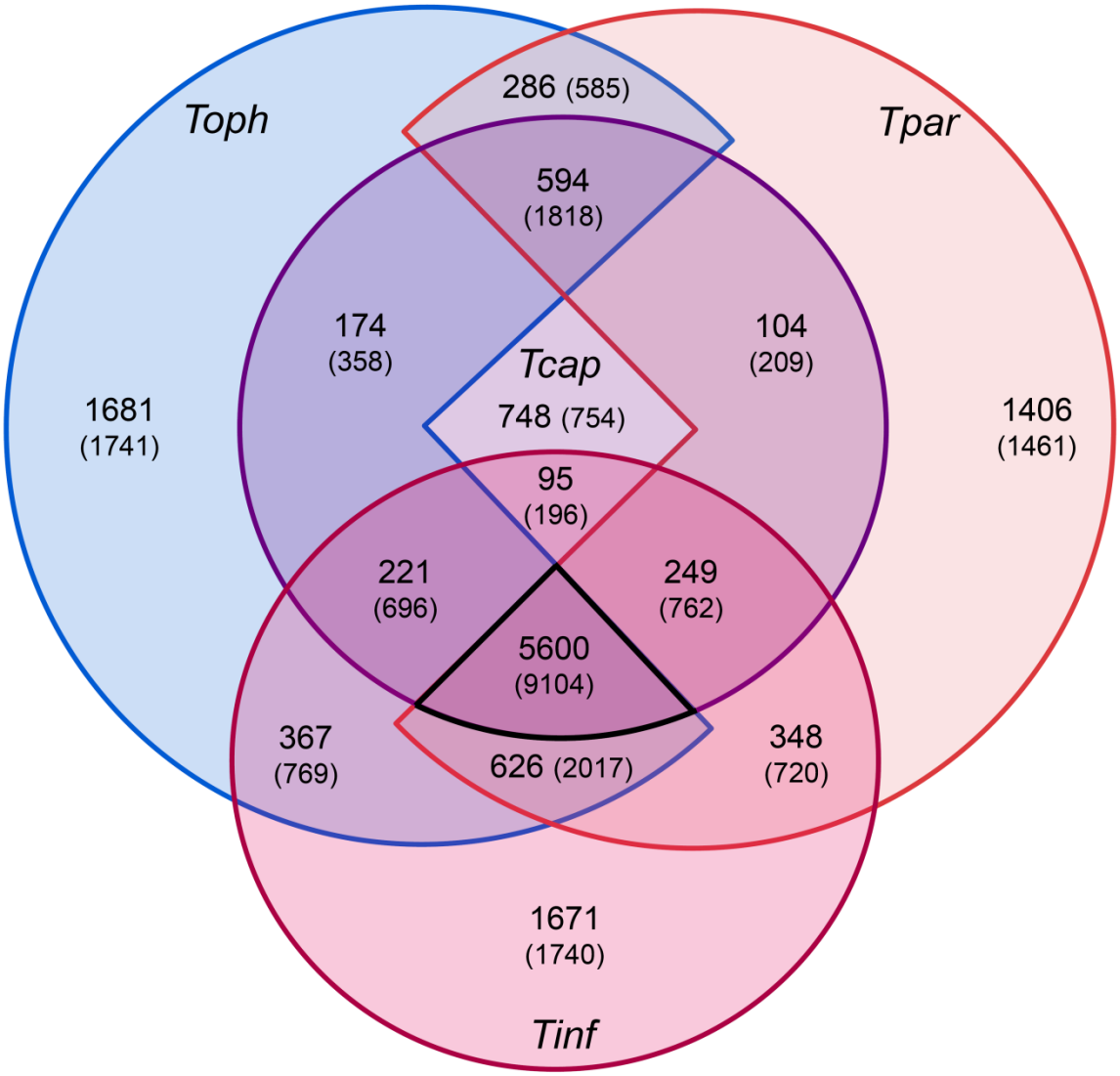


Figure 5.1. Venn diagram of orthologous protein clusters shared and unique to *T. capitatum* (Tcap), *T. inflatum* (Tinf), *T. ophioglossoides* (Toph), and *T. paradoxum* (Tpar), with numbers of total number of protein-coding genes in parentheses. The clusters and genes shared between all four taxa is outlined in black.

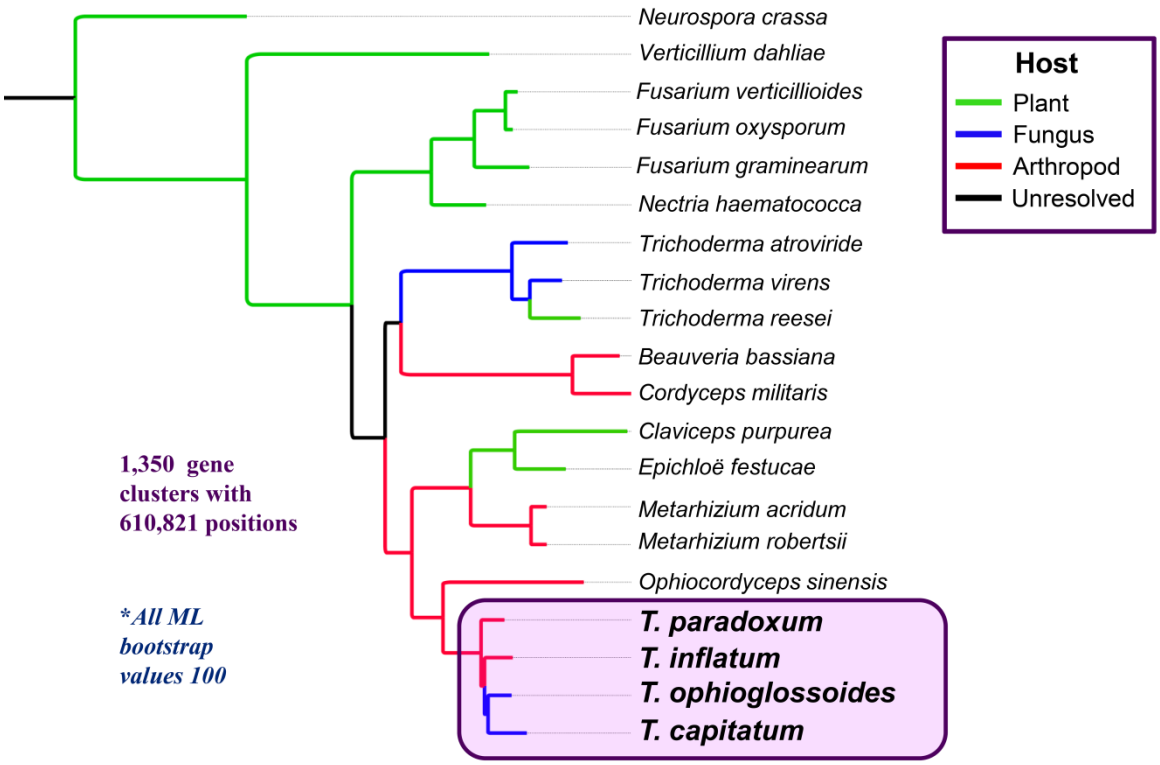


Figure 5.2. RAxML phylogeny of Hypocreales with outgroup taxa, *Verticillium dahliae* and *Neurospora crassa*, based on 1,350 orthologous protein clusters which included 610,821 positions in the alignment. Genus *Tolypocladium* is highlighted in the purple box, and host is mapped onto the branches with colors according to the key.

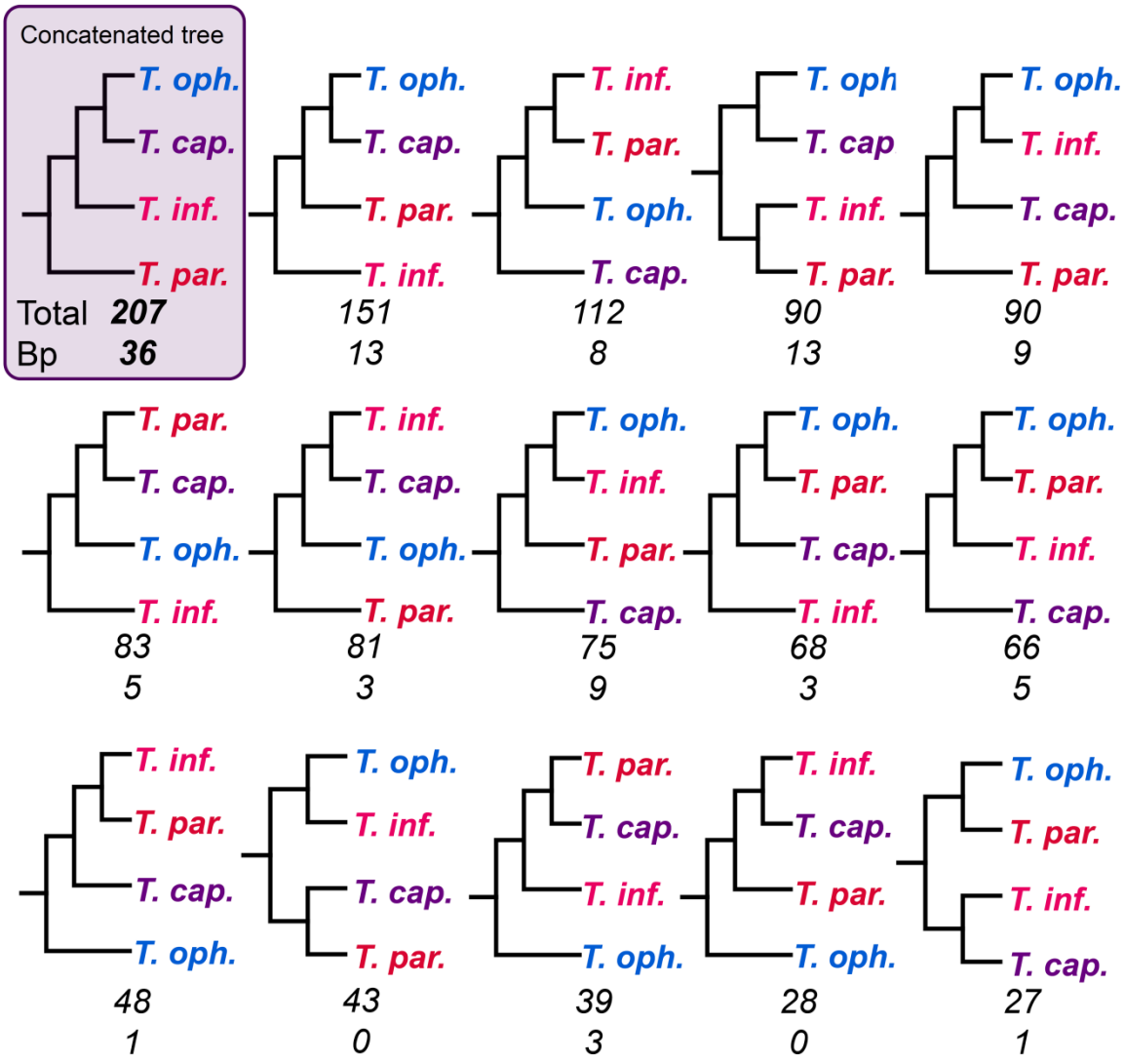


Figure 5.3. Total number of gene tree topologies recovered for the 15 possible *Tolypocladium* topologies from the 1,208 gene trees that resolve the genus as monophyletic (Top number for each tree; Total), and the subset of those total gene trees that resolved all three nodes within the genus with bootstrap support MLBP ≥ 70 (Bottom number; Bp). Abbreviations as in Figure 5.1.

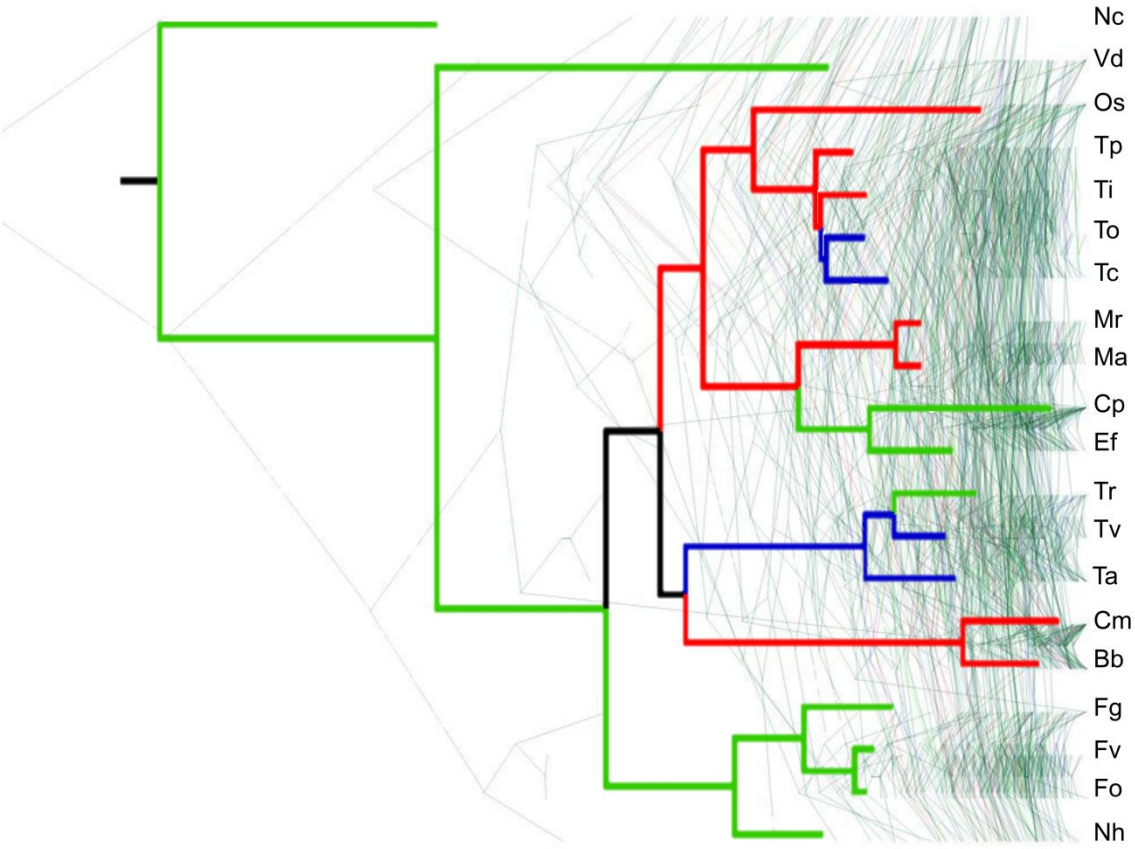
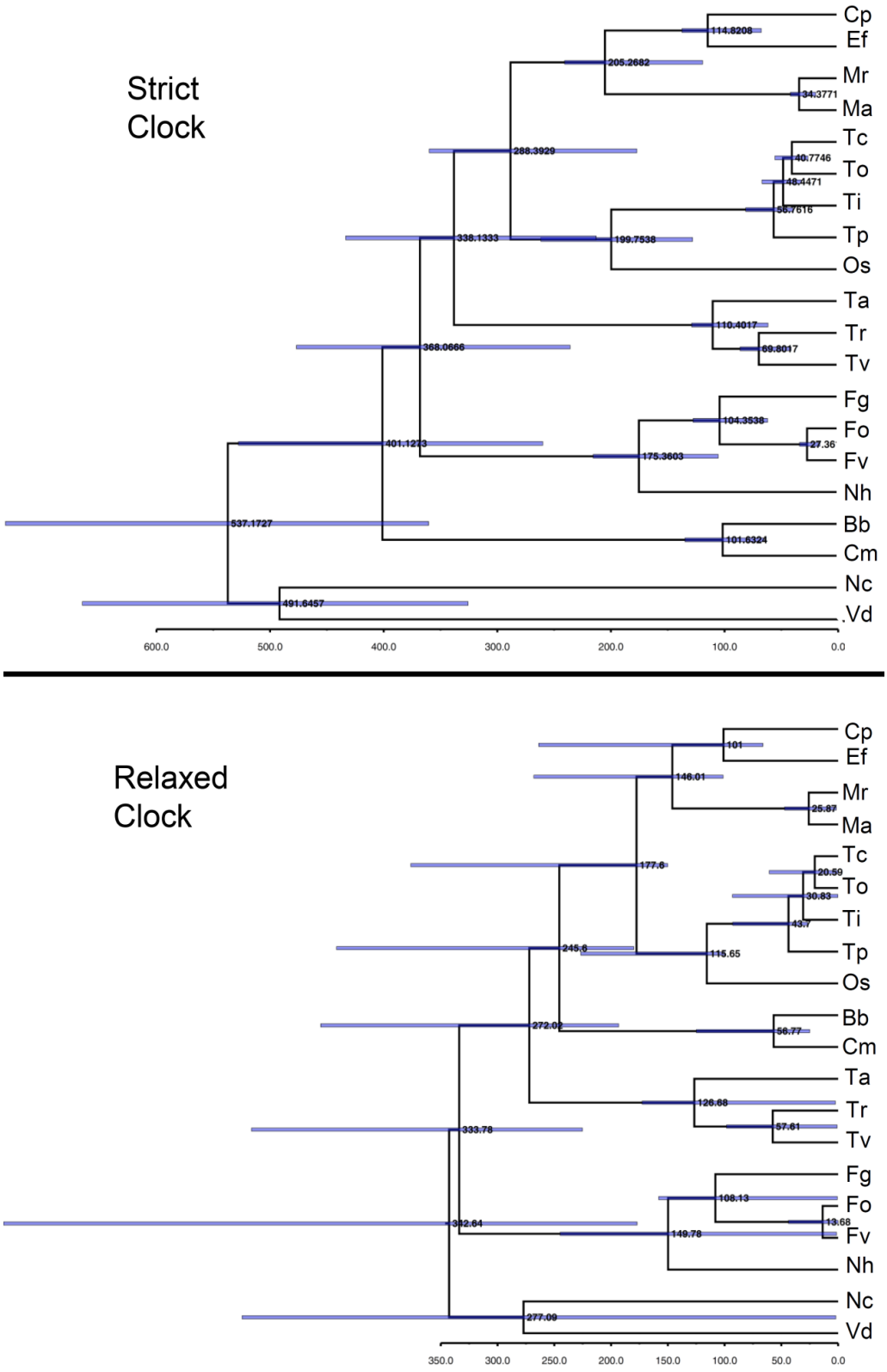


Figure 5.4. Densitree visualization of the 109 Hypocreales gene trees that have bootstrap support (MLBP ≥70) for all nodes in a monophyletic *Tolypocladium* with the concatenated tree overlay. Large differences in gene tree branch length can be seen. Abbreviations: *Neurospora crassa* (Nc), *Verticillium dahliae* (Vd), *Ophiocordyceps sinensis* (Os), *T. paradoxum* (Tp), *T. inflatum* (Ti), *T. ophioglossoides* (To), *T. capitatum* (Tc), *Metarhizium robertsii* (Mr), *M. acridum* (Ma), *Claviceps purpurea* (Cp), *Epichloë festucae* (Ef), *Trichoderma reesei* (Tr), *Tr. virens* (Tv), *Tr. atroviride* (Ta), *Cordyceps militaris* (Cm), *Beauveria bassiana* (Bb), *Fusarium graminearum* (Fg), *F. verticillioides* (Fv), *F. oxysporum* (Fo), and *Nectria haematococca* (Nh).

Figure 5.5. Comparison of the strict and relaxed log normal molecular clock BEAST analysis of 10 randomly selected clusters (of the 36 that resolved with support the concatenated topology of *Tolypocladium*; see Figure 5.3). Blue bars represent the 95% credibility intervals around the mean node ages (in millions of years). Scales are in millions of years. Abbreviations as in Figure 5.4.

Figure 5.5


Strict
Clock
Cp
Ef
Mr
Ma
Tc
To
Ti
Tp
Os
Ta
Tr
Tv
Fg
Fo
Fv
Nh
Bb
Cm
Nc
Vd
114.8208
205.2682
34.3771
288.3929
40.7746
48.4471
56.7616
338.1333
199.7538
110.4017
69.8017
368.0666
104.3538
27.36
401.1223
175.3603
537.1727
101.6324
491.6457
600.0
500.0
400.0
300.0
200.0
100.0
0.0
Relaxed
Clock
Cp
Ef
Mr
Ma
Tc
To
Ti
Tp
Os
Bb
Cm
Ta
Tr
Tv
Fg
Fo
Fv
Nh
Nc
Vd
101
146.01
25.87
177.6
20.59
30.83
43.7
245.8
115.65
272.02
56.77
126.68
57.61
333.78
108.13
13.68
392.64
149.78
277.09
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0

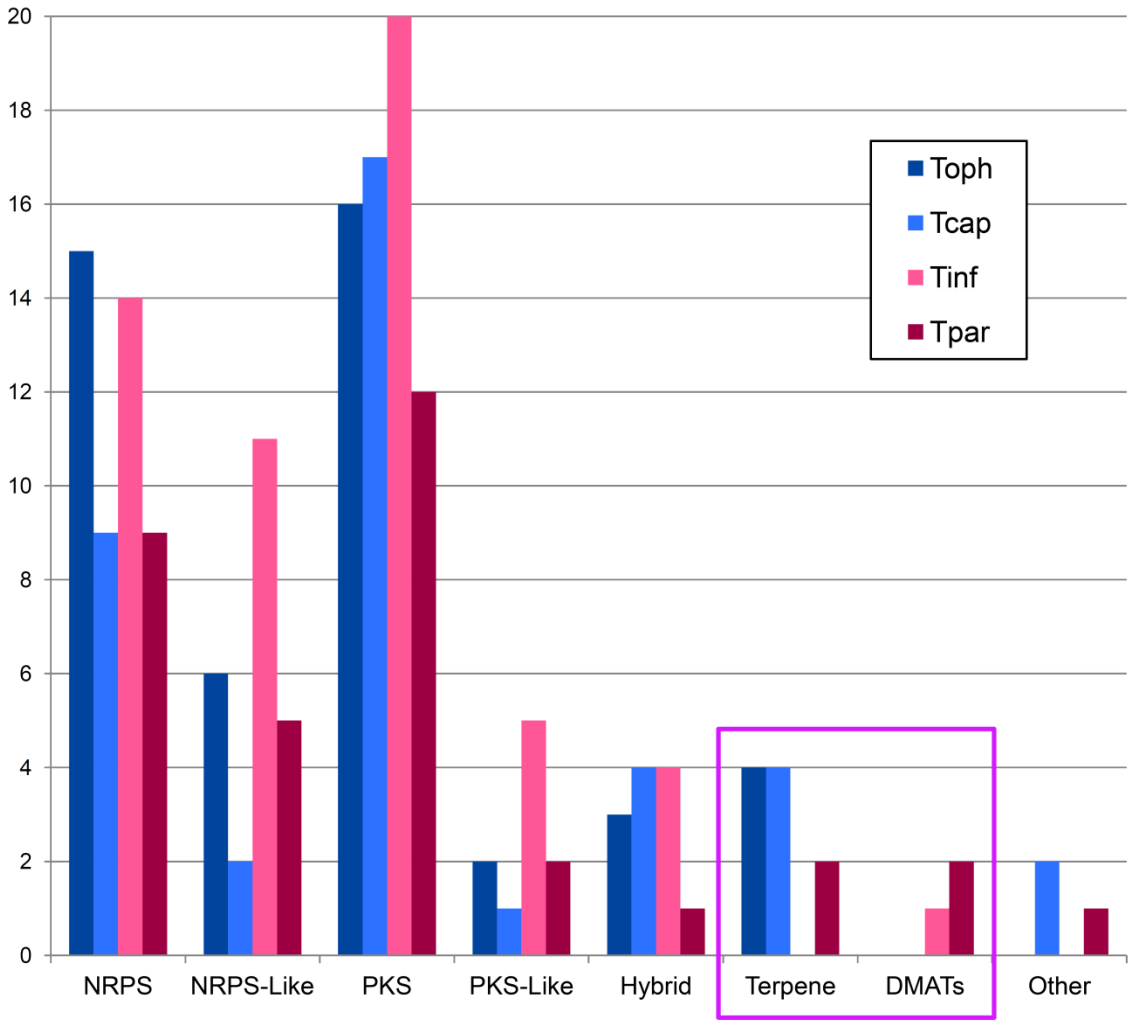


Figure 5.6. Total number of secondary metabolite core genes per category per *Tolypocladium* genome (Abbreviations as in Figure 5.1). Purple box highlights the differences in terpene and DMAT synthases in the mycoparasitic species compared to the insect pathogens.

Table 5.1. Genome statistics for the sequenced genomes of *T. ophioglossoides* (Toph), *T. capitatum* (Tcap), *T. inflatum* (Tinf), and *T. paradoxum* (Tpar). Mycoparasites are highlighted in blue, insect pathogens are in pink. **First published in this study.

	Toph	***Tcap***	***Tinf***	***Tpar***
Genome Size (Mb)	31.27	22.9	30.3	27.6
# of Scaffolds	173	1345	101	1139
N50	668,222	34,296	1,509,745	79,523
GC %	57.3	59.7	58	58
Longest scaffold (bp)	2,309,933	167,703	3,562,345	260,449
L50	12	202	8	116
CEGMA %	97	98	98	96
# Protein coding genes	10,135	8,106	9,998	9,773

Table 5.2. Secreted protein clusters specific to *T. inflatum* and *T. paradoxum* and their putative annotations. Gray boxes represent orphan protein models.

Cluster Name	*T. inflatum* and *T. paradoxum* ID	Putative annotation
ORTHOMCL1977	TINF05694 TINF02675 TINF05743 TPAR_05485	LysM domain-containing protein
ORTHOMCL3176	TINF03188 TINF03233 TPAR_02033 TPAR_02438	serine protease
ORTHOMCL5763	TINF00106 TPAR_02281 TPAR_04178	hypothetical protein
ORTHOMCL6180	TINF05702 TINF06745 TPAR_04183	candidapepsin-4 precursor
ORTHOMCL6395	TINF07751 TPAR_04933 TPAR_02792	GPI anchored serine-threonine rich protein
ORTHOMCL7402	TINF06858 TPAR_08472	Bromodomain associated domain protein
ORTHOMCL7411	TINF07546 TPAR_03963	alkaline phosphatase
ORTHOMCL7421	TINF09114 TPAR_09309	
ORTHOMCL7423	TINF00842 TPAR_09276	
ORTHOMCL7435	TINF06706 TPAR_03234	
ORTHOMCL7448	TINF08834 TPAR_01208	N2227-like protein
ORTHOMCL7486	TINF09860 TPAR_01928	hypothetical protein
ORTHOMCL7493	TINF08766 TPAR_07352	elastinolytic metalloproteinase Mep
ORTHOMCL7515	TINF02062 TPAR_06098	ABC-type Fe3+ transport system
ORTHOMCL7520	TINF09915 TPAR_04518	major allergen Asp f 2-like protein
ORTHOMCL7523	TINF06082 TPAR_05265	glycoside hydrolase family 37 protein
ORTHOMCL7545	TINF09879 TPAR_07102	beta-1,3-glucanosyltransferase
ORTHOMCL7549	TINF09927 TPAR_04519	Aminopeptidase merops family M28E
ORTHOMCL7551	TINF03566 TPAR_09166	hypothetical protein
ORTHOMCL7574	TINF00108 TPAR_06828	cell wall protein
ORTHOMCL7579	TINF01516 TPAR_00196	L-amino acid oxidase
ORTHOMCL7586	TINF04478 TPAR_00677	hypothetical protein
ORTHOMCL7592	TINF02814 TPAR_06230	hypothetical protein
ORTHOMCL7628	TINF02012 TPAR_08877	
ORTHOMCL7638	TINF08292 TPAR_04376	extracellular serine-rich protein

Table 5.2. (Continued)

Cluster Name	*T. inflatum* and *T. paradoxum* ID	Putative annotation
ORTHOMCL7649	TINF09784 TPAR_00175	
ORTHOMCL7701	TINF02758 TPAR_08213	Kazal domain-containing protein
ORTHOMCL7830	TINF04116 TPAR_06084	hypothetical protein
ORTHOMCL7837	TINF04948 TPAR_04249	glycoside hydrolase family 27
ORTHOMCL7840	TINF02180 TPAR_01046	hypothetical protein
ORTHOMCL7848	TINF09191 TPAR_03952	penicillin-binding protein
ORTHOMCL7863	TINF00415 TPAR_05619	hypothetical protein
ORTHOMCL7883	TINF01527 TPAR_01515	phospholipase D
ORTHOMCL7921	TINF08617 TPAR_04187	Triacylglycerol lipase
ORTHOMCL7923	TINF07573 TPAR_01705	
ORTHOMCL7929	TINF02880 TPAR_00645	inorganic pyrophosphatase
ORTHOMCL7933	TINF02208 TPAR_08880	hypothetical protein
ORTHOMCL7937	TINF04768 TPAR_02718	hypothetical protein
ORTHOMCL7939	TINF06693 TPAR_09664	
ORTHOMCL7943	TINF02848 TPAR_07765	phospholipase D
ORTHOMCL7957	TINF07505 TPAR_04459	heat-labile enterotoxin alpha chain
ORTHOMCL7960	TINF00757 TPAR_06085	protein-tyrosine phosphatase
ORTHOMCL7977	TINF00517 TPAR_04988	N,O-diacetyl muramidase
ORTHOMCL7982	TINF00581 TPAR_06521	Cyclin-like F-box
ORTHOMCL7990	TINF03544 TPAR_09305	herpesvirus latent membrane protein-like
ORTHOMCL7995	TINF01169 TPAR_00699	
ORTHOMCL8036	TINF03259 TPAR_07429	extracellular dioxygenase
ORTHOMCL8038	TINF04379 TPAR_08119	IdtS
ORTHOMCL8039	TINF05403 TPAR_08989	hypothetical protein
ORTHOMCL8042	TINF04699 TPAR_08125	P450 monooxygenase
ORTHOMCL8043	TINF03163 TPAR_05411	hypothetical protein

Table 5.2. (Continued)

Cluster Name	*T. inflatum* and *T. paradoxum* ID	Putative annotation
ORTHOMCL8065	TINF02991 TPAR_07303	hypothetical protein
ORTHOMCL8081	TINF05087 TPAR_03043	ankyrin repeats & 6-phosphofructo-2-kinase
ORTHOMCL8083	TINF02615 TPAR_08243	hypothetical protein
ORTHOMCL8088	TINF02263 TPAR_03803	Fum15
ORTHOMCL8110	TINF05389 TPAR_00996	IDI-2 precursor
ORTHOMCL8118	TINF05693 TPAR_01300	WSC domain protein
ORTHOMCL8172	TINF02201 TPAR_07753	
ORTHOMCL8185	TINF03017 TPAR_05409	hypothetical protein

Table 5.3. Secreted clusters specific to *T. ophioglossoides* and *T. capitatum* and their putative annotations. Boxes in gray are orphans.

Cluster Name	*T. ophioglossoides* and *T. capitatum* ID	Putative annotation
ORTHOMCL8224	TOPH_06753 TCAP_08088	Peptidoglycan-binding Lysin subgroup
ORTHOMCL8249	TOPH_09677 TCAP_07769	WD40/YVTN repeat-like-containing domain protein
ORTHOMCL8280	TOPH_09355 TCAP_00961	
ORTHOMCL8318	TOPH_01209 TCAP_02412	Peptidase S10, serine carboxypeptidase
ORTHOMCL8393	TOPH_08942 TCAP_05778	siderophore biosynthesis enzyme
ORTHOMCL8495	TOPH_02171 TCAP_06931	FAD/FMN-containing isoamyl alcohol oxidase MreA-like
ORTHOMCL8513	TOPH_06263 TCAP_02792	hypothetical protein
ORTHOMCL8516	TOPH_06479 TCAP_07174	FAD-binding domain-containing protein
ORTHOMCL8517	TOPH_03279 TCAP_03541	V8-like Glu-specific endopeptidase
ORTHOMCL8534	TOPH_07042 TCAP_07772	hypothetical protein
ORTHOMCL8552	TOPH_03786 TCAP_04715	

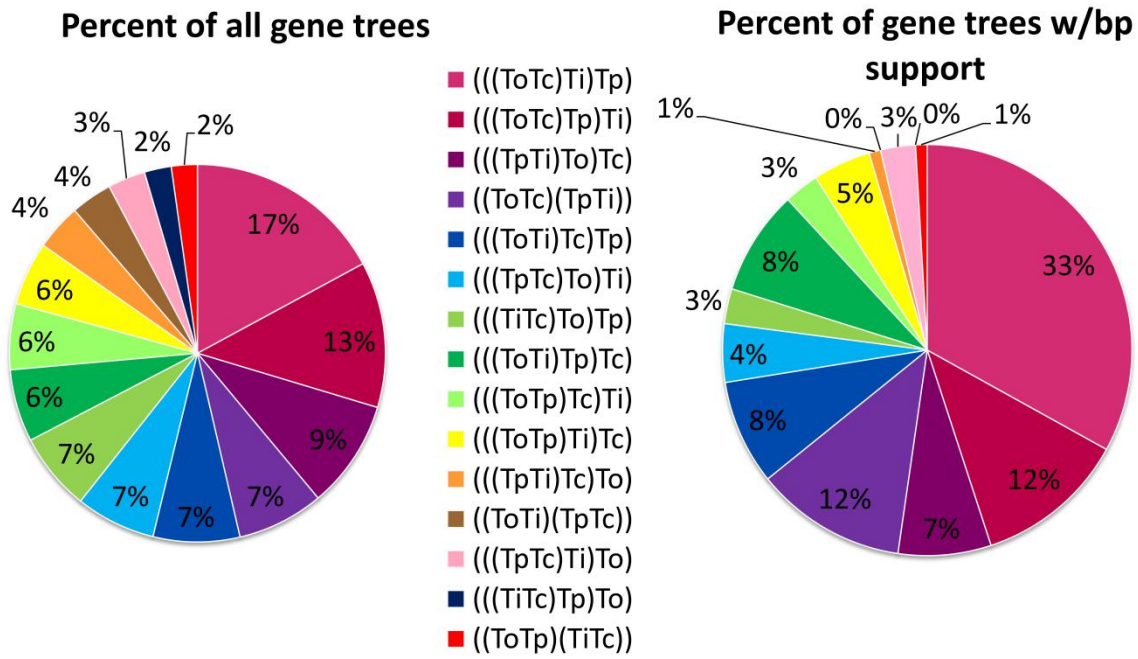


Figure 5.S1. Percentages of topologies and topologies with support. Of the gene trees that resolve a monophyletic *Tolypocladium*, the proportion of each topology recovered, and the proportion of each topology receiving bootstrap support (MLBP≥70) for all *Tolypocladium* internodes out of all the gene trees that receive support.

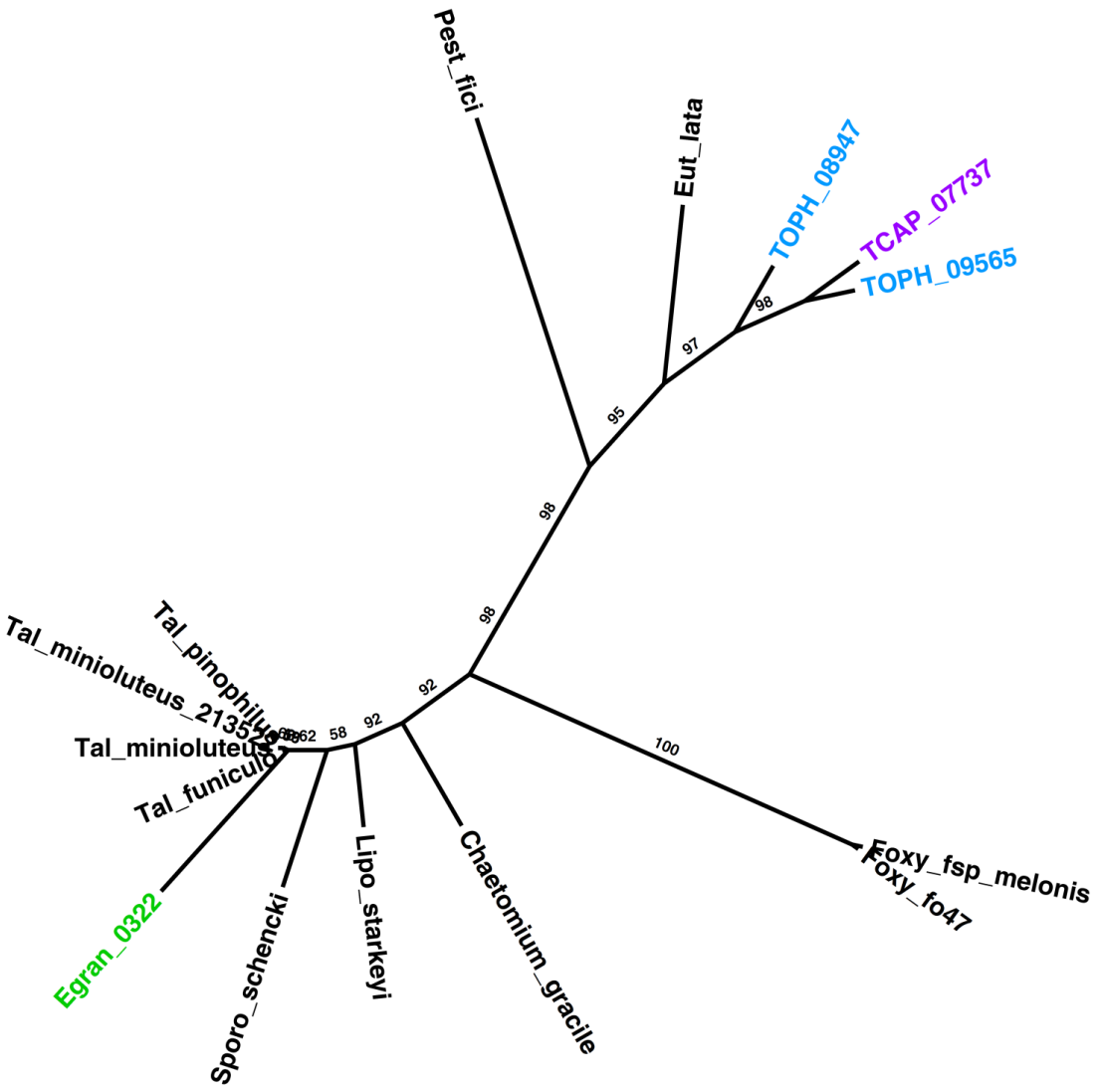


Figure 5.S2. Phylogeny of dextranse genes specific to *T. ophioglossoides* and *T. capitatum* with closest homologs in the Soradariomycetes including *Eutypa lata* (Eut_lata), *Pestalotiopsis fici* (Pest_fici), *Fusarium oxysporum* strains (Foxy_fsp_melonis and Foxy_fo47) , and *Chaetomium gracile*, the Eurotiales including *Talaromyces pinophilus* (Tal_pinophilus), *Ta. minioluteus* (Tal_minioluteus), *Ta. funiculo* (Tal_funiculo), and *Elaphomyces granulatus* (Egran_0322), and other lineages of fungi [Lipo_starkeyi=*Lipomyces starkeyi*, Sporo_schencki=*Sporothrix schencki*].

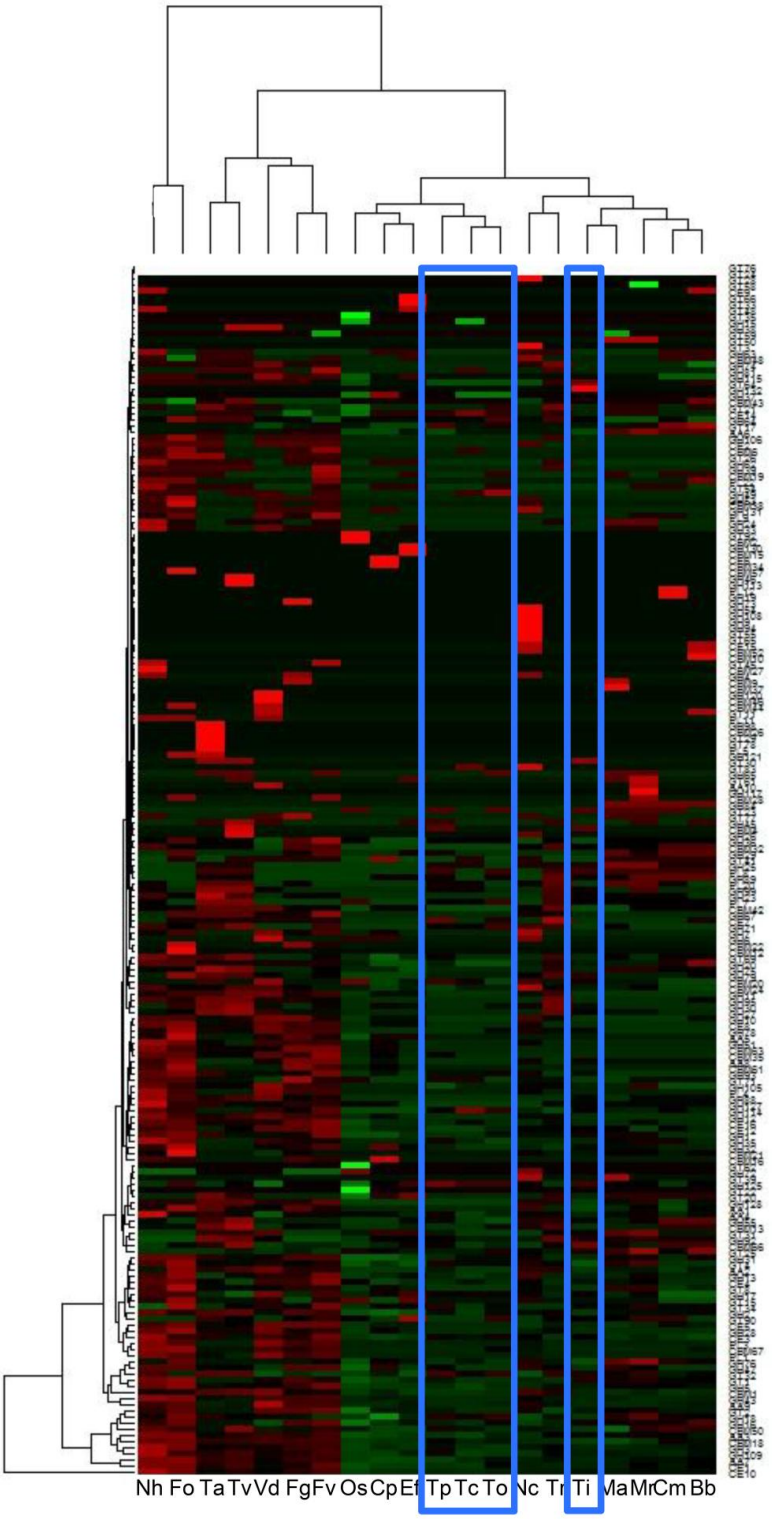


Figure 5.S3. Heatmap of CAZyme family counts for *Tolypocladium* species (blue boxes) compared to other hypocrealean taxa and outgroups (abbreviations as in Figure 5.4).

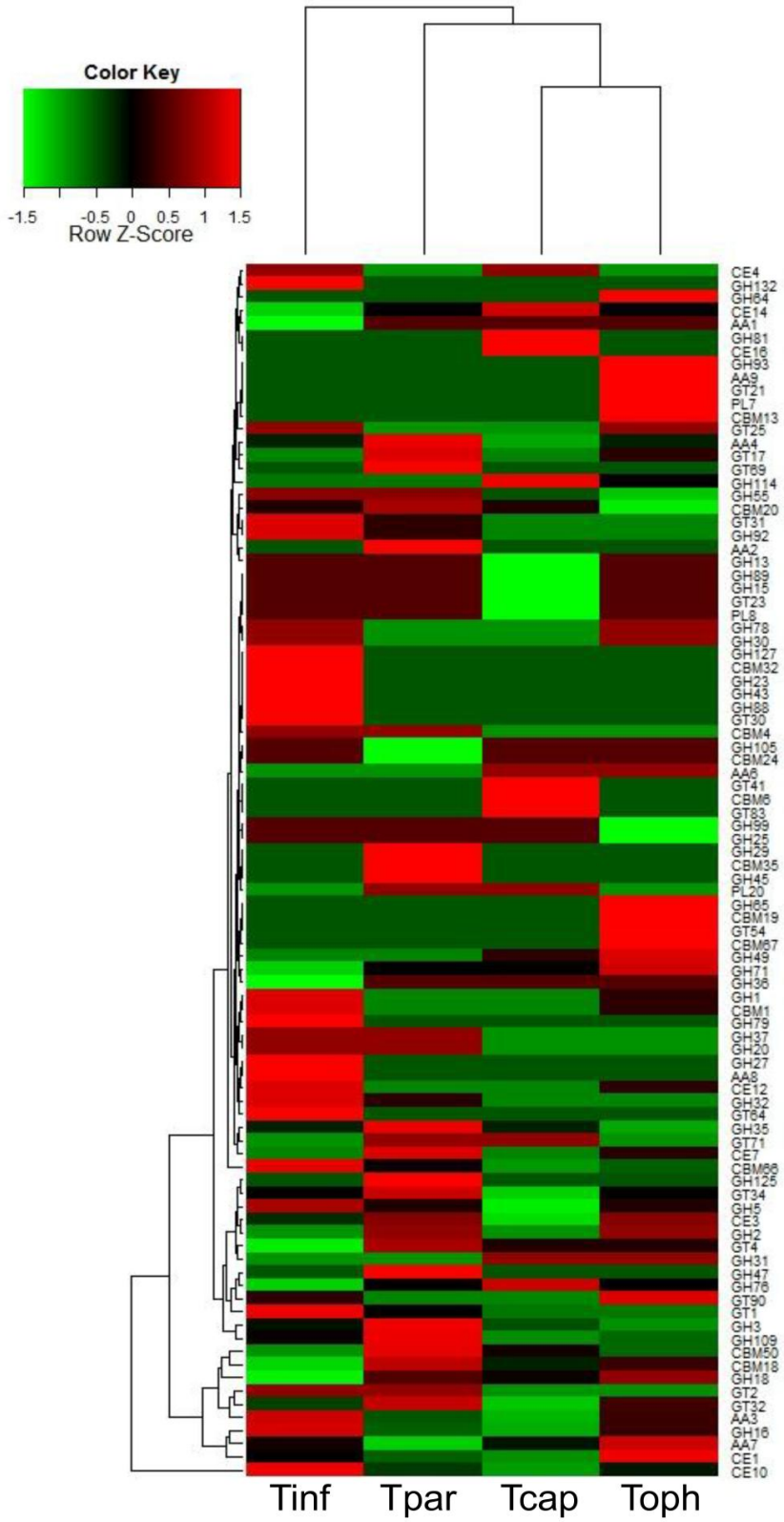


Figure 5.S4. Heatmap comparison of CAZyme family counts for *Tolypocladium* species.

Figure 5.S5. Phylogeny of known fungal Adenylation domains and those mined from *Tolypocladium* genomes. Supported branches (MLBP ≥70) are colored green, and domains belonging to *T. capitatum* are colored light blue, *T. inflatum* (pink), *T. ophioglossoides* (dark blue), and *T. paradoxum* (red). Major clades are labeled based on known products from other fungi and *T. inflatum*.

Figure 5.S5

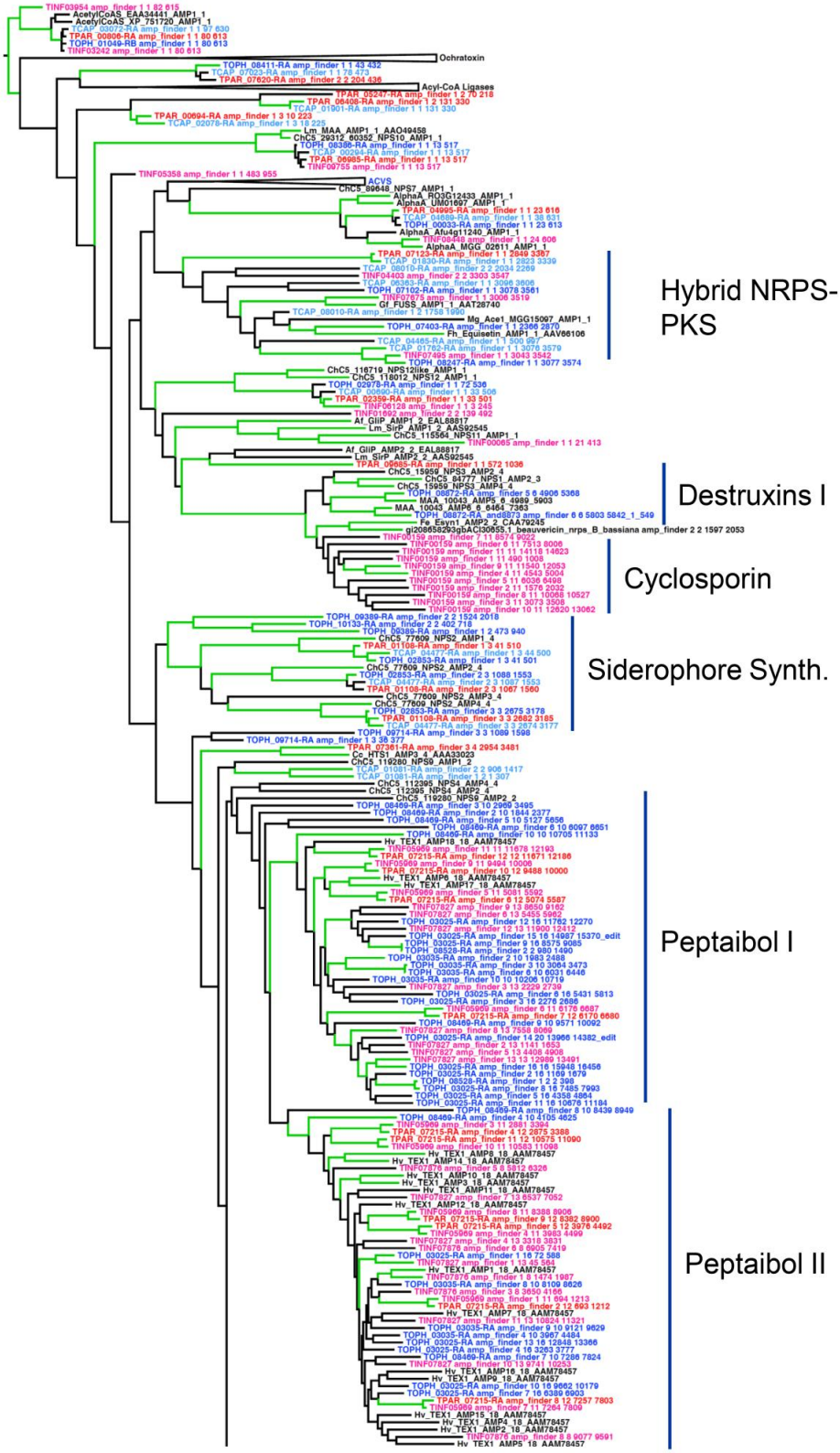

Hybrid NRPS-PKS
Destruxins I
Cyclosporin
Siderophore Synth.
Peptaibol I
Peptaibol II

Figure 5.S5 (Continued)

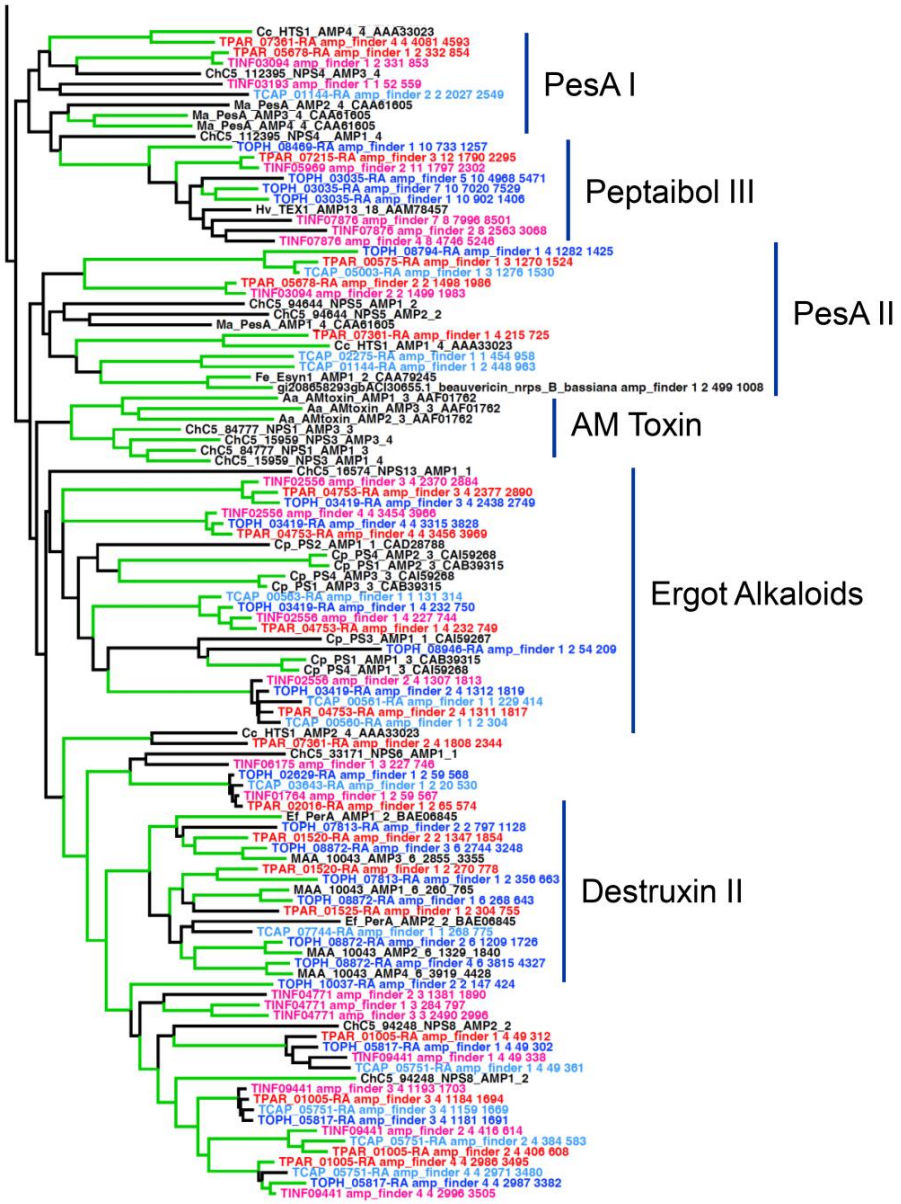


Table 5.S1. Core secondary metabolite predictions for *T. capitatum*. Multiple genes predicted as part of the same cluster are highlighted in the same color.

Core Gene	Scaffold	Description	anti-SMASH	SMURF
TCAP_07926	NODE_20_length_16623_cov_92.796188	Terpene	x	
TCAP_00294	NODE_79_length_43736_cov_96.827850	NRPS-Like	x	x
TCAP_00428	NODE_114_length_42349_cov_95.855675	PKS	x	x
TCAP_00559	NODE_144_length_78432_cov_89.380219	NRPS	x	
TCAP_00648	NODE_166_length_33084_cov_91.434738	PKS	x	x
TCAP_01018	NODE_233_length_94954_cov_93.630409	PKS	x	x
TCAP_01055	NODE_246_length_15696_cov_90.827156	PKS	x	x
TCAP_01081	NODE_254_length_16918_cov_81.966782	NRPS		x
TCAP_01144	NODE_285_length_10735_cov_100.028503	NRPS	x	x
TCAP_01163	NODE_293_length_59719_cov_92.936836	Terpene	x	
TCAP_01539	NODE_380_length_27253_cov_85.736214	Terpene	x	
TCAP_01762	NODE_450_length_84126_cov_97.154831	Hybrid	x	x
TCAP_01806	NODE_455_length_23211_cov_97.313171	PKS	x	x
TCAP_01809	NODE_455_length_23211_cov_97.313171	PKS	x	x
TCAP_01830	NODE_457_length_102010_cov_94.237976	Hybrid	x	x
TCAP_02274	NODE_570_length_40528_cov_96.272331	PKS	x	x
TCAP_02275	NODE_570_length_40528_cov_96.272331	NRPS-Like	x	x
TCAP_08059	NODE_768_length_12066_cov_73.482101	Terpene	x	
TCAP_03123	NODE_845_length_95458_cov_90.981308	PKS-Like		x
TCAP_03221	NODE_868_length_47867_cov_93.513527	PKS	x	x
TCAP_03390	NODE_906_length_25544_cov_92.614197	PKS	x	x
TCAP_03404	NODE_918_length_17662_cov_99.096931	PKS	x	x
TCAP_03643	NODE_1002_length_12413_cov_84.266975	NRPS	x	x
TCAP_04465	NODE_1531_length_23681_cov_97.023941	NRPS	x	x
TCAP_04466	NODE_1531_length_23681_cov_97.023941	PKS	x	x
TCAP_04477	NODE_1542_length_65483_cov_94.608849	NRPS	x	x

Table 5.S1 (Continued)

Core Gene	Scaffold	Description	anti-SMASH	SMURF
TCAP_04689	NODE_1743_length_25549_cov_93.212845	Other	x	
TCAP_04976	NODE_1982_length_64769_cov_90.175224	PKS	x	x
TCAP_05003	NODE_1984_length_79077_cov_92.388504	NRPS	x	x
TCAP_05457	NODE_2311_length_106035_cov_92.998680	PKS	x	x
TCAP_05751	NODE_2726_length_72463_cov_93.302841	NRPS	x	x
TCAP_05990	NODE_3319_length_24352_cov_86.359192	Hglks-PKS	x	
TCAP_07744	NODE_4063_length_37213_cov_90.404297	Other	x	
TCAP_06363	NODE_4186_length_66294_cov_92.988533	Hybrid	x	x
TCAP_06767	NODE_5234_length_60471_cov_96.740417	Hybrid	x	x
TCAP_07023	NODE_6224_length_43091_cov_87.391960	NRPS	x	
TCAP_07124	NODE_6981_length_79953_cov_88.345314	PKS	x	x
TCAP_07314	NODE_9290_length_33434_cov_100.429207	PKS	x	x
TCAP_07319	NODE_9290_length_33434_cov_100.429207	PKS	x	x

Table 5.S2. Secondary metabolite predictions for *T. paradoxum*. Multiple genes predicted as part of the same cluster are highlighted in the same color.

Core Gene	Scaffold	Description	anti-SMASH	SMURF
TPAR_00575	NODE_390_length_53150_cov_75.627998	NRPS-Like	x	x
TPAR_00873	NODE_554_length_26394_cov_79.396263	PKS	x	x
TPAR_01005	NODE_679_length_59636_cov_79.653450	NRPS	x	x
TPAR_01108	NODE_771_length_24286_cov_80.307297	NRPS	x	x
TPAR_01520	NODE_1106_length_110920_cov_78.881119	NRPS	x	x
TPAR_01525	NODE_1106_length_110920_cov_78.881119	NRPS-Like	x	x
TPAR_01585	NODE_1190_length_45953_cov_80.320328	NRPS-Like	x	x
TPAR_02016	NODE_1435_length_92676_cov_77.662468	NRPS	x	x
TPAR_02367	NODE_1830_length_98209_cov_78.840118	PKS-Like	x	x
TPAR_02393	NODE_1890_length_152105_cov_80.674706	Other	x	
TPAR_02498	NODE_1987_length_18371_cov_77.396767	Terpene	x	
TPAR_02978	NODE_2475_length_80250_cov_85.215889	PKS	x	x
TPAR_03202	NODE_2730_length_33995_cov_82.740639	PKS	x	x
TPAR_03745	NODE_3423_length_34887_cov_85.782616	PKS	x	x
TPAR_03807	NODE_3533_length_147468_cov_79.679886	PKS	x	x
TPAR_04492	NODE_5097_length_147358_cov_79.809937	NRPS	x	x
TPAR_04493	NODE_5097_length_147358_cov_79.809937	PKS	x	x
TPAR_04684	NODE_5560_length_19641_cov_75.153305	PKS	x	x
TPAR_04753	NODE_5637_length_134795_cov_79.314178	NRPS	x	x
TPAR_05168	NODE_6344_length_60288_cov_80.321709	Terpene	x	
TPAR_05297	NODE_6901_length_221959_cov_82.272202	PKS	x	x
TPAR_05678	NODE_9151_length_116873_cov_81.953735	NRPS	x	x
TPAR_06055	NODE_12488_length_220493_cov_80.158409	PKS	x	x
TPAR_06058	NODE_12488_length_220493_cov_80.158409	PKS	x	x

Table 5.S2 (Continued)

Core Gene	Scaffold	Description	anti-SMASH	SMURF
TPAR_06530	NODE_15248_length_99165_cov_79.136955	PKS	x	x
TPAR_06653	NODE_15847_length_110007_cov_84.814651	PKS	x	x
TPAR_06961	NODE_18335_length_54632_cov_76.764885	PKS-Like		x
TPAR_06985	NODE_18586_length_75015_cov_78.107483	NRPS-Like	x	x
TPAR_07123	NODE_20578_length_151764_cov_80.355194	HYBRID	x	x
TPAR_07215	NODE_21367_length_71547_cov_87.168045	NRPS	x	x
TPAR_07361	NODE_22835_length_90426_cov_81.684059	NRPS	x	x
TPAR_07414	NODE_24413_length_68987_cov_77.316612	DMATs	x	x
TPAR_07620	NODE_25460_length_86066_cov_80.073929	NRPS-Like	x	
TPAR_08126	NODE_36682_length_119547_cov_81.818863	DMATs	x	x

Chapter 6. Metagenome sequence of *Elaphomyces granulatus* from sporocarp tissue reveals Ascomycota ectomycorrhizal fingerprints of genome expansion and a Proteobacteria rich microbiome.

C. Alisha Quandt, Annegret Kohler, Cedar Hesse, Emmanuelle Morin, Claude Murat

Aurelie Deveau, Thomas Sharpton, Francis Martin, Joseph W. Spatafora

Target: *Environmental Microbiology*

Introduction

Elaphomyces Nees (Elaphomycetaceae, Eurotiales) is an ectomycorrhizal genus of fungi with broad host associations that include both angiosperms and gymnosperms (Trappe 1979). As the only family to include mycorrhizal taxa within class Eurotiomycetes, Elaphomycetaceae represents one of the few independent origins of the mycorrhizal symbiosis in Ascomycota (Tedersoo *et al.* 2010). Other ectomycorrhizal Ascomycota include several genera within Pezizomycetes (*e.g.*, *Tuber*, *Otidea*, *etc.*) and *Cenococcum* in Dothideomycetes (Tedersoo *et al.* 2006, 2010). The only other genome sequence published from an ectomycorrhizal ascomycete is *Tuber melanosporum* [Pezizales, Pezizomycetes], the black perigord truffle (Martin *et al.* 2010). The *T. melanosporum* genome exhibited a significant expansion in size as compared to other filamentous Ascomycota (125 megabases [Mb] versus an average of 30-40 Mb), and contained a high number of transposable elements. Conversely, it was characterized by fewer protein-coding genes than a typical Ascomycota with a reduced number of genes encoding for secondary metabolites and Carbohydrate Active Enzymes (CAZymes) (Cantarel *et al.* 2009), including those involved in degradation of plant material. Recently, the genome of a saprobic pezizomycete, *Pyronema confluens*, has been sequenced (Traeger *et al.* 2013), providing the ability to compare *T. melanosporum* with a non-mycorrhizal relative. The genome of *P. confluens* was found to be much smaller (50 Mb) than *T. melanosporum* and contain several thousand more protein models than its ectomycorrhizal relative including a large suite of CAZymes.

Elaphomyces is one of two genera in the family Elaphomycetaceae, Eurotiales (Dictionary of the Fungi 10th Ed. Kirk *et al.* 2008). The type species, *E. granulatus* Fr.,

has a broad distribution across both Europe and North America, but there has been debate about whether this represents a single species or a species-complex based on morphological variation (see Hawker *et al.* 1967). The genus includes more than 90 described species that collectively have a near global distribution with species described from all continents except Africa and Antarctica (Castellano *et al.* 1989, 2011, 2012). *Elaphomyces* spp. produce subglobose, hypogeous "truffle" fruiting bodies, which have an organized outer layer of tissue called a peridium that encloses the gleba or spore bearing tissue (Trappe 1979). "Truffles" have evolved independently numerous times in multiple clades of Kingdom Fungi (Hibbett & Thorn 2001, Hosaka *et al.* 2006), presumably under selection pressure for spore dispersal by animals (Maser *et al.* 1978, Theirs 1984). Like most mycorrhizal symbionts, *Elaphomyces* spp. are inherently difficult to culture and maintain in the laboratory, although there are reports of successful culturing of some species (Miller & Miller 1984).

Despite striking differences in ecology and morphology, *Elaphomyces* is a member of Eurotiales, an order that mostly comprises soil-inhabiting and medicinally important molds, including *Penicillium* and *Apsergillus*. This evolutionary relationship was first hypothesized by Korf (1973), but it was the use of molecular data that solidified the placement of *Elaphomyces* within Eurotiales (Landvik *et al.* 1996, LoBuglio *et al.* 1996). Two studies (Miller *et al.* 2001, Geiser *et al.* 2006), using varying amounts of genetic data and taxonomic sampling, provided some support for Elaphomycetaceae as an early diverging lineage within Eurotiales, but its exact relationship to the other major clades of the order remains unclear. One uniting character for the order is the production of completely enclosed ascomata, called cleistothecia, with a few exceptions (*i.e.*

Trichocoma spp.). Several genome sequences are available for *Penicillium* and *Aspergillus* spp., in addition to other eurotialean genomes including *Monascus ruber*, used in rice fermentation, and *Talaromyces* spp., many of which were traditionally classified as *Penicillium* because they often possess biverticillate penicillium-like conidiophores (Samson *et al.* 2011). Sister to Eurotiales is the order Onygenales, of which many species are animal-associated either as dermatophytes (*e.g.*, *Trichophyton* spp.), pathogens (*e.g.*, *Coccidioides* spp.), or coprophiles (*e.g.*, *Histoplasma* spp.). Finally, sister to both of these orders is the Coryneliales which includes plant pathogens, such as *Caliciopsis orientalis*. Within Coryneliales, both ascus morphology and ascomatal ontogeny differ from that of species in Eurotiales and Onygenales, and their placement within the fungal tree of life was debated until the use of molecular data (Inderbitzin *et al.* 2004, Geiser *et al.* 2006).

Recent studies are just beginning to shed light on the microbiome of fungi and their fruiting bodies (Antony-Babu *et al.* 2013, Desirò *et al.* 2014). The microbial community associated with the fruiting bodies of *Elaphomyces* represents an unknown aspect of its biology, but this community could play an important ecological role in the ectomycorrhizal symbiosis (Deveau *et al.* 2007, Frey-Klett *et al.* 2007, Lehr *et al.* 2007, Kurth *et al.* 2013). Ectomycorrhizal helper bacteria (EHB) of Streptomycetaceae have been shown to promote hyphal growth and formation of mycorrhizae, the latter through the production of analogs of the plant growth hormone, auxin (Riedlinger *et al.* 2006). These EHB may also function in defining the bacterial community that may exist in the ectomycorhizosphere through the production of antibiotics (Frey-Klett *et al.* 2011). To date, the most dominant bacteria commonly found within the tissues of various fungi and

plants include Streptomycetaceae, Rhizobiaceae, and Pseudomonadaceae. These families are known to generally interact with hosts (*e.g.*, nitrogen fixation, anti-pathogenic secondary metabolite production, *etc.*), but are also known to include pathogenic members (Crawford *et al.* 1993, Parke *et al.* 2001, Offre *et al.* 2008, Hayat *et al.* 2010). Bacteria are either extra- or endohyphal in association with fungi; the endohyphal bacteria are often vertically transmitted, whereas it is presumed many of the extrahyphal symbionts are derived from the soil microbial community (Mondo *et al.* 2012). Endohyphal bacteria are far less characterized, which is likely due to difficulties in culturing, but they have been identified in multiple lineages of fungi (Hoffman *et al.* 2013, Desirò *et al.* 2014), including several from arbuscular mycorrhizal fungi of Glomeromycota (Bianciotto *et al.* 2004, Mondo *et al.* 2013, Desirò *et al.* 2014). Studies examining the microbiome of truffles of *Tuber* spp. have shown that both the gleba and peridium harbor complex bacterial communities (Antony-Babu *et al.* 2013) and that they are enriched in bacteria of Bradyrhizobiaceae (Barbieri *et al.* 2005, Gryndler *et al.* 2013, Antony-Babu *et al.* 2013).

Several questions about the biology of *Elaphomyces* remain unanswered including: (1) How is *Elaphomyces* related to other lineages of Eurotiales? (2) What similarities, if any, exist between the *E. granulatus* genome and that of the other sequenced ectomycorrhizal ascomycete *T. melanosporum*? (3) What genes are expanded and contracted in *Elaphomyces* as compared to saprobic relatives in Eurotiales? (4) What is the bacterial community composition in an *E. granulatus* truffle and is it similar to communities found in soil and that of *Tuber* truffles? A draft genome of *E. granulatus* was sequenced to obtain insight into these questions. Due to the absence of an isolated

culture, *E. granulatus* DNA was extracted from field-collected sporocarp tissue and subject to shotgun metagenomic sequencing. Analysis of these metagenomic data characterized the gene space of *E. granulatus*, which in turn enabled its comparison to a broad representation of eurotialean fungi, and quantified the taxonomic and functional diversity of the *Elaphomyces* microbiome.

Materials & Methods

Tissue extraction and nucleic acid preparation

A single collection of several *E.granulatus* fruiting bodies (OSC 145934) was made from Cummins Creek area of Siuslaw National Forest, Lane County, Oregon, (44°16’03.9”N, 124°5’59.8”W) and transferred to the laboratory for nucleic acid extraction. The largest of the fruiting bodies was cleaned of debris and surface sterilized for 30 seconds using 85% ethanol and a sterilized spatula was used to physically remove the warty outer peridium. The fruiting body was then cut in half and as much of the gleba as possible was removed from the peridium using a sterilized spatula. The entire peridium was lyophilized and half of the peridium was ground in liquid nitrogen with a mortar and pestle. DNA was extracted from ground tissue using a Qiagen Plant DNeasy kit using the “fungi” protocol without modifications. Paired-end Illumina libraries were then constructed with ~5 µg DNA using New England Biomedicals NEBNext kits and protocols, with gel-selected small insert sizes of 322 and 354 base pairs (bp), from genomic DNA prepared from two separate extractions. Each library was sequenced for 101-cycles for each paired-end on one full lane of the Illumina HiSeq2000 at the Oregon

State University Center for Genome Research and Biocomputing (CGRB); the 322 bp library with Illumina version 2 chemistry and the 354 bp library with version 3 chemistry. RNA was extracted using the Qiagen Plant RNeasy kit and prepared for sequencing and sequenced at Beckman Coulter Genomics for improvement of gene model annotation.

Bioinformatic analyses

Reads were initially assembled and filtered for quality in CLC Genomics Workbench v. 6.5.1. BLASTx searches of contigs to the NCBI database initially identified several most commonly encountered bacterial taxa from the fruit body microbiome, and the subsequent methods were employed to separate bacterial data from the target genome and are summarized in Figure 6.S1. A database of the 35 most commonly hit genera of bacteria was created (Table 6.S1) and raw reads were aligned to this custom database using Bowtie 2 v. 2.0.6. (Langmead & Salzberg 2012) with default alignment settings. Reads that aligned to the bacterial database were excluded, and the remaining reads were reassembled in CLC Genomics Workbench as before. Subsequent BLASTx revealed that many bacterial sequences had not been removed due to significant divergence from sequenced reference strains. A binning procedure based on e-values ($\leq 1e^{-5}$) was then used to separate contigs into three categories: 1) those contigs with hits to sequenced Eurotiomycetes genomes available on the MycoCosm website (Grigoriev *et al.* 2014), 2) contigs with hits to the NCBI bacterial database, and 3) those contigs without hits to either. This third category was sorted based on coverage, either greater or less than 200x coverage. Using the contigs in category one and category three which had >200x coverage, *ab initio* protein modeling was performed using AUGUSTUS (Stanke *et*

al. 2006) with *A. fumigatus* as the reference species. Many (14,760) of the 27,436 predicted protein models were found to be bacterial using BLASTx. Manual curation of the contigs containing these bacterial hits was completed to mitigate the inclusion of chimeric contigs in the assembly. Most of these contigs with metaches to bacteria contained proteins which had matches exclusively to bacterial proteins and none to Eurotiomycetes proteins, and these were deemed non-target and excluded from the genome assembly. Contigs containing one protein with bacterial matches and one or (as was the most commonly occurring case) multiple proteins with matches to Eurotiomycetes were kept in the final assembly.

Genome size estimation was performed using the kmer counting software Jellyfish v 2.0 with a kmer size of 21 (Marcais & Kingford 2011) and the equation presented in Li *et al.* (2010). The Core Eukaryotic Mapping Genes Approach (CEGMA) was used to estimate the completeness of the *E. granulatus* genome (Parra *et al.* 2007, 2009). Whole scaffold alignments were performed in Mauve (Darling *et al.* 2010) with default progressiveMauve alignment settings.

Quality filtering of RNA sequences was performed based on Illumina quality flagging. Trimming of the first five and last 15 bp of RNA sequences was performed using the fastx toolkit (Gordon 2011), and RNA sequences were assembled in Trinity (Grabherr *et al.* 2011) using default settings. The Trinity RNA assembly and the final *E. granulatus* genome assembly were imported into the Maker annotation pipeline (Cantarel *et al.* 2008) for protein model prediction. Within the Maker pipeline, RepeatMasker v 3.2.8 was used to identify repetitive elements with “fungi” as the species database (Smit *et al.* 1996), and a custom repeat library was created and provided to Maker using

RepeatScout v 1.0.3 and other filtering scripts distributed with that package (Price *et al.* 2005). Other inputs into Maker included an hidden Markov model (HMM) created in Genemark-ES v 2.0 (Ter-Hovhannisyan *et al.* 2008) trained on the *E. granulatus* assembly, protein models from all Eurotiales genomes used in this study, and *ab initio* predictions from AUGUSTUS (Stanke *et al.* 2006).

Selection of taxa used for phylogenomic analysis was based on breadth of Pezizomycotina diversity and depth of diversity from publically available Eurotiomycetes. The phylogenomic pipeline Hal (Robbertse *et al.* 2011) was used to identify single copy orthologous clusters of proteins for use in phylogenetic analyses. Briefly, orthologous clusters of proteins were identified in MCL (Enright *et al.* 2002) across inflation parameters 1.2, 3 and 5. Orthologous clusters were filtered for retention of clusters with one sequence per genome and removal of any redundant clusters. The resulting unique, single-copy orthologous clusters of proteins were aligned in MUSCLE (Edgar 2004) with default settings; poorly aligned regions were identified using Gblocks (Talavera & Castresana 2007; gap removal setting = liberal) and excluded from subsequent analyses. Aligned and trimmed alignments were concatenated into a superalignment and maximum likelihood phylogenetic analysis was performed in RAxML v 7.2.6 using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates (Stamatakis 2006).

Hmmer3 package v 2.4 (Eddy 2011) was used to create a custom HMM based on adenylation domains from a wide variety of characterized nonribosomal peptide synthetases (NRPSs) (Bushley & Turgeon 2010) and then used to identify adenylation domains of NRPS-like genes and clusters. Manual curation of these genes and clusters

was completed using the NCBI database and pFAM domain annotation (Punta *et al.* 2012). Prediction of polyketide synthases (PKSs) and other types of secondary metabolite genes and clusters was performed using antiSMASH (Blin *et al.* 2013) and SMURF (Khaldi *et al.* 2010). Hmms from dbCAN were used to annotate CAZymes in the *E. granulatus* genome (Yin *et al.* 2012). Heatmap clustering in R was based on default Euclidean distances as a part of the heatmap.2 program.

Bowtie 2 v. 2.0.6 alignment (Langmead & Salzberg 2012) of raw reads to the *E. granulatus* specific contigs was used to identify those reads that did not align and were used for downstream analysis of the microbial community. These more than 131 million reads were uploaded onto the MG-RAST webserver version 3.3.9 (Meyer *et al.* 2008). Because the read length was not long enough to assemble the paired reads across the insert size, paired-end reads were uploaded as separate files to MG-RAST server (MG-RAST IDs 4526250.3, 4526252.3, 4526251.3, and 4526253.3). MG-RAST is a web-based tool for analyzing metagenomic sequence data which assigns taxonomic and functional ranks (where possible) to all sequences uploaded, regardless of whether the DNA sequenced was obtained from a particular portion of the genome (e.g. rDNA, mtDNA, *etc*.). All downstream analyses of microbial community were performed using data output from MG-RAST annotations of the data, and the M5NR non-redundant database was selected. Principal coordinate analyses (PCoA) were conducted using Bray-Curtis dissimilarity distances within MG-RAST. Rarefaction of enzyme commission (EC) annotations was conducted in the Vegan program in R (Dixon 2003).

Results

Genome structure and content

The initial assembly of all ~321 million reads from the fruiting body resulted in 61,028 contigs with a total of 286,657,254 base pairs. Many contigs had an average coverage of approximately 288x. These represented the core contigs of the *E. granulatus* genome (Figure 6.1a), whereas the vast majority of fruiting body-associated bacterial contigs had coverage of less than 100x. Contigs with coverage >288x were also part of the *E. granulatus* genome (with top BLAST hits to Eurotiomycetes sequences), and represent repetitive sequences (see below).

After the conservative removal of non-target contigs, the final version of the *E. granulatus* draft genome is composed of 8,675 contigs encompassing 54.2 Mb with an N50 of 11,830 bp (Table 6.1). There are 460 contigs with an average coverage of over 3000x, and 102 of these high coverage contigs were annotated as simple repetitive sequences or transposable elements (Figure 6.1b), in addition to the *E. granulatus* rDNA and mitochondrial DNA. The *E. granulatus* rDNA had a median coverage of 4626x, and when compared to the 279x coverage of single copy genes (*e.g.*, TEF and RPB1) resulted in an estimated 16 rDNA tandem repeat copy number. The contig with the single highest coverage (15,432x) is only 1,350 bp, but is annotated by RepeatMasker as a Gypsy-like long terminal repeat (LTR). Using the custom repeat library generated in RepeatScout, RepeatMasker estimates that 13.16% of the assembled 54 Mb is repetitive, and RepeatMasker also identified 496 Gypsy-like LTRs totaling 245,503 bp. Due to the high coverage of many repetitive sequences and transposable elements (Figure 6.1), a kmer sampling method of raw reads was used to more accurately measure the genome size,

which was estimated to be approximately 71 Mb. If approximately 7 Mb (13%) of the assembled 54 Mb is repetitive and the difference between assembly size and estimated genome size (17 MB) is primarily due to repetitive sequence, then repetitive sequence (24 MB) accounts for approximately 34% of the *E. granulatus* genome.

The genome of *E. granulatus* is slightly AT-rich, with an estimated GC content of 45.9% (Table 6.1). This lower GC content is driven by the non-coding regions of the genome, which is approximately 44.7% GC, whereas the coding regions have a GC content of 50%. Almost the entire mitochondrial genome lies on a single contig of 51,855 bp, and it shares a high degree of synteny with the mitochondrial assembly found of *A. fumigatus* (Figure 6.S2). For the nuclear genome, there are 7,184 protein models supported by RNA evidence, and another 1,293 *ab initio* predicted proteins that had significant (e-value $\leq 1e^{-5}$) hits to one or more protein models in the Eurotiales genomes sampled in this study. CEGMA analysis of the *E. granulatus* protein models identified 220 full length alignments (88.71%) and 226 full or partial length alignments (91.13%) out of the 248 core eukarotic genes, indicating that despite the metagenomic nature of the data, the assembly sufficiently captures a high proportion of the *E. granulatus* gene space.

Phylogenetic placement of *Elaphomyces*

Using the conservative set of *E. granulatus* protein models with RNA evidence, a total of 320 single copy orthologous clusters of proteins with 100% representation of all 31 taxa analyzed were identified. This resulted in a final super alignment of 123,356 amino acid positions. The maximum likelihood analysis provided support for all of the

nodes in Pezizomycotina (MLBP=100), except the node joining classes Pezizomycetes and Orbiliomyctes (MLBP=59) (Figure 6.2). *Elaphomyces granulatus* is reconstructed as sister to the two *Talaromyces* spp. sampled, and *M. ruber* is placed as sister to *Aspergillus* and *Penicillium* spp. The root of Eurotiales lies between the *Elaphomyces*/*Talaromyces* clade and the remaining taxa of Eurotiales.

In this analysis Eurotiales and Onygenales are reconstructed as sister clades. The single representative of the Coryneliales, *Caliciopsis orientalis*, is sister to both of these orders, and is resolved as the earliest diverging sampled taxon within Eurotiomycetidae. The *C. orientalis* genome was sequenced as part of the Assembly the Fungal Tree of Life 2 (AFTOL 2) project. Sequencing was performed for a 310 bp insert library on a single lane of Illumina using version 2 chemistry. This plant pathogen has an estimated genome size of 28.2 Mb assembled onto 643 scaffolds (Table 6.S2).

Comparative genomics

In contrast to sequenced Eurotiales, *E. granulatus* is reduced overall in CAZymes (Figure 6.3). Of particular interst, *E. granulatus* possesses no cellulases acting on crystalline cellulose in CAZy families GH6 and GH7, which were also absent in the *T. melanosporum* genome (Martin *et al.* 2010). This absence of GH6 and GH7 proteins is in contrast to *Talaromyces* spp. and *P. confluens*, the closest sequenced relatives of *E. granulatus* and *T. melanosporum* respectively, which both contain copies of GH6 and GH7 cellulases (Table 6.S3). Analysis of CAZymes in Eurotiomycetes and Pezizomycetes resulted in *E. granulatus* and *T. melanosporum* clustering with the animal associated taxa of Onygenales due to the reduction of CAZymes in these taxa (Figure

6.3). *Talaromyces* spp. cluster with *Aspergillus* and most *Pencillium* spp., while *P. confluens* is grouped with taxa containing an intermediate amount of CAZymes. *Elaphomyces granulatus* is expanded in only four CAZy families compared to the other Eurotiomyces (Figure 6.3), including having five copies of the endoglucanse/xyloglucanase family GH74 (the average Eurotiomycete has one and the range is one to three), and an extra copy of CBM6 which is responsible for binding either cellulose or xylose. Family GH114 is also expanded in *E. granulatus*, which has 4 protein models in this family of polygalactosaminidases. Another CAZyme expansion in *E. granulatus* is the presence of two GH37 proteins, trehalases, compared to the class wide average of one for Eurotiomycetes and zero in *T. melanosporum*.

The 50,097 orthologous clusters produced by the HAL pipeline at liberal MCL inflation parameter of 1.2 were analyzed for *E. granulatus* lineage specific expansions. A large group of *Elaphomyces*-expanded protein clusters were identified (Table 6.S4), including proteins involved in regulating sulfur metabolism, ATP-dependent DNA helicases, PIF1-like helicases, mitochondrial helicases, sugar transport, nitrogen metabolism, and radical SAM domain containing proteins; however, the majority of clusters that are expanded in copy number have unknown functions and no putative functional domains including signal peptides or transmembrane helices. By far, the most expanded clusters were helicases (Table 6.2). The largest protein family of helicases in *E. granulatus* contains 31 protein models (cluster 1200611), compared to a subphylum wide average of 3 and a range of zero to 24 (*Talaromyces stipitatus* has 24 helicases in this cluster), and all have a best BLAST matches to fungal helicases. Two of the proteins in this cluster (Egran_06015 and Egran_06016) are located on the same contig and within

2000 bp of one of the Gypsy LTRs. Another helicase cluster (120055) contains 24 *E. granulatus* protein models, while the average number per species is seven and the range is three to 17. Many of these helicases are also tandemly located within the genome. *Elaphomyces granulatus* has five protein models in the SkpA sulfur metabolism regulator cluster (12001782), which has an average representation of one per species and a range of zero to two.

Within the *E. granulatus* draft genome, three NRPSs, 18 PKS or PKS-like genes, and three Hybrid NRPS-PKS gene clusters were identified. Of the NRPS core genes, one (Egran_00898) is a homolog of α-aminoadipate reductase involved in lysine biosynthesis and one (Egran_00889) is likely an intracellular siderophore synthetase, which produces iron chelaters and is present in most Ascomycota. Based on the phylogeny of adenylation domains (A-domains) (Figure 6.S3), the third NRPS gene (Egran_01550) appears to be the product of lineage specific duplications in *E. granulatus* as the five A-domains are all most closely related to each other. The next closest related A-domains are from an NRPS in *Pe. chrysogenum* (Pc13g14330) and *pes1* in *A. fumigatus* (AFUB_078070) which is responsible for production of the ergoline alkaloid, fumigaclavine C (Reeves *et al.* 2006, O'Hanlon *et al.* 2012).

E. granulatus peridium metagenomic community analysis

While approximately 60% (190 million reads) of the sequenced DNA was from the host, *E. granulatus*, 131 million Illumina sequences represented the metagenomic community of the peridium. This was a more than sufficient depth to characterize taxonomic diversity (Figure 6.S4). Twenty-four percent of the unfiltered, untrimmed raw

reads did not pass the MG-RAST quality control steps. (Read 1 (forward mate-pair) of the second library [4526251.3] had a higher percentage of reads that received a classification of Unknown Protein [53%] than the other reads, which had an average of 31.6% Unknown Protein.) Of the remaining reads, 88% were characterized as bacterial and 11% as eukaryotic, some of which could be residual host (*E. granulatus*) sequences (Figure 6.S5). The remaining 1% of annotated reads was classified as either Virus or Archaea.

At the level of class, the peridium community is dominated by Alphaproteobacteria accounting for over 33% of the reads sequenced (Figure 6.4). Rhizobiales accounts for the majority of this class, but other major orders of Alphaproteobacteria represented in the peridium sample include Rhodobacteriales and Rhodospirillales. Other classes of Proteobacteria were dominant portions of the peridium community as well, including Beta- (8.8%), Gamma- (7.9%), and Deltaproteobacteria (3.4%). Major non-Proteobacteria classes included Actinobacteria (11.8%) and Sphingobacteria (2.9%).

Several genera were found to be common members of the peridium community (Figure 6.5). *Bradyrhizobium*, for example, was the genus most commonly attributed to the sequenced reads. The third most common genus identified, *Catenulispora*, is a genus of acidophilic members of Actinomycetes that have only been isolated from forest soils (Busti *et al.* 2006, Copeland *et al.* 2009). The other most common genera were *Rhodopseudomonas*, *Burkholderia*, and *Streptomyces*, all of which are dominant taxa in soil communities. A large percentage of the community was identified to the candidate genera of Candidatus *Koribacter* and Candidatus *Solibacter* (both are isolates from soil

and members of the phylum Acidobacteria) (Ward *et al.* 2009). Although none of the contigs >100,000 bp assembled from the metagenomic data have an exact match to previously sequenced bacteria, the longest contig (>890 Kb) assembled from the raw data had a best hit to Candidatus *Koribacter versatilis* Ellin345 (Figure 6.5). On this contig, antiSMASH predicted two secondary metabolite clusters; one terpene cluster that matches a cluster in Ca. *Koribacter versatilis* and one hybrid NRPS-PKS cluster that is similar to a cluster in *Burkholderia phymatum*, a species described from legume root nodules (Moulin *et al.* 2001, 2014). This contig also contains proteins involved in the Type IV secretion system. Other bacterial contigs >100,000 bp (59 contigs) contain other secondary metabolite clusters including 2 PKSs, a large NRPS-lantipeptide, two bacteriocins, one phosphonate, another terpene, and an ectoine, which are used as osmolytes that confer resistance to salt and temperature stress. The percentage of per-taxon abundance (within the community) did not correlate with contig length, however (Figure 6.5). The longest *Bradyrhizobium* contig assembled, for example, was less than 200 kb, despite accounting for more than 10% of all reads.

In order to determine how the *E. granulatus* microbial community might be related to or differ from other communities, publically-available similarly-sequenced datasets from the Human microbiome project (HMP) and forest soil were chosen for comparison (MG-RAST ID 4508942.3, 002007_ATTCCT_filtered_merged.fastq Leafy wood soil, and 4508937.3, 002002_Airport_GTGGCC_filtered_merged.fastq agricultural soil). A PCoA of the taxonomic composition of these communities revealed that *E. granulatus* was similar to but distinct from soil communities (Figure 6.6a). While PC1 groups soil and *E. granulatus* peridium communities together accounting for 58% of the

variation, there is also significant difference between these communities, as seen along PC2 which separates soil and *E. granulatus* peridium and explains over 31% of the variation between these samples.

Functional annotation of the peridium microbiome

Due to the shotgun sequencing approach taken (as opposed to targeted sequencing of ribosomal markers), there exists the opportunity to characterize the physiology of the community via functional annotation of the bacterial component of the metagenome. In comparison with publically available HMP and soil communities, *E. granulatus* is roughly equally distant from both forest soil and HMP samples interms of its functional composition according to PC1 accounts for nearly 70% of the variation (Figure 6.6b). PC2 accounts for ~24% of the variation, and separates *E. granulatus* from both of the other datasets. While nearly 1.9 million reads from the *E. granulatus* metagenome received KEGG annotations, these reads were only distributed across 368 Enzyme Commission numbers (ECs). By comparison, the representative shotgun sequenced soil microbiome (4508937.3) was found to have 5.13 million KEGG-annotated sequences distributed across 1920 ECs. Rarefaction analysis of the *E. granulatus* metagenome indicates that this reduced functional diversity is not a product of insufficient sequence depth (Figure 6.S6). In particular, ECs from pathways involved in glycan biosynthesis and metabolism, secondary metabolism, lipid metabolism, and metabolism of cofactors and vitamins are reduced or not detected at all (Figure 6.7). Pathways that are present in the peridium community include nucleotide biosynthesis, the citric acid cycle, amino acid metabolism, fatty acid biosynthesis, and oxidative phosphorylation.

Discussion

Fungal genome sequencing from complex sporocarp tissues

Every published fungal genome sequenced to date was generated from tissue growing in axenic culture. While this is the most straightforward approach to generating genome sequences, it has created a significant bias against understanding genome content and organization in uncultured organisms. Many fungi cannot be readily cultured either due to obligate biotrophy or other nutritional or abiotic conditions not met by standard culture conditions. In bacteria, the analysis of metagenomes has resulted in the recovery of assembled genomes in uncultured organisms (Wrighton *et al.* 2012). This study demonstrates that analysis of a metagenome generated from *E. granulatus* sporocarp tissue reveals both the genomic content of this uncultured fungus as well as the structure and physiology of the microbiome that associates with it. This approach to the study of uncultured fungal genomes will help address systematic sampling biases in phylogenomic analyses and inform biological inquiry into the nature of obligate symbioses and the complex ecological relationships between fungi and other microorganisms. In addition, generating draft fungal genomes through metagenomic analysis may yield improvements in culturing fungi as these draft genomes can provide information that can be used to design media that are better suited to culture organisms of interest. For example, the genome sequence of one species of bacteria has also facilitated the development of culture-media that is suited to pure isolation of that species (Renestro *et al.* 2003).

Genome expansion and repetitive element proliferation

E. granulatus has a highly repetitive (13-34%) and large genome (54 Mb to 71 Mb) compared to the closest sequenced relatives and any other sequenced species in Eurotiales, which range in approximate size from 24 to 37 Mb (Grigoriev *et al.* 2014). This pattern of a relatively expanded genome size is similar to what is seen in the ectomycorrhizal species, *T. melanosporum* (Martin *et al.* 2010). *Tuber melanosporum*, like *E. granulatus*, also contains a large percentage of transposable elements, the most common of which were Gypsy-like LTRs. While there are a few Eurotiomycetes with large Gypsy content including *Paracoccidioides lutzii* (Dejardins *et al.* 2011) and *Pe. chrysogenum* (Braumann *et al.* 2007), it is not to the degree seen in both *E. granulatus* and *T. melanosporum*. Since these two species are the only ectomycorrhizal ascomycete genomes sequenced to date, it is difficult to know if this is a consistent pattern for this ecology in Ascomycota. In Agaricomycotina there is a large variation in genome size among ectomycorrhizal species (*e.g.* 65 Mb in *L. biocolor*, 175 Mb in *Tricholoma matsutake* [MycoCosm, Grigoriev *et al.* 2014]). A key difference, though, is that there are fewer Ascomycota that are ectomycorrhizal and there have been fewer transitions to the mycorrhizal symbiosis than in Agaricomycotina (Hibbett & Mattheny 2009, Tedersoo *et al.* 2010).

The low GC content (45.9%) of *E. granulatus* is a character shared with *Talaromyces* spp. (Table 6.1), and differs from the nucleotide content observed in *Aspergillus*, *Penicillium*, and also *Tuber* (Pel *et al.* 2007, Fedorova *et al.* 2008, van den Berg *et al.* 2008, Martin *et al.* 2010). When these observations are placed in a phylogenetic context, the lower GC content might be best explained as a phylogenetic

character state inherited from their last common ancestor. However, the coding regions of *E. granulatus* have a higher GC content (50%) than non-coding (44.7%), suggesting that the non-coding regions are AT-rich and driving the overall low genome-wide GC content. Based on this observation, genome expansions resulting from independent invasions of AT-rich repetitive elements would result in parallel divergences of genome nucleotide content. Further sampling of other taxa within the *Talaromyces* s. str. and *Elaphomyces* lineage (*e.g. Trichocoma*, *Rasamsonia*, *Thermomyces*) is necessary to differentiate between the lower GC content in the *Elaphomyces*-*Talaromyces* lineage versus a pattern of convergent evolution stemming from increased repetitive DNA content of the *E. granulatus* genome.

Phylogenetic relationships

This is the first study to find definitive statistical support for the sister relationship between *Talaromyces* s. str. and *Elaphomyces* (Figure 6.2) and informs a number of evolutionary hypotheses for the order. Many *Talaromyces* produce a penicillium-like asexual state, and thus, it appears that the absence of asexual reproduction in Elaphomycetaceae represents a derived (or unobserved) character state, and not the ancestral state for Eurotiales. In addition, these genome scale data provide support for some of the intra-ordinal classifications advanced by Houbraken & Samson (2011). They describe and emend the large, paraphyletic family Trichocomaceae, which they restrict to *Talaromyces* and other taxa that are closely related to the type genus, *Trichocoma* (for which no genomic sequence is currently available). Traditionally, Trichocomaceae included all the eurotialean taxa included in this study's analysis except *Monascus ruber*

and *E. granulatus*, members of the Monascaceae and Elaphomycetaceae, respectively. Aspergillaceae sensu Houbraken & Samson (2011) includes *Aspergillus* s. str. and *Penicillium* s. str. along with *Monascus* and other related genera. Houbraken & Samson (2011) did not include any data from Elaphomycetaceae in their study, but the few studies (with very limited taxon and molecular sampling) to include data from *Trichocoma* and Elaphomycetaceae (Geiser *et al.* 2006, Tedersoo *et al.* 2010, Morgenstern *et al.* 2012), show a close relationship between these genera.

As anticipated, the sister relationship between Eurotiales and Onygenales (Geiser et al 2006, Sharpton *et al.* 2009, Desjardins *et al.* 2011) is reconfirmed in the phylogenomic analysis presented here. Within Onygenales these results are congruent with findings from other genome scale studies, specifically, the sister relationship between Onygenaceae and Ajellomyceteaceae (Desjardins *et al.* 2011). This is the first study to analyze genomic data for the plant pathogen, *C. orientalis*. Supporting what was seen in multigene phylogenies (Geiser *et al.* 2006), the concatenated genome scale analysis resolves the Coryneliales as the earliest diverging lineage of the Eurotiomycetidae, one of two subclasses of Eurotiomycetes.

Comparative genomics

The significant reduction in CAZymes in *E. granulatus* as compared to closely related taxa is similar to what was observed in *T. melanosporum*, as well as the ectomycorrhizal basidiomycetes *Laccaria bicolor* and *Amanita muscaria* (Martin *et al.* 2008, 2010, Wolfe *et al.* 2012). Many of the CAZymes involved in degradation of plant cell wall materials including cellulases, hemicellulases, and xylases are reduced or absent

in the *E. granulatus* genome (Table 6.S3). The reduction in these genes is proposed to facilitate symbiosis with plant hosts, by not eliciting a plant immune response. This reduction may also be due to the abundance of simple carbohydrates from the host plant due to the intimate contact between plant and mycorrhizae. Reduction of these same plant-tissue-degrading CAZymes is also seen in taxa of Onygenales (Figure 6.3), as these species are animal associated fungi. *Caliciopsis orientalis*, the only plant pathogen sampled within Eurotiomycetes, is quantitatively similar in plant degrading CAZymes to the saprobic taxa of Eurotiales (*e.g. Aspergillus*, *Penicillium*). The expansion in trehalases of CAZy family GH37 is perplexing, as most plants do not produce trehalose, but it is commonly produced by most fungi. Trehalose is a disaccharide of two glucose units and functions in prevention of desiccation, as a signaling compound, and as a source of energy (Elbein *et al.* 2003). Fungi absorb monosaccharides like glucose, the primary form of carbon transferred from plants to mycorrhizal fungi, and convert them to trehalose, which has decreased permeability across the cell membrane. This expansion of trehalases suggests ectomycorrhizal *Elaphomyces* may use trehalose to a greater physiological degree than closely saprobic relatives.

The largest lineage specific expansions in the *E. granulatus* genome are numerous helicase-type protein families (Table 6.2). Helicases are responsible for separating the strands of nucleic acids (DNA and RNA) during DNA replication, transcription of RNA, and many other processes (Patel & Donmez 2006). It is possible that these are involved in the proliferation of transposable elements and the expansion of the *E. granulatus* genome, although only one of these helicases is located next to one of the many Gypsy LTRs in the genome.

A large number of secondary metabolites have been described from species of *Aspergillus*, *Penicillium*, and related taxa using both genome-mining and traditional chemical isolation approaches (Keller *et al.* 2005, Nierman *et al.* 2005, Khaldi *et al.* 2010, Inglis *et al.* 2013). In contrast, the genome of *E. granulatus* is depauperate in, though not without, secondary metabolic core genes compared to other eurotialean taxa. There are several predicted PKS clusters in the *E. granulatus* genome, but relatively few NRPS cluster compared to other Eurotiales (Khadi *et al.* 2010). Because NRPSs are large and modular, this may be a result of short read sequence technologies, however, this phenomenon was not observed in other genome sequencing projects (*e.g.* Bushley *et al.* 2013, Quandt *et al.* Ch. 3, Ch. 5). Limited presence of secondary metabolism was seen in the genome of *T. melanosporum* (Martin *et al.* 2010), in the Onygenales (Khaldi *et al.* 2010), and in the obligate biotrophic plant-pathogen, *Blumeria graminis* (Spanu *et al.* 2010). One known function of secondary metabolites is as pathogenicity factors (Fox & Howlett 2008), and as *E. granulatus* is a mycorrhizal species, it may not require a large repertoire of secondary metabolites as that of the other, saprobic eurotialean taxa.

The *E. granulatus* peridium microbiome

Proteobacteria dominated the sequences of the *E. granulatus* microbiome, especially those of the class Alphaproteobacteria which accounted for more than 33% of the sequences annotated (Figure 6.4). Other abundant non-proteobacteria classes included Actinobacteria and Sphingobacteria. The largest order of sequences present was Rhizobiales, with the greatest proportion of these belonging to Bradyrhizobiaceae.

The dominance of *Bradyrhizobium* in the *E. granulatus* peridium (Figure 6.5) mirrors results from the other study using high-throughput data examining *Tuber* fruiting bodies (Antony-Babu *et al.* 2013). In contrast, *Catenulispora* has not been found as a common component of the communities associated with other fungal fruiting bodies. Since adjacent soil cores were not sampled in addition to the truffle, it is unknown if *Catenulispora* is also present in similar abundance in the surrounding soil, or if this species is more abundant within the *E. granulatus* peridium. *Rhodopseudomonas*, *Burkholderia*, and *Streptomyces* were other dominant taxa in the community. *Streptomyces* and Burkholderiales have been identified as EHB in numerous systems (Frey-Klett & Garbaye 2005). These findings suggest that these bacteria may be selected for in the intra-fruit body niche. Acidobacteria, Candidatus *Koribacter versatilis* and Candidatus *Solibacter usitatus* were other abundant members or the community, and several of the longest bacterial contigs assembled had hits to these taxa (Figure 6.5). These are candidate species that are difficult to grow in culture, but due to their common presence in soil have recently been targeted for genome sequencing (Ward *et al.* 2009).

The principal coordinates analysis suggests that *E. granulatus* peridium community is similar to but distinct from soil communities (Figure 6.6a). *Bradyrhizobium* and *Rhodopseudomonas*, for instance, are common in soil communities, but constitute a large percentage of the peridium community. The large percentages of the community derived from Acidobacteria (*e.g.* Ca. *Koribacter*, Ca. *Solibacter*, and *Acidobacterium*), Sphingobacteria (*e.g. Chitinophaga*), and *Catenulispora* are likely driving the taxonomic differentiation along PC2 (Figure 6.6a). Although they require further confirmation and sampling of co-occuring *E. granulatus* truffles and soil samples,

these findings are consistent with the *E. granulatus* microbiome being a taxonomically reduced community that is the product of environmental filtering of the neighboring soil community.

The abundance of reads annotated as a particular taxon did not correlate with the ability to assemble genomes as measured by longest contig (Figure 6.5). This was unexpected and three factors could be contributing to this. One, the community could contain several related taxa (*e.g.*, many species from a genus) that are diverged to an extent that assembly into longer contigs is compromised. Two, large genome size and therefore low coverage could be inhibiting assembly of long contigs. This could explain the relatively poor assembly of the most abundant genus, *Bradyrhizobium*, species of which can have relatively large genome sizes (>9 Mb) (Boussau *et al.* 2004). Three, some of the genomes could be repetitive, limiting the ability to assemble long contigs.

Comparison of the functional diversity based on KEGG annotations, placed the *E. granulatus* community as intermediate between the HMP and soil communities (Figure 6.6), but at least some of the variation (24%) separated *E. granulatus* from both of the communities. This observation suggests that while the *E. granulatus* microbiome is taxonomically most similar to soil bacterial communities, it may be functionally diverged from soil bacterial communities through reduction in metabolic potential. Very few EC numbers were annotated from the *E. granulatus* peridium sequences (Figure 6.7), and rarefaction analyses of these annotations suggested this was not due to under sampling (Figure 6.S6). This reduction is indicative of a microbiome that may rely on its *Elaphomyces* host for some metabolic functions (*e.g.* nutrition) or that is specialized in what metabolic functions it can provide to its host fungus.

Conclusions

Given the vast number of fungi that are difficult if not impossible to grow in culture, it is important to develop culture-independent techniques for genome – and therefore metagenome – sequencing and analysis. The ectomycorrhizal truffle *E. granulatus* was successfully sequenced from sporocarp tissue resulting in a robust assessment of the majority of gene content. Bioinformatic analyses support that there are many more similarities in genome size, structure, and gene content between *E. granulatus* and *T. melanosporum* than between either of these and non-mycorrhizal species. When compared to their closest phylogenetic relatives, both genomes are characterized by a reduction in overall gene content, especially those associated with secondary metabolism and CAZymes, and both genomes are characterized by an increase in overall genome size due to the presence of a high number of Gypsy-like LTRs. These findings suggest a similar pattern of evolution between these independent lineages (Eurotiales and Pezizales) of ectomycorrhizal symbionts of Ascomycota. A diverse but distinct community of bacteria, which is dominated by Proteobacteria, exists within the peridium of *E. granulatus*. Bacterial sequences of the metagenome had best hits to genera of Rhizobiales, Actinomycetales, and Burkholderiales, but the number of sequences per taxon did not correlate with the longest assembled contig. The ability to assemble longer bacterial contigs was interpreted as some genera of the community being represented by fewer taxa, smaller genomes, or both. Functional annotation of this community indicates it may be reduced in function, but further work is needed to understand how this community may contribute to the complex symbiosis between *E.*

granulatus and its plant hosts, and if any species have evolved as endohyphal symbionts. This study provides a framework on which to further examine the evolution, symbiosis, and microbiome of one of the most broadly distributed and evolutionarily unique lineages of ectomycorrhizal fungi.

Acknowledgements

The authors acknowledge the contributions made by following: Mark Desanko, Alija Mujic, and the North American Truffling Society. Funding was provided from the following sources: NSF grants DEB-0732993 to JWS, and the Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ANR-11-LABX-0002-01). A Laboratory of Excellence Visiting Scientist Award and an NSF Graduate Research Fellowship supported CAQ.

References

Antony-Babu S, Deveau A, Van Nostrand JD, Zhou J, Tacon FL, Robin, C, Frey-Klett P, Uroz S (2013) Black truffle-associated bacterial communities during the development and maturation of *Tuber melanosporum* ascocarps and putative functional roles. *Environmental Microbiology* DOI: 10.1111/1462-2920.12294.

Barbieri E, Bertini L, Rossi I, Ceccaroli P, Saltarelli R, Guidi C, Zambonelli A, Stocchi V (2005) New evidence for bacterial diversity in the ascoma of the ectomycorrhizal fungus *Tuber borchii* Vittad. *FEMS Microbiology Letters* **247**: 23–35.

Bianciotto V, Lumini E, Bonfante P, Vandamme P (2003) Candidatus '*Glomeribacter gigasporarum*' gen. nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. *International Journal of Systematic and Evolutionary Microbiology* **53**: 121–124.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* **41**: W204-W212.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE (2004) Computational inference of scenarios for α-proteobacterial genome evolution. *Proceedings of the National Academy of Sciences* **101**: 9722-9727.

Braumann I, van den Berg M, Kempken F (2007) Transposons in biotechnologically relevant strains of *Aspergillus niger* and *Penicillium chrysogenum*. *Fungal Genetics and Biology* **44**: 1399-1414.

Bushley KE, and BG Turgeon (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology* **10**: 26.

Busti E, Cavaletti L, Monciardini P, Schumann P, Rohde M, Sosi M, Donadio S (2006) *Catenulispora acidiphila* gen. nov., sp. nov., a novel, mycelium-forming actinomycete, and proposal of Catenulisporaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology* **56**: 1741-1746.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* **37**: D233-238.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188-196.

Castellano MA, Trappe JM, Maser Z, Maser C (1989) *Key to spores of the genera of hypogeous fungi of North Temperate forests with special reference to animal mycophagy*. Arcata, Calif: Mad River Press.

Castellano MA, Trappe JM, Vernes K (2011) Australian species of *Elaphomyces* (Elaphomycetaceae, Eurotiales, Ascomycota). *Australian Systematic Botany* **24**: 32-57.

Castellano MA, Beever RE, Trappe JM (2012) Sequestrate fungi of New Zealand: *Elaphomyces* and *Rupticutis* gen. nov. (Ascomycota, Eurotiales, Elaphomycetaceae). *New Zealand Journal of Botany* **50**: 423-433.

Copeland A, Lapidus A, Del Rio TG, Nolan M, Lucas S, Chen F, Tice H, Cheng J-F, Bruce D, Goodwin L, Pitluck S, Mikhailova N, Pati A, Ivanova N, Mavromatis K, Chen A, Palaniappan K, Chain P, Land M, Hauser L, Chang Y-J, Jeffries CD, Chertkov O, Brettin T, Detter JC, Han C, Ali Z, Tindall BJ, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H-P (2009) Complete genome sequence of *Catenulispora acidiphila* type strain (ID 139908). *Standards in Genomic Sciences* **1**: 119-125.

Crawford DL, Lynch JM, Whipps JM, Ousley MA (1993) Isolation and characterization of Actinomycete antagonists of a fungal root pathogen. *Applied Environmental Microbiology* **59**: 3899–3905.

Desirò A, Salvioli A, Ngonkeu EL, Mondo SJ, Epis S, Faccio A, Kaech A, Pawlowska TE, Bonfante P (2014) Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. *The ISME Journal* **8**: 257–270.

Desjardins CA, Champion MD, Holder JW, Muszewska A, Goldberg J, Bailão AM, Brigido MM, da Silva Ferreira ME, Garcia AM, Grynberg M, Gujja S, Heiman DI, Henn MR, Kodira CD, León-Narváez H, Longo LVG, Ma L-J, Malavazi I, Matsuo AL, Morais FV, Pereira M, Rodríguez-Brito S, Sakthikumar S, Salem-

Izacc SM, Sykes SM, Teixeira MM, Vallejo MC, Walter MEMT, Yandava C, Young S, Zeng Q, Zucker J, Felipe MS, Goldman GH, Haas BJ, McEwan JG, Nino-Vega G, Puccia R, San-Blas G, Soares CMdA, Birren BW, Cuomo CA (2011) Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. *PLoS Genetics* **10**: e1002345.

Deveau A, Palin B, Delaruelle C, Peter M, Kohler A, Pierrat JC, Sarniguet A, Garbaye J, Martin F, Frey-Klett P (2007) The Mycorrhiza Helper *Pseudomonas Fluorescens* BBc6R8 has a specific priming effect on the growth, morphology and gene expression of the ectomycorrhizal fungus *Laccaria bicolor* S238N. *The New Phytologist* **175**: 743–55.

Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**: 927-930.

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology* **7**: e1002195.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. *Glycobiology* **13**: 17R-27R.

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575-1584.

Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, Angiuoli S, Bussey H, Bowyer P, Cotty PJ, Dyer PS, Egan A, Galens K, Fraser-Liggett CM, Haas BJ, Inman JM, Kent R, Lemieux S, Malavazi I, Orvis J, Roemer T, Ronning CM, Sundaram JP, Sutton G, Turner G, Venter JC, White OR, Whitty BR, Youngman P, Wolfe KH, Goldman GH, Wortman JR, Jiang B, Denning DW, Nierman WC (2008) Genomic Islands in the Pathogenic Filamentous Fungus *Aspergillus fumigatus*. *PLoS Genetics* **4**: e1000046.

Fox EM, Howlett BJ (2008) Secondary metabolism: regulation and role in fungal biology. *Current Opinion in Microbiology* **11**: 481-487.

Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A. 2011. Bacterial-Fungal Interactions: Hyphens between Agricultural, Clinical, Environmental, and Food Microbiologists. *Microbiology and Molecular Biology Reviews* **75**: 583–609.

Frey-Klett P, Garbaye J (2005) Mycorrhiza helper bacteria: a promising model for the genomic analysis of fungal–bacterial interactions. *New Phytologist* **168**: 4-8.

Frey-Klett P, Garbaye J, Tarkka M (2007) The mycorrhiza helper bacteria revisited. *New Phytologist* **176**: 22–36.

Geiser DM, Gueidan C, Miadlikowska J, Lutzoni F, Kauff F, Hofstetter V, Fraker E, Schoch CL, Tibell L, Untereiner WA, Aptroot A (2006) Eurotiomycetes: Eurotiomycetidae and Chaetothyriomycetidae. *Mycologia* **98**: 1053-1064.

Ghignone S, Salvioli A, Anca I, Lumini E, Ortu G, Petiti L, Cruveiller S, Bianciotto V, Piffanelli P, Lanfranco L, Bonfante P (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *The ISME Journal* **6**: 136–145.

Gordon A (2011) "FASTX-Toolkit" FASTQ/A short-reads pre-processing tools. Available at http://hannonlab.cshl.edu/fastx_toolkit/

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644-652.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699-704.

Gryndler M, Soukupová L, Hršelová H, Gryndlerová H, Borovička J, Streiblová E, Jansa J (2013) A quest for indigenous truffle helper prokaryotes. *Environmental Microbiology Reports* **5**: 346-352.

Hawker LE, Fraymouth J, de la Torre M. 1967. The identity of *Elaphomyces granulatus*. *Transactions of the British Mycological Society* **50**: 129-136.

Hayat R, Ali S, Amara U, Khalid R, Ahmed I (2010) Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology* **60**: 579–598.

Hibbett DS, Matheny PB (2009) The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC biology* **7**: 13.

Hibbett DS, Thorn RG (2001) Basidiomycota: homobasidiomycetes. In *Systematics and Evolution* pp. 121-168. Berlin Heidelberg: Springer.

Hoffman MT, Gunatilaka MK, Wijeratne K, Gunatilaka L, Arnold AE (2013) Endohyphal bacterium enhances production of indole-3-acetic acid by a foliar fungal endophyte. *PLoS ONE* 8: e73132.

Hosaka K, Kasuya T, Reynolds HT, Sung GH (2010) A new record of *Elaphomyces guangdongensis* (Elaphomycetaceae, Eurotiales, Fungi) from Taiwan. *Bulletin of the National Museum of Natural Science* Ser B **36**: 107-115.

Houbraken J, Samson RA (2011) Phylogeny of *Penicillium* and the segregation of Trichocomaceae into three families. *Studies in Mycology* **70**: 1-51.

Inderbitzin P, Lim SR, Volkmann-Kohlmeyer B, Kohlmeyer J, Berbee ML (2004) The phylogenetic position of *Spathulospora* based on DNA sequences from dried herbarium material. *Mycological Research* **108**: 737-748.

Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology* **3**: 937-947.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736-741.

Kirk PM, Cannon PF, Minter DW, Stalpers JA (Eds.) (2008) *The Dictionary of the Fungi*, 10th Edition. Wallingford, UK: CAB International.

Korf RP (1973) Discomycetes and tuberales. *The fungi*: 249-319.

Kurth F, Zeitler K, Feldhahn L, Neu TR, Weber T, Krištůfek V, Wubet T, Herrmann S, Buscot F, Tarkka MT (2013) Detection and quantification of a mycorrhization

helper bacterium and a mycorrhizal fungus in plant-soil microcosms at different levels of complexity. *BMC Microbiology* **13**: 205.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Landvik S, Shaller NFJ, Eriksson OE (1996) SSU rDNA sequence support for a close relationship between the Elaphomycetales and the Eurotiales and Onygenales. *Mycoscience* **37**: 237-241.

Lehr NA, Schrey SD, Bauer R, Hampp R, Tarkka MT (2007) Suppression of plant defense response by a mycorrhiza helper bacterium. *New Phytologist* **174**: 892–903.

Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, *et al.* 2010. The sequence and de novo assembly of the giant panda genome. *Nature* **463**: 311-317.

LoBuglio KF, Berbee ML, Taylor JW (1996) Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the Ascomycetes. *Molecular Phylogenetics and Evolution* **6**: 287-294.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764-770.

Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buee M, Brokstein P, Canback B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbe J, Lin YC, Legue V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculity-Hirzel H, Oudot-Le Secq MP, Quesneville H, Rajaskekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kues U, Lucas S, Van de Peer Y, Podila GK, Polles A, Pukkila PJ, Richardson PM, Rouze P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88-93.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, Porcel B, Rubini A, Amicucci A, Amselem J, Anthouard V, Arcioni S, Artiguenave F, Aury J-M, Ballario P, Bolchi A, Brenna A, Brun A, Buée M, Cantarel B, Chevalier G, Couloux A, Da Silva C, Denoeud F, Duplessis S, Ghignone S, Hilselberger B, Iotti M, Marçais B, Mello A, Miranda M, Pacioni G, Quesneville H, Riccioni C, Ruotolo R, Splivallo R, Stocci V, Tisserant E, Roberto Viscomi A, Zambonelli A, Zampieri E, Henrissat B, Lebrun M-H, Paolocci F, Bonfante P, Ottonello S, Wincker P (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033-1038.

Maser C, Trappe JM, Nussbaum RA (1978) Fungal-small mammal interrelationships with emphasis on Oregon coniferous forests. *Ecology* **59**: 799-809.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Miller OK Jr, Henkel TW, James TY, Miller SL. 2001. *Pseudotulostoma*, a remarkable new volvate genus in the Elaphomycetaceae from Guyana. *Mycological Research* **105**: 1268-1272.

Miller SL, Miller OK Jr (1984) Synthesis of *Elaphomyces muricatus*+ *Pinus sylvestris* ectomycorrhizae. *Canadian journal of botany* **62**: 2363-2369.

Mondo SJ, Toomer KH, Morton JB, Lekberg Y, Pawlowska TE (2012) Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66**: 2564–2576.

Morgenstern I, Powlowski J, Ishmael N, Darmond C, Marqueteau S, Moisan M-C, Quenneville G, Tsang A (2012) A molecular phylogeny of thermophilic fungi. *Fungal Biology* **116**: 489-502.

Moulin L, Agnieszka A, Caroline B, Booth K, Vriezen JA, Melkonian R, Riley M. (2014). Complete Genome sequence of *Burkholderia phymatum* STM815, a broad host range and efficient nitrogen-fixing symbiont of Mimosa species. *Standards in Genomic Sciences* **9**.

Moulin L, Munive A, Dreyfus B, Boivin-Masson C (2001) Nodulation of legumes by members of the β-subclass of Proteobacteria. *Nature* **411**: 948-950.

Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulsen R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblyum TV, Fischer R, Fosker N, Fraser A, García JL, García MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones S, Gwilliam R, Haas B, Haas H, Harris D, Horiuchi H, Huang J, Humphray S, Jiménez J, Keller N, Khouri H, Kitamoto K, Kobayashi T, Konzack S, Kulkarni R, Kumagai T, Lafton A, Latgé J-P, Li W, Lord A, Lu C, Majoros WH, May GS, Miller BL, Mohamoud Y, Molina M, Monod M, Mouyna I, Mulligan S, Murphy L, O'Neil S, Paulsen I, Peñalva MA, Pertea M, Price C, Pritchard BL, Quail MA, Rabbinowitsch E, Rawlins N, Rajandream M-A, Reichard U, Renauld H, Robson GD, Rodriguez de Cordoba S, Rodríguez-Peña, Ronning CM, Rutter S, Salzberg SL, Sanchez M, Sánchez-Ferrero JC, Saunders D, Seeger K, Squares R, Squares S, Takeuchi M, Tekaia F, Turner G, Vazquez de Aladan CR, Weidman J, White O, Woodward J, Yu J-H, Fraser C, Galagan JE, Asai K, Machida M, Hall N, Barrell B, Denning DW (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**: 1151-1156.

Parke JL, Gurian-Sherman D (2001) Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annual Review of Phytopathology* **39**: 225–58.

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

Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 298-297.

Patel SS, Donmez I (2006) Mechanisms of helicases. *Journal of Biological Chemistry* **281**: 18265-18268.

Pel HJ, De Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JAE, van den

Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EGJ, Debets AJM, Dekker P, van Dijck PWM, van Dijk A, Dijkhuizen L, Driessen AJM, d'Enfert C, Geysens S, Goosen C, Groot GSP, de Groot PWJ, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JPTW, van den Hondel CAMJJ, van der Heijden RTJM, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJEC, Meulenberg R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NNME, Ram AFJ, Rinas U, Roubos JA, Sagt CMJ, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de Vondervoort PJJ, Wedler H, Wösten HAB, Zeng A-P, van Ooyen AJJ, Visser J, Stam H (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nature Biotechnology* **25**: 221-231.

Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351-i358.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Research* **40**: D290-D301.

Offre P, Pivato B, Mazurier S, Siblot S, Berta G, Lemanceau P, Mougel C (2008) Microdiversity of Burkholderiales associated with mycorrhizal and nonmycorrhizal roots of *Medicago truncatula*. *FEMS Microbiology Ecology* **65**: 180–92.

O'Hanlon KA, Gallagher L, Schrettl M, Jöchl C, Kavanagh K, Larsen TO, Doyle S (2012) Nonribosomal peptide synthetase genes pesL and pes1 are essential for fumigaclavine C production in *Aspergillus fumigatus*. *Applied and Environmental Microbiology* **78**: 3166-3176.

Reeves EP, Reiber K, Neville C, Scheibner O, Kavanagh K, Doyle S (2006) A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in *Aspergillus fumigatus*. *FEBS Journal* **273**: 3038-3053.

Renesto P, Crapoulet N, Ogata H, La Scola B, Vestris G, Claverie JM, Raoult D (2003) Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *The Lancet* **362**: 447-449.

Riedlinger J, Schrey SD, Tarkka MT, Hampp R, Kapur M, Fiedler H-P (2006) Auxofuran, a novel metabolite that stimulates the growth of Fly Agaric, is produced by the mycorrhiza helper bacterium *Streptomyces* strain AcH 505. *Applied and Environmental Microbiology* **72**: 3550–3557.

Robbertse B, Yoder RJ, Boyd A, Reeves J, Spatafora JW (2011) Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLoS Currents* **3**: RRN1213.

Samson RA, Yilmaz N, Houbraken J, Spierenburg H, Seifert KA, Peterson SW, Varga J, Frisvad JC (2011) Phylogeny and nomenclature of the genus *Talaromyces* and taxa accommodated in *Penicillium* subgenus *Biverticillium*. *Studies in Mycology* **70**: 159-183.

Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, Maiti R, Kodira CD, Neafsey DE, Zeng Q, Hung C-Y, McMahan C, Muszewska A, Grynberg M, Mandel MA, Kellner EM, Barker BM, Galgiani JN, Orbach MJ,

Kirkland TN, Cole GT, Henn MR, Birren BW, Taylor JW (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Research* **19**: 1722-1731.

Smit AF, Green P (1996) RepeatMasker. *Published on the web at http://www. Repeatmasker.org*.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, Lebrun M-H, Ridout CJ, Schulze-Lefert P, Talbot NJ, Ahmadinejad N, Ametz C, Barton GR, Benjdia M, Bidzinski P, Bindschedler LV, Both M, Brewer MT, Cadle-Davidson L, Cadle-Davidson MM, Collermare J, Cramer R, Frenkel O, Godfrey D, Harriman J, Hoede C, King BC, Klages S, Kleemann J, Knoll D, Koti PS, Kreplak J, López-Ruiz FJ, Lu X, Maekawa T, Mahanil S, Micali C, Milgroom MG, Montana G, Noir S, O'Connell RJ, Oberhaensli S, Parlange F, Pedersen C, Quesneville H, Reinhardt R, Rott M, Sacristán S, Schmidt SM, Schön M, Skamnioti P, Sommer H, Stephans A, Takahara H, Thordal-Christensen H, Vigouroux M, Weßling R, Wicker T, Panstruga R (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543-1546.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC bioinformatics* **7**: 62.

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564-577.

Tedersoo L, Hansen K, Perry BA, Kjøller R (2006) Molecular and morphological diversity of pezizalean ectomycorrhiza. *New Phytologist* **170**: 581-596.

Tedersoo L, May TW, Smith ME (2010) Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Myocorrhiza* **20**: 217-263.

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

Thiers HD (1984) The secotioid syndrome. *Mycologia* **76**: 1-8.

Traeger S, Altegoer F, Freitag M, Gabaldon T, Kempken F, Kumar A, Marcet-Houben M, Pöggeler S, Stajich JE, Nowrousian M (2013) The genome and development-dependent transcriptomes of *Pyronema confluens*: a window into fungal evolution. *PLoS Genetics* **9**: e1003820.

Trappe JM (1979) The orders, families, and genera of hypogeous Ascomycotina (truffles and their relatives). *Mycotaxon* **9**: 297-340.

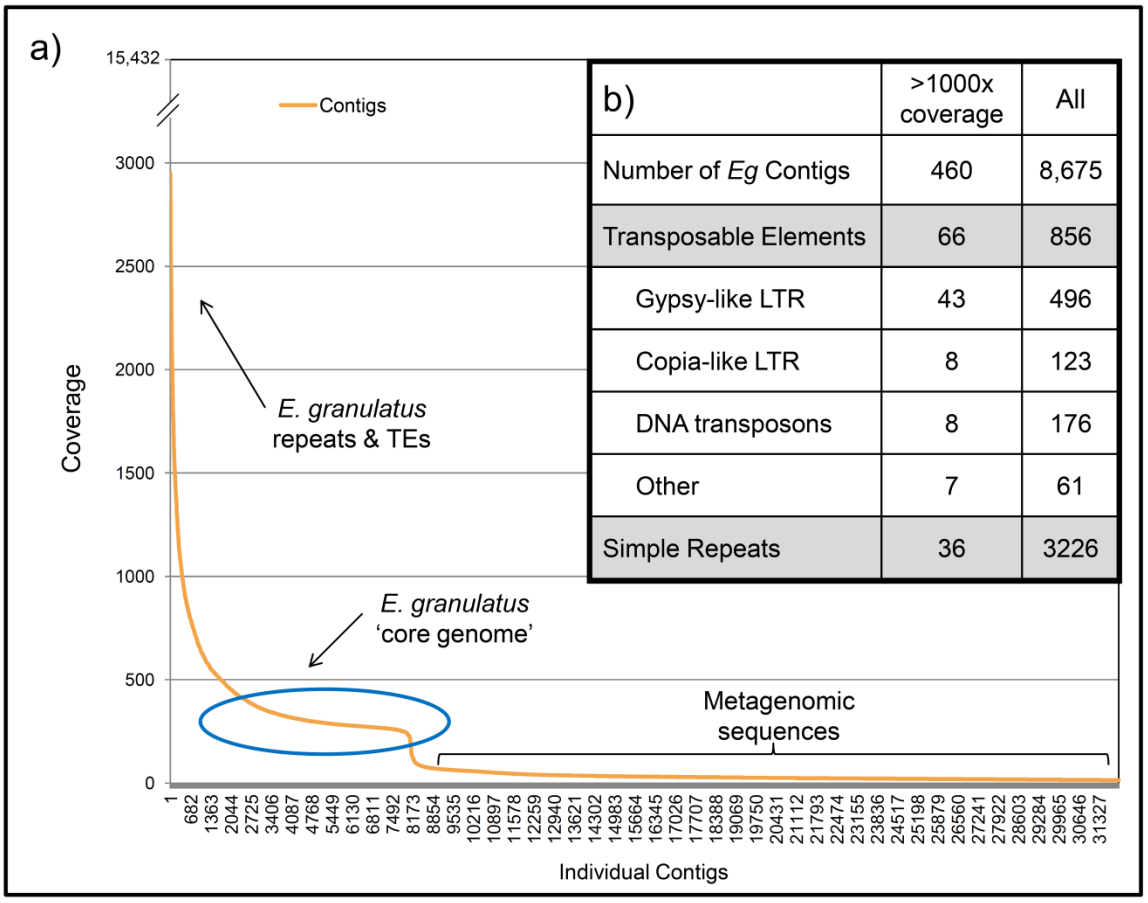
van den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, Driessen AJM, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WHM, Joardar V, Kiel JAKW, Kovalchuk A, Martin JF, Nierman WC, Nijland JG, Pronk JT, Roubos JA, van der Klei IJ, van Peij NNME, Veenhuis M, von Döhren H, Wagner C,

Wortman J, Bovenber RAL (2008) Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nature Biotechnology* **26**: 1161-1168.

Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter C, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Hothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, Kuske CR (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology* **75**: 2046-2056.

Wrighton KC, Thomas BC., Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC,Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF (2012) Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**: 1661-1665.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: W445-W451.



b)

	>1000x coverage	All
Number of *Eg* Contigs	460	8,675
Transposable Elements	66	856
Gypsy-like LTR	43	496
Copia-like LTR	8	123
DNA transposons	8	176
Other	7	61
Simple Repeats	36	3226

Figure 6.1. Coverages of initial contigs and their associations. a) Contigs are plotted as individual points along the x axis, with their respective coverages along the y axis. For visualization purposes contigs with coverage over 3000x and less than 20x have been omitted from the graph. The average coverage for most *E. granulatus* contigs is approximately 270x, while contigs with higher coverage are repetitive portions of the *E. granulatus* genome. Contigs representing the microbial community of the peridium had coverages much lower than the *E. granulatus* 'core genome.' b) RepeatMasker annotation of repetitive sequences and transposable elements in the 460 contigs with coverage higher than 1000x.

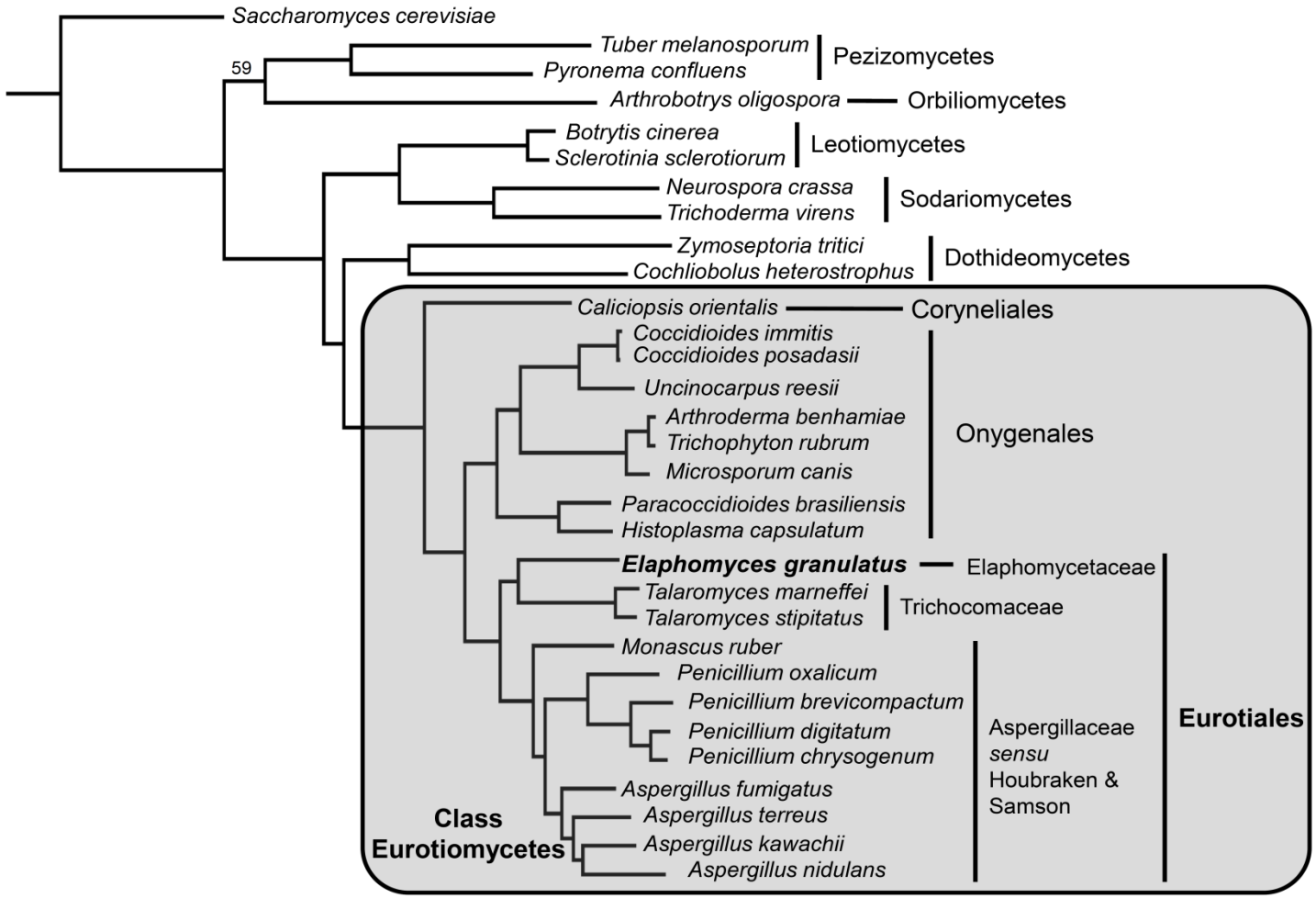


Figure 6.2. RAxML phylogeny created using 320 orthologous clusters from the selected Pezizomycotina taxa plus the outgroup *S. cerevisiae*. Bootstrap support for all nodes is 100, except for the node joining the Pezizomycetes and Orbiliomycetes (MLBP=59)

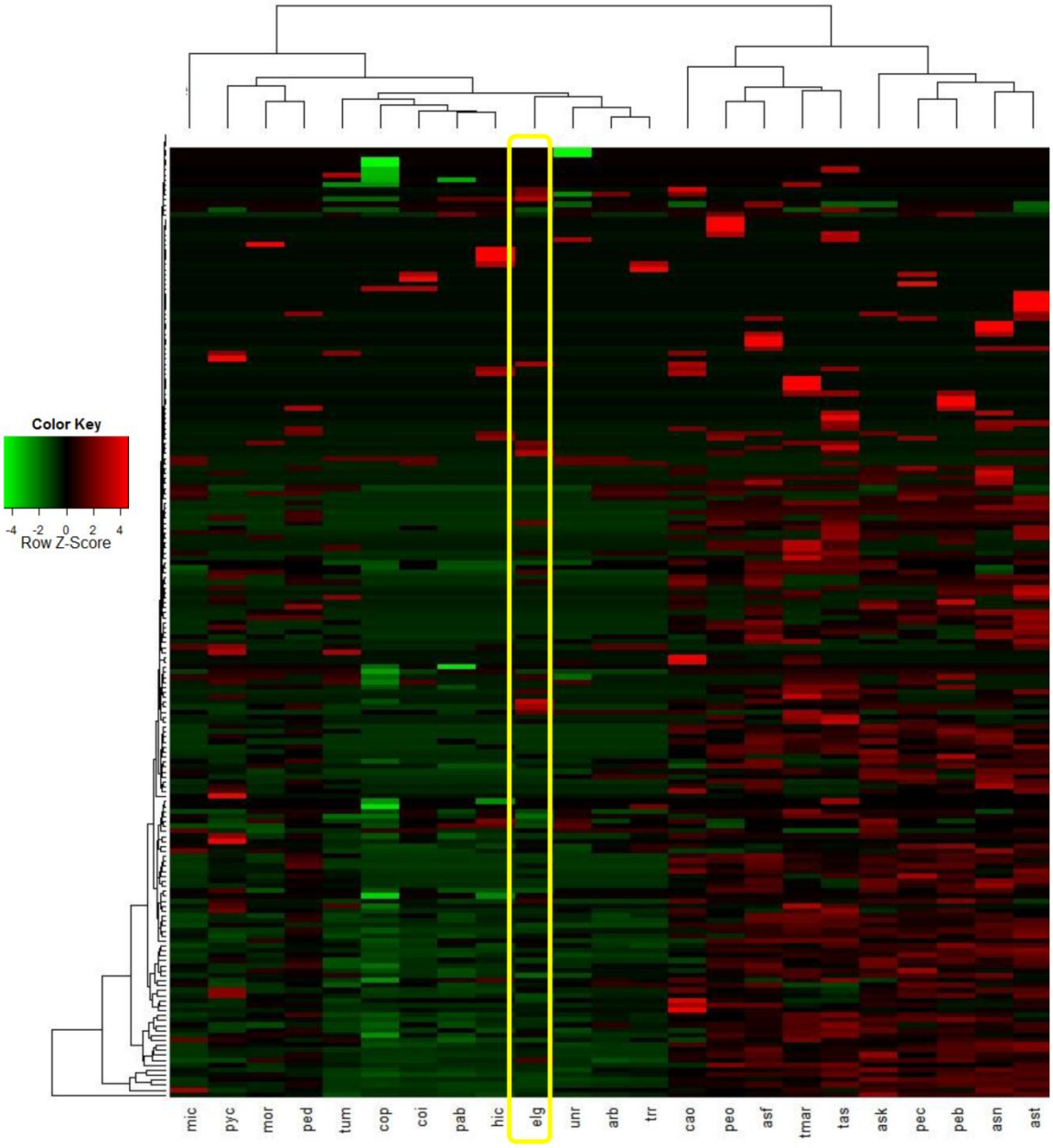


Figure 6.3. Heatmap of CAZymes for Eurotiomycetes and Pezizomycetes sampled in this study including *Arthroderma benhamiae* (arb), *Aspergillus terreus* (ast), *A. nidulans* (asn), *A. fumigatus* (asf), *A. kawachii* (ask), *Caliciopsis orientalis* (cao), *Coccidiodes immitis* (coi), *Co. posadasii* (cop), *E. granulatus* (elg), *Histoplasma capsulatum* (hic), *Microsporum canis* (mic), *Monascus ruber* (mor), *Paracoccidiodes brasiliensis* (pab), *Penicillium brevicompactum* (peb), *Pe. chrysogenum* (pec), *Pe. digitatum* (ped), *Pe. oxalicum* (peo), *Talaromyces marneffei* (tmar), *Ta. stipitatum* (tas), *Trichophyton rubrum* (trr). Dendrograms calculated using Euclidean distances.

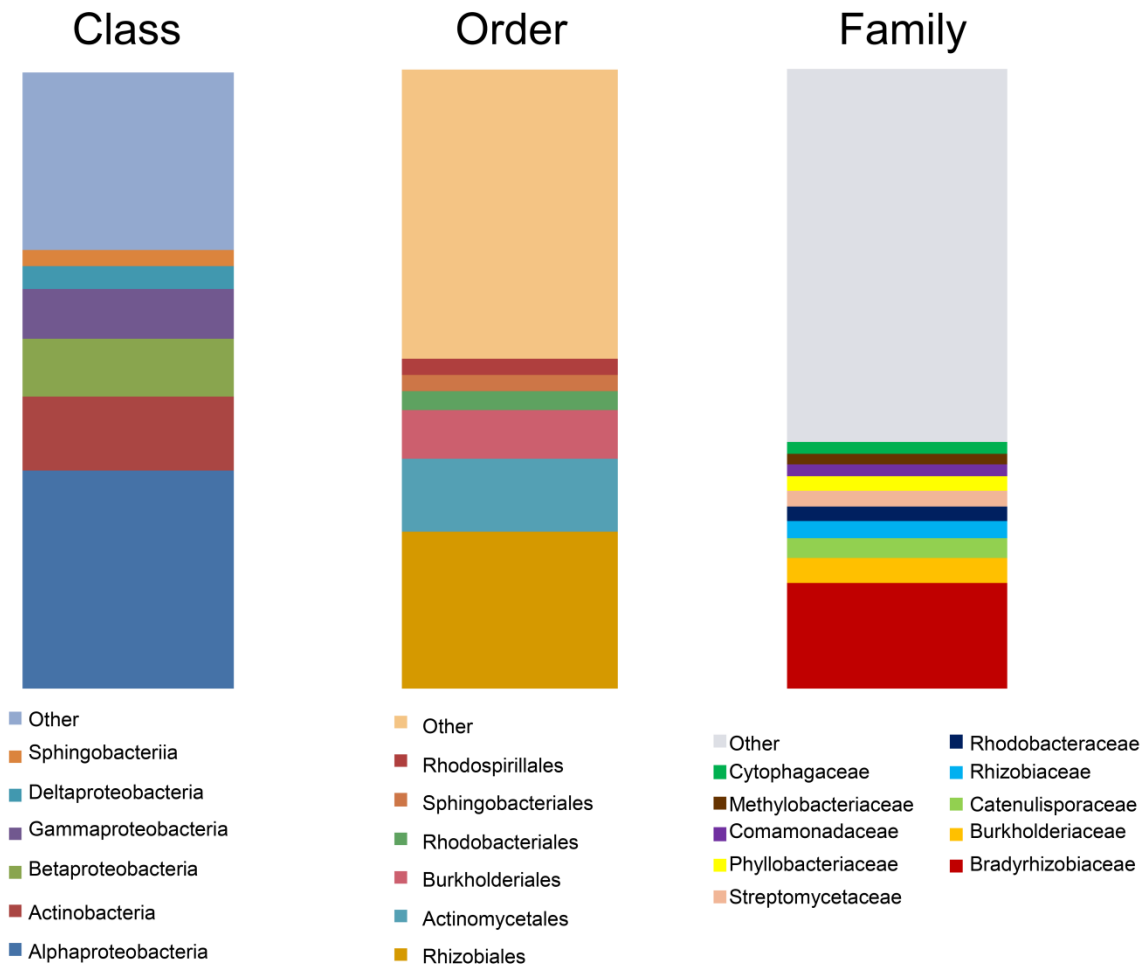


Figure 6.4. Bar charts showing percentages of bacterial taxa for total annotations at given taxonomic rank found in *E. granulatus* peridium.

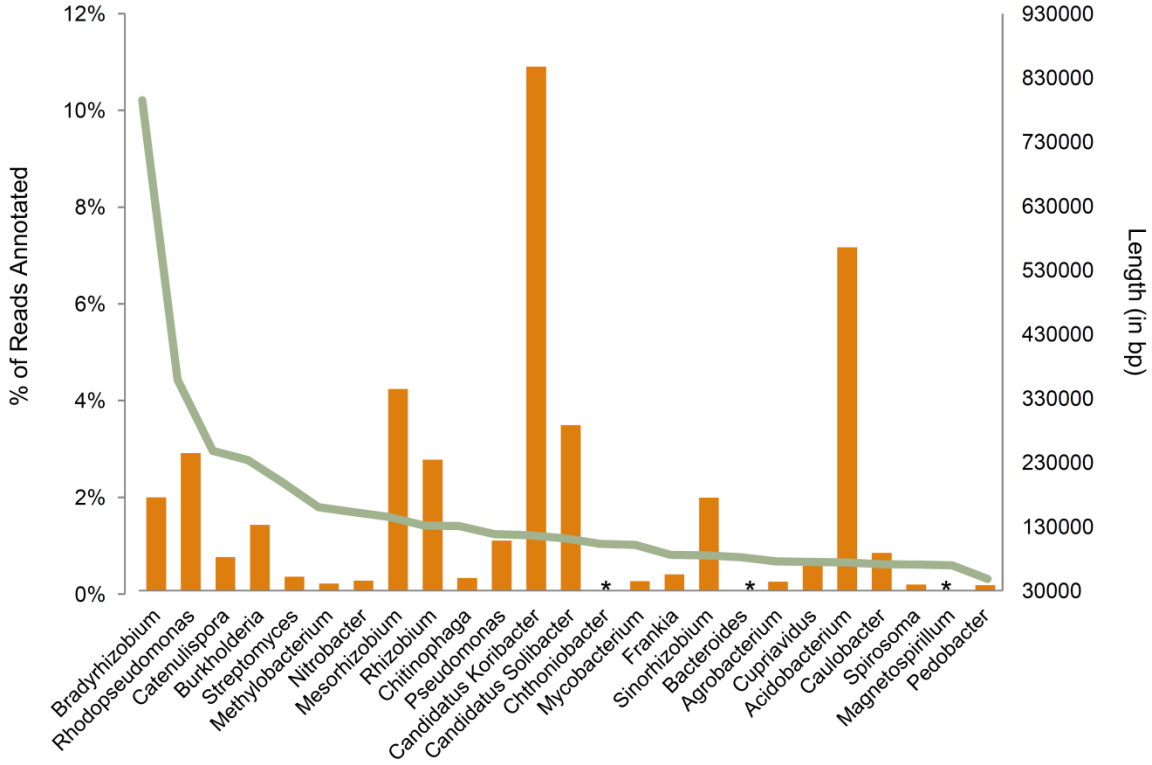


Figure 6.5. The 25 most common genera in the *E. granulatus* microbiome. Abundances of the top 25 most common genera in the *E. granulatus* peridium community are represented on the left y-axis. The longest assembled contig for each of those taxa is denoted by bar charts based on the right y-axis.

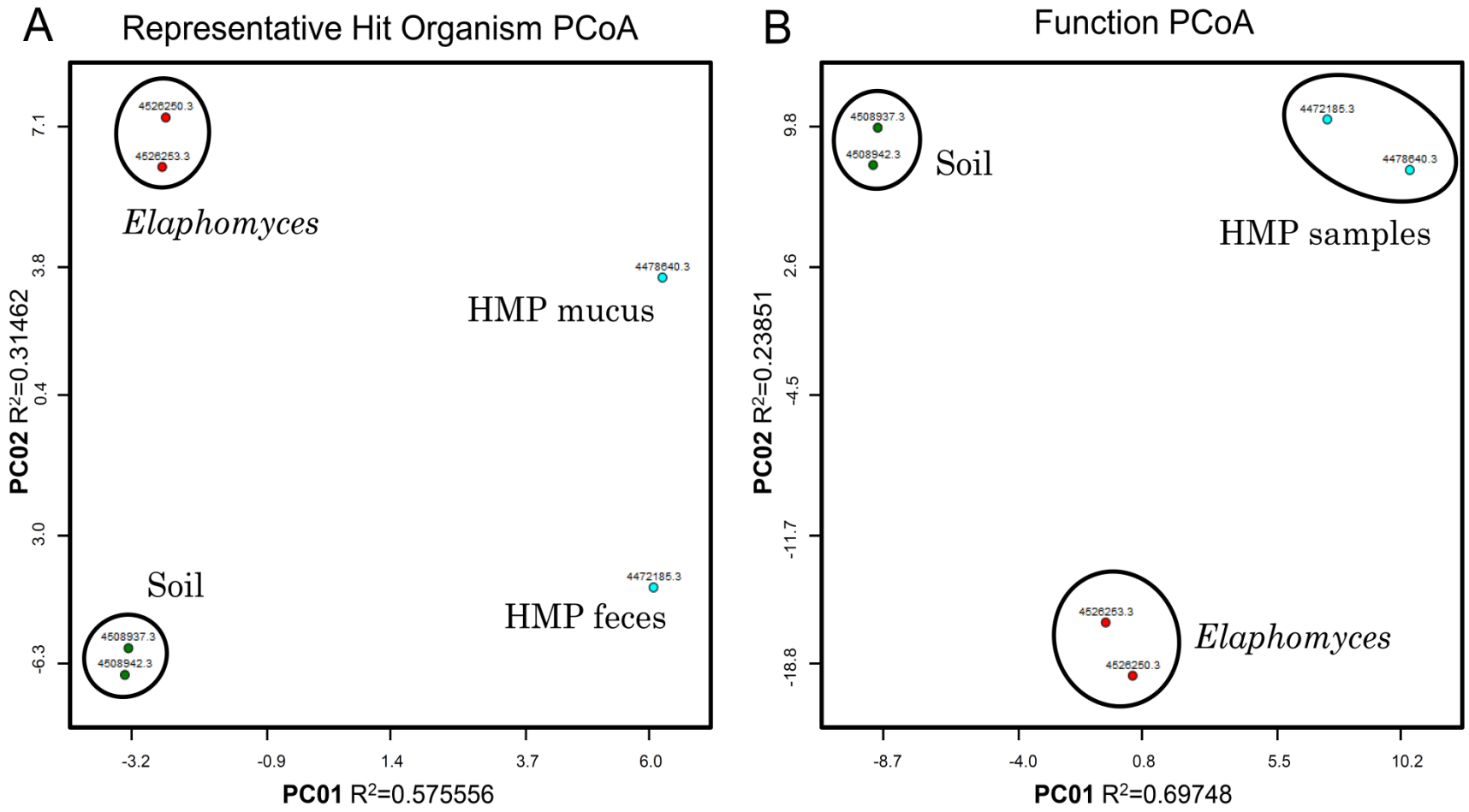


Figure 6.6. Principal coordinate analysis of *E. granulatus* samples, similarly shotgun sequenced comparable read count soil samples, and Human Microbiome Project (HMP) mucus and feces samples analyzing A) best representative hit classification and B) the KEGG functional annotations from MG-RAST.

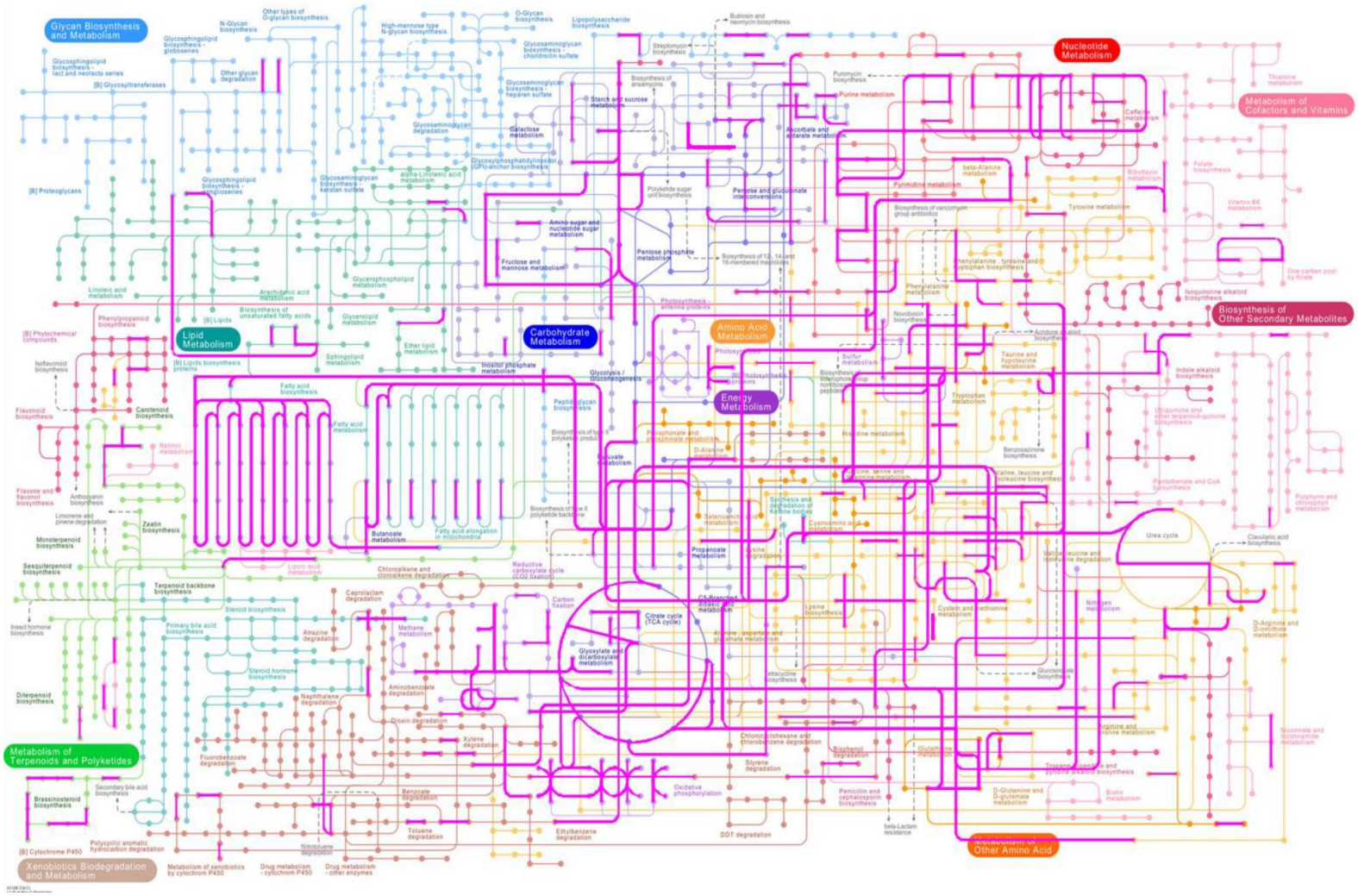


Figure 6.7. KeggMapper visualization of the ECs present in the *E. granulatus* peridium microbiome highlighted in purple.

Table 6.1. Genome assembly statistics and comparison with close and selected distant relatives in the Eurotiales and *Tuber melanosporum*. Estimates of *E. granulatus* genome size and protein model number is given, with the actual size of the assembly and protein models with RNA evidence given in parentheses.

	E. granulatus	*T. stipitatus*[1]	*T. marneffei*[2]	*Pe. chrysogenum*[3]	*A. fumigatus*[4]	*T. melanosporum*[5]
Size (MB)	71.5 (54.2)	35.7	28.64	32.2	29.2	124.9
Longest scaffold	90,200	5,640,589	6,407,042	6387817	4,896,001	2,785,000
GC %	45.9	46.1	46.7	49.0	49.5	52.0
N50 (bp)	11,830	4,363,329	3,339,384	3,889,175	3,791,214	638,000
Total scaffolds	8,675	820	452	49	55	398
Predicted Proteins	8,477 (7,184)	13,252	10,638	13,671	9,916	7,496

[1]T. stipitatus ATCC 10500, JCVI
[2]*T. marneffei* ATCC 18224, JCVI
[3] *P. chrysogenum* Wisconsin54-1255, van den Berg *et al.* 2008.
[4] *A. fumigatus* A1163, Fedorova *et al.* 2008
[5]*T. melanosproum* Mel28, Martin *et al.* 2010

Table 6.2. Expanded helicase clusters in *E. granulatus* and the counts per cluster for Eurotiales and *Tuber melanosporum*. Other taxa included in the analysis (see Figure 2) have been excluded from the table due to space limitations. Abbreviations: Elg, *E. granulatus*; Tam, *Talaromyces marneffei*; Tas, *Talarmomyces stipitatus*; Mor, *Monascus ruber*; Asf, *Aspergillus fumigatus*; Ask, *Aspergillus kawachii*; Asn, *Aspergillus nidulans*; Ast, *Aspergillus terreus*; Peb, *Penicillium brevicompactum*; Pec, *Pencillium chrysogenum*; Ped, *Penicillium digitatum*; Peo, *Penicillium oxalicum*; Tub, *T. melanosporum*.

Cluster #	Annotation	Elg	Tam	Tas	Mor	Asf	Ask	Asn	Ast	Peb	Pec	Ped	Peo	Tub	Avg.
1200611	helicase	31	0	24	0	0	0	0	0	1	6	0	1	0	3
120055	helicase	23	8	9	7	6	6	6	7	5	6	5	6	5	7
120067	Mito DNA helicase	13	6	9	6	6	6	7	6	4	8	6	7	3	6

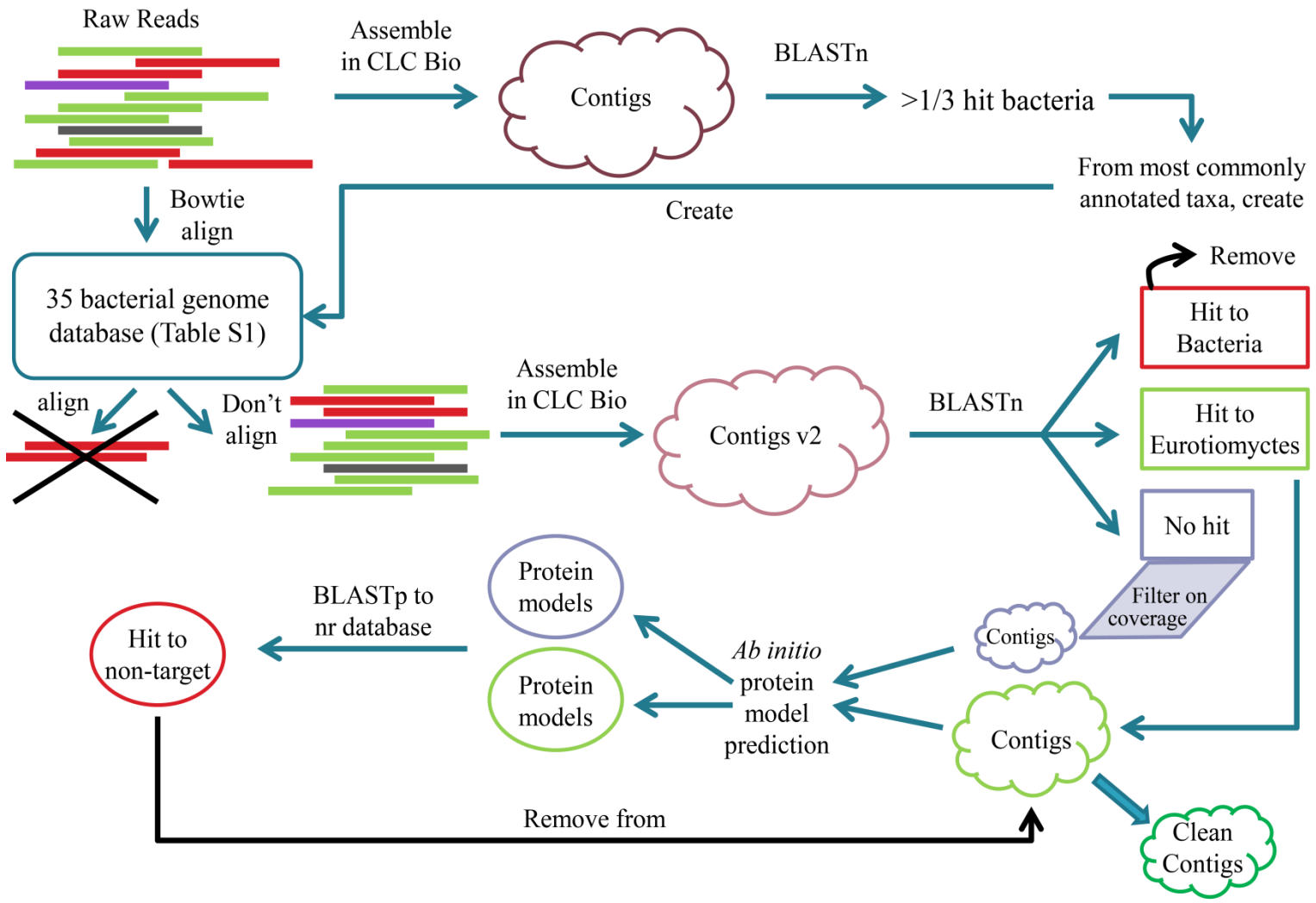


Figure 6.S1. Graphical representation of the workflow for obtaining the contigs of the target genome, *E. granulatus*.

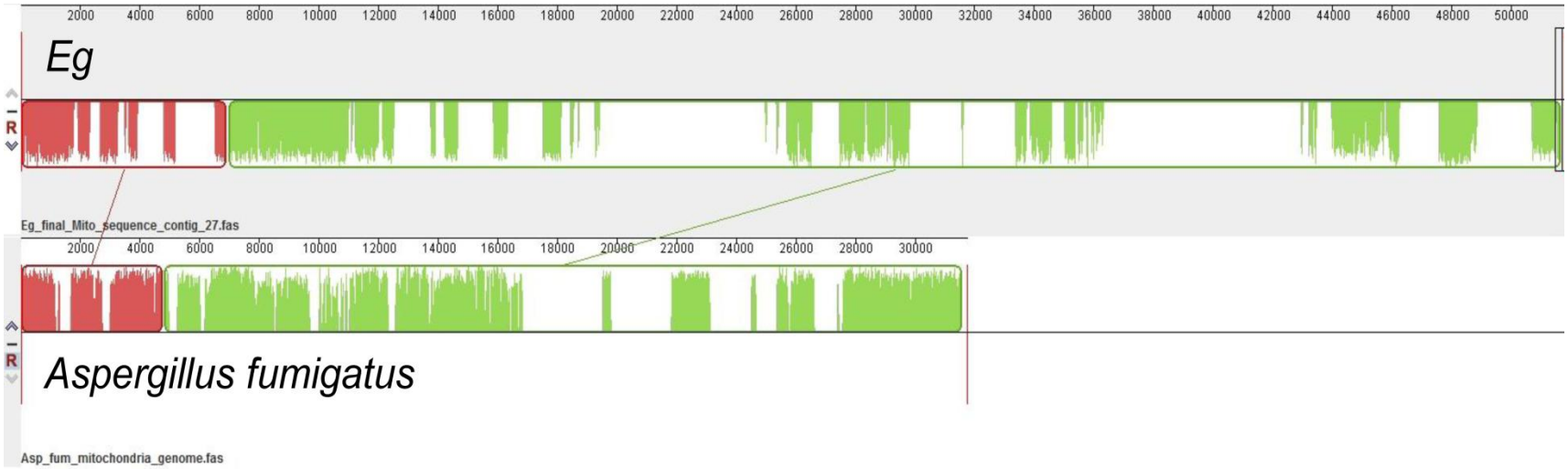


Figure 6.S2. Alignment of mitochondrial genome assemblies of *E. granulatus* and *Aspergillus fumigatus*.

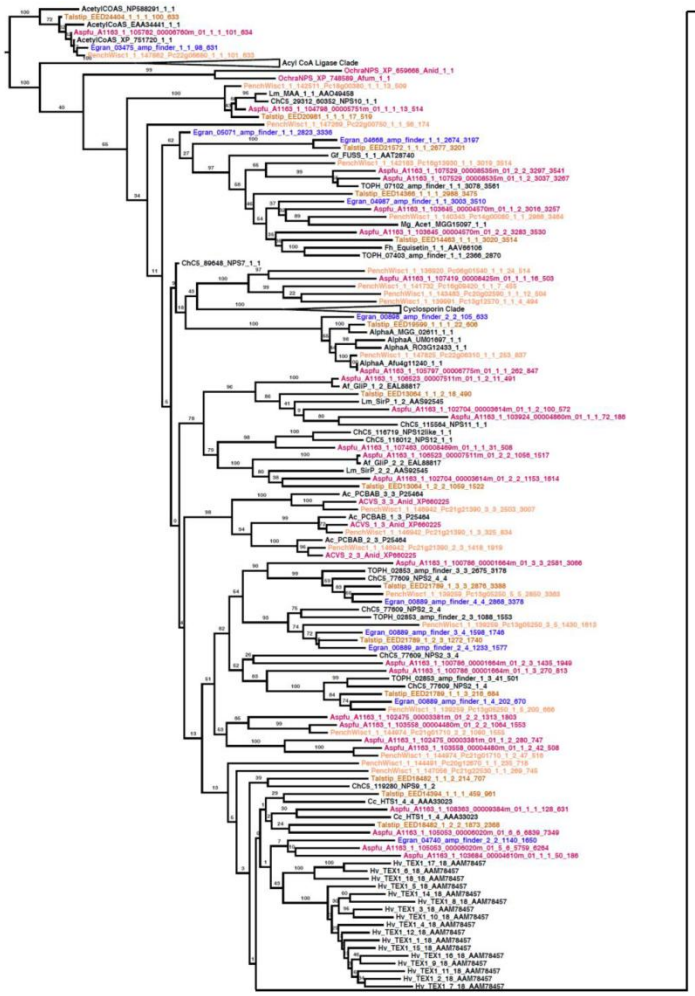


Figure 6.S3. Adenylation domain phylogeny with broad sampling of NRPS A-domains from various Ascomycota (in black text) with whole genome sampling from *E. granulatus* (Egran) in blue, *Aspergillus fumigatus* (Aspfu) in pink, *Penicillium chrysogenum* (Pench) in peach, and *Talaromyces stipitatus* (Talstip) in ochre. The red box highlights the *E. granulatus* specific NRPS A-domains that form a monophyletic clade.

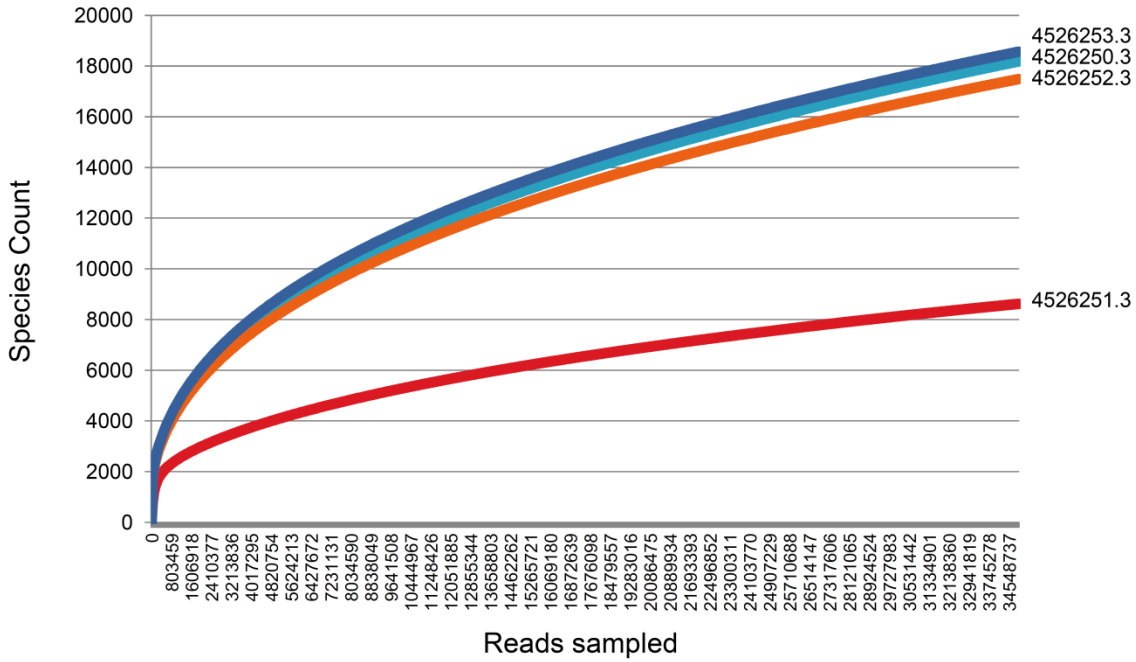


Figure 6.S4. Rarefaction curves for species richness versus sampling depth for the two peridium microbiome libraries with paired-end reads uploaded separately (2 reads per library). MG-RAST IDs are given to the right of the curves.

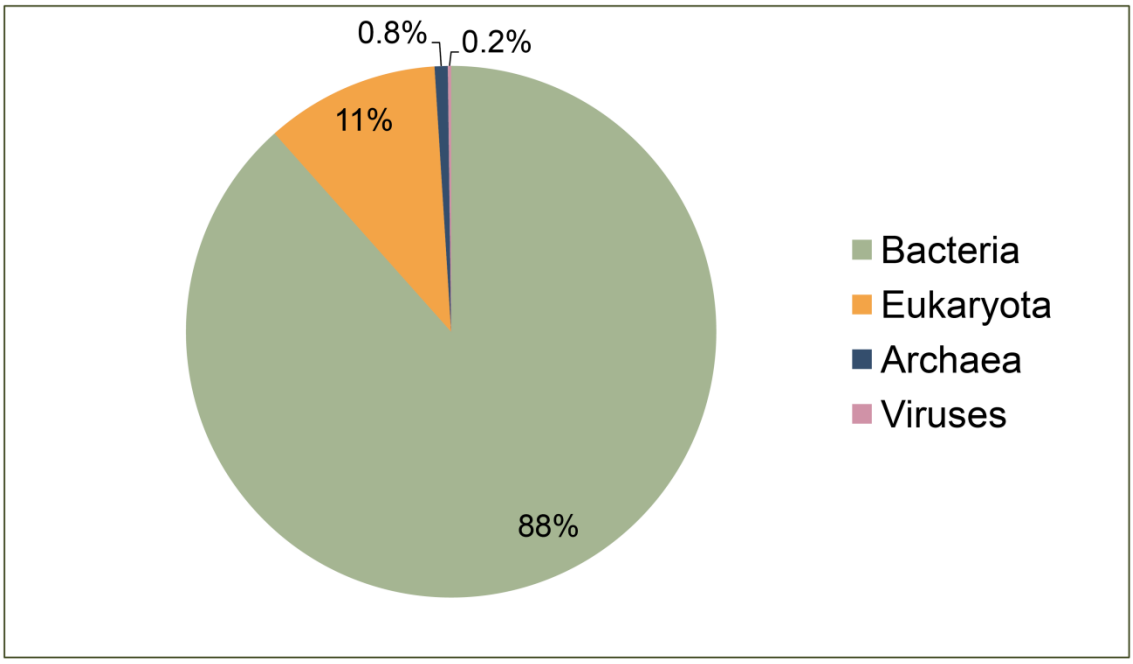


Figure 6.S5. Taxonomy pie chart. Percentages of the reads in the *E. granulatus* peridium microbiome annotated at the phylum level.

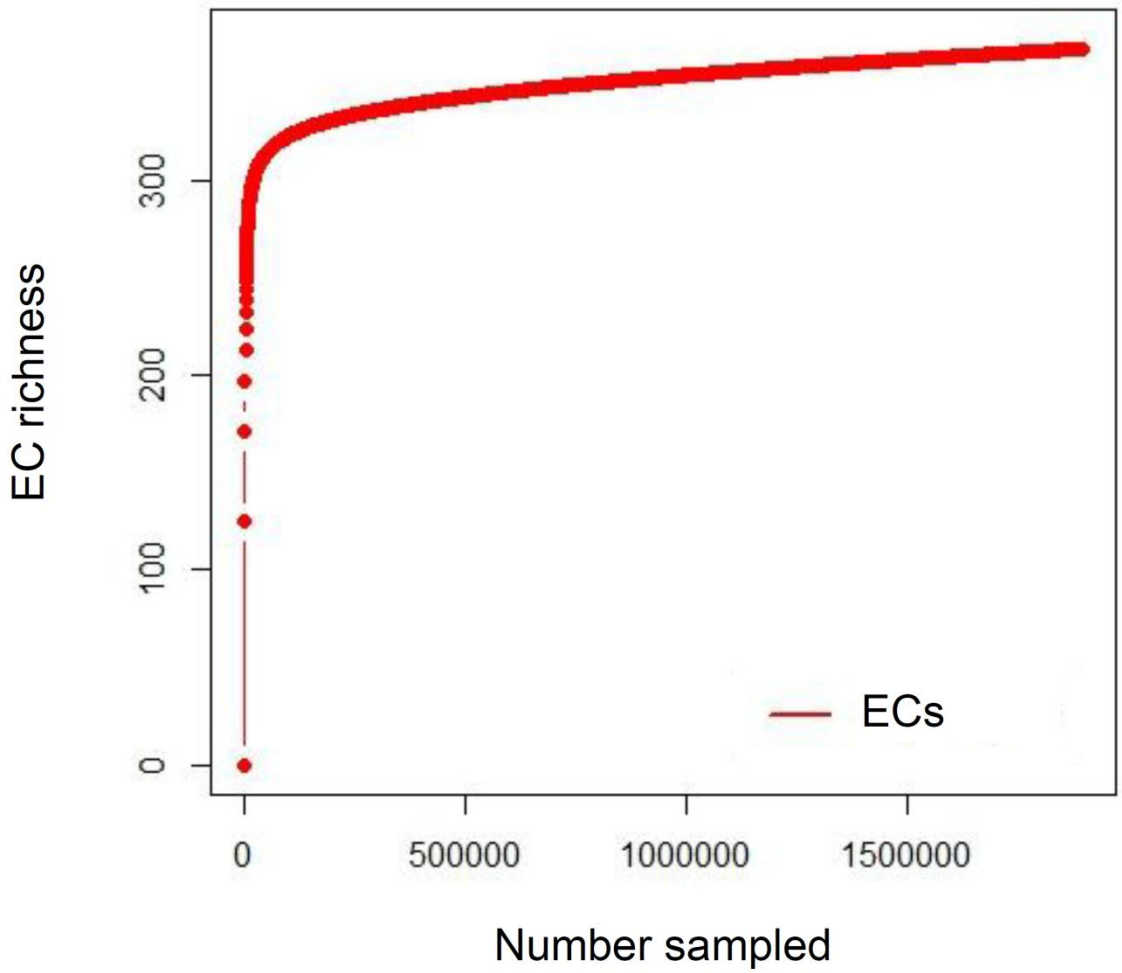


Figure 6.S6. Rarefaction of the Enzyme Commission numbers (ECs) for *E. granulatus* peridium microbiome with subsample size of 250.

Table 6.S1. List of the 35 bacterial genomes used in the bacterial database to remove raw reads using Bowtie 2.

Class	Strains
Acidobacteriia	*Acidobacterium capsulatum* ATCC 51196 Candidatus *Koribacter versatilis* Ellin345 *Granulicella mallensis* MP5ACTX8
Alphaproteobacteria	*Azospirillum basilense* Sp245 *Beijerinckia indica* subsp. *indica* ATCC 9039 *Bradyrhizobium japonicum* USDA110 *Chelativorans* sp. BNC1 *Mesorhizobium ciceri* bv. *biserrulae* WSM1271 *Methylobacterium nodulans* ORS 2060 *Rhizobium leguminosarum* bv. *trifolii* WSM1325 *Sinorhizobium fredii* HH103
Betaproteobacteria	*Azoarcus* sp. KH32C *Burkholderia cepacia* GG4 Candidatus *Accumulibacter phosphatis* clade IIA str. UW-1 *Herbaspirillum seropedicae* SmR1 *Methylibium petroleiphilum* PM1
Deltaproteobacteria	*Bdellovibrio exovorus* JSS *Haliangium ochraceum* DSM 14365 *Stigmatella aurantiaca* DW4/3-1 *Syntrophus aciditrophicus* SB
Gammaproteobacteria	*Pseudomonas aeruginosa* DK2 *Rhodanobacter* sp. 2APBS1 *Thioalkalivibrio sulfidophilus* HL-EbGr7 *Thiocystis violascens* DSM 198
Actinobacteria	*Catenulispora acidiphila* DSM 44928 *Conexibacter woesei* DSM 14684 *Streptomyces bingchenggensis* BCW-1
Deinococci	*Deinococcus maricopensis* DSM 21211
Cytophagia	*Spirosoma linguale* DSM 74
Flavobacteriia	*Robiginitalea biformata* HTCC2501
Sphingobacteriia	*Niastella koreensis* GR20-10 *Pedobacter saltans* DSM 12145
Oscillatoriophycideae (subclass)	*Microcoleus* sp. PCC 7113 *Oscillatoria nigro-viridis* PCC 7112

Table 6.S2. Genome assembly statistics for *Caliciopsis orientalis* and *Monascus ruber*, both of which are being published for the first time here. * Sequenced and assembled at Oregon State University using methods described in text. **Sequenced and assembled at the Joint Genomes Institute as a part of 1KFG CSP. ******Ab initio* models created in AUGUSTUS – no RNA sequenced.

	*C. orientalis**	*M. ruber***
Strain	CBS 138.64	NRRL 1597
Size (MB)	28.17	24.80
Longest scaffold	392,255	498,644
GC%	49.9	48.9
N50 (bp)	86,295	153,832
L50 (contig #)	111	49
Total scaffold number	643	320
Predicted protein models	8,494***	9,650

Table 6.S3. Counts of CAZymes for the selected Eurotiomycetes and Pezizomycetes analyzed in this study, including *E. granulatus* (highlighted in yellow), other Eurotiales (green), Onygenales (red), *Caliciopsis orientalis* (olive), Pezizales (blue).

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
AA1	2	3	2	1	3	2	5	2	7	6	2	2	2	0	0	0	3	1	4	0	4	2	1
AA10	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
AA2	1	4	4	3	3	3	3	2	3	3	3	3	3	3	1	3	3	2	3	3	2	2	1
AA3	3	12	17	10	31	25	22	13	27	25	14	14	4	6	4	7	5	7	5	3	16	15	4
AA4	2	6	6	6	4	5	3	1	4	4	4	3	1	3	2	2	1	1	2	2	4	2	1
AA5	3	4	2	0	0	0	0	0	0	2	0	0	2	1	1	1	1	0	2	2	2	0	0
AA6	1	4	4	2	2	1	1	2	4	2	2	2	1	1	0	1	1	0	1	1	1	1	1
AA7	13	27	34	23	43	38	33	10	29	29	15	18	18	12	8	11	17	9	16	7	21	13	10
AA8	0	0	0	2	2	1	4	0	2	1	1	1	0	0	0	0	0	0	0	0	5	3	1
AA9	4	3	1	6	6	10	10	1	6	9	2	4	2	1	0	0	1	0	2	2	11	14	5
CBM1	1	17	20	17	7	6	15	0	6	6	4	23	0	0	0	0	0	0	0	0	0	7	3
CBM10	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0
CBM11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CBM12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
CBM13	0	0	2	0	1	0	2	0	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0
CBM14	0	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	13	0
CBM16	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	7	1	0
CBM18	3	20	12	9	11	15	13	10	15	10	7	13	6	5	4	4	19	1	9	8	4	13	13
CBM19	1	2	2	1	0	0	1	0	1	1	0	0	1	0	0	0	1	0	1	0	0	0	0
CBM20	1	3	4	5	2	4	4	1	2	3	2	4	2	2	0	1	1	1	2	2	20	1	0
CBM21	1	2	2	1	1	1	1	1	1	2	1	1	1	2	0	1	1	1	1	2	1	2	2
CBM22	0	1	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM23	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM24	1	1	2	7	4	1	4	3	1	4	3	7	1	0	0	0	0	0	1	0	10	0	0
CBM27	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
CBM3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6.S3 (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
CBM32	0	1	0	1	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	2
CBM35	0	1	1	2	1	1	3	0	0	0	0	2	0	0	0	0	0	0	0	0	1	2	0
CBM37	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM38	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM4	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CBM41	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM42	0	4	2	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0
CBM43	1	1	1	4	4	3	3	2	3	3	3	2	2	1	0	1	2	1	2	1	1	1	1
CBM45	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
CBM46	0	0	1	1	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
CBM48	2	2	2	1	1	1	1	2	3	2	1	1	1	1	0	1	2	1	1	0	1	2	2
CBM49	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CBM50	3	24	7	12	27	36	23	6	9	19	11	13	12	3	3	1	48	0	15	13	8	1	1
CBM52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4
CBM53	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CBM6	2	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CBM61	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0
CBM63	2	1	2	1	1	1	1	0	1	1	1	2	0	0	0	0	0	0	0	0	1	0	0
CBM64	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CBM66	0	4	0	3	1	1	4	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0
CBM67	0	2	0	4	5	7	3	1	4	3	2	3	0	0	0	0	0	0	0	0	1	0	1
CBM9	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	0
CE1	11	24	30	25	22	26	28	21	27	18	15	26	15	10	8	11	16	9	16	13	19	17	11
CE10	26	44	48	43	81	55	65	36	54	53	32	37	23	23	21	15	29	20	26	20	43	28	13
CE11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CE12	1	3	3	3	4	3	3	1	3	3	3	3	1	2	0	1	2	0	2	1	4	6	1

Table 6.S3 (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
CE14	3	4	3	3	3	2	4	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
CE15	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
CE16	0	5	5	4	3	5	5	2	3	3	1	4	0	1	1	1	0	1	0	0	4	3	1
CE2	0	0	0	1	1	1	3	2	1	2	2	0	0	0	0	0	0	0	0	0	1	0	0
CE3	2	2	2	1	1	6	3	3	3	3	2	0	2	0	0	1	5	0	3	0	0	1	0
CE4	1	7	4	6	7	8	4	4	7	7	5	5	4	3	3	3	3	2	4	2	5	4	6
CE5	0	3	2	4	5	4	8	3	8	7	4	4	2	1	1	2	2	1	2	1	3	10	2
CE7	0	2	2	2	2	1	2	1	1	1	1	1	1	1	0	0	1	0	1	0	2	1	0
CE8	0	2	2	4	3	3	1	0	3	2	3	3	0	0	0	0	0	0	0	0	4	1	0
CE9	1	2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
dockerin	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH1	0	3	2	5	3	3	3	0	4	3	3	4	0	0	0	0	0	0	0	0	3	1	2
GH10	0	0	1	3	1	2	3	0	1	2	0	2	0	0	0	0	0	0	0	0	1	1	1
GH105	0	3	2	3	2	4	2	1	3	2	1	1	0	0	0	0	0	0	0	0	1	2	0
GH106	0	0	0	1	2	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	1	0	0
GH109	5	16	14	11	15	15	8	8	19	16	8	9	7	4	3	6	6	4	6	4	9	8	5
GH11	0	2	3	3	4	2	2	0	1	1	1	3	0	0	0	0	0	0	0	0	1	0	0
GH113	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH114	4	1	1	1	2	2	1	1	2	2	2	2	0	0	0	0	0	0	0	0	1	0	0
GH115	0	0	1	1	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
GH12	0	3	3	4	4	1	6	1	3	3	2	3	0	0	0	0	0	0	0	0	6	1	1
GH125	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1
GH127	0	2	3	1	1	1	3	0	1	1	0	1	0	1	0	0	0	0	0	0	2	1	0
GH128	1	6	4	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1
GH13	14	7	10	13	14	10	10	12	15	13	9	7	1	3	3	5	2	5	2	4	10	2	5
GH131	1	0	0	2	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	2	0	1
GH132	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	2	2	2	10	1
GH15	2	2	2	2	1	0	1	2	1	2	1	1	0	0	0	0	0	0	0	0	1	2	1

Table 6.S3 (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
GH16	9	15	17	12	12	12	8	10	13	14	14	10	9	8	5	7	7	6	8	8	12	8	6
GH17	3	4	4	5	5	5	4	3	5	4	4	5	4	4	3	3	4	3	4	4	4	5	4
GH18	6	13	16	10	12	10	12	4	10	7	7	11	11	8	3	6	9	4	7	8	10	5	5
GH2	2	5	5	6	6	9	10	3	7	6	3	6	2	1	1	1	2	1	2	2	3	2	2
GH20	1	2	2	2	3	2	2	1	2	2	2	1	2	2	1	3	1	3	2	3	2	1	2
GH22	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GH23	0	1	0	1	0	0	0	0	2	1	0	2	0	0	0	1	0	2	0	1	1	0	1
GH24	0	0	0	0	0	1	2	0	0	0	0	0	2	0	0	1	2	0	2	0	0	3	1
GH25	1	1	1	3	0	3	2	1	0	1	0	1	1	0	0	0	1	0	1	0	0	0	1
GH26	0	0	0	0	1	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0
GH27	0	2	2	3	3	3	1	0	2	1	0	1	0	0	0	0	0	1	0	0	2	1	0
GH28	2	7	8	12	19	9	8	2	12	5	8	11	0	0	0	0	0	0	0	0	8	2	2
GH29	0	0	2	0	1	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
GH3	4	13	22	18	16	20	21	7	20	17	14	14	6	5	4	4	6	4	6	5	13	8	5
GH30	0	5	6	1	1	0	1	1	1	1	1	3	0	0	0	0	0	0	0	0	0	0	0
GH31	6	8	7	6	7	10	11	3	9	12	6	6	2	2	1	3	2	1	2	1	5	5	5
GH32	1	0	4	3	3	2	5	0	4	8	3	2	0	0	0	0	0	0	0	0	1	1	1
GH33	0	0	0	1	1	1	2	0	1	2	0	0	1	0	0	0	1	0	1	0	1	0	0
GH35	1	3	2	4	3	2	3	1	4	4	4	2	0	0	0	0	0	0	0	0	3	1	0
GH36	0	0	1	2	1	3	3	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
GH37	3	1	1	1	1	1	1	1	1	1	1	1	1	1	0	2	1	2	1	1	1	1	0
GH38	1	1	2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
GH39	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
GH42	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
GH43	4	6	10	17	10	16	19	3	15	14	3	10	1	1	1	2	1	2	1	1	6	4	1
GH45	0	2	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
GH47	6	7	9	5	5	7	6	6	6	7	6	6	7	6	3	8	6	7	6	6	8	5	5

Table 6.S3 (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
GH49	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH5	6	6	8	9	7	14	16	6	12	9	7	8	3	2	1	4	3	3	3	2	12	8	7
GH51	0	1	2	2	4	2	3	1	3	3	1	3	0	0	0	0	0	0	0	0	3	2	0
GH53	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0
GH55	3	7	7	7	3	4	4	5	4	3	3	6	1	1	0	3	2	4	1	1	1	1	2
GH6	0	0	0	0	1	2	1	0	1	2	1	0	0	0	0	0	0	0	0	0	1	0	0
GH62	0	1	1	1	1	2	2	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
GH63	1	2	2	2	2	2	2	2	2	2	2	2	1	1	0	2	1	2	1	1	2	2	2
GH64	1	3	5	0	0	0	0	0	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0
GH65	0	0	2	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0
GH67	0	2	2	1	1	1	2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
GH7	0	0	0	2	1	2	2	0	1	1	1	1	0	0	0	0	0	0	0	0	2	2	0
GH71	2	6	5	6	5	4	5	1	3	3	4	1	0	1	0	1	0	0	0	1	2	1	0
GH72	2	4	3	3	3	2	3	3	3	3	3	3	2	3	2	3	2	2	2	3	3	2	3
GH74	5	2	2	3	2	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2	1	1
GH75	1	1	1	3	2	2	2	2	3	1	1	3	1	1	1	0	1	0	2	1	0	1	1
GH76	2	6	7	8	11	7	10	4	10	8	7	8	3	3	2	4	3	5	3	3	2	7	1
GH78	0	1	2	2	3	3	1	0	4	2	2	1	0	0	0	0	0	0	0	0	3	0	1
GH79	1	0	1	0	3	1	2	1	5	2	1	2	0	0	0	0	0	0	0	0	1	0	0
GH81	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	2
GH82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
GH84	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0
GH85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
GH88	0	0	0	2	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GH89	0	2	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GH92	1	5	4	4	5	5	6	1	5	4	4	4	2	1	0	1	2	1	2	2	2	1	1
GH93	0	2	3	3	1	2	4	1	2	2	2	3	0	0	0	1	0	0	0	0	1	0	0
GH95	0	1	2	2	2	3	3	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	0

Table 6.S3 (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
GH99	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0
GT1	5	4	8	6	7	8	8	2	11	7	4	6	2	2	1	1	1	4	1	4	3	3	4
GT11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
GT14	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GT15	3	7	5	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	5	5
GT17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
GT2	13	17	19	14	14	11	12	14	14	13	13	16	9	11	3	11	9	8	9	9	11	13	11
GT20	4	7	6	7	6	4	5	6	5	10	5	5	4	3	2	3	4	2	4	4	3	3	3
GT21	1	2	2	2	4	2	2	3	1	2	2	1	1	2	1	4	2	2	2	3	3	1	1
GT22	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	2	4	4	4	4	4	4	4
GT23	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	1
GT24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
GT25	0	3	2	2	4	4	4	0	1	3	1	1	2	0	0	0	2	0	1	0	2	1	0
GT26	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GT28	1	1	0	2	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
GT3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GT30	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GT31	2	4	4	4	4	5	7	1	3	5	3	1	1	1	1	1	2	0	1	3	4	0	0
GT32	6	12	13	8	13	8	10	7	10	9	9	9	5	3	2	2	5	7	5	2	3	3	3
GT33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
GT34	2	5	7	2	2	3	3	3	5	4	3	3	2	2	1	2	2	2	2	3	5	6	2
GT35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GT39	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	4	3	3	3	3
GT4	5	5	5	8	8	6	7	4	6	6	7	7	5	4	2	4	5	4	5	4	5	4	3
GT40	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GT41	0	1	1	1	0	1	0	0	1	1	1	1	1	0	0	0	1	0	1	0	1	0	1
GT45	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GT48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1

Table 6.S3. (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
GT5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
GT50	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	0	2	1	1
GT54	0	1	1	1	3	3	1	0	0	3	0	1	2	0	0	0	1	0	2	1	0	0	0
GT57	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	0	2	2	2	2	2
GT58	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GT59	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
GT61	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
GT62	3	3	4	3	3	3	3	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3	3
GT66	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
GT69	3	1	1	3	5	3	4	1	2	2	2	4	4	4	4	2	5	2	4	4	2	2	2
GT71	0	1	1	3	4	2	4	3	2	3	2	4	2	3	2	3	2	3	2	3	12	1	0
GT76	1	2	2	2	2	0	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1
GT78	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GT8	5	4	3	4	5	5	2	3	7	5	3	4	2	3	3	2	2	2	2	2	6	5	2
GT81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
GT90	4	2	2	9	5	4	6	3	9	9	5	7	7	5	4	3	6	2	7	5	6	2	1
GT91	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL1	0	2	0	6	7	8	7	1	4	5	3	3	0	0	0	0	0	0	0	0	2	2	2
PL10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
PL11	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
PL17	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL20	0	1	1	1	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL22	0	0	0	0	2	0	1	0	4	1	1	0	0	0	0	0	0	0	0	0	1	1	0
PL3	0	0	0	3	0	5	3	0	1	1	1	0	0	0	0	0	0	0	0	0	2	2	0
PL4	0	0	0	3	2	4	3	0	2	3	2	3	0	0	0	0	0	0	0	0	1	1	1
PL5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL7	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6.S3. (Continued)

cazymes	Elg	Tam	Tas	Asf	Ask	Asn	Ast	Mor	Peb	Pec	Ped	Peo	Arb	Coi	Cop	Hic	Mic	Pab	Trr	Unr	Cao	Pyc	Tum
PL9	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table 6.S4. MCL clusters, generated through the Hal pipeline at 1.2 inflation parameter, in which *E. granulatus* is overrepresented compared to species in Figure 2 and the number of proteins per cluster with *E. granulatus* (highlighted in yellow). To conserve space, non- Eurotiomycetes have been removed, except for *Tuber melanosporum* (Tub). The average number of protein models per cluster per taxon for all taxa in Figure 2 is presented in the last column. Putative annotations are given, where known. Helicases are highlighted in green. Abbreviations as in Figure 6.3.

cluster	Annotation	Elg	Tam	Tas	Mor	Asf	Ask	Asn	Ast	Peb	Pec	Ped	Peo	Tub	Avg.
1200611	helicase	31	0	24	0	0	0	0	0	1	6	0	1	0	3
120075	Zinc finger	24	7	16	17	10	7	1	1	3	6	2	3	0	6
120055	Helicase	23	8	9	7	6	6	6	7	5	6	5	6	5	7
1200322	Unknown	14	7	6	8	4	7	4	7	2	5	0	4	1	4
120067	Mito DNA helicase	13	6	9	6	6	6	7	6	4	8	6	7	3	6
1200703	Unknown	10	1	2	1	1	1	1	2	1	3	2	2	1	2
1200711	Unknown	10	2	1	1	1	2	1	5	0	3	1	1	0	2
1200217	Acetyltransferase	8	8	6	5	6	4	4	7	6	7	1	6	1	4
1200590	Sugar Transporter	8	2	3	1	3	2	4	4	4	4	3	2	0	3
1200599	Unknown/ Phosphorylase	8	5	3	2	3	3	3	4	4	2	2	2	2	3
12001782	Sulfer metabolism regulator SkpA	5	2	2	1	1	1	1	1	1	1	1	1	1	1
12006319	Unknown	4	0	0	0	0	0	0	0	0	0	0	0	0	0
12001149	N2,N2-dimethylguanosine tRNA methyltransferase	3	1	1	2	2	2	2	1	2	2	2	2	2	2
12001691	non classical export protein NCL2	3	1	1	1	1	1	2	1	2	1	1	1	0	1
12001946	cell polarity protein, repeat	3	1	1	1	1	1	1	1	1	1	1	1	2	1
12002140	Serine/Threonine protein Kinase	3	1	2	1	1	1	1	1	1	1	1	1	1	1
12002348	Lipoic acid Synth precursor	3	1	1	1	1	1	1	1	1	1	1	1	1	1
12002863	GTPase activating protein (Evi5), repeat	3	1	1	1	1	1	1	1	1	1	1	1	1	1
12003050	RNA-binding protein	3	1	1	1	1	1	1	1	1	1	1	1	1	1
12003919	Glutamine Synthetase	3	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 6.S4. (Continued)

cluster	Annotation	Elg	Tam	Tas	Mor	Asf	Ask	Asn	Ast	Peb	Pec	Ped	Peo	Tub	Avg.
12004058	endo alpha-1,4 polygalactosaminidase	3	1	1	1	1	2	2	1	2	2	2	2	0	1
12004382	potassium transporter	3	1	1	1	0	1	0	1	0	0	0	0	0	1
12004880	Unknown	3	1	1	1	1	1	1	1	1	1	1	1	0	1
12005399	Unknown no signal peptide, short	3	1	1	0	0	0	1	1	1	0	1	1	0	0
12005403	Radical SAM domain containing protein	3	0	0	0	0	0	0	0	0	0	0	0	0	0
12005406	Unknown	3	1	0	1	0	0	0	0	0	2	0	0	0	0
12006126	Unknown	3	0	1	0	0	0	0	0	0	0	0	0	0	0
12006314	Ubiquitin specific peptidase with Zinc finger binding domain	3	0	0	0	0	0	0	0	0	0	0	0	0	0
12007330	AcetylCoA synth	3	0	0	0	0	0	0	0	0	0	0	0	0	0
12007336	Unknown	3	0	0	0	0	0	0	0	0	0	0	0	0	0
12003973	NUDIX family hydrolase	2	1	1	1	1	1	1	0	1	1	1	1	1	1
12004175	Serine-threonine protein kinase	2	1	1	1	1	1	1	1	1	1	1	1	0	1
12004355	coatomer subunit epsilon	2	1	1	0	1	1	1	1	1	1	1	1	1	1
12004367	Unknown short	2	1	1	1	1	0	1	1	1	1	0	1	1	1
12005543	Metallopeptidase	2	0	0	0	0	1	0	1	0	0	0	0	0	0
12006536	Unknown	2	0	0	0	1	1	1	0	0	0	0	0	0	0
12006827	Unknown	2	0	0	0	0	0	1	0	0	0	0	0	0	0
12006843	Unknown	2	0	0	0	0	0	1	0	0	0	0	0	0	0
12006846	Hyp prot specific to L. bicolor	2	0	0	0	0	0	0	0	0	0	0	0	1	0
12007328	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12007371	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008294	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008295	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6.S4 (Continued)

cluster	Annotation	Elg	Tam	Tas	Mor	Asf	Ask	Asn	Ast	Peb	Pec	Ped	Peo	Tub	Avg.
12008297	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008300	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008303	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008312	Transposable element	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008313	Zn2/Cys6 DNA-binding protein	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008321	reverse transcriptase domain protein	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008325	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008330	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008337	Unknown short	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008340	Unknown short	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008341	DNA breaking-rejoining enzyme, partial	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008342	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12008361	Unknown	2	0	0	0	0	0	0	0	0	0	0	0	0	0

Chapter 7. Conclusion

The 18 described species of *Tolypocladium* that parasitize *Elaphomyces* truffles are the only fungi known to regularly infect and fruit from these hosts. They are descended from other cordyceps-like species that were insect pathogens, and this close relationship between the insect pathogenic and mycoparasitic *Tolypocladium* spp. has been perplexing to scientists for some time.

The results of Chapter 2 establish a taxonomic framework in which there is a single monophyletic genus, *Tolypocladium*, which encompasses a diversity of pathogens of insects and other invertebrates, parasites of *Elaphomyces* fruiting bodies, and saprobic species isolated from soil and on lichens. A number of asexual morphs are known for members of this genus, including the *Tolpocladium* morphology itself, verticillium-like, and chaunopycnis-like, but many species, including most of the parasites of *Elaphomyces*, have no known or associated asexual state. Another impact of this chapter is the providence of a set of six names given to the well-supported monophyletic clades of Ophiocordycipitaceae. This nomenclature will provide a framework for future systematic and taxonomic studies within the family and to the wider research community as a whole.

A large repertoire of secondary metabolite genes and clusters were identified by sequencing the genome of *T. ophioglossoides*. The majority of these have no putative product. Unlike *T. inflatum*, neither the gene, *simA*, which is responsible for the production of cyclosporin A, nor any genes in that cluster were identified in the *T. ophioglossoides* genome. Several peptaibiotic gene clusters were identified. Despite the presence of petaibiotic clusters in *T. inflatum*, they are not orthologous to those found in *T. ophioglossoides*. Many of the adenylation domains of *Tolypocladium* peptaibiotics

deeply coalesce at the common ancestor of the four most derived families, suggesting a complex pattern of duplications and losses likely resulted in the divergent peptaibiotics present in *T. ophioglossoides* and *T. inflatum*. This is in contrast to the peptaibiotic genes in species of the genus *Trichoderma*, which are highly conserved between its species, both in their modular synteny and (for the most part) with respect to their arrangement within the genome.

The RNA-Seq experiment resulted in several findings. First, it resulted in the development of a novel media for simulating the *Elaphomyces* hosts of *Tolypocladium* spp. parasitic to them. Next, it provided support for the hypothesis that PTH11-related GPCRs are involved in host recognition in hypocrealean fungi and that differential expression of these proteins is correlated with growth on different host-containing substrates. During growth in media containing *Elaphomyces*, *T. ophioglossoides* upregulates a Class V Subgroup C chitinase that is phylogenetically related to a chitinase that is upregulated during mycoparasitism in *T. virens*. This suggests that, despite independent origins of mycoparasitism, similar chitinolytic mechanisms may be employed by these distant relatives. This study also identified an adhesin, *Mad1*, to be one of the most highly expressed genes; it is also upregulated on media containing *Elaphomyces*. Of the *T. ophioglossoides* proteins upregulated in *Elaphomyces* peridium, those involved with oxidation and reduction were identified as statistically overrepresented based on GO categories, suggesting that these may be involved in breakdown of host tissue as well.

In Chapter 5, a comparative genomics analysis of four *Tolypocladium* species, provided the first study where insect pathogens are concluded to be the earliest diverging

lineages of the genus with bootstrap support; mycoparasitism is therefore interpreted as the most derived ecological character state. Rapid evolution and speciation in the genus resulted in genes with a high frequency of phylogenetic conflict that is likely the product of incomplete lineage sorting during the early evolution of the genus. Protein clusters specific to the insect pathogenic species, *T. inflatum* and *T. paradoxum*, have a higher percentage of signal peptides, indicating that the mycoparasites may be reduced in secreted proteins that are involved in interaction with insect hosts. It is likely due to these rapid speciation events that large quantitative differences in primary metabolism are not detected, although this does not inform how these genes may be differentially expressed. All sequenced species of *Tolypocladium* have diverse array of secondary metabolites, but they each have a unique profile. *Elaphomyces* parasites do not possess DMATs and have more terpene synthases, whereas *simA* is limited to *T. inflatum* and peptaibols are found in all species except *T. capitatum*.

The genome of *Elaphomyces granulatus* (Chapter 6), the host of mycoparasitic species of *Tolypocladium*, revealed it is most closely related to *Talaromyces* and Trichocomaceae s.s. The genome assembly was difficult, and it remains on many contigs, but gene space is estimated to be mostly complete even though it was sequenced from sporocarp tissue. This is one of the first genomes sequences produced from fungal sporocarp tissue. *Elaphomyces granulatus* is reduced in CAZyme genes, despite a large expansion in genome size, both of which are consistent with what is seen in *T. melanosporum*, the other sequenced ectomycorrhizal ascomycete. The *E. granulatus* microbiome is a complex community that is dominated by bacteria, especially those of Bradyrhizobiaceae. There is evidence suggesting the community may be reduced in

functional capacity. The number of reads per generic classification in the bacterial community did not correlate with ability to assemble long bacterial contigs, suggesting that some genera may have greater species diversity than others present in the community, or that the genomes may be difficult to assemble for other reasons such as genome size and/or GC content.

Future studies could expand upon this dissertation in several ways. With regard to the systematics of Ophiocordycipitaceae, three issues stand out. First, as in studies before (Sung *et al.* 2007, Kepler *et al.* 2012), unambiguous molecular support for the internal nodes of *Ophiocordyceps* is marginal. Two approaches could be taken to resolve this issue: either sampling more taxa or more data within taxa. Both are necessary, but given the fact that the analysis in Chapter 2 added more than 222 new taxa to that clade than previous studies, more taxa may not help resolve this issue most immediately. Genome-scale sampling of taxa within the major clades of *Ophiocordyceps*, however, may provide the depth of necessary to determine these intrageneric relationships. The second intriguing hypothesis put forth in Chapter 2 is the relationship between the two largest nematophagous lineages within the family. As noted in that chapter, sequence data were not available for many of the species described within those two clades, but there are shared characteristics between some species in *Drechmeria* (*i.e. D. harposporioides*) and *Harposporium*. *Purpureocillium*, which is well supported as the most closely related lineage to this clade, also contains species capable of parasitizing nematodes, and a study examining the evolution of nematophagy within the early diverging lineages of Ophiocordycipitaceae would be informative, especially given the economic impact of plant pathogenic nematodes. Finally, the chaunopycnis-like species

within *Tolypocladium* possess an unusual morphology and possibly represent a monophyletic lineage within the genus, and therefore should be further studied within a phylogenetic and morphological context. Of the three described species, their habitats are varied but suggest they are saprobes, although this remains largely unknown. Given the results of Chapter 2 and Chapter 5 (examining intrageneric relationships at a genome scale level), the chaunopycnis-like species of *Tolypocladium* should be targeted for genome-scale sequencing to further elucidate the relationships within the genus and evolution of ecology within the clade.

An obvious continuation of the work done in Chapter 3 would be the production of metabolomic data to support the presence of numerous secondary metabolite clusters in the *T. ophioglossoides* genome. Description of the peptaibiotic structures predicted from the results of Chapter 3 could link genome to metabolome. This was in fact pursued during the course of this dissertation, but conclusive data on the structure of the peptaibiotics described from *T. ophioglossoides* were elusive, probably due to the unique composition of these compounds and the small amounts of these compounds produced under the laboratory conditions studied. These products are likely to be bioactive; from a natural products and drug discovery point of view, species like *T. ophioglossoides*, which possess so many novel gene clusters, are good candidates to discover multiple, new compounds. Only briefly mentioned in the dissertation is the pan-Northern hemisphere distribution of *T. ophioglossoides*. Five gene molecular phylogenies show some amount of diversity among isolates collected from different parts of the world. Future population genomics or phylogeographic studies examining *T. ophioglossoides* diversity and

distribution would be interesting to determine if there is any evidence for cryptic speciation.

The RNA-Seq study could be augmented in several ways. The simulation of the host, by the addition of *Elaphomyces* tissue into media was a good first step, but likely misses much of the interaction between host and parasite *in vivo*. Future studies would seek to obtain axenic cultures of *Elaphomyces* or its close relative and model organsism, *Aspergillus*, and performing confrontation assays to look for signs of antagonism, which if present could be examined using scanning electron microscopy (SEM) of the hyphal interactions, RNA sampling of both the host and parasite could further resolve their responses to each other, or metabolomics to identify what, if any, compounds are being produced by *Tolypocladium* or the host. It would also be useful to study confrontation assays and RNA-Seq analysis of other mycoparasitic species of *Tolypocladium* that do not grow as readily in culture as *T. ophioglossoides* (*e.g.*, the strain of *T. capitatum* sequenced in this dissertation), because, as discussed in Chapter 1, the ability to readily and quickly grow in culture suggests that some members of this genus may not be obligate parasites, whereas the other truffle parasites may be.

The phylogenetic results of Chapter 5 lead the way to a new set of questions about the intrageneric relationships within this clade. Specifically, is cicada pathogenesis the ancestral ecology of the genus? Given the results of Chapter 2, which had a different topology for the genus, how are the chaunopycnis-like *Tolypocladium* spp. related to the rest of the genus? Were they a part of this rapid radiation? Ideally, more genome sequences from cicada pathogens (*T. ineogoense* and *T. toriharamontanum*), soil-isolated saprobes (*e.g. T. album*, *T. geodes*), and more (rarely collected) *Elaphomyces* parasites

(*e.g. T. rouxii*, *T. fractum*) could help tease apart the evolution of host association within this small, but diverse genus. Due to the large amount of incomplete lineage sorting, traditional approaches to comparative genomics (including examining quantitative differences in genome or proteome size, overall expansions in CAZymes or proteases) may not be as fruitful as in other systems. Targeted approaches, examining specific groups of proteins or clusters specific to species with certain ecologies (as taken in this dissertation) may reveal more about the genes essential to a particular ecology and the changes that take place within the context of host-jumping.

The genome sequence of *E. granulatus* from sporocarp has supported and produced many hypotheses about the evolution of ectomycorrhizae in Ascomycota. Future studies would seek to produce a better version of the genome, possibly based on material grown in pure culture. This could also help in parsing the 'non-target' community sequences out of the metagenomic data collected. Future studies examining the metagenomic community associated with *Elaphomyces* sporocarps, should also take corresponding soil samples to analyze what portion of the sporocarp community is derived from surrounding soil community and what taxa, if any, are candidates for vertical transmission within the fungus. Fluorescence microscopy could also be employed to identify the precise location of target members of the microbial community within the peridium tissue.

Understanding of the biology and systematics of organisms greatly increases the utility of tools such as genomics and transcriptomics in addressing evolutionary biology questions. This dissertation was guided by this principal while still seeking to advance the technologies available to answer the questions that have arisen. Hopefully, this

dissertation can inform future work on both *Tolypocladium* and *Elaphomyces* and their interactions with each other.

References

Kepler RM, Sung GH, Harada Y, Tanaka K, Tanaka E, Hosoya T, Bischoff JF, Spatafora JW (2012) Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilaginoidea* (Clavicipitaceae). *American Journal of Botany* **99**: 552-561.

Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW (2007) Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Studies in Mycology* **57**: 5-59.