UNIVERSITY OF CALIFORNIA
RIVERSIDE

Investigating the Evolution of Environmental and Biotic Interactions in Basal Fungal Lineages Through Comparative Genomics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

Steven Robert Ahrendt

August 2015

Dissertation Committee:

Dr. Jason E. Stajich , Chairperson
Dr. Chia-en A. Chang
Dr. Katherine A. Borkovich

Copyright by
Steven Robert Ahrendt
2015

The Dissertation of Steven Robert Ahrendt is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

There are many people to whom I am deeply indebted for their contribution to this dissertation and to whom I would like to extend my appreciation. First and foremost, to my research advisor and mentor Dr. Jason Stajich, who allowed me both the freedom and guidance necessary to see this research to completion. To my co-advisor, Dr. Chia-en Angelina Chang, whose input and assistance helped provide a fresh perspective on aspects of this research. And to Dr. Katherine Borkovich, for her support, guidance, and technical expertise. To my undergraduate internship research advisor Dr. Jamie Foster at the Kennedy Space Center and her graduate student Jen, without whose inspiration, advice, and continued support I would never have pursued graduate school at all. I would like to extend a huge thank you to the members of the Stajich and Chang labs, and especially to the undergraduate students I mentored for their assistance; also to the members of the Borkovich lab, whose equipment and expertise I borrowed early and often. To my parents, for their constant support and continued patience. And finally to Kat, who joined me on the journey and stayed with me until the end, for which I am eternally grateful.

The text of this dissertation, in part or in full, is a reprint of the material as is appears in "Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia", *Current Biology*, 2013. The co-author, Dr. Jason Stajich, listed in that publication directed and supervised the research which forms the basis for this dissertation.

ABSTRACT OF THE DISSERTATION

Investigating the Evolution of Environmental and Biotic Interactions in Basal Fungal Lineages Through Comparative Genomics

by

Steven Robert Ahrendt

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, August 2015
Dr. Jason E. Stajich , Chairperson

Species belonging to the basal fungal lineages (Blastocladiomycota, Chytridiomycota, Cryptomycota, and Neocallimastigomycota) reproduce via motile zoospores and are found in both aquatic and terrestrial environments. These organisms, colloquially referred to as "chytrids", are traditionally understudied, despite being active decomposers, parasites, and symbionts with other organisms in the ecosystem. This dissertation research uses a comparative genomics approach to answer questions about these fungi and their interactions with their environment and other fungi. Chapters 2 and 3 examine the details of putative rhodopsin-mediated photoreception in basal lineages using structural mechanics and genome-wide gain-loss analyses. Chapter 4 provides a transcriptome analysis of one member of the genus *Coelomomyces*, the only known entomopathic chytrid genus. Chapter 5 describes surprising observations regarding competitive and inhibitory behavior in one member of the Chytridiomycota. Finally, Appendix A briefly looks at gain-loss analysis of molecular aspects of the evolutionary transition from aquatic motile single cells to terrestrial multicellullar organisms.

Contents

List of Figures viii

List of Tables ix

1 Introduction to the Basal Fungal lineages 1
- 1.1 Overview of fungal phylogenetics and the importance of the basal lineages 1
- 1.2 History of bioinformatic resources for basal fungi 5
- 1.3 Hypotheses and Objectives 8

2 Structural characteristics of opsin-like proteins found in basal fungal lineages 15
- 2.1 Introduction 15
- 2.2 Methods 19
- 2.3 Results 22
- 2.4 Discussion 26

3 Rhodopsin related signaling pathways in basal fungi 35
- 3.1 Introduction 35
- 3.2 Methods 39
- 3.3 Results 45
- 3.4 Discussion 49

4 Transcriptome analysis of the anopholean pathogenic fungus *Coelomomyces lativittatus* 65
- 4.1 Introduction 65
- 4.2 Methods 68
- 4.3 Results 73
- 4.4 Discussion 83

5 Growth suppression properties of the Chytridiomycete *Homolaphlyctis polyrhiza* JEL142 102
- 5.1 Introduction 102
- 5.2 Methods 105
- 5.3 Results 111
- 5.4 Discussion 116

6 Conclusions 126

Bibliography **131**

A Comparative genomics analysis of flagellar motility **154**
A.1 Introduction 154
A.2 Results and Discussion 155
A.3 Methods 156

B Datasets and scripts **158**

List of Figures

2.1 Structural details of the *S. punctatus* chytriopsin homology model. . . . 30
2.2 Docking screen for *Bd* and *Am* models 31
2.3 MD summary plots 32

3.1 Photosensory survey 52
3.2 Gα MSA 52
3.3 Gα tree, group 1 53
3.4 Gα tree, group 2 53
3.5 Gα tree, group 3 54
3.6 Gα tree, group 4 55
3.7 Gβ WD40 repeats 56
3.8 Gβ tree 57
3.9 Gγ MSA 58
3.10 Gγ tree 59
3.11 RGS proteins 59
3.12 PDEα/β tree 60
3.13 *P. pastoris* heterologous opsin expression 61

4.1 *C. lat* transcriptome GO term distribution 90
4.2 Opsin-GC fusion proteins 91
4.3 β-carotene enzyme presence / absence 91
4.4 PF00112 RAxML tree 92
4.5 PF00089 RAxML tree 93
4.6 20HE alignment 94
4.7 20HE and Nuclear Receptor RAxML tree 95
4.8 BCMO1 history 96
4.9 PAS alignment of WC1 homologs 97

5.1 Inhibition appears unique to *Hp* among other chytrids 119
5.2 Inhibition is not specific to *N. crassa* 120
5.3 Conditioned media inhibits *N. crassa* growth 120
5.4 *Hp* is more tolerant of environmental stresses than *Bd* 121
5.5 *N. crassa* displays no avoidance behavior of solid agar blocks. 122
5.6 VM dilution series demonstrates efficacy at various concentrations . . . 122
5.7 *S. cerevisiae* and *E. coli* are inhibited by *Hp* filtrate 123

A.1 Heatmap cluster analysis of flagellar proteins from *Naegleria gruberi* . . 157

List of Tables

2.1 Quality metrics for rhodopsin structures 33
2.2 Backbone RMSD measurements for rhodopsin structures 33
2.3 Locations of key residues in rhodopsin structural motifs 34
2.4 Interaction energies for covalent docking 34

3.1 Gα subunit comparison 62
3.2 Gβ subunit comparison. 63
3.3 Gγ subunit comparison. 64

4.1 Top 20 PFAM domain counts 98
4.2 orthoMCL clusters for transmembrane proteins 99
4.3 *C. lativittatus* photosensing proteins 99
4.4 FAADB predictions for chytrid proteins 100
4.5 BLASTP hits for *C. lat* m.10080 101

5.1 List of *N. crassa* mutants screened for sensitivity 124
5.2 Chytrid secretome predictions 125
5.3 Putative *Hp* terpene synthases 125

B.1 List of proteomes used in comparative analyses. 162
B.2 PDB IDs and descriptions of structures used in comparative analyses . . 163

Chapter 1

Introduction to the Basal Fungal lineages

1.1 Overview of fungal phylogenetics and the importance of the basal lineages

The Fungi are one of the major kingdoms of eukaryotic life on Earth. Various studies have attempted to estimate the date of the emergence of the Kingdom Fungi [1], when it diverged from the metazoan lineages. These studies place this occurance at approximately 1 billion years ago, with a range of around ±500 MYA: 600 MYA [2], 965 MYA [3], and 1.6 BYA [4]. These loose approximations are based on correlation between evolutionary events in fungi and in other organisms, and are under continued re-evaluation and refinement [5].

A comprehensive review of a collection of 21st century phylogenetic studies [6] proposes that the Fungal Kingdom comprises seven phyla: Microsporidia, Chytridiomycota, Blastocladiomycota, Neocallimastigomycota, Glomeromycota, Basidiomycota, and

Ascomycota, with the most recent inclusion of the phylum Cryptomycota [7].

The most recently diverged fungal phyla are the Ascomycota and Basidiomycota, and together make up the subkingdom Dikarya. This subkingdom is so named because its members undergo cell fusion (plasmogamy) without nuclear fusion (karyogamy) during sexual development, resulting in cells with nuclei from individual parents ("dikaryons"). These organisms are muticellular, with terrestrial habitats, sexual and asexual life cycle components, and filamentous growth structures. Collectively, the Dikarya is the most widely studied group and is home to several model and non-model organisms of research interest, including the model filamentous Ascomycete *Neurospora crassa*, the economically critical *Saccharomyces cerevisiae*, and the numerous medically relevant *Aspergillus* spp.

Going further back in time is the phylum Glomeromycota. This group contains mycorrhizal fungi (arbuscular and ento-) which associate with approximately 90% of all plant species and are thus of great ecological importance. This phylum also contains four subphyla *incertae sedis*: Mucormycotina, Entomophthoromycotina, Zoopagomcotina, and Kickxellomycotina [8]. These subphyla were previously classified in the phylum Zygomycota and contain primarily terrestrial fungi of various medical and industrial research interest. It is important to point out here that the definitive phylogenetic relationship between the Glomeromycota and Zygomycota *incertae sedis* lineages is unresolved and is a current focus of research [6]. Therefore in this text, the nonflagellated members of the Glomeromycota, Mucormycotina, Entomophthoromycotina, Zoopagomycotina, and Kickxellomycotina will, for convenience, be referred to collectively as Zygomycota.

Closest to the fungal-animal evolutionary divergence are the basal fungal lineages: the Microsporidia, Cryptomycota, Chytridiomycota, Blastocladiomycota, and Neocallimastigomycota. It is these lineages, particularly the Blastocladiomycota and

Chytridiomycota, on which this dissertation will primarily be regarding. These groups are sometimes collectively referred to as "chytrids", although this is not to be confused with the formal phylum Chytridiomycota. Broadly speaking, these organisms have asexual life cycles which progress through development as motile, flagellated zoospores, followed by sessile, non-flagellated, spore-producing sporangia. During the motile stage, the zoospore seeks an appropriate environmental substrate, encysts upon it, retracts its flagellum, and develops a cell wall. Many species will stay dormant in this stage as a thick-walled resting spore, and only develop into a thin-walled zoosporangia after a certain time period [9]. Other chytrid species will instead progress directly to the zoosporangia stage, undergo several rounds of mitotic cell division, and ultimately produces and releases hundreds of new zoospores [9].

While generally being described as asexual, certain species within the Blastocladiomycota, such as *Allomyces reticulatus* and *Coelomomyces punctatus*, have demonstrated sexual reproductive cycles utilizing zoospores of different mating types [10].

The basal fungal lineages are characterized as true Fungi and distinct from other water molds, like Oomycetes, and fall sister to both Metazoan lineages as well as the other fungal lineages (Zygomycota, Ascomycota, and Basidiomycota). While these lineages only represent less than 2% of all described fungi [11], they serve as a unique system in which to infer characteristics presumed to have been present in the fungal-animal common ancestor.

Additionally, basal fungal lineages are presumed to have a nearly cosmopolitan distribution [12]. Members of these lineages are found in widespread environments [13], and in some biomes represent the dominant member of the soil fungal community [14]. Chytrids fulfill nearly all varieties of ecological niches, primarily decomposition in

terrestrial and aquatic environments, but also including pathogenic interactions with a wide variety of hosts: arbuscular mycorrhizal fungi (*Spizellomyces punctatum* (Chytridiomycota) [15]), insects (*Coelomomyces psorophorae* (Blastocladiomycota) [16]), plants (*Olpidium brassicae* (Chytridiomycota) [17]), vertebrates (*Batrachochytrium dendrobatidis* (Chytridiomycota) [18]), nematodes (*Catenaria anguillulae* (Blastocladiomycota) [19]), algae (*Zygorhizidium plantonicum* [20]), and even intracellular symbioses with other chytrids (*Rozella allomyces* (Cryptomycota) [21]). This distribution of life styles speaks to the vast biological challenges they must face and therefore suggests a number of novel mechanisms which have yet to be fully studied and explored.

Attempts at formal description of early-diverging fungi, based primarily on collection and observation, began as early as 1858 and proceeded through the latter half of 19th century with pioneering work of researchers such as Schroter, Fischer, Zopf, Lowenthal, Nowakowski, and Woronin [22]. A primarily systematic approach allowed for the establishment of (among others) the order Chytridiales, defined broadly as lacking mycelium and having an unknown sexual cycle, and the order Blastocladiales, defined as having mycelium and a sexual reproductive component.

Significant microscopy work was carried out on chytrid species as early as 1953 by William Koch. This allowed for the discussion of zoosporic ultrastructure characters [23, 24] and the description of 6 major cell morphologies [25], which demonstrated a high degree of structural diversity among the zoosporic lineages.

With the advent of nucleic acid-based molecular phylogenetic techniques at the turn of the 21st century, several researchers addressed revisions of fungal phylogenetics and attempted to reconcile traditional ultrastructure-based phylogenetic ideas with these modern techniques. Whereas earlier, characters such as zoospore discharge, thallus development, and ultrastructural features were used to place basal lineages, modern

SSU rDNA techniques corroborated these placements. Modern molecular phylogenetics also supported the monophyletic nature of the Fungi and metazoan lineages, as well as the inter-kingdom relationships between the fungal phyla. This and other work helped firmly establish the Chytridiomycota as one of the four major fungal phyla alongside the Zygomycota, Basidiomycota, and Ascomycota [26, 27, 28, 29, 9, 6].

In 1998, the Chytridiomycete *Batrachochytrium dendrobatidis* emerged as a global pathogen and the accepted causative agent of worldwide amphibian decline [30]. This emergence renewed interest in chytrids as a general system of study as emphasis was placed on understanding the distribution, diversity, and pathogenicity of *Bd*.

Despite widespread distribution in both geography and ecology, and the existence of species pathogenic in a wide range of hosts, chytrids remain understudied as a whole, driven in part because few species are of substantial economic importance [12, 31] as well as moderate difficulty in collection and culturing methods. However, a greater molecular understanding of the phylogeny of the basal fungi is a current subject of study. Within the past decade, formal descriptions for two new genera, the Irineochytrium (Chytridiomycota; Chytridiales) [32] and Fayochytriomyces (Chytridiomycota; Chytridiales) [33], and one new order, the Loulomycetales (Chytridiomycota) [34] have been published.

1.2 History of bioinformatic resources for basal fungi

Chytrid bioinformatic resources represent a small but growing proportion of all available fungal resources due to the increasing accessibility of high throughput sequencing capability. The first available such resource was an expressed sequence tag

(EST) dataset published in 2005 for the Blastocladiomycete *Blastocladiella emersonii* [35]. This collection of 16,984 high-quality ESTs provided a first approach to understanding gene complexity in chytrids. In 2006, a draft assembly for the genome of *B. dendrobatidis* strain JEL423 was made publically available through the Broad Institute Fungal Genome Initiative (`http://www.broadinstitute.org`), representing the first whole genome assembly for any chytrid. The resulting assembly, using Sanger sequencing reads, is 23.72 Mb and represents 7.4X coverage of this diploid strain. In 2008, a second draft genome was released for *B. dendrobatidis* strain JAM81 through the Joint Genome Institute [36]. This assembly is 24.3 Mb and represents 8.74X coverage.

As part of the push to understand the Origins of Multicellularity [37], the genomes for *Allomyces macrogynus* (Blastocladiomycete) and the exclusively terrestrial *Spizellomyces punctatus* (Chytridiomycete) were sequenced by the Broad Institute in 2009.

In 2011, the non-pathogenic species *Homolaphlyctis polyrhiza* strain JEL142, the closest identified relative to *Bd* was sequenced for comparision to try to identify aspects of pathogenicity in *Bd* [38]. This aquatic chytrid has been isolated only once and while it had been used in previous phylogenetic [31, 29, 39] studies was only recently provided with a formal name [40]. The resulting assembly generated from 454 sequencing technology was 26.7 Mb (haploid) from 11.2X coverage. In 2014, with the help of a postdoctoral researcher in the Stajich Lab, Dr. Peng Liu, I generated Illumina sequencing libraries for *H. polyrhiza* and assisted in assessing the assembly and annotation with Dr. Stajich; the results of which are described in Chapter 5.

Gonapodya prolifera (Monoblepharidomycota) is an aquatic fungus with both sexual and asexual reproductive schemes, and encompassing both hyphal and zoosporic growth stages. In the environment, *G. prolifera* is an active degrader of plant material

[41]. A draft genome assembly was made available in 2011 through JGI, with the goal of identifying potentially novel degredation-related enzymes for biofuels applications [42].

Catenaria anguillulae (Blastocladiomycota) is a facultative parasite of nematodes [19]. A draft genome was made available in 2010 by the JGI and was the second Blastocladiomycete genome (after *A. macrogynus*). Genomic resources for *C. anguillulae* allow for research into monitoring and potential remediation strategies as the nematodes upon which it parasitizes are themselves parasites of agriculturally important crops.

Members of the Neocallimastigomycota lineage, first isolated and described in 1975, are found in the anaerobic environment of mammalian rumen [43]. These fungi are uniquely adapted to degradation of the high fiber content of the typical diets of cattle and sheep. Thus they are important models for potential manipulation to not only improve digestion in these livestock sources [44] but also potential biofuels applications [45, 46]. *Piromyces* and *Orpinomyces* are two members of this group and were sequenced in 2011 and 2013, respectively, in the hopes that understanding the genomic content would provide starting points for these applications.

The genome of Cryptomycete *Rozella allomycis*, the intracellular parasite of *Allomyces*, was sequenced in 2013 [47] and the analysis used to propose a unification of the Cryptomycota and Microsporidian lineages. I assisted in this work by analysing a comparative search of flagellar-associated proteins and conservation across the fungal lineages. A summary of this contribution is provided in Appendix A.

In 2014 the first transcriptome of the mosquito pathogen *Coelomomyces lativitattus* (*Cl*) was generated by isolating RNA from gametes emerging from copepods. The RNA extraction and Illumina library preparation was performed by Rob Hice, a researcher in Dr. Brian Federici's lab at UCR. The sequencing was performed at the UCR IIGB Genomics core. The resulting sequence data was assembled and annotated

using scripts provided by Dr. Jason Stajich, and my analysis is described in Chapter 4.

The near future of bioinformatics resources for the basal lineages is promising due to the efforts of the 1000 Fungal Genomes Project (`http://1000.fungalgenomes.org`) with plans to sequence the genomes of additional fungi from the Chytridiomycota and Blastocladiomycota lineages. A total of 16 Chytridiomycota species have been nominated for sequencing, including *Operculomyces laminatus* JEL223, *Rhizoclosmatium hyalinus* JEL800, and *Obelidium mucronatum* JEL802, which are being prepared by the Stajich lab and were used in the experiments described in Chapter 5. *Coelomomyces lativittatus*, the mosquito pathogen for which a transcriptome analysis is presented in Chapter 4 is being prepared for sequencing by Dr. Brian Federicis' lab at UCR and will be the third Blstocladiomycte genome produced after *A. macrogynus* and *C. anguillulae*.

1.3 Hypotheses and Objectives

Rapid advances in the feasability of genome sequencing have yielded and will continue to yield an increasing number of fungal genomes, especially those in the early-diverging lineages, for comparative analyses. While incorporating comparisons to already well-characterized representative fungal groups, this dissertation work gives specific focus to members of the early-diverging lineages, and in particular to the ones for which genomic resources are available and which have obvious economic or ecological importance: the amphibian pathogen *Batrachochytrium dendrobatidis* and its closest, non-pathogenic relative, *Homolaphlyctis polyrhiza*; the saprobic *Spizellomyces punctatus*; and the aquatic Blastocladiomycete *Allomyces macrogynus* and related *Catenaria anguillulae*. This thesis is presented in four chapters comprising three aspects of the

biology of members of the basal fungal lineages. The primary focus of this work is on sensing and interpretation of biotic and environmental signals by these organisms.

Mechanics and evolution of Fungal rhodopsin-based photosensing in the basal lineages

During the course of a given day, an organism experiences a multitude of environmental stimuli, with light being one of the most prominent. The biochemical ability to appropriately process and respond to these signals is an incredibly complex and involved task, and understanding the underlying mechanisms of these responses is an ongoing scientific challenge.

Previous work has shown that some of the basal fungi are phototaxic (see [48] and [49]), however the full extent of photosensing in zoosporic fungi has not been fully explored. A recent review of fungal photobiology suggests a sporadic distribution of photosensory proteins among the non-flagellated fungal lineages (i.e., Zygomycota and Dikarya), with little emphasis placed on the basal lineages [50]. There are many classes of photoreceptor proteins in fungi capable of producing a cellular response from an environmental light signal, all of which have different mechanisms of action and specializations: phytochromes, cryptochromes, the white-collar complex (WCC), and opsins [50]. In plants, phytochromes function as day-night sensors to regulate the circadian rhythm and flowering response. This is accomplished through a conformational shift between the red and far-red sensitive forms of the protein structure [51]. While relatively little is known about fungal phytochrome function, research on *A. nidulans* suggests that the phytochrome protein is a member of an elaborate complex with regulatory functions involved with the asexual-sexual transition and secondary metabolite biosynthesis [50].

Cryptochromes, found predominantly in plants, animals, and insects, are blue-light sensitive proteins involved in circadian rhythm regulation and light activated DNA damage repair [50]. Additional evidence suggests cryptochrome proteins play a role in mediating the phototactic behavior of sponge larvae [52].

First studied in the model filamentous Ascomycete fungus *Neurospora crassa*, the well-characterized white-collar complex assembles as a heterodimer comprising White-collar 1 and 2 proteins. This complex functions to sense blue and near UV wavelengths, and, when active, directly interacts with DNA to regulate the circadian clock machinery, sporulation, pigmentation, and phototropism [53, 54, 55].

The largest family of membrane receptors by far is that of the seven-transmembrane (7TM) receptors, comprising upwards of 800 genes [56]. This family includes various receptors for a wide range of ligands, including hormones, neurotransmitters, odorants, and photons. While there are three distinct subfamilies (A, B, and C), they share very little sequence similarity. Opsins, examples of which can be found in bacteria, archaea, and eukaryotes, are part of the largest family of 7TM proteins. One subclass of opsin, the Type 2 rhodopsins, are G protein-coupled receptor (GPCR) proteins which function via photoisomerization of a covalently bound retinyledene chromophore, typically 11-*cis*-retinal [57].

The retinal chromophore utilized in Type 2 rhodopsin-mediated photoreception is biosynthesized from β-carotene [58]. Photoisomerization of this chromophore results in a conformational shift to the all-*trans* isomer [59] and activation of the coupled heterotrimeric G protein. A comparative analysis of auxillary proteins in basal fungi involved in this downstream signalling cascade is given in Chapter 3, and a description of findings dealing with structural and functional analyses of putative homologs of Type 2 rhodopsin in several species of basal fungi is provided in Chapter 2.

Biosynthesis of β-carotene begins with phytoene cyclase converting two molecules of geranylgeranyl pyrophosphate to one molecule of phytoene. Phytoene desaturase then acts in a five-step pathway to convert phytoene into lycopene [60]. Lycopene cyclase finally acts to convert lycopene to β-carotene [61]. Subsequently, two different cleavage enzymes can potentially act on β-carotene. The enzyme β,β-carotene 15,15'-monooxygenase 1 (BCMO1) cleaves β-carotene into two all-*trans*-retinal molecules, and is considered a key enzyme for retinoid metabolism [62]. The structurally related enzyme β,β-carotene 9',10'-dioxygenase (BCDO2) also acts on β-carotene to produce β-apo-10'-carotenal and β-ionone, however its physiological role is less well-characterized [63]. Comparative analysis and discussion of retinal biosynthesis enzymes is provided in Chapter 4.

Towards the development of bioinformatic resources for entomopathogenic Blastocladiomycete *Coelomomyces lativittatus*

While fungi which invade insects have been observed since antiquity (by some estimates, approximately 900 AD in Japan), the exact nature of the fungal-insect relationship was not cleanly determined until around the 1880s [64]. After this relationship was established, research during the following several decades primarily focused on development of applications related to control of agricultural pests. During the mid-20th century, interest in these sorts of applied pest control strategies waned, but taxonomic knowledge greatly increased. In the late 20th century renewed interest in alternative pest control strategies picked up, and due to enhanced technology (microbial, genetic, genomic techniques) this is ongoing [64].

In basal lineages, members of the genus *Coelomomyces* in the Blastocladiomy-

cota are the only known chytrid entomopathogens. There are more than 70 described species of *Coelomomyces*, though the true diversity is estimated be be several hundred [65]. While initially studied as a promising avenue for mosquito control, and an alternative to traditional pesticides, difficulties in culturing *Coelomomyces* have lead to a decline in its research. However its specific host range and continued search for pesticide alternatives have allowed it to persist as an interesting avenue of research.

Chapter 4 presents a preliminary analysis of transcriptome data obtained from *C. lativittatus*. This analysis serves two purposes. First, it lays the groundwork for future RNASeq and proteomic studies of this organism. And secondarily, it attempts to assign molecular detail to previously published observational research about certain biochemical mechanisms (eg β-carotene production and photoreception).

Competition-based secondary metabolism and anti-fungal properties of *Homolaphlyctis polyrhiza*

Interactions between microorganisms are facilitated by biological signals. These include proteins, small molecules, and various chemical compounds, either bound to the cell surface or secreted into the environment. Many of these compounds can be classified as secondary metabolites: chemicals not required for growth or development of the organism.

Resource competition likely plays a role in the evolution of natural antifungal production [66]. Secretion by an organism in a resource-limited environment of secondary metabolites which also happen to negatively impact neighboring organisms would confer a selective advantage upon the producer.

Comparative genomic analyses have identified a host of degradation enzymes

in basal fungi, suggesting saprotrophic and sometimes pathogenic associations with other organisms. There are few explored examples of secreted or secondary metabolite molecules produced by any of the zoosporic fungi.

Secondary metabolite production as it applies to basal fungi is discussed in more detail in Chapter 5 using the non-pathogenic Chytridiomycete *Homolaphlyctis polyrhiza* JEL142. In this chapter, I address three major questions regarding an initial observation I made in the Stajich lab of *Hp* inhibition of the vegetative hyphal growth of *N. crassa* via an unknown secreted compound. Namely, "Is Hp unique among the chytrids in this behavior?", "Is this behavior specific to *N. crassa*?", and "What is the underlying biochemical mechanism by which this behavior is accomplished?". These questions are addressed using observational assays with the sporangia of related Chytridiomycetes, and probing the breadth of non-Chytridiomycete fungi whose growth is susceptible to *Hp*, encompassing Ascomycete, Basidiomycete, and Zygomycete species, and including both temperature and proteinase screens. Finally, to better explore *Hp* gene content, I produced an improved genome assembly and annotation, by assisting Dr. Peng Liu and Dr. Jason Stajich in the collection of fungal material and the assembly and annotation of the resulting genome sequence.

Eukaryotic Flagellar motility

One of the defining characteristics of the early-diverging fungal lineages is the presence of a posterior flagellum, which is used by the zoospores for motility [23]. The chytrid flagellar apparatus is composed of the flagellar stalk (axoneme), the kinetosome (basal body), and the rootlet system [67]. Microscopy analyses from Koch and others [23] describe nine fibril doublets surrounding a paired central core, which is character-

istic of the "9+2" arrangement of microtubule stalks found in other eukaryotic flagella, such as those of human and *Ciona intestinalis* sperm cells [68], and *Chlamydomonas reinhardtii* [69]. The use of these morphological characteristics as phylogenetic markers is supported by modern nucleic acid-based phylogenetic techniques [31].

During the course of fungal evolution, there was a transition from flagellated motile aquatic single celled organisms to terrestrial multicellular organisms [1, 11]. There is support for anywhere from a single flagellar loss event [70] to at least four different such events [29] prior to the divergence of the Zygomycota.

The chytrid flagellum is the primary method of zoospore motility. In most cases, the chytrid flagellum exists as a posteriorly oriented appendage, with a few exceptions. The zoospores of the Neocallimastigomycota lineages, species most commonly found in the anaerobic environment of the mammalian rumen, are posteriorly multiflagellated [44]. In the Blastocladiomycota, *Coelomomyces* species are biflagellate during a part of their life cycle after the uniflagellate gametes of opposing mating types fuse to form a biflagellate zygote [71].

A comparative genomic study of the chytrid flagellar apparatus is presented in Appendix A. Included is a collection of genes which serve as a "core chytrid" flagellar geneset, which may prove useful in future assessments of chytrid and other basal fungal genomes.

Chapter 2

Structural characteristics of opsin-like proteins found in basal fungal lineages

2.1 Introduction

During the course of a given day, an organism experiences a multitude of environmental stimuli, including chemicals, gravity, the Earth's magnetic field, pressure, and light. The biochemical ability to appropriately process and respond to these signals is an incredibly complex and involved task, and understanding the underlying mechanisms of these responses is an ongoing scientific challenge.

The presence or absence of light is perhaps one of the easiest sources of stimuli to comprehend and observe. The sun and rotation of the planet has had such a profound influence on the development of life that it comes as no surprise to find some form of photoreception in every major lineage on the planet. The widespread occurrence of such

an ability, however varied in its implementation, speaks to its importance during the earliest stages of development of life.

In Fungi, there are several classes of proteins capable of photoreception, all of which have different mechanisms of action and specializations. These include phytochrome, cryptochrome, white-collar, and opsin [50]. In plants, phytochromes function as day-night sensors to regulate the circadian rhythm and flowering response. This is accomplished through a conformational shift between the red and far-red sensitive forms of the protein structure [51]. While relatively little is known about fungal phytochrome function, research on *A. nidulans* suggests that the phytochrome protein is a member of an elaborate complex with regulatory functions involved with the asexual-sexual transition and secondary metabolite biosynthesis [50]. The white-collar complex (WCC), on the other hand, is very well characterized in Fungi. First studied in the model filamentous Ascomycete fungus *Neurospora crassa*, WCC functions as a heterodimer comprising White-collar 1 and 2 proteins to sense blue and near UV wavelengths, and, when active, directly interacts with DNA to regulate the circadian clock machinery, sporulation, pigmentation, and phototropism [53, 54, 55]. Cryptochromes are photoreceptors which belong to a large group of flavoproteins. Initial observations of blue-light sensitive photoreception in plants, without concurrent description of the responsible photoreceptor protein, led to the name "cryptochrome": because of their "cryptic" nature [72]. These proteins can be found in plants, animals, and insects, and are involved with circadian rhythm regulation and light activated DNA damage repair [50].

Rhodopsin is a broadly defined term used to describe a large class of seven-transmembrane proteins which use retinylidene compounds for photoreception. This class can be subdivided into two types based on sequence similarity and function, despite similarities in structure (ie seven helical transmembrane domains) and mechanism

of activation (ie photoisomerization of a retinaldehyde chromophore) [73].

The ion transporter rhodopsins ("Type 1") are activated by the photoisomerization of all-*trans*-retinal to 13-*cis*-retinal. These function as membrane channels and are typically used for light-driven membrane depolarization via proton or chloride ion pumping. Examples of this group can be found in bacteria, archaea, and eukaryotes, and include the bacterial sensory rhodopsins, channelrhodopsins, bacteriorhodopsins, halorhodopsins, and proteorhodopsins.

The G-protein coupled receptor (GPCR) rhodopsins ("Type 2") are activated by the photoisomerization of 11-*cis*-retinal to all-*trans*-retinal. These function as visual receptors, and are the largest class of an even larger GPCR superfamily found only in eukaryotes. The general class of photosensitive GPCRs in animals are often referred to as "opsins", with the visual opsins being one of the distinct subfamilies and known as "rhodopsins". In animals, the other subfamiles are the melanopsins, peropsins, neuropsins, and encephalopsins. The exact nature of the evolutionary relationship between the Type 1 and Type 2 rhodopsins has not been clearly established and is currently the subject of discussion [74, 75].

Nonetheless, the rhodopsin pigment of both types is generated when the retinaldehyde chromophore is covalently joined to an opsin apoprotein via a Schiff-base linkage to a conserved lysine residue. While 11-*cis*-retinal is the most common chromophore observed in vertebrates and invertebrates, others are found elsewhere in nature. For example, 3,4-dehydroretinal is observed in fish, amphibia, and reptiles. Switching between the 11-*cis* and 3,4-dehydro- chromophores can be employed as a light adaptation strategy in certain freshwater fish [75]. 3-hydroxyretinal is found in insects, while 4-hydroxyretinal is observed in the firefly squid. In addition to the 11-*cis* conformation, retinal can adopt a number of different isomers, including all-*trans*, 13-*cis*, and 9-*cis*

[75]. Molecular mechanics simulations suggest that the 11-*cis*-retinal isomer has been selected for evolutionarily as the optimal chromophore due to the energetic stability of the resulting chromophore-opsin construct [76].

Previous work has demonstrated that certain basal fungal species are phototaxic. For example, the phototactic capabilities of the marine fungus *Rhizophydium littoreum* were quantified in 1987. This fungus demonstrated responses to light at a variety of wavelenths, with the most rapid response occuring at 400 nm. While the evidence strongly suggests a blue-light sensitive photoreceptor, the researchers were unable to specifically characterize the active photoreceptor [49].

A decade later, Saranak and Foster described their work on the phototactic capabilities of the Blastocladiomycete *Allomyces reticulatus* [48]. This fungus has a visible, red-pigmented eyespot in which the photosensitive proteins are localized. Careful analysis determined that action spectrum of the phototactic zoospores peaks at 536$\pm$ 4 nm, simlar to that of the human green-sensitive cone. Furthermore, the researchers were able to destroy and subsequently restore the phototactic phenotype by reversibly inhibiting the biosynthesis of β-carotene, the molecular precursor to retinal. Taken together, these results suggested the presence of a rhodopsin protein of the Type 2 subfamily.

The increasing availability of genomes from the traditionally understudied basal fungal lineages, coupled with a fairly well understood and important environmental sensing system, yields an opportunity to expand on known information about the photosensory response in fungi. In this chapter, I describe a computational approach toward understanding the structural mechanisms involved with rhodopsin-specific photoreception in basal fungi.

2.2 Methods

Sequence identification and homology modeling

Putative rhodopsin sequences in basal fungal lineages were identified based on sequence similarity to the Profile Hidden Markov model from the Pfam database [77], accession PF00001 ("7tm_1"). Identified fungal sequences were aligned with a subset of animal rhodopsin sequences from GenBank using Expresso [78], a modified version of T-Coffee [79], which incorporates protein structural information to guide the sequence alignment. The MAFFT [80, 81] and Muscle [82] multiple alignment modules were added to the default Expresso alignment. Based on the multiple sequence alignment, the structure model of the *B. dendrobatidis* protein was built using Modeller (v9.9) [83] with explicit loop refinement and refined with OPUSRota (v1.0) [84]. The *S. punctatus* and *A. macrogynus* models were constructed using iTASSER against the provided GPCR specific library [85]. The *Bd* sequence was initially modeled using iTASSER along with the *Sp* and *Am* sequences. However, the Modeller-produced model was selected for further analyses as a conserved structural motif in EL2 of the iTASSER best-scoring model for *Bd* was modeled incorrectly when compared to crystal structures 2Z73 and 1U19, and the *Sp* and *Am* iTASSER models. For the *Sp* model, manual correction of the K320 orientation was performed by energy minimization using the general Amber force field (GAFF) [86] in Avogadro [87] after automatic refinement with OpusROTA.

For the structures generated using Modeller, the output consisted of five potential models and corresponding Discrete Optimized Protein Energy ("DOPE") [88] and MODELLER objective function ("molpdf") scores. Optimal models were therefore selected which had the lowest DOPE and molpdf values. For the iTASSER structures, the optimal model was selected using the iTASSER provided "c-score", a confidence value

based on the significance of threading template alignments.

The quality of these selected models was assessesd using PROCHECK (v3.5) [89, 90] and Verify3D [91]. The melatonin model was constructed using the human melatonin sequence (UniProt ID: P48039) and subjected to homology modeling with Modeller (v9.9) [83] using the *T. pacificus* rhodopsin crystal structure (2Z73) as a template. The Modeller-generated homology model was of better stereochemical quality than the iTasser generated Melatonin model using the GPCR database. As such, the highest quality Modeller-generated model was selected for further side chain refinements with OPUSRota [84], similar to the *Bd* model generation.

In Figure 2.1, panels featuring structural diagrams were created using the PyMOL Molecular Graphics System, Version 1.7.4 [92]. The membrane topology figure panel was drawn using the TEXtopo package [93].

Docking

Automated protein-ligand docking was accomplished using Autodock 4 [94] which implements a Lamarckian genetic algorithm approach for calculating the minimum free energy of binding of small molecules. Small molecule files were obtained from PubChem [95] for the following isomers of retinal: 11-*cis* (A1), all-*trans*, 9-*cis*, 13-*cis*, 3,4-dehydro- (A2), 3-hydroxy- (A3), and 4-hydroxy- (A4) used in the covalent docking screen. A covalent linkage was formed by manually specifying the presence of a bond between the terminal carbon atom in retinal and terminal nitrogen atom in the lysine side chain. The specific lysine predicted to be involved in functional photoreception was inferred through multiple sequence alignment.

For the non-covalent docking screen, the 11-*cis*-retinal (A1) isomer was used

as a search query in ZINC (v 12) [96] with a cutoff value of 0.9. 83 compounds were retrieved and used in addition to 11-*cis*-retinal in Autodock 4.

RMSD calculation

The loop regions in all models were removed such that the models contained only the seven transmembrane helix regions. The helix-only structures were then aligned using the STAMP Structural alignment method [97] of VMD (v1.9.1) [98], and the RMSD values of the backbones of the aligned helix-only structures were computed using the $rmsd()$ function in the Bio3D R package [99].

Molecular Dynamics

Molecular dynamics simulations were performed using the Amber14 suite of programs [100]. For MD simulations of the squid structure, PDBID 2Z73 was used along with the structure of 11-*cis*-retinal crystalized with it. For the *S. punctatus* structure, simulations were performed using 9-*cis*-retinal ligand in the lowest energy conformation. 9-*cis*-retinal was chosen based on the covalent docking screen results in Table 2.4. Initial minimization for 1ns, followed by three equilibration steps for 50ps progressing from 200K to 250K to 298K. The final production simulation was run for 10ns at 298K. Due to the computational expense of an explicit solvation model for simulating water molecules, an implicit solvation model [101] (modified from the generalized Born solvation model [102]) was implemented in AMBER by the $igb = 2$ flag. Backbone atoms were kept rigid while binding pocket residues (as identified in Table 2.3) were made flexible. Trajectory visualization was accomplished using VMD (v1.9.1) [98]. RMSD values and potential

energy of the system were summarized using *cpptraj* and *process_mdout.perl* script, respectively, provided with the AMBER package.

2.3 Results

Homology models are of reasonable quality

Ramachandran plots were generated for all structure models using PROCHECK [89, 90]. These plots graphically display the backbone dihedral angles ψ and ϕ of each amino acid residue in a protein and are indicative of model quality, and are summarized in Table 2.1. For *B. dendrobatidis*, *S. punctatus*, and *A. macrogynus*, the percentage of model residues after refinement which fell within the most favorable regions was 86.1%, 84.2%, and 66.4%, respectively. For comparison, the *T. pacificus* (2Z73) and *B. taurus* (1U19) published crystal structures have scores of 90.9% and 79.9%, respectively. A model with a score of >90% in this category is considered to be of good quality.

3D profile scores, computed using Verify3D [91], are provided in Table 2.1. The *B. dendrobatidis* model has a score of 55.41, and the models for *S. punctatus* and *A. macrogynus* have scores of 73.60 and 131.62, respectively. For comparison, the 3D profile scores for *T. pacificus* (2Z73) and *B. taurus* (1U19) published crystal structures are 87.85 and 109.14, respectively.

Structural conservation reveals *S. punctatus* structure to be most likely functional as photoreceptor

After generating models for the chytropsin sequences, the best-scoring models were selected, representing the most computationally and chemically ideal configura-

tions. A number of structural features provide support for the relationship between chytriopsins and other members of the opsin family. Based on sequence similarity, the chytropsin sequences are expected to have seven transmembrane architecture. Table 2.2 displays the pairwise backbone RMSD calculations which display high degree of agreement with rhodopsin crystal structures. Coupled with the overall structure alignment presented in Figure 2.1A, this expected architecture is confirmed.

These values are from a Ramachandran plot [103] generated using PROCHECK [89], a method for checking the stereochemical quality (both overall and residue-by-residue geometry) of a protein structure. The results represent the percentage of residues which, based on their ϕ and ψ angles, fall within specific stereochemical regions as defined by analysis of experimentally solved structures. A good quality model would be expected to have over 90% in the "most favored" regions. Since our models have reasonably high percentages in the "most favored" regions, and reasonably low percentages in the "disallowed" regions, this table suggests that our chytriopsin and melatonin homology models are of reasonable quality.

Generally speaking, the root-mean-square deviation (RMSD) is a measure of the difference between values predicted by a model and those actually observed. In the context of protein structure prediction, the RMSD value is a measure of the average distance between backbone atoms of superimposed protein structures. The RMSD measurement can be used as a quantitative comparison between two aligned structures, and similar structures will have lower RMSD values.

In our case, these values describe the pairwise similarity for our chytriopsin and melatonin homology models and the experimentally-verified animal rhodopsin crystal structures. Since low RMSD values correspond to similar structures, and since the RMSD values for Melatonin against solved structures are much higher than those for our

chytriopsin models, this table suggests that our chytriopsin models are more structurally similar to animal rhodopsins than to melatonin receptors.

There is a defined sequence of events in the activation mechanism initiated by photoisomerization of 11-*cis*-retinal: protonation of Glu113 (typically by proton transfer mediated by protonated Schiff-base and Lys296), outward rotation of H6 (breaking the ion lock), and protonation of Glu134 (re-stabalization in active state).

Binding pocket, lysine, and counterion- The photoisomerization process involves light interacting with a retinal chromophore producing a conformation change and proton transfer cascade [104, 59]. The most critical residues (*B. taurus* numbering) in this cascade are Lys296, responsible for formation of the protonated Schiff-base covalent linkage to 11-*cis*-retinal, Glu113, the counterion responsible for proton transfer during photoisomerization, and the H-bond network required for dark-state stability, centered around His211 and Glu122, including Glu181, Tyr192, Tyr268, Ser186, Glu113, Cys187, and Thr94.

The *S. punctatus* structure posesses both the conserved lysine (K320) and a suitable counterion (D94) in positions favorable for proper function. The structures of *B. dendrobatidis* and *A. macrogynus*, on the other hand, lack the conserved lysine and counterion residues in analagous positions. Binding pocket residues, and lysine, counterion, and H-bond network residues are compared in Figure 2.1C.

Ion lock- The (E/D)RY and NPXXY motifs function together as the "ionic lock": a structural motif responsible for stabilizing the protein in the inactive conformation and which is broken upon receptor activation [59]. The (E/D)RY motif of *B. taurus* consists of the Glu134-Arg135-Tyr136 residues. A salt bridge between Arg135 on H3 and Glu247 on H6 stabilizes the lock in this inactive state. Upon receptor activation, the NPXXY motif, specifically Tyr306 rotates toward Arg135 to break the lock. The ERY

motif in the *S. punctatus* structure comprises Glu115-Arg116-Tyr117, and the NPXXY motif is functionally conserved with Asn326-Pro327-Val328-Leu329-Phe330. The *B. dendrobatidis* motifs are slightly less conserved with Asn104-His105-Tyr106 for ERY, and Asn353-Pro354-Ile355-Val356-Phe357 for NPXXY. The *A. macrogynus* motifs are much more conserved: Glu155-Arg156-Tyr157 for ERY, and Asn504-Pro505-Leu506-Leu507-Ser508 for the NPXXY motif. Comparisons are displayed in Figures 2.1E.

Salt bridge / disulfide bond- In Bovine rhodopsin, the extracellular loop region (EL2) between Trp175 on H4 and Thr198 on H5 contains two linkages that are critical for correct rhodopsin folding: the conserved disulfide bond between residues Cys110 and Cys187 and a conserved salt bridge between Arg177 and Asp190 [59]. The residues that correspond to the disulfide bond are conserved in the three chytrid structures: Cys80-Cys155 in *B. dendrobatidis*, Cys91-Cys166 in *S. punctatus*, and Cys131-Cys220 in *A. macrogynus*. The salt bridge residues are relatively conserved in *B. dendrobatidis*, with Lys145 and Asp158. However they are somewhat less conserved in *S. punctatus* (Ala156 and Asp169) and *A. macrogynus* (Thr203 and Ala223). Comparisons are displayed in Figure 2.1D.

in silico chemical screen

Computational protein-ligand docking was accomplished using Autodock 4 with 11-*cis*-retinal, all-*trans*-retinal, 9-*cis*-retinal, 13-*cis*-retinal, 3-dihydroretinal, and 4-dihydroretinal (Table 2.4). When docked against the squid crystal structure (PDB ID 2Z73), 11-*cis*-retinal had the lowest free energy of binding. This was to be expected as 11-*cis*-retinal is the functional chromophore for the squid rhodopsin protein. Additionally, all-*trans*-retinal had the highest free energy of binding.

The lowest energy conformation for the *S. punctatus* modeled structure were observed when bound to 9-*cis*-retinal isomer, with the next lowest conformation observed with the 11-*cis*-retinal isomer.

Molecular Dynamics simulations

In order to assess how the stability of the predicted *S. punctatus*+9-*cis*-retinal complex compares to that of the canonical squid+11-*cis*-retinal complex, I performed molecular dynamics simulations using AMBER 14. An overview of the potential energy of two systems during the 10ns simulation is given in Figure 2.3A. While the potential energy of the *S. punctatus* complex is much lower than that of the squid, both complexes are extremely stable over the long term. The average structure was generated using *cpptraj* by RMS fitting backbone atom coordinates from 2000 snapshots at 5ps intervals and averaging the coordinates. For both complexes, these results are given in Figure 2.3B. The squid complex achieves equilibrium starting from 1 ns of the trajectory period, and the deviation from the starting structure is about 3Å. Similarly, the *Sp* complex achieves equilibrium starting from 3 ns, while the deviation from the starting structure is close to 8Å.

2.4 Discussion

The opsin class of visual receptors can be divided into two subtypes, based on sequence similarity and function. While both types have similar tertiary structure (eg seven transmembrane helices), the Type 2 rhodopsins, which act as GPCR proteins, have

thus far only been identified in metazoan lineages, while the Type 1 rhodopsins, which function as ion channels, are typically found in bacteria and archaea. The rhodopsin-like proteins identified in recently sequenced, early-diverging flagellated fungi are most similar to these Type 2 proteins, and thus are in an excellent position to add to the expanding knowledge base of the evolution of vision.

The work described in this chapter seeks to address questions related to the structure and function of these identified chytrid rhodopsins, namely I) these proteins are structurally similar to visual rhodopsins, and II) functional characteristics can be determined through *in silico* chemical ligand and molecular dynamics simulations.

In support of this hypothesis are the results of a number of comparative analyses. Across the fungi, there are different types of photosensitive proteins, each with different structures and regulatory mechanisms.

The rhodopsin-like proteins identified in the chytrid lineages have notable similarities and differences relative to well described rhodopsins. The expected seven transmembrane structure is conserved in every sequence identified, with proper orientation of N- and C-termini. The β-sheet motif at the top of the structure is conserved, as is the cysteine bridge and ion lock motifs important for structural stability.

The lysine residue involved in retinal binding is conserved in the *S. punctatus* sequence, but is absent in the *B. dendrobatidis* and *A. macrogynus* sequences. This is notable for the potential functional and evolutionary implications, especially in light of its presence in the *S. punctatus* structure. However, experimental evidence suggests that the covalent linkage facilitated by the lysine residue, while highly desired and most evolutionarily favorable [76], is not necessary for activation of the light-driven cascades in bacteriorhodopsin [105] and rhodopsin [106]. In the case of bacteriorhodopsin specifically, a K216A mutant was generated and homologously expressed in *Halobacterium*

salinarium L33 and provided with retinylidene-n-alkylamines to achieve a Schiff-base construct withought the covalent linkage. As this mutation mirrors the *B. dendrobatidis* protein (alanine present at the critical position), this suggests that perhaps a non-traditional chromophore is required for rhodopsin function in *B. dendrobatidis*.

A broad, non-covalent docking screen using ligands similar to 11-*cis*-retinal was performed to assess the capacity of the *B. dendrobatidis* and *A. macrogynus* pockets to accomodate, structurally, a ligand of this shape. The results of this screen suggest that these pockets are indeed sufficiently large to accomodate a molecule of that size. Furthermore, several 11-cis-retinal analogous molecules present lower binding affinities for these pockets, suggesting functionality is present despite the absence of the conserved lysine residue. Thus, the lack of the lysine in these two structures does not necessarily imply that they are non-functional. Future work in the form of *in vitro* functional assays or *in silico* chemical screens may be necessary to further understand the exact functional nature of this protein.

Protein models known to be correct have higher 3D profile scores [91] compared to incorrectly modeled structures. As indicated in Table 4, the models for *B. dendrobatidis* and *S. punctatus* using the Type 2 rhodopsin structures 2Z73 and 1U19, respectively, had scores nearly double those of the same sequences modeled against the Type 1 sensory rhodopsin II structure 1H68. As could be expected, the experimentally determined crystal structures used as templates had scores 1.5-2 times larger than the modeled structures (80-100). Longer proteins tend to have higher scores in general. All protein models scored were approximately equal in size (approx. 350 aa), with the exception of AMAG00698 (536 aa). As previous work has shown that Type 1 and Type 2 opsin proteins have similar but not quite identical structures [73], this finding supports the hypothesis that the chytrid sequences are Type 2 and not Type 1 rhodopsins.

In silico docking screens were performed to assess how the *S. punctatus* homology model binds to known ligands, as it is the only Type 2 rhodopsin identified in chytrids which posesses the conserved lysine and counterion residues. Based on this screen, 9-*cis*-retinal and 11-*cis*-retinal appeared to be the most favorable ligands for use by *S. punctatus*. As such, 9-*cis* isomer was used in subsequent refinement by molecular dynamics. When compared to the squid crystal structure (PDB ID: 2Z73) and its canonical 11-*cis*-retinal ligand, the *S. punctatus*+9-*cis*-retinal complex reaches a plateau after more time and at a greater resolution. However both complexes are highly stable. Thus, the *S. punctatus*+9-*cis*-retinal complex after MD simulations is a good candidate for future work involving refined docking screens, and supports the hypothesis that this GCPR is a functional photoreceptor.

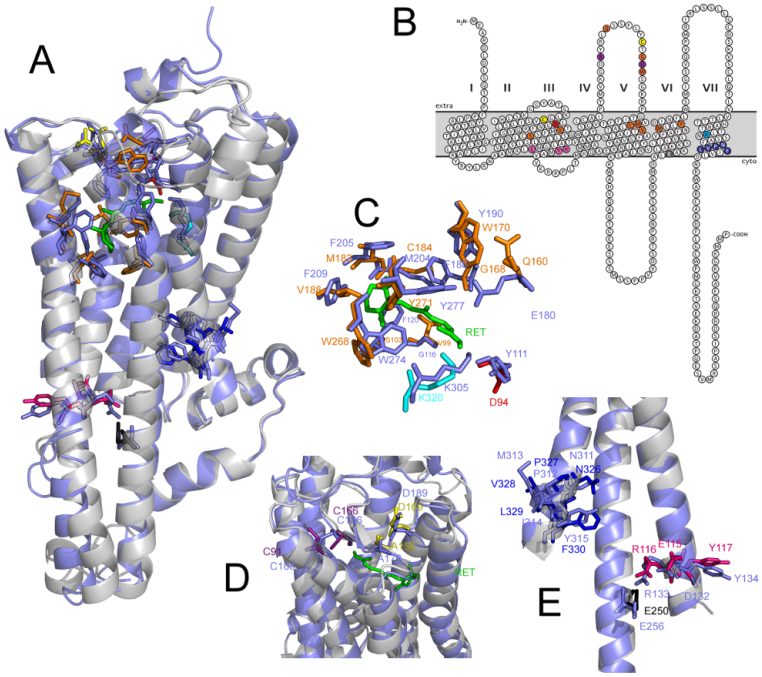


Figure 2.1: *S. punctatus* residues are colored according to function: orange (binding pocket residues), red (putative counterion), purple (disulfide bond), yellow (salt bridge), dark blue (NPxxY motif), and pink & black (ion lock). Light purple functional and backbone residues belong to *T. pacificus*, while grey backbone residues belong to *S. punctatus*. The ideal position of the 11-*cis*-retinal ligand, taken from the *T. pacificus* crystal structure, is shown in green. A) Cα backbone structural alignment of *S. punctatus* homology model and *T. pacificus* x-ray crystal structure. B) Topography plot of membrane spanning regions of *S. punctatus* homology model. C) Detail of *S. punctatus* binding pocket residues aligned with those of *T. pacificus*. D) Detail of *S. punctatus* disulfide bond (purple) and salt bridge (yellow) regions aligned with those of *T. pacificus*. The view is from the top (extracellular side) of the protein, into the 11-*cis*-retinal (green) binding pocket. E) Detail of *S. punctatus* ERY and NPxxY regions aligned with those of *T. pacificus*.

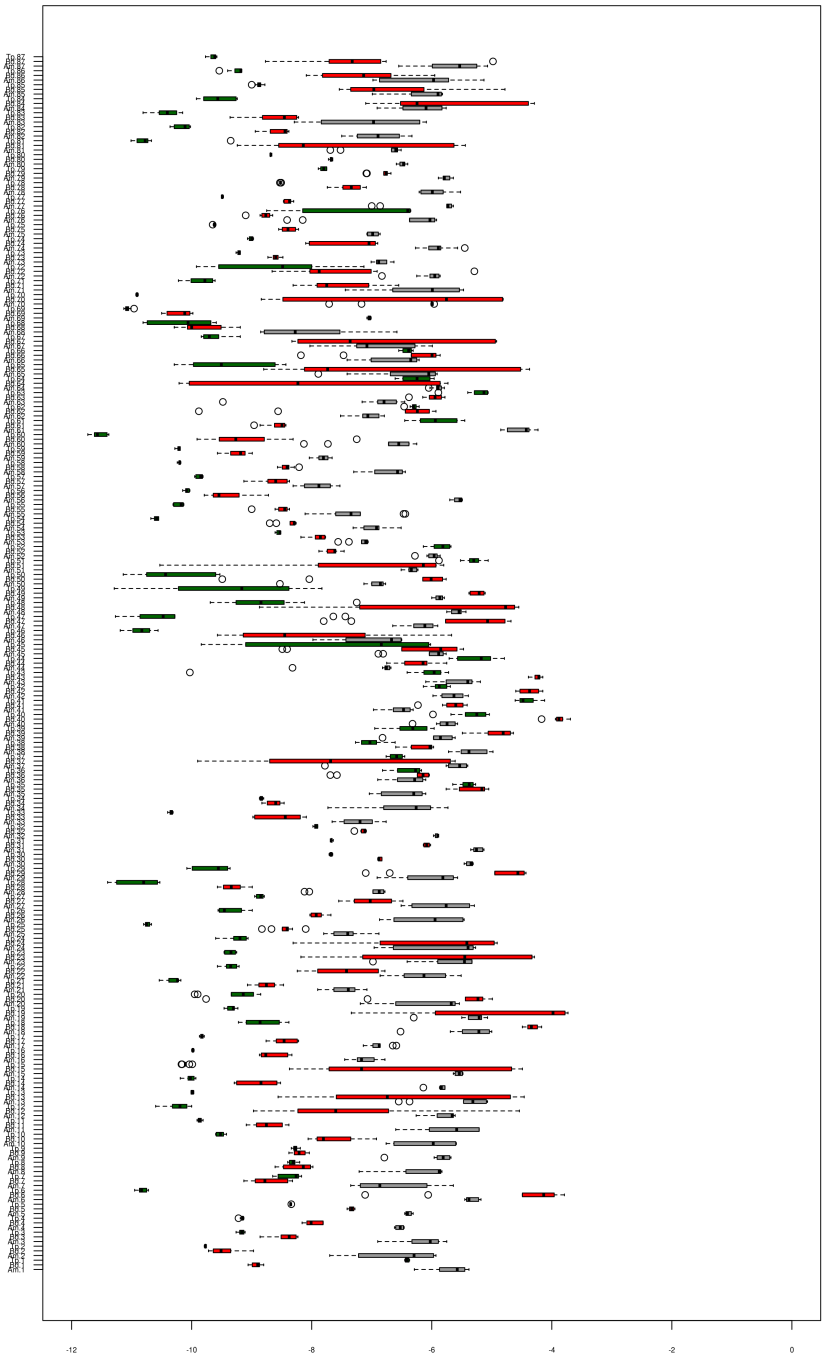


Figure 2.2: A non-covalent docking screen using Autodock 4 and ligands obtained from ZINC against homology models for *B. dendrobatidis* (red), *A. macrogynus* (gray), and *T. pacificus* (squid; 2Z73) (green). Histograms show distribution of ten lowest binding energies for protein-ligand conformations. Binding energy given on X-axis; ligand given on Y-axis. Ligands were obtained from ZINC database based on similarity to 11-*cis*-retinal.

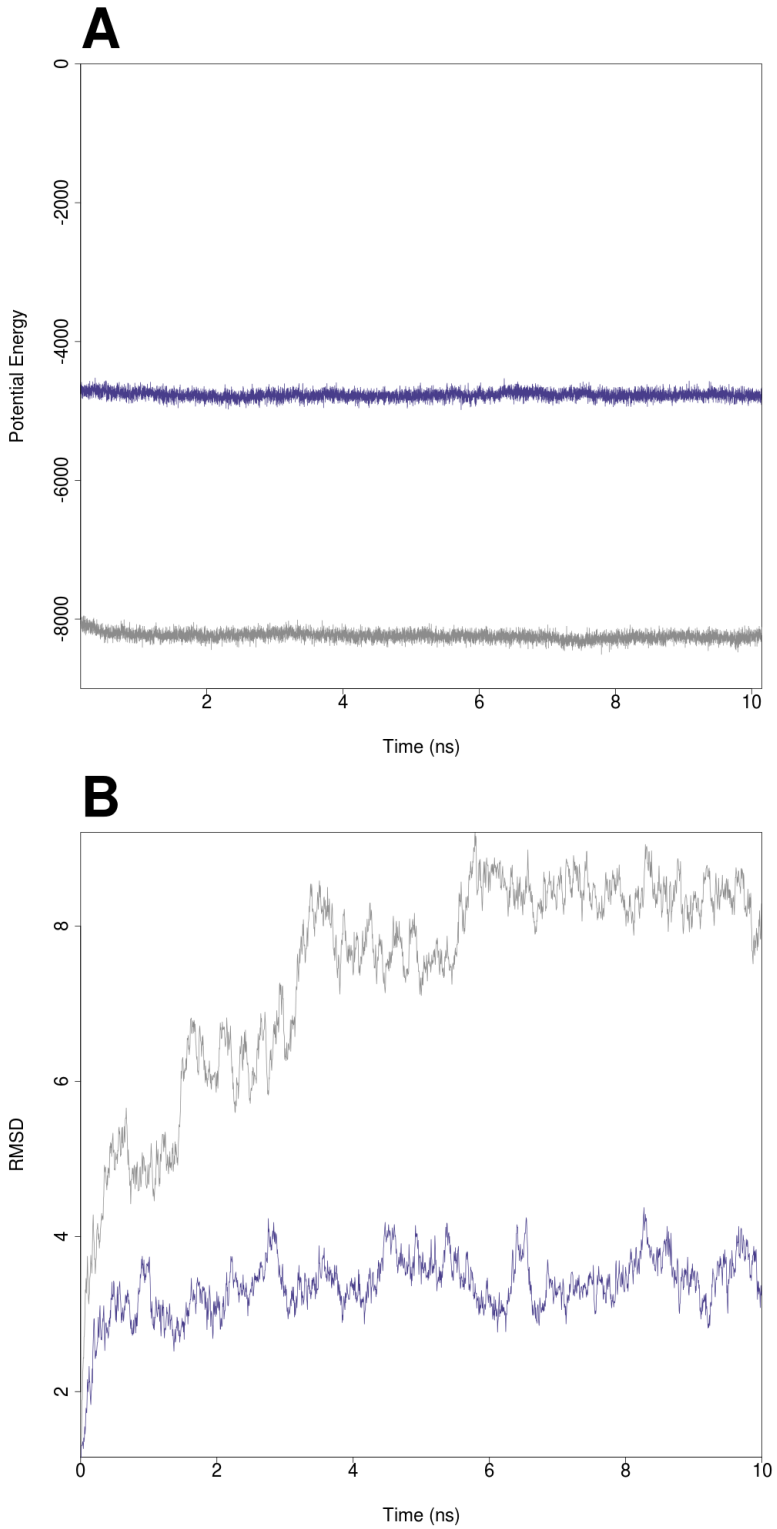


Figure 2.3: Overview plots of MD simulation runs of squid 2Z73 crystal structure with 11-*cis*-retinal (purple) and *S. punctatus* model with 9-*cis*-retinal (gray). A) Potential energy over course of simulation. During the simulation, both structures remain relatively stable. The *S. punctatus* stucture has substantially lower potential energy than the squid structure. B) RMSd fits over course of simulation. The RMSd fit of the *S. punctatus* structural model increases much more rapidly than that of the squid structure, but ultimately reaches a plateau after approximately 6 ns.

Table 2.1: PROCHECK Ramachandran plot results for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures. Spun, *Spizellomyces punctatus*. Bden, *Batrachochytrium dendrobatidis*. Amac, *Allomyces macrogynus*. Tpac, *Todarodes pacificus* (PDBID 2Z73). Btau, *Bos taurus* (PDBID 1U19). Hsap, *Homo sapiens*.

	Spun	Bden	Amac	Tpac	Btau	Hsap
Most favored regions	90.7%	86.1%	66.4%	90.0%	79.2%	91.1%
Additional allowed regions	7.5%	11.6%	24.1%	9.1%	15.8%	7.6%
Generously allowed regions	1.2%	1.4%	6.0%	0.0%	3.8%	0.6%
Disallowed regions	0.6%	0.9%	3.4%	0.0%	1.2%	0.6%
Verify3D	73.60	55.41	131.62	87.85	109.14	00.0

Table 2.2: Pairwise backbone RMSD measurements for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures, calculated using the Bio3D package. Spun, *Spizellomyces punctatus*. Bden, *Batrachochytrium dendrobatidis*. Amac, *Allomyces macrogynus*. Tpac, *Todarodes pacificus* (PDBID 2Z73). Btau, *Bos taurus* (PDBID 1U19). Hsap, *Homo sapiens*.

	Spun	Bden	Amac	Tpac	Btau	Hsap
Spun	0.00	2.54	3.05	2.52	3.12	4.92
Bden	2.54	0.00	3.33	2.55	2.70	5.11
Amac	3.05	3.33	0.00	3.71	2.91	5.81
Tpac	2.52	2.55	3.71	0.00	3.09	4.96
Btau	3.12	2.70	2.91	3.09	0.00	5.50
Hsap	4.92	5.11	5.81	4.96	5.50	0.00

Table 2.3: Conserved Rhodopsin motifs in X-ray crystal structures and chytropsin homology models. Btau, *Bos taurus* (PDBID 1U19). Tpac, *Todares pacificus*. Bden, *Batrachochytrium dendrobatidis*. Spun, *Spizellomyces punctatus*. Amac, *Allomyces macrogynus*

	Description	Btau	Tpac	Bden	Spun	Amac
1	Salt Bridge	R177	A176	K145	A156	T203
2	Salt Bridge	D190	D189	D158	D169	A223
3	Binding pocket	T118	G116	Q88	V99	Q139
4	Misc	E122	F120	A92	S103	V143
5	Binding pocket	W265	W274	W317	W268	W468
6	H-bond core/Photoisomerase counterion	E181	E180	R149	Q160	–
7	Misc	I189	F188	Y157	G168	–
8	Binding pocket/H-bond network w/Y268	Y191	Y190	Y159	W170	–
9	Misc	M207	M204	L173	M183	L237
10	Misc	F208	F205	I174	C184	A238
11	Misc	F212	F209	V178	V188	L242
12	Binding pocket	W265	W274	W317	W268	W468
13	Binding pocket/H-bond network w/Y191	Y268	Y277	T320	Y271	Y471
14	Conserved Lysine	K296	K305	A347	K320	S498
15	Counterion	E113	Y111	V83	D94	N134
16	[L]AxAD	L79	L76	V48	L60	L97
17	L[A]xAD	A80	A77	V49	S61	–
18	LAx[A]D	A82	S79	S51	T63	–
19	LAxA[D]	D83	D80	D52	D64	D101
20	DisulfideBond	C110	C108	C80	C91	C131
21	DisulfideBond	C187	C186	C155	C166	C220
22	[E]RY	E134	D132	N104	E115	E155
23	E[R]Y/IonicLock w/E247	R135	R133	H105	R116	R156
24	ER[Y]	Y136	Y134	Y106	Y117	Y157
25	[N]PxxY	N302	N311	N353	N326	N504
26	N[P]xxY	P303	P312	P354	P327	P505
27	NP[x]xY	V304	M313	I355	V328	L506
28	NPx[x]Y	I305	I314	V356	L329	L507
29	NPxx[Y]	Y306	Y315	F357	F330	S508
30	IonicLock w/E247	V138	V136	V108	A119	R159
31	IonicLock w/R135	E247	E256	L299	E250	A450

Table 2.4: Interaction energies for covalent docking with Autodock. Average of five lowest energy conformations using various retinal isomers covalently bound to rhodopsin models and crystal structures. Tpac, *Todarodes pacificus* (PDBID 2Z73). Spun, *Spizellomyces punctatus*. SRII, *Natronomonas pharaonis* (PDBID 1H68).

Structure	11-cis	all-trans	3,4-dehydro	3-hydroxy	4-hydroxy	9-cis	13-cis
Squid	-5.52	-4.75	-4.33	-2.51	-3.10	-5.41	-2.55
Spun	16.68	24.18	24.78	26.06	25.58	9.56	26.24
SRII	-4.67	-5.72	-5.83	-5.42	-5.69	-3.70	-5.86
Nop-1	-4.30	-5.70	-5.20	-4.01	-5.27	-5.27	-4.67
Beme-BR	-4.81	-1.28	-2.95	-2.07	-2.10	-2.50	-4.13

Chapter 3

Rhodopsin related signaling pathways in basal fungi

3.1 Introduction

G-protein coupled receptors (GPCRs) are a broad class of seven-transmembrane proteins which receive extracellular signals and initiate an intracellular response [107]. This process is accomplished via signal transduction pathways which link certain effector proteins with the receptors using heterotrimeric GTP-binding and hydrolysing proteins (G proteins) [108]. There are two major pathways which GPCRs are principally associated with: the cAMP signaling pathway and the phosphatidylinositol pathway [109].

G protein complexes associate with a transmembrane GPCR and are loosely coupled to the intracellular side of the plasma membrane [110]. These GTPases are composed of an α, a β, and a γ subunit. While the Gα subunit was historically thought to be solely responsible for signal transduction [109], it has been demonstrated that the Gβ/γ subunit is capable of producing a response in yeast [111]. Currently the number of effectors regulated by one or both subunits (α or β/γ) is equivalent [110].

The Gα subunit is bound to a molecule of GDP in its inactive state. Upon receptor stimulation, for example photoisomerization of 11-*cis*-retinal in rhodopsin, the bound GDP is exchanged for GTP and the now active Gα subunit dissociates from both the receptor and the Gβ/γ subunit [112].

Gα proteins of different types interact with difference downstream effectors. Four major Gα protein subfamilies have been identified: G_s, G_i, G_q, and G_{12} [108]. The type of G-protein to which the receptor is coupled can help classify the Type 2 rhodopsins. The G_s group contains the G_s and G_{olf} subunits, the latter being found in the olfactory neuroepithelial cells. These enzymes enhance the rate of cAMP synthesis by stimulating adenylate cyclase. Additionally, the $G_s\alpha$ subunit regulates Na^+ and Ca^{2+} channels [108]. The G_i group contains four subclasses: G_i, G_o, G_t, G_z. All G_i subclasses possess a consensus sequence for ADP-ribosylation by pertussis toxin and function to inhibit adenylylcyclase [113]. G_o proteins are abundant in the brain and implicated in membrane trafficking. G_t, or transducin, is activated by rhodopsin, activates cGMP-phosphodiesterase (cGMP-PDE) and closes cGMP-gated sodium channels. G_z activity is relatively unknown, but evidence suggests that it inhibits adenylylcyclase activity as well [113]. The G_q group contains members G_q, G_{11}, G_{14}, G_{15}, and G_{16}. These are widely expressed and involved in signal transduction through activation of phospholipase C-β1 [113]. Finally, the G_{12} group contains the G_{12} and G_{13} subunits.

Four distinct subgroups have been previously identified from fungal Gα proteins [114]. Only Group I and III proteins have discernible similarity to mammalian families: the inhibitory (Gαi) and stimulatory (Gαs) families, respectively. Group I proteins contain amino-terminal myristoylation consensus sequences (MGxxxS) as well as carboxy-terminal pertussis toxin ADP-ribosylation sites (C[GAVLIP]2x), both of which are conserved in the Gαi superfamily [115]. Group III proteins are highly conserved and

also possess a myristoylation motif at their amino-termini. While there is evidence that many Group III proteins influence cAMP levels, suggesting placement in the Gαs family, GNA-3 in *N. crassa* is much more similar to Gαi2 [115]. Group II Gα proteins have no homologous mammalian counterpart, and a biological function has only been observed for a few members. In *S. cerevisiae*, the Gpa1p protein functions as a mating factor, while the Gpa2p protein functions in intracellular cAMP regulation [113]. The *magC* gene from *Magnaporthe grisea* has been demonstrated to be involved in ascospore development [116]. The fourth and arguably minor Gα subgroup has little to no identified function, and was initially characterized as one unusual protein identified in *Ustilago maydis* which clustered separately from other fungal Gα proteins and for deletion of which there was no obvious phenotype [114].

GTP-bound Gα proteins are themselves acted on by a regulator of G protein signaling (RGS) protein [117]. RGS proteins are responsible for the rapid return of active GTP-bound Gα-subunits to their inactive state, and function to tightly regulate the signal generated by a GPCR.

Activated Gα proteins also interact with the γ-subunit of cGMP phosphodiesterase. Phosphodiesterases (PDEs) are a superfamily of proteins which enzymatically cleave phosphodiester bonds. In mammals, this superfamily is comprised of 12 families based on sequence similarity, tissue distribution, and substrate specificity. PDEs act on cAMP (PDE families 4, 7, and 8), cGMP (PDE families 5, 6, and 9), or both (PDE families 1, 2, 3, 10, and 11) [118]. Due to this activity, they are importance regulators of signal transduction. In mammalian systems, for example, they associate with activeated Gα subunits in order to close cGMP-gated cation channels located in the plasma membrane and regulate the influx of Ca^{2+} and Na^{+}. One such interaction is the association between transducin (G_t) and cGMP phosphodiesterase in the rhodopsin visual signaling

cascasde [119]. The transducin α subunit (Tα), activated by rhodopsin interaction with light, interacts with the PDE complex to disrupt the inactivation caused by the PDEγ subunit, resulting in hyperpolarization of the cell membrane [119].

The Gβ/γ subunit is responsible for a number of regulatory functions during signal transduction [110]. Previous predictions suggest that Dikarya fungi have one Gβ subunit, while the zygomycete *Rhizopus oryzae* has four [115]. Structurally, the Gβ subunit contains a 20aa α-helix region and a large domain composed of repeated WD40 motifs. While the WD-repeat region is unknown, it is presumed to be related to assembly of the complex. [110].

All Gγ subunits posess specific CaaX motifs at their C-termini. These motifs are subject to postranslational modification, the nature of which both targets the Gβ/γ complex to the plasma membrane, and governs interactions between Gγ and Gα, receptors, and/or effector proteins [120].

As discussed in Chapter 2, rhodopsins are a broad class of photosensitive seven-transmembrane proteins which respond to light through photoisomerization of a retinaldehyde chromophore, typically 11-*cis*-retinal. Type 2 rhodopsins are GPCRs which function in, among other things, metazoan visual pathways.

The work in this chapter deals with presence and absence of components of the rhodopsin signaling pathway in basal fungi, including the distribution of heterotrimeric G protein subunits and potential effector proteins.

3.2 Methods

Identification of homologous photosensory proteins

To get a sense for the distribution of photosensory proteins in fungi, the fungal proteomes listed in Table B.1 were searched using the classes of photosensory proteins from previous analyses described in by Idnurm et al. The white collar protein complex, comprising WC-1 and 2, were searched for using the *N. crassa* proteins NCU02356 and NCU00902, respectively, as queries using *ssearch*36 from the FASTA package [121] with an e-val cutoff of 1e-10. Type 1 and Type 2 opsins were searched for using HMMER (v3.0) *hmmsearch* [122] with HMM models PF01036 and PF00001 (obtained from the Pfam database [77]), respectively, using an e-val cutoff of 1e-20. Cryptochrome homologs were searched using HMMER (v3.0) *hmmsearch* with an HMM generated from the 20 seed sequences in the Pfam domain PF12546 (Cryptochrome_C). Hits were retained above a threshold of 1e-20. Similarly, phytochrome homologs were searched using *hmmsearch* with an HMM profile generated from the 80 seed sequences in PF00360 (PHY). Hits were retained above a threshold of 1e-20.

G protein analysis

Homologs of Gα proteins were identified using HMMER *hmmsearch* with an HMM profile generated from the seed set of Pfam domain family PF00503. Fungal hits above a threshold of e-20 were kept for subsequent analysis. A maximum likelihood tree was constructed using RAxML (v7.5.4) [123] with 100 bootstrap replicates using representative Fungal hits from basal fungi, Zygomycete, Ascomycete, and Basidiomycete lineages, along with representative outgroup eukaryotic Gα sequences from the

G_s (IPR000367), G_q (IPR000654), G_i (IPR001408), and G_{12} (IPR000469) families.

Homologs of Gβ proteins were identified using HMMER *hmmsearch* with an HMM profile generated from the seed set of PRINTS domain family PR00319. Fungal hits above a threshold of e-100 were kept for subsequent analysis. A maximum likelihood tree was constructed using RAxML (v7.5.4) with 100 boostrap replicates using fungal hits from basal lineages, Zyogomycete, Ascomycete, and Basidiomycete lineages, along with representative sequences for Gnb-1, Gnb-2, Gnb-3, Gnb-4, and Gnb-5 from mouse and human. Also included were other outgroup sequences from *Phytophthora sojae*, *Chlamydomonas reinhardtii*, and *Drosphila melanogaster*.

Homologs of Gγ proteins were identified using HMMER *hmmsearch* with an HMM profile built using *hmmbuild* from 43 sequences matching the InterPro IPR001770 domain: 1 Choanoflagellate, 2 Porifera, 1 Placozoa, 1 Ctenophora, 6 Cnidaria, 1 Nematoda, 6 Arthropoda, and 25 Chordata. Fungal hits above a threshold of e-5 were kept for phylogenetic analysis, along with a representative set from the original IPR001770 dataset: 1 Choanoflagelatte, 12 Chordata, 1 Placozoa, and 2 Porifera.

Homologs of RGS proteins were identified using HMMER *hmmsearch* with an HMM profile genearated from the seed set of Pfam domain family PF00615.

Multiple sequence alignments were drawn and annotated using the TEXshadepackage [124], and protein domain figures were drawn using the pgfmolbio package (`http://www.ctan.org/pkg/pgfmolbio`).

Phosphodiesterase analysis

Homologs of phospohodiesterase (PDE) subunits were identified using sequences from KEGG families K08718 (α), K13756 (β), and K13759 (γ). Each KEGG family con-

tained a set of phylogenetically diverse sequences from which HMM models were built using T-coffee (v8.97_101117) [79] and HMMER. K08718 (PDEα) contained 18 chordata sequences, K13756 (PDEβ) contained 24 chordata sequences, and K13759 (PDEγ) contained 25 chordata sequences. These sequences correspond to PDE-6, a member of the cGMP-specific family of PDEs.

Maintenance of chytrid cultures

B. dendrobatidis JEL423 cultures were grown on 1% Tryptone plates [tryptone (10 g/L), glucose (3.2 g/L), and 1% agar] and maintained at 23°C. *S. punctatus* SW-1 cultures were grown on PmTG agar plates [peptonized milk (0.5 g/L), tryptone (1 g/L), glucose (5 g/L), and 1% agar]. All cultures were maintained at room temperature (23°C) under an unregulated lighting scheme. Motile zoospores, were collected from actively growing (2-4 day old) plates by flooding with 2-4 mls of sterile di H_2O, waiting 30-45 minutes, and collecting the liquid.

Phototaxis

To examine the extent of phototaxis in *B. dendrobatidis* and *S. punctatus*, I followed a protocol established previously to observe phototaxis in the marine Chytridiomycete *Rhizophydium littoreum* [49]. Briefly, light was projected upwards through the bottom of a 60mm plastic petri dish containing a concentrated (approx. 10^6 cells/ml) suspension of freshly harvested, motile (approx. 75%) zoospores. An "X" pattern is cut in a piece of dark cardboard which is placed between the light source and the petri dish. Phototaxis is determined by zoospore aggregation, avoidance, or ignorance of the "X"-

shaped light source, and is scored as attraction, aversion, or absence, respectively. The light source was a Kodak slide projector with 300W white light bulb and was redirected using a stainless steel cosmetic mirror. The light source was placed at a distance such that a light intensity of 950 lux was achieved. Phototaxis experiments were carried out in a darkroom at 25°C.

Pichia pastoris heterologous expression

Genomic DNA was extracted using a modified bead-beating procedure. Briefly, approximately 100 mg of material containing both zoospores and sporangia was scraped from actively-growing, zoospore-rich *B. dendrobatidis* and *S. punctatus* plates. The material was added to approximately 100 mg of silicon beads (0.5mm dia.) and mixed with 600 μl of Cell lysis solution (Qiagen, Germantown, MD) and 3 μl of proteinase K. The solution was homegenized with a bead beater using a 30s pulse at 4°C and subsequently incubated for 2h at 55°C. 200 μl of protein precipitation solution (Qiagen, Germantown, MD) was added to the mixture and iced for 15 min. After centrifugation at 14000xg for 3 min at room temperature, the supernatant was collected, mixed with 600 μl isopropanol, and spun at 1400xg for 1 min at room temperature. Pellet was washed with 600 μl ice cold 70% EtOH and spun at 1400xg for 1 min at room temperature. Finally the pellet was air dried for 15 min at room temperature, resuspended in 50 μl H_2O, incubated at 65°C for 1 hr, and stored at -20°C.

For plasmid construction, a sequence comprising the *B. dendrobatidis* rhodopsin gene (BDEG_04847), along with an additional 78 upstream bases to account for the first transmembrane helix, was amplified from *B. dendrobatidis* JEL423 and cloned into the pHIL-S1 *P. pastoris* expression plasmid using a strategy described previously

[125]. Modifications to the *B. dendrobatidis* gene include the addition of 5' and 3' EcoRI restriction sites, as well as a C-terminal hexahistidine epitope. These modifications were accomplished using the following PCR primers (provided in Appendix B): "$Bden_EcoRI_F$" (forward) and "$Bden_EcoRI_His_R$" (reverse). PCR conditions were as follows: 10 μl 5X buffer, 27.5 μl H2O, 5 μl of each primer (10 μM), 1 μl of 10 mM dNTPs, 1 ng DNA, and 0.5 μl Phusion Polymerase in each 50 μl reaction. Cycle parameters used were 98°C for 30s, 30 rounds of: 95°C for 5s, 58°C for 20s, 72°C for 30s, and a final 72°C for 10 min. The resulting fragment was purified using the Qiagen PCR purification kit, digested with EcoRI, purified again, and inserted into the EcoRI site of the pHIL-S1 plasmid. The resulting BdpHIL-S1 plasmids were transformed into chemically competent *E. coli* JM109 cells. Transformants were checked for proper insertion / orientation by colony PCR.

The *S. punctatus* gene (SPPG_00350) was synthesized and inserted into the pHIL-S1 vector by GenScript (GenScript USA Inc. Piscataway, NJ 08854). Two versions were constructed: a wild type, which codes for the conserved lysine residue, and a mutant, which replaces the lysine residue with alanine.

The NoppHIL-S1 vector containing the Nop-1 protein previously described in *Neurospora crassa* [125] was generously loaned from Dr. Katherine Borkovich at the University of California, Riverside. The availability of sequences used for these analyses is provided in the appendix.

For expression and membrane preparation, the BdpHIL-S1, SppHIL-S1, NoppHIL-S1, and pHIL-S1 vectors were cloned into *P. pastoris* strain GS115 using the strategy described previously by Bieszke et al. for expression of Nop-1. The transformation vectors were linearized overnight with StuI. A 500 ml culture of *P. pastoris* was grown in Medium A [yeast extract (10 g/L), proteose peptone (20 g/L), and dextrose (20 g/L)]

shaking at 250rpm and at 30°C until A_{600} = 2.1. The culture was split into two 250 ml samples and spun at 1500g for 5 minutes at 4°C. The pellets were resuspended in 250 ml ice-cold H_2O, spun again, resuspended in 100 ml ice-cold H_2O, spun again, and resuspended in 1 ml ice-cold 1M sorbitol. The cells were transferred to 1.5 ml microfuge tube on ice and used in transformation by electroporation. 80 μl of cells were added to 10 μl of linearized vector, iced for 5 minutes, and electroporated using 2 mm gap cuvettes and the following parameters: voltage gradient: 7.5 kV/cm, resistance: 600 Ω, capacitance: 25 μF. Transformants were screened for integration of the plasmids by PCR. One of each transformant was grown on a large scale as described previously [125].

Harvested cell pellets were washed in 1 pellet volume of ice-cold, sterile water and centrifuged at 1500xg for 5 min at 4°C. The pellets were subsequently washed and resuspended in 1 pellet volume of Buffer A (7 mM NaH2PO4, pH 6.5, 7 mM EDTA, 7 mM dithiothreitol, 1 mM PMSF). Acid-washed 0.5mm glass beads were used to disrupt 1 ml aliquots using three 1-min pulses and two 90-s pulses with a minibead beater at 4°C. The supernatants were collected and pooled to yield the cell lysate. The lysate was layered on a 70% sucrose cushion (w/v in Buffer A) and centrifuged with no braking at 92000xg for 1 hour at 4°C in a SW27 swinging bucket rotor. The membrane layer, located at the interface between the lysate and sucrose cushion, was collected, stored at 4 °C, and used for the membrane preparation.

Immunoblot analysis on the membrane preparation was conducted using 50 μg of protein by PAGE. Mouse anti-6X His monoclonal antibody (Fisher Scientific, Pittsburgh, PA) was used at both 1:1000 and 1:3000 dilution for the primary antibody. Goat anti-mouse (Bio-Rad, Hercules, CA) was used as the secondary antibody at a 1:5000 dilution.

3.3 Results

Photosensory

In order to expand on previous summaries (eg [50]) of the extent of photosensing in fungi, I searched for known photosensory proteins within proteomes (Table B.1) of sequenced fungi from across the kingdom, with a focus on the recently sequenced basal lineages. This search included White-Collar complex proteins, Phytochromes, Cryptochromes, and opsins (Type 1 and 2). The results (Figure 3.1) are consistent with previous reviews but provided a higher level of resolution in the basal lineages, particularly the Chytridiomycota and Blastocladiomycota.

It is worth noting here that the basal fungal genomes surveyed have different capacities for photosensing. In *Spizellomyces punctatus*, both a Type 2 rhodopsin and White collar complex members are identified, the structure of the former of which is elaborated upon in Chapter 2. This rhodopsin protein is the only known so far in chytrids which posesses the critical lysine residue important for proper rhodopsin function [59]. In *Batrachochytrium dendrobatidis* however, the only photosensory protein identified is the Type 2 rhodopsin, which does not appear to posess the critical lysine residue. Yet the proteome of its closest relative, the non-pathogenic *Homolaphlyctis polyrhiza*, contains an apparent opsin-GC fusion protein, the likes of which have only recently been described in the Blastocladiomycete *Blastocladiella emersonii* [126]. This architecture is also observed in the three other Blastocladiomycete organisms, *Allomyces macrogynus*, *Catenaria anguillalae*, and the transcriptome of *Coelomomyces lativittatus* (described in more detail in Chapter 4). Structural features of these opsin-GC fusion proteins are explored further in Chapter 2.

G protein analysis

Multiple Gα proteins were identified at an e-val $< 1e^{-20}$ in all surveyed chytrid genomes. Counts of predicted proteins are presented in Figure 3.1.

One of the *S. punctatus* Gα proteins, SPPG_05404, contains a C-terminal pertussis toxin sensitivity motif (C[GAVLIP]2X) and is 76.8% identical to *N.crassa* GNA-1 (NCU06493). Similarly, *A. macrogynus* contains two predicted Gα proteins, AMAG03583 and AMAG04903, both of which have the same motif. These proteins are approximately 70% identical to *N. crassa* GNA-1 (71.2% and 69.6%, respectively). Additionally, all of these proteins contain an N-terminal myristoylation motif (MGXXXS), consistent with members of the G_i subfamily. *R. allomycis*, possesses one protein with a pertussis toxin motif and has 69.77% identity to *N. crassa* GNA-1. *Piromyces sp.* possesses two proteins (18092 and 48456) with pertussis motifs. Only the former, however, also posesses an N-terminal myristoylation motif. It is 73.7% identical to *N. crassa* GNA-1. The latter appears have a large portion of its N-terminus truncated relative to the former, and is more similar to *N. crassa* GNA-3 than GNA-1 (65.7% vs 55.2% identity).

Two proteins from *B. dendrobatidis*, BDET_07008 and BDEG_07009, have high similarity to GNA-1 at 66.29% and 70.68% identity, respectively. However, only BDET_07008 contains an N-terminal myristoylation motif. Neither of them contain the C-terminal pertussis motif.

SPPG_05884 from *S. punctatus* has 75.6% identity to *N. crassa* GNA-1, but lacks the C-terminal pertussis motif (however it contains the N-terminal myristoylation motif). Similarly, SPPG_01130 contains the N-terminal myristoylation motif and is 66.85% identity to GNA-3.

G. prolifera posesses two predicted proteins with high (approx. 74%) similarity to *N. crassa* GNA-1. Both contain N-terminal myristoylation motifs, however only one also contains the C-terminal pertussis motif. A third predicted protein contains the myristoylation motif and is 62.6% identical to *N. crassa* GNA-3.

The similarities of all identified chytrid Gα proteins to identified *N. crassa* Gα proteins are presented in Table 3.1, and a multiple sequence alignment, highlighting the pertussis and myristoylation motifs shared among all fungal Gα proteins containing such motifs is presented in Figure 3.2.

A phylogenetic analysis was performed on all identified fungal Gα proteins using RaxML (Figures 3.3, 3.4, 3.5, and 3.6). All characterized Gα groups contain chytrid members, suggesting an ancient origin for these families. Additionally, the fungal Gα proteins cluster separately from the metazoan proteins. The Group II Gα family contains only a single chytrid sequence from *R. allomycis*. This protein is 360aa long and has the highest similarity to the Gα-9 subunit from *Dictyostelium discoideum* (Uniprot: Q54R41.1; e-val 4e-52; 32% identity). The majority of identified chytrid Gα proteins were most closely associated with the Group IV family. This group remains largely uncharacterized, though there is evidence to suggest that the *Ustilago maydis* homolog is induced during pathogenic development [114].

All surveyed chytrids were predicted to possess one or more Gβ proteins. Counts are presented in Tables 3.1, and similarities to *N. crassa* GNB-1 are presented in Table 3.2. A phylogenetic tree is given in Figure 3.8 illustrating sequences from basal lineages clustering among each other. All identified fungal sequences possess multiple WD40 repeat domains typical of Gβ proteins identified with InterproScan (Figure 3.7).

A. macrogynus, *B. dendrobatidis*, *C. anguillulae*, *H. polyrhiza*, *Orpinomyces*, *Piromyces*, and *R. allomycis* were each predicted to possess a single Gγ protein (e-val $<$

1e-5). No Gγ subunits were predicted in *S. punctatus* or from the *C. lativittatus* transcriptome. Of these identified Gγ proteins, only those from *B. dendrobatidis* and *Orpinomyces* contained a classic pertussis toxin sensitivity motif of the form (C[GAVLIP]{2}X), with 51.4% and 35.0% identity to NCU00041 (*N. crassa* GNG-1). Counts are presented in Figure 3.1, and similarities to *N. crassa* GNG-1 are presented in Table 3.3. Multiple sequence alignment of fungal sequences recovered which possess the pertussis motif is provided in Figure 3.9. A maximum likelihood phylogenetic tree of recovered fungal sequences and representative metazoan Gγ sequences places the basal proteins expectedly at the base of the fungal group, which forms a distinct cluster from the metazoan outgroup sequences.

In chytrids, RGS homologs were predicted in *R. allomycis*, *S. punctatus*, and *B. dendrobatidis*. The two *Sp* proteins shared 22.1% and 18.8% identity with the *N. crassa* RGS protein NCU08319 [127], while the *Bd* and *R. allomycis* proteins shared 17% and 13.4% identity, respectively, with NCU08319. One of the *Sp* proteins (SPPG07577) contained a Gγ binding region motif (GGL domain: PF00631), while SPPG04061 did not. The *Bd* protein (BDEG00728) contained the GGL domain as well.

Phosphodiesterase proteins

A search of phosphodiesterase protein complement, associated with G protein signaling pathways, is presented in Figure 3.1). PDE α and β subunit homologs were recovered in the basal lineages. In most cases, the recovered protein had significant similarity to both α and β queries, which belong to the cGMP-specific PDE-6 subfamily. A maximum likelihood phylogenetic analysis (Figure 3.12) places recovered fungal proteins as distinct from metazoan lineages, with the exception of one sequence in *Hp*.

Phototaxis and Heterologous Protein expression

In order to determine the phototactic abilities of *B. dendrobatidis* and *S. punctatus*, I followed the procedure outlined in [49] for *Rhizophydium littoreum*. However, no phototaxis was observed in either *B. dendrobatidis* or *S. punctatus*. Furthermore, zoospores of *H. polyrhiza* were not collected at sufficiently high quantities for meaningful phototaxis observation.

Heterologous protein expression of the chytrid opsin proteins was not observed using either BdpHIL-S1 or SppHIL-S1 vectors, despite successful expression using NoppHIL-S1 (Figure 3.13).

3.4 Discussion

The goals for the research in this chapter were to assess, in the basal fungi, the complement of proteins which are secondarily involved in the rhodopsin-mediated photosignaling cascade (ie proteins other than the rhodopsin GCPR protein, including the intermediate heterotrimeric G proteins, and the effectors involved in the cAMP signaling and phosphotidylinositol pathways. From a functional perspective, this comparative study will demonstrate which of the basal fungi have a complete pathway, and would be helpful in understanding at which point in evolution these components were lost.

Gα subunits posessing N-terminal myristoylation and C-terminal pertussis motifs are members of the G_i subfamily. Transducin is one member of this family, and is known to associate with rhodopsin to function to activate cGMP-PDE. *S. punctatus*

and *A. macrogynus* posess Gα proteins which contain these motifs and which have high similarity to *N. crassa* Gna-1. In *N. crassa*, Δ*gna-1* strains are deficient in multiple pathways during both vegetative and sexual development, one of which (macroconidiation and mass accumulation) seems to be related to photosensing [128].

The Gα group in fungi can be classified into four families [114]. Chytrid Gα proteins are present in all four of these families. Multiple Gα proteins in *Bd* and *Hp* appear to fall within the fungal Group IV (for which there is no *N. crassa* homolog, however there is an *Ustillago maydis* homolog implicated in pathogenicity).

Phosphodiesterase is an enzyme which interacts with Gα. PDE subunits were found accross the fungal lineages, but in the basal lineages they appear in higher numbers relative to the non-flagellated dikarya. PDEγ subunits were found exclusively in the Blastocladiomycete lineages. In the Zygomycete lineages, some species have similar presence patterns to chytrids, while others have presence patterns similar to the dikarya lineages. Transducin is one class of Gα protein which is used in the mammalian visual cascade and which interacts with both rhodopsin and phosphodiesterase. This interaction activates PDE to lower the concentration of cGMP causing hyperpolarization of the cell and a decrease in calcium levels. The PDE subunits α / β subunits found in the fungi are distinct from those found in the metazoan lineages.

Despite the presence of presumably active photoreceptor proteins in *S. punctatus* and *B. dendrobatidis*, phototactic behavior was not observed in experiments similar to those described for *Rhizophydium littoreum* [49]. There is no prior evidence of phototaxis in these species, only observed in *Allomyces reticulatus* [48] and *R. littoreum* [49]. One hypothesis as to the lack of phototaxis is that perhaps the GPCR type 2 rhodopsins are not used in phototaxis, but govern another light-regulated response mechanism. While the receptors are present, the downstream associated effectors and

response-related protein components are not. This interpretation is supported by the fact that *Bd* and *Sp* do not posess PDE-γ subunit homologs, which are known to have functions related to hyperpolarization and which could thus be related to flagellar beating.

The lack of heterologous protein expression using the *Pichia pastoris* cloning system was unexpected given the successes enjoyed with expression of the Bovine rhodopsin [129] and an opsin protein identified in *N. crassa* [125]. This lack of expression for chytropsins from *B. dendrobatidis* and *S. punctatus* using the BdpHIL-S1 and SppHIL-S1 vectors, respectively, can potentially be attribituted to the long intracellular loop regions found in these proteins introducing sufficient disorder so as to impede proper folding and membrane integration. For comparision, the cytoplasmic loop 3 (CL3) regions in Bovine rhodopsin (1U19) and Nop1 are only 6 and 8 amino acids long, respectively, whereas the CL3 regions in *Bd* and *Sp* are 99 and 36 amino acids long, respectively.

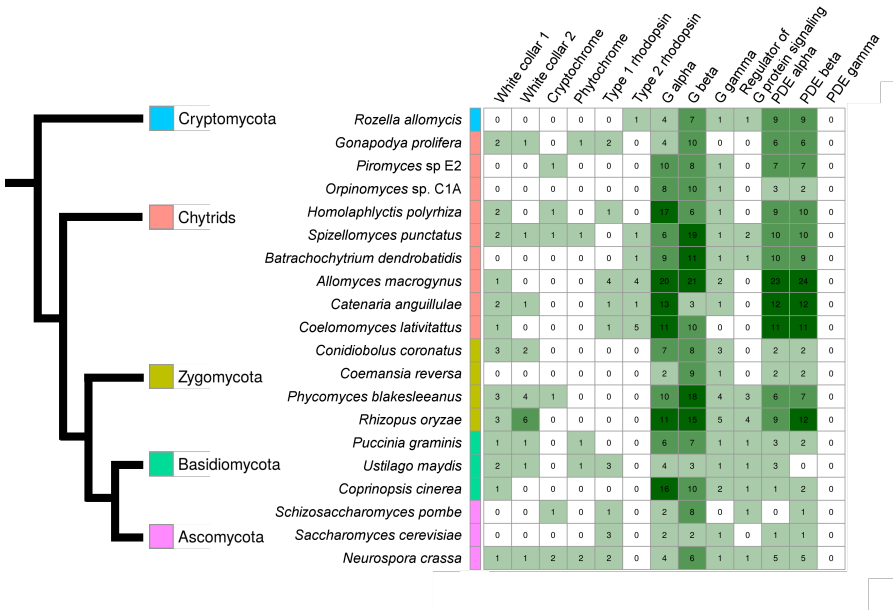


Figure 3.1: Photosensory protein distribution and G protein signaling pathway component distribution in representative subset of fungi.

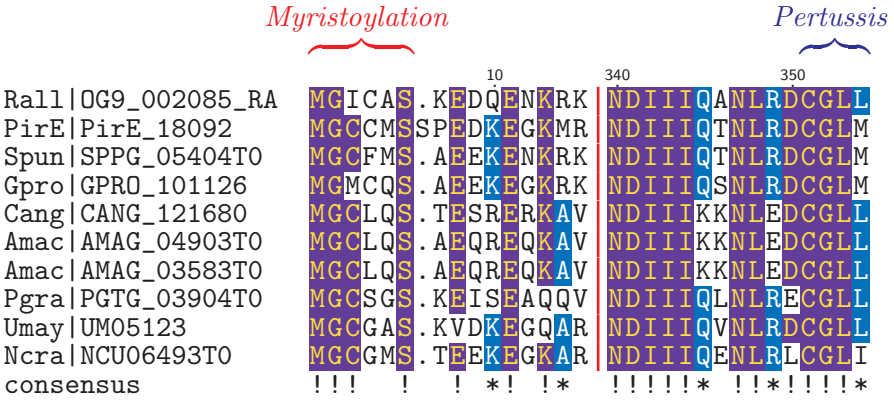


Figure 3.2: Multiple sequence aligment of N-termini and C-termini. Gα proteins identified in fungi which posessed both N-terminal myristoylation (MGXXXS) and C-terminal pertussis (C[GAVLIP]{2}X) motifs were aligned using T-coffee.

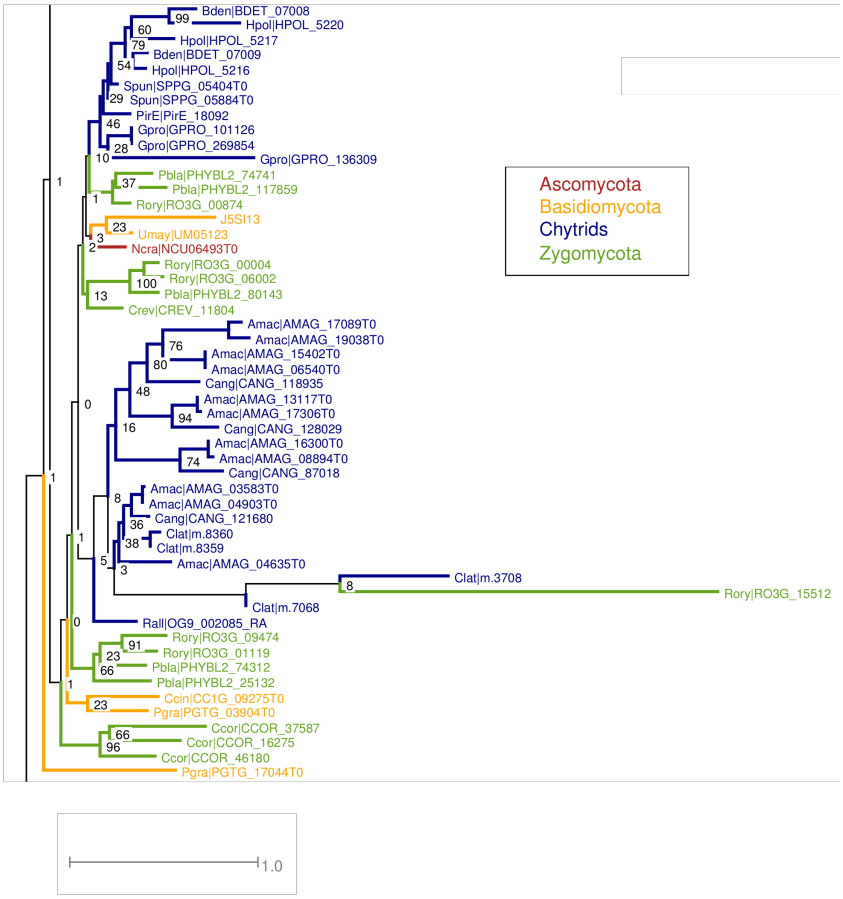


Figure 3.3: Maximum likelihood tree of identified G-α subunits in fungi (group I, as defined by inclusion of NCU06493)

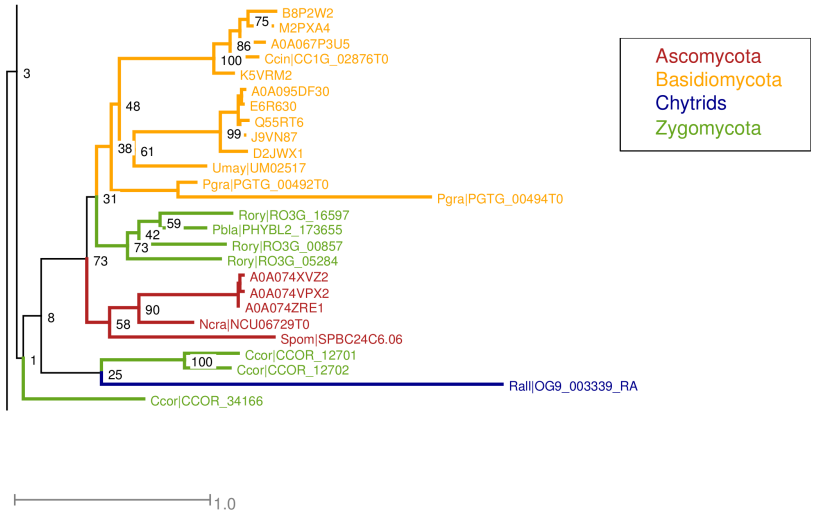


Figure 3.4: Maximum likelihood tree of identified G-α subunits in fungi (group II, as defined by inclusion of NCU06729)

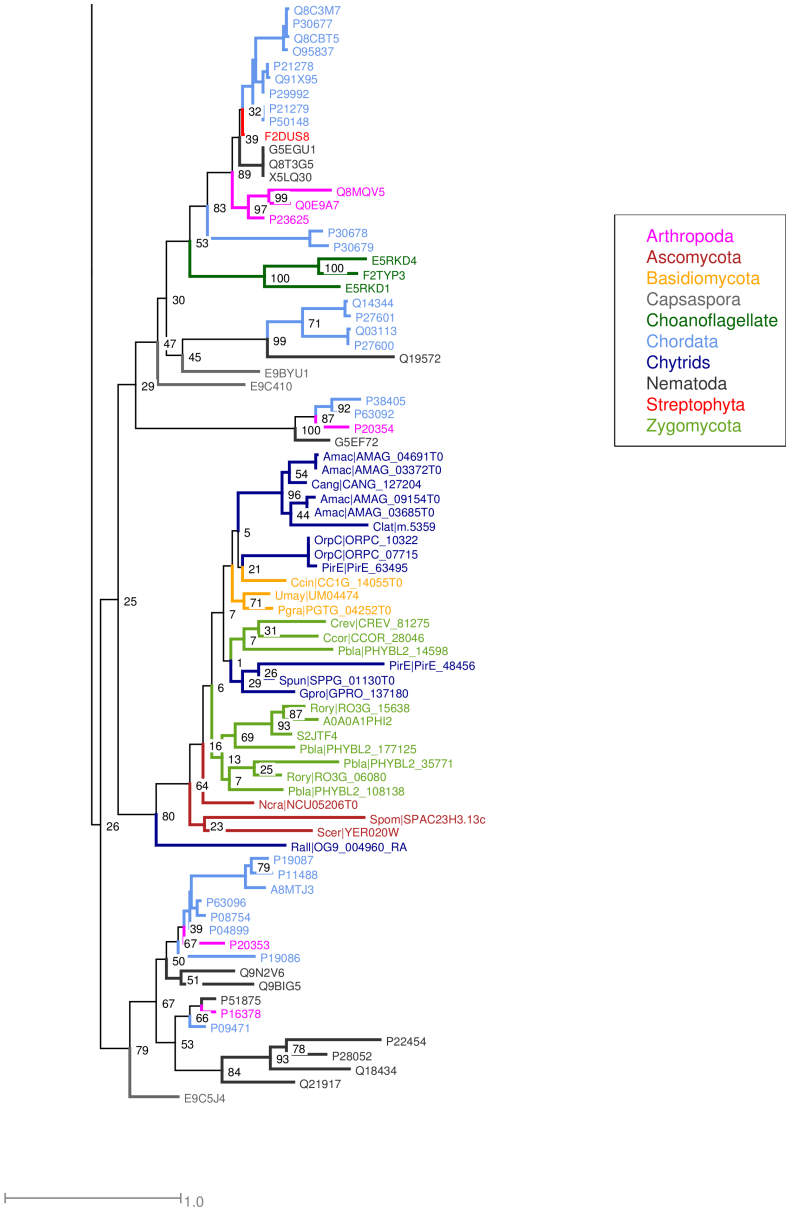


Figure 3.5: Maximum likelihood tree of identified G-α subunits in fungi (group III, as defined by inclusion of NCU05206) and outgroups

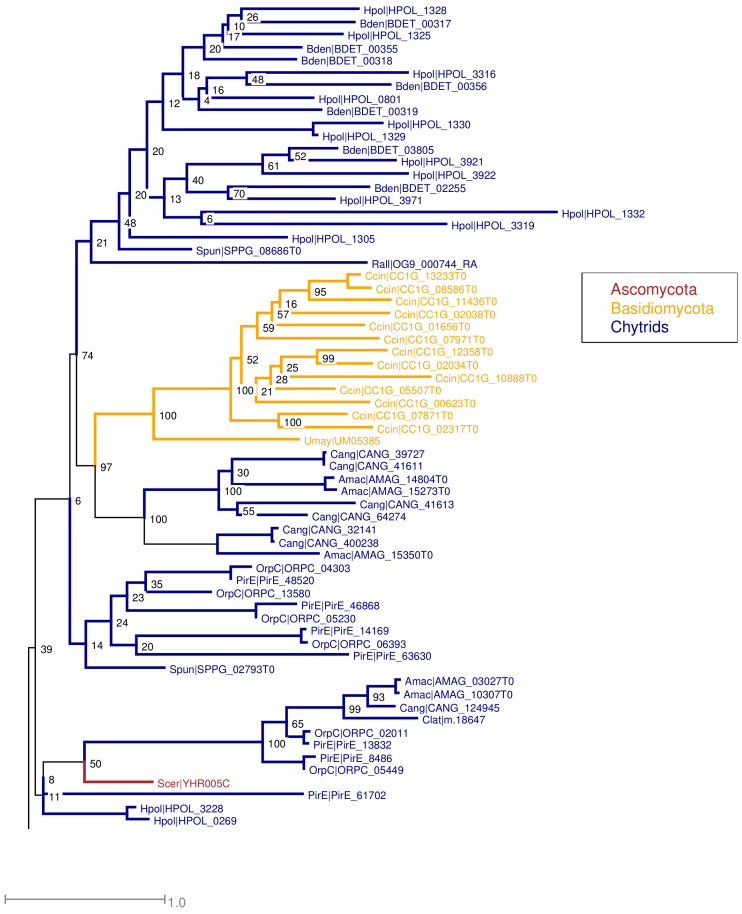


Figure 3.6: Maximum likelihood tree of identified G-α subunits in fungi (group IV, as defined by absence of *N. crassa* homologs and inclusion of UM05385 from *Ustilago maydis*)

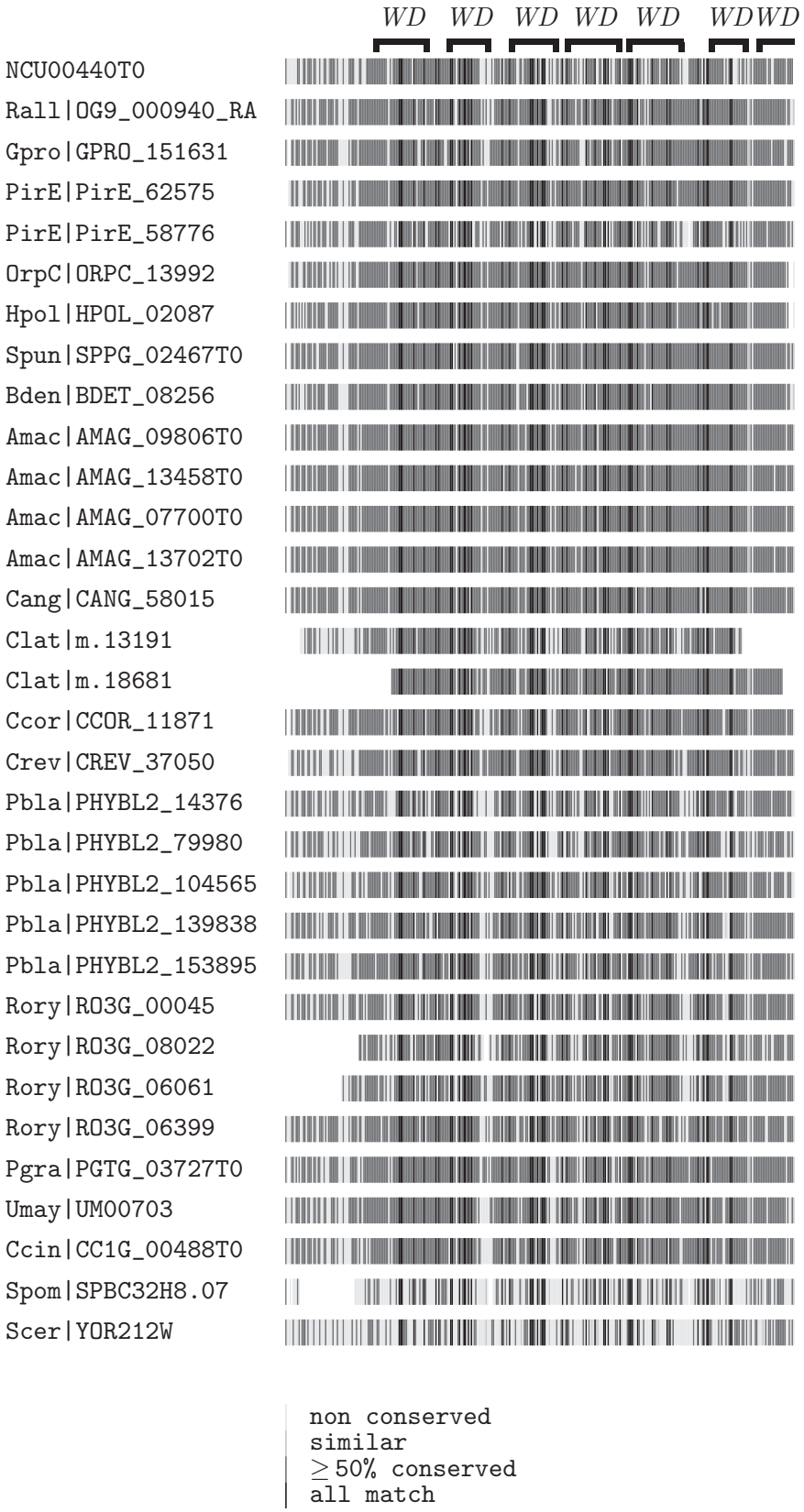


Figure 3.7: Schematic of identified fungal Gβ proteins highlighting conservation and location of multiple WD40 repeat domains

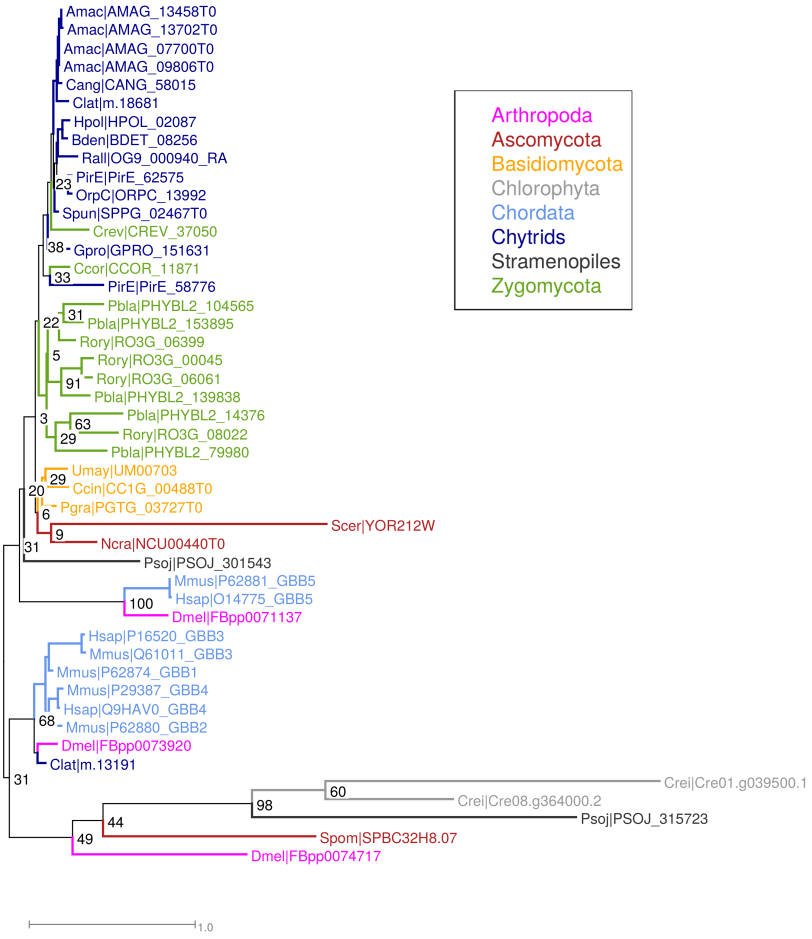


Figure 3.8: Maximum likelihood tree of identified Gβ subunits in Fungi and Metazoan outgroups

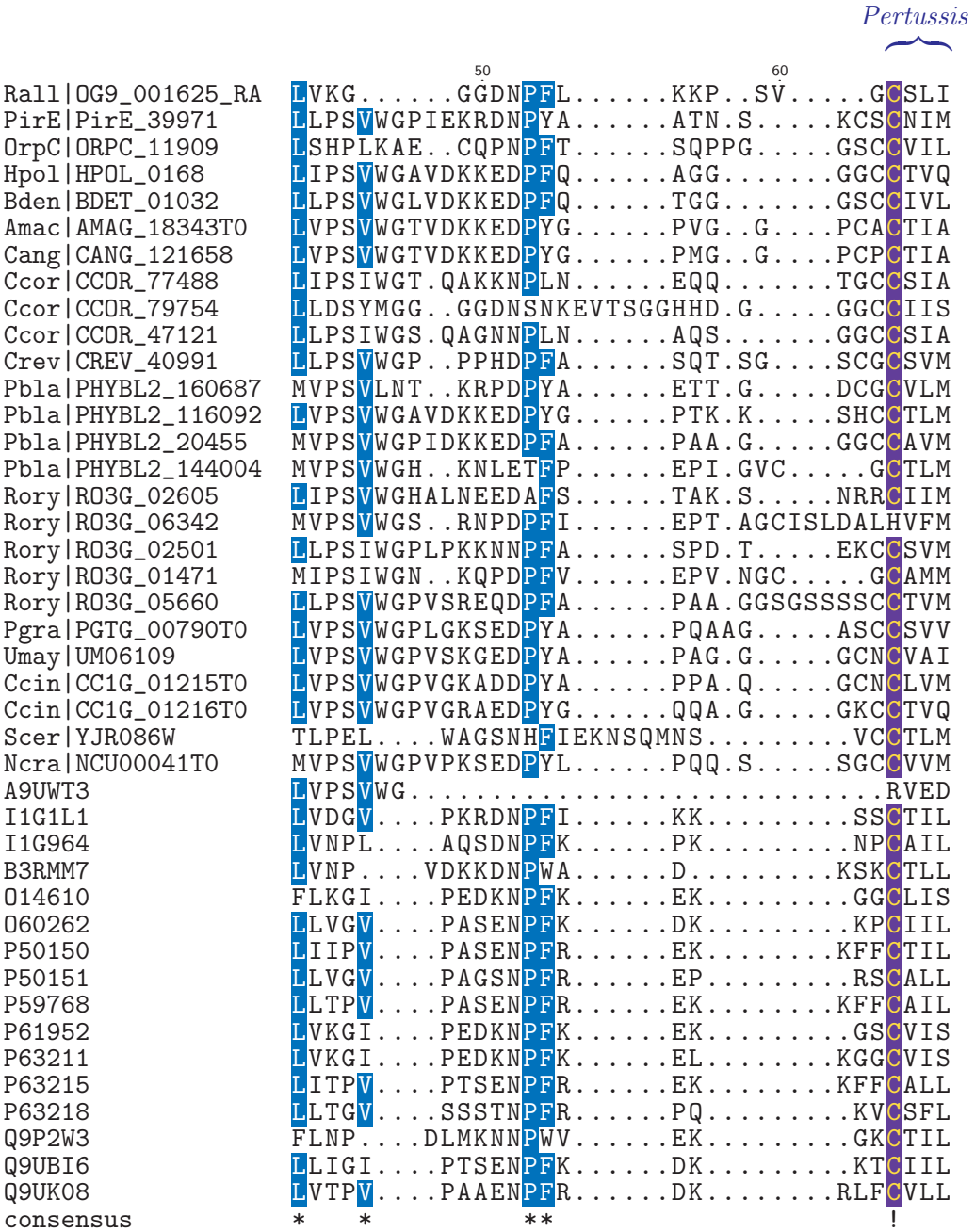


Figure 3.9: Multiple sequence aligment of C-termini of Gγ proteins identified in fungi which posessed C-terminal pertussis (C[GAVLIP]{2}X) motifs aligned using T-coffee.

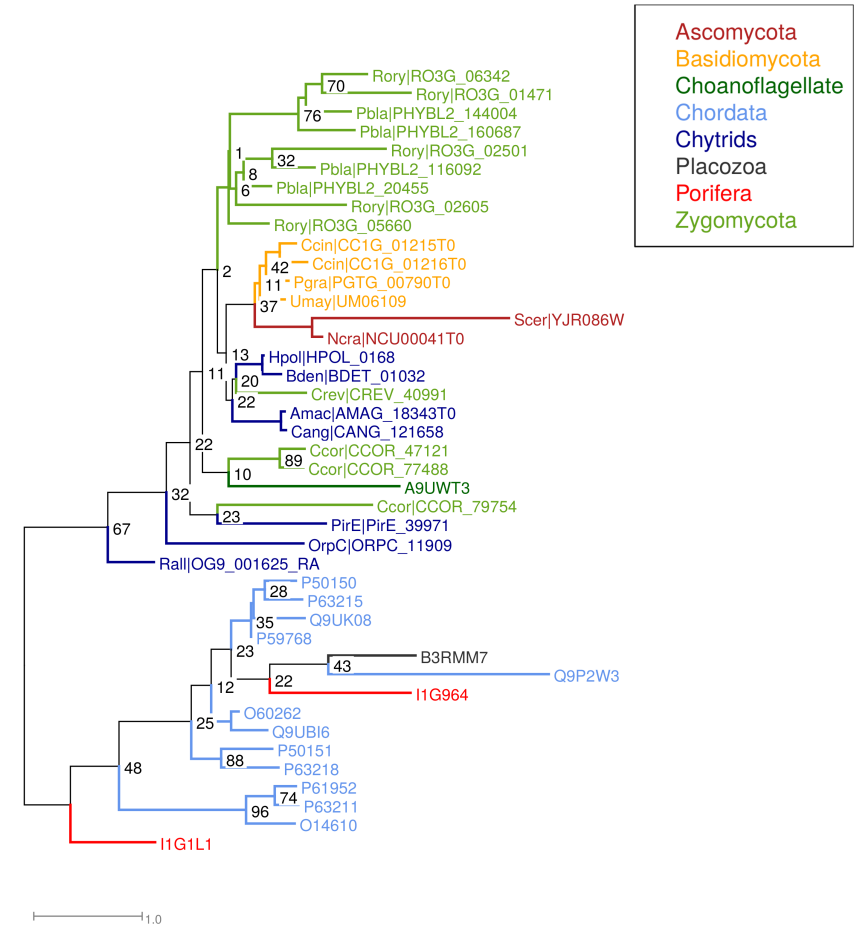


Figure 3.10: Maximum likelihood tree of identified Gγ subunits in fungi and animal outgroups

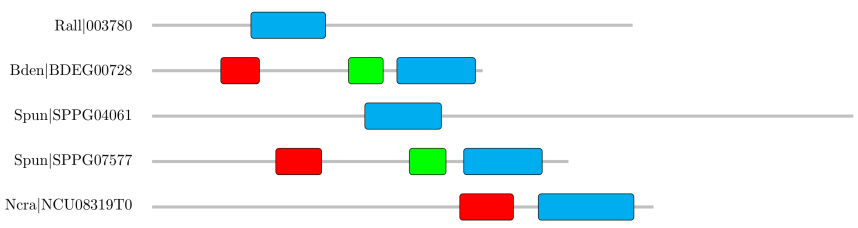


Figure 3.11: Structural domains of RGS proteins identified from HMM search in chytrids. Fungal hits contain RGS-domains (PF00615; blue). The *Bd* hit and one *Sp* hit additionally contain GGL (PF00631; green) and DEP (PF00610; red) domains. A second *Sp* hit, and the hit from *Rozella* only contain the RGS domain.

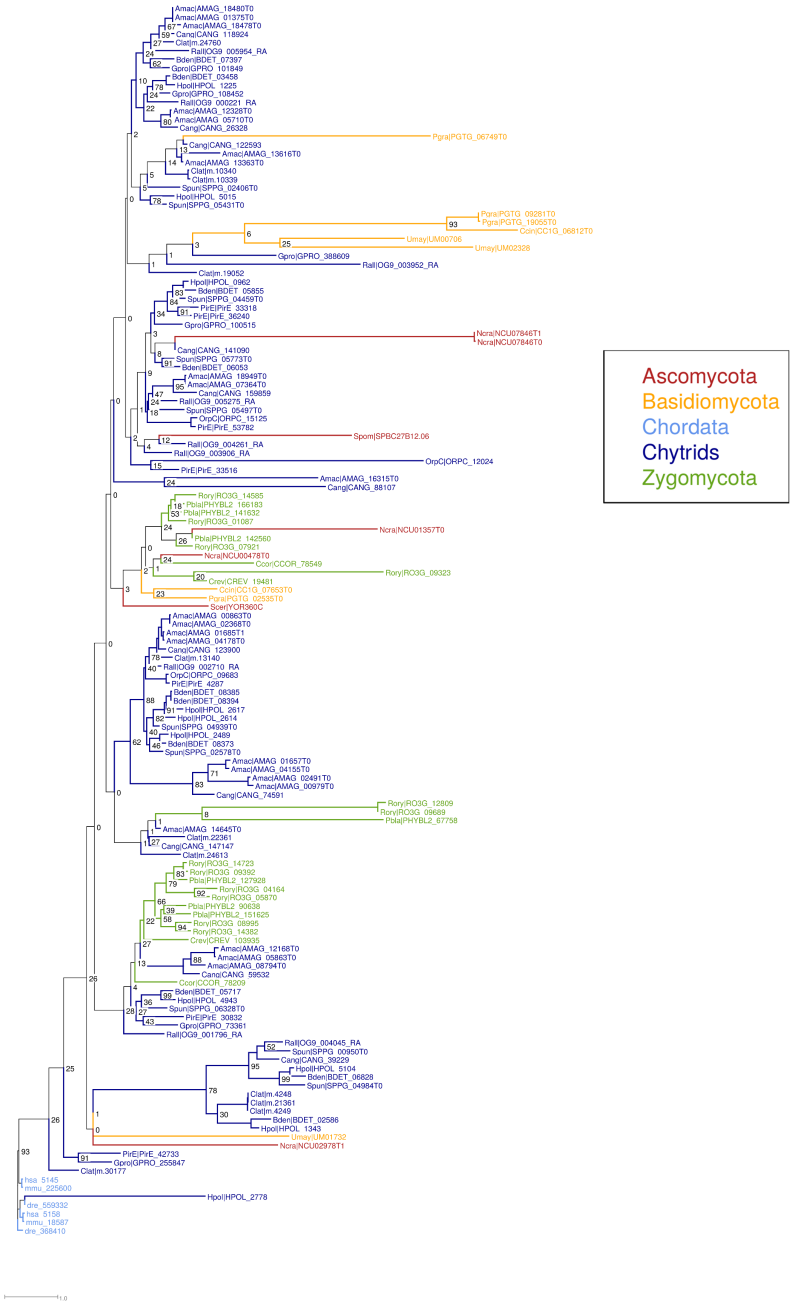


Figure 3.12: Maximum likelihood tree of phosphodiesterase α and β subunits recovered from HMM search in representative fungal groups. Representative metazoan sequences obtained from Uniprot are provided as outgroups.

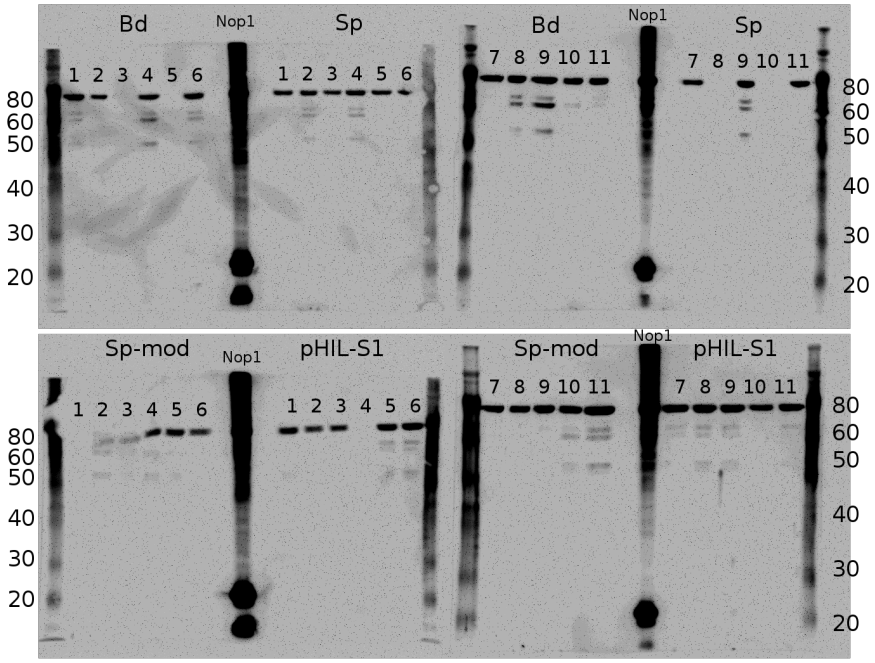


Figure 3.13: Western blots for heterologous *Pichia pastoris* expression of opsin constructs from *Bd* and *sp*, both native and missing lysine ("*Sp*-mod"). Expression of the expected Nop-1 protein from *Neurospora crassa* can be seen, and empty vector pHIL-S1 is included as a negative control.

Table 3.1: Comparision of recovered Gα subunits from basal fungal lineages to known Gα proteins in *Neurospora crassa* and *Ustilago maydis*. [a] Percent identity, when compared to NCU06493 (Group I), NCU06729 (Group II), NCU05206 (Group III), and UM05385 (Group IV); [p] Contains pertussis motif sequence; [m] Contains myrstoylation motif sequence.

Species	Class	Name	Protein ID	Identity[a]
Neurospora crassa	I	GNA-1	NCU06493^{pm}	100.00
-	II	GNA-2	NCU06729	100.00
-	III	GNA-3	NCU05206^{m}	100.00
Ustilago maydis	IV	GPA4	UM05385	100.00
Allomyces macrogynus	I	-	AMAG_03583^{pm}	67.51
-	I	-	AMAG_04903^{pm}	69.58
-	I	-	AMAG_19038	51.69
-	I	-	AMAG_04635^{m}	60.45
-	I	-	AMAG_17306^{m}	62.57
-	I	-	AMAG_13117^{m}	62.02
-	I	-	AMAG_15402^{m}	61.52
-	I	-	AMAG_06540^{m}	61.24
-	I	-	AMAG_16300^{m}	56.86
-	I	-	AMAG_08894^{m}	56.46
-	I	-	AMAG_17089	54.12
-	III	-	AMAG_03685^{m}	62.42
-	III	-	AMAG_09154^{m}	62.50
-	III	-	AMAG_03372^{m}	61.28
-	III	-	AMAG_04691^{m}	61.28
-	IV	-	AMAG_14804	31.25
-	IV	-	AMAG_10307	31.25
-	IV	-	AMAG_03027	30.24
-	IV	-	AMAG_15273	33.73
-	IV	-	AMAG_15350	31.82
Batrachochytrium dendrobatidis	I	-	BDET_07008^{m}	66.29
-	I	-	BDET_07009	70.68
-	IV	-	BDET_00356	32.02
-	IV	-	BDET_00319	32.92
-	IV	-	BDET_00318	35.94
-	IV	-	BDET_02255	32.64
-	IV	-	BDET_03805	36.65
-	IV	-	BDET_00317	33.87
-	IV	-	BDET_00355	34.31
Catenaria anguillulae	I	-	CANG_87018^{m}	56.42
-	I	-	CANG_118935^{m}	63.46
-	I	-	CANG_121680^{pm}	68.93
-	I	-	CANG_128029^{m}	58.50
-	III	-	CANG_127204^{m}	60.00
-	IV	-	CANG_41613	25.12
-	IV	-	CANG_124945	33.51
-	IV	-	CANG_39727	25.64
-	IV	-	CANG_64288	25.95
-	IV	-	CANG_64274	34.31
-	IV	-	CANG_400238	30.92
-	IV	-	CANG_41611	32.41
-	IV	-	CANG_32141	35.48
Coelomomyces lativittatus	I	-	Clat|m.7068^{m}	58.33
-	I	-	Clat|m.8359^{p}	74.52
-	I	-	Clat|m.3708	42.97
-	I	-	Clat|m.8360	71.79
-	IV	-	Clat|m.18647	31.85
Gonapodya prolifera	I	-	GPRO_136309	42.31
-	I	-	GPRO_269854^{m}	74.28
-	I	-	GPRO_101126^{pm}	74.22
-	III	-	GPRO_137180^{m}	61.80
Homolaphlyctis polyrhiza	I	-	HPOL_5216	74.09
-	I	-	HPOL_5217	59.70
-	I	-	HPOL_5220^{m}	59.52
-	IV	-	HPOL_3316	41.84
-	IV	-	HPOL_3319	30.77
-	IV	-	HPOL_0801	35.60
-	IV	-	HPOL_0269	30.69
-	IV	-	HPOL_3228	32.93
-	IV	-	HPOL_3922	37.25
-	IV	-	HPOL_1325	34.76
-	IV	-	HPOL_1328^{p}	41.30
-	IV	-	HPOL_3921	40.34
-	IV	-	HPOL_1332	32.81
-	IV	-	HPOL_1330	25.26
-	IV	-	HPOL_1305	45.78
-	IV	-	HPOL_3971	39.64
-	IV	-	HPOL_1329	38.68
Orpinomyces sp.	III	-	ORPC_07715^{m}	49.02
-	III	-	ORPC_10322^{m}	50.98
-	IV	-	ORPC_05449	27.72

-	IV	-	ORPC_06393	32.31
-	IV	-	ORPC_02011	28.79
-	IV	-	ORPC_05230	36.25
-	IV	-	ORPC_13580^{m}	34.76
-	IV	-	ORPC_04303	34.85
Piromyces sp.	I	-	PirE_18092^{pm}	73.73
-	III	-	PirE_48456^{p}	65.69
-	III	-	PirE_63495	42.50
-	IV	-	PirE_61702	27.51
-	IV	-	PirE_8486	34.12
-	IV	-	PirE_48520	37.44
-	IV	-	PirE_63630	31.03
-	IV	-	PirE_13832	29.92
-	IV	-	PirE_46868	44.05
-	IV	-	PirE_14169^{m}	33.05
Rozella allomycis	I	-	OG9_002085-RA^{pm}	69.77
-	II	-	OG9_003339-RA^{m}	30.11
-	IV	-	OG9_000744-RA	28.02
-	IV	-	OG9_004960-RA	36.80
Spizellomyces punctatus	I	-	SPPG_05404^{pm}	76.77
-	I	-	SPPG_05884^{m}	75.64
-	III	-	SPPG_01130^{m}	66.85
-	IV	-	SPPG_02793^{m}	40.65
-	IV	-	SPPG_08686	37.12

Table 3.2: Comparison of recovered Gβ subunits from basal fungal lineages to known Gβ protein from *Neurospora crassa*. [a] Percent identity when compared to NCU00440.

Species	Name	Protein ID	Identity[a]
Neurospora crassa	GNB-1	NCU00440	100.00
Allomyces macrogynus	-	AMAG_07700	73.10
-	-	AMAG_09806	73.10
-	-	AMAG_13458	72.22
-	-	AMAG_13702	72.22
Batrachochytrium dendrobatidis	-	BDET_08256	71.76
Catenaria anguillulae	-	CANG_58015	72.22
Coelomomyces lativitattus	-	Clat|m.18681	72.63
Gonapodya prolifera		GPRO_151631	71.88
Homolaphlyctis polyrhiza	-	HPOL_02087	71.68
Orpinomyces sp.	-	ORPC_13992	71.93
Piromyces sp.	-	PirE_58776	65.51
-	-	PirE_62575	71.35
Rozella allomycis	-	OG9_000940-RA	69.91
Spizellomyces punctatus	-	SPPG_02467	73.10

Table 3.3: Comparison of recovered Gγ subunits from basal fungal lineages to known Gγ protein from *Neurospora crassa*. [a] Percent identity when compared to NCU00041.

Species	Name	Protein ID	Identity[a]
Neurospora crassa	GNG-1	NCU00041	100.00
Allomyces macrogynus	-	AMAG_18343	50.70
Batrachochytrium dendrobatidis	-	BDET_01032	47.95
Catenaria anguillulae	-	CANG_121658	50.70
Homolaphlyctis polyrhiza	-	HPOL_0168	49.30
Orpinomyces sp.	-	ORPC_11909	33.85
Piromyces sp.	-	PirE_39971	43.24
Rozella allomycis	-	OG9_001625	42.25

Chapter 4

Transcriptome analysis of the anopholean pathogenic fungus *Coelomomyces lativittatus*

4.1 Introduction

Species of *Coelomomyces* belong to the phylum Blastocladiomycota, one of the basal fungal lineages. These species in general are obligate parasites which cycle between insect and crustacean hosts [130]. The lifecycle initiates begins when biflagellate zygotes encounter mosquito larvae. The spore settles on and attaches to the host cuticle, a process facilitated by the secretion of adhesion vesicles which contain a glue-like substance [131]. After secretion of a thin cell wall, the encysted spore develops an appressorium and penetration tube which breaks through the host cuticle [16]. Once inside the larval hemocoel, the spore develops into a sporangia. Host death liberates these sporangia. Meiosis within the sporangia produces haploid uniflagellate meiospores

of opposing mating types, which are subsequently released to individually infect crustacean hosts (typically copepods, though ostracods can serve as hosts as well [132]). The penetration of copepods is thought to occur in a manner similar to that of the mosquito larvae [16]. Gametangia develop from these meiospores within the copepod hemocoel, which are ultimately cleaved into gametes and released upon crustacean host death. In the environment once again, opposing gametes fuse to create biflagellate zygotes, which propagate the cycle by infecting new mosquito larvae [130].

Coelomomyces species have been studied previously as a potential system for biocontrol of mosquito populations [133]. While the potential for use as a biological control agent has been explored, the exact biochemical nature of mosquito infection, including descriptions of all enzymes and pathways involved, has not. However, the advent and development of genomic tools facilitates the study of these questions.

There are many examples of entomopathogenic organisms specializing in mosquito hosts, covering 13 genera across 2 kingdoms (Fungi and Chromista) [133]. The Ascomycete fungus *Metarhizium anisopliae* is one of the well-studied fungal models for investigations into this specialized group. Early research looked at the range of enzymes produced by pathogenic isolates of this fungus, and identified a variety including proteases, amino-/carboxy- peptidases, lipases, esterases, chitinases, NAGases, catalases, polyphenol oxidases, and deoxy- and ribonucleases [134]. Later studies added to this repertoire the production of toxic cyclic peptides known as destruxins [135].

The dual-host, multistage life cycle of *Coelomomyces*, which passes through a number of chemically distinct environments, suggests the presence of an elaborate sensory repertoire. For instance, experimental evidence demonstrates that gametes of some *Coelomomyces* species are specifically attracted to mosquito ovaries, and that this attraction is, at least in part, mediated by the hormone 20-hydroxyecdysone (20HE)

[136]. Other evidence demonstrates a species-specific, photoperiod-dependent periodicity of gamete release from the copepod host [137], strongly implying that *Coelomomyces* has the molecular capacity for some manner of circadian rhythm regulation.

Coelomomyces are known producers of β-carotene [138], the production of which is indicative of mating type, resulting in gametangia and gametes that are either strong orange (arbitrarily "male") or colorless/amber (arbitrarily "female"). β-carotene is ubiquitous in nature and exists primarily as a precursor for the biosynthesis of Vitamin A.

The total number of species of *Coelomomyces* worldwide is estimated to be several hundred, yet little is known about the more detailed aspects of biochemistry and genomics. Therefore, the work described in this chapter is motivated by an ongoing effort toward the assembly and annotation of a *Coelomomyces* transcriptome, which will not only add to the growing collection of knowledge about chytrid fungi broadly, but will also provide new insights into the underlying mechanisms that govern the alternating life cycle of *Coelomomyces* and can help further its development as a biological agent of mosquito control.

This research represents the first exploratory investigation of *Coelomomyces* genomics using the transcriptome of *C. lattivitatus*. In this chapter, I compare protein functions of expressed transcripts relative to other zoosporic fungi, biochemically reconstruct known pathways of carotenoid and retinal biosynthesis, and identify potential members of what is presumed to be a vast and complicated sensory network.

4.2 Methods

Mosquito and copepod cultures

Mosquito larvae and copepods used for maintenance of *C. lativittatus* cultures were *Anopheles quadrimaculatus* and *Acanthocyclops vernalis*, respectively. Cultures were maintained by the members of Dr. Brian Federici's entomology lab at UCR according to methods described previously [137].

RNA extraction and library preparation

The following RNA extraction and library preparation protocol was performed in its entirety by Rob Hice of the Federici lab. Infected copepods were cleaned with water, and placed in tubes with distilled water. After the fungal gametes emerged from the copepod larvae, the copepod carcasses were allowed to settle to the tube bottom and the fungal supernatant was transferred to a new tube. Gametes/zygotes were spun down at 6000xg for 3 minutes, and the supernatant was removed. Pellets were snap-frozen in liquid nitrogen. RNA was extracted with Trizol (Life Technologies, Grand Island, NY) as per the manufacturer's protocol. 1.2 μg of RNA was used as the starting material for the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, MA). Poly-A RNA was purified as per instructions and converted to adapter-ligated, size-selected cDNA. An aliquot of the library was cloned into pJet1.2 (Thermo Fisher Scientific, Waltham, MA) and clones sequenced with standard methods to check library quality. An aliquot was also run on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) to check average size, which was 371 bp (including adapters). The resulting library was sequenced by the Institute for Integrative Genome

Biology Core facility at the University of California at Riverside using the MiSeq instrument (Illumina, San Diego, CA).

Transcriptome assembly and annotation

Transcriptome assembly was carried out using Trinity (v. r2014-02-14) [139] with quality trimming performed using Trimmomatic [140], normalization using Trinity's default built-in normalization process, and reconstruction using PasaFly, an implementation of the PASA assembly algorithm [141]. ORF prediction was carried out using Transdecoder [141] using the Pfam database. Annotation was performed using Trinotate, also part of the Trinity package.

Pfam distribution

Pfam domain distribution for the *C. lativittatus* transcriptome was predicted using an HMM search against the Pfam database. Hits above a threshold of 1e-05 were retained. The top 20 most abundant domains were then queried in the *Allomyces macrogynus, Catenaria anguilullae, Batrachochytrium dendrobatidis, Spizellomyces punctatus*, and *Homolaphlyctis polyrhiza* proteomes to determine relative abundance and putative expansion of domains.

Insect virulence survey

Enzymes predicted to be related to insect virulence were used as search queries against the *C. lativittatus* transcriptome. To search for peptidase proteins, an HMM

profile constructed from the 157 seed sequences for the Pfam domain family PF00112 (Peptidase_C1) was used as a query in HMMER (v3.0) *hmmsearch* against basal fungal proteomes (Table B.1), and the resulting hits were filtered according to presence of the functional catalytic diad residues Cys25 and His159. Identified hits were trimmed such that only the mature peptide (defined as the C-terminal 200 residues, lacking the signal and precursor regions) were used in the subsequent analyses. A maximum likelihood tree using the seed sequences and the filtered fungal sequences was generated using RAxML (v7.5.4) using 100 bootstrap replicates and the GTR+Γ+WAG substitution model. Similarly for the Trypsin proteins, an HMM profile constructred from the 71 seed sequences for the Pfam domain family PF00089 (Trypsin) was used in an HMMsearch. The resulting hits were filtered according to presence of the functional catalytic triad residues His57, Asp102, and Ser195. A maximum likelihood tree was constructed using RAxML (v7.5.4) using 100 bootstrap replicates and the GTR+Γ+WAG substitution model. The Ecdysone receptor from *Drosophila melanogaster* (GI:157318) was used as a query using *ssearch*36 at a threshold of 1e-5. A maximum likelihood tree using human nuclear receptors from all major families and EcR receptors identified in the majority of insect orders was generated using RAxML (v7.5.4) [123] using 100 boostrap replicates and the GTR+Γ+Dayhoff protein model.

To predict potential adhesion related proteins, I obtained a dataset from the Fungal Adhesin and Adhesin-like Database (FaaDB; `http://bioinfo.icgeb.res.in/faap/faap.html`) containing experimentally verified fungal adhesins from predominantly Dikarya. This positive dataset was then used as a query in a FASTA search (using *ssearch*36) against the *C. lativittatus* transcriptome. Hits matching an e-val threshold of 1e-10 were submitted to the FAApred SVM-based prediction method, which was trained on both positive and negative adhesin datasets. Positive matches had an SVM

score greater than -0.8.

TMHMM (v2.0) [142] was used to predict transmembrane proteins in the basal fungal proteomes and *C. lativittatus* transcriptome. Those with 6-9 domains were retained and compared using OrthoMCL [143].

β-carotene survey

To assess the completeness of the β-carotene pathway in *C. lativittatus*, I queried the transcriptome using the three β-carotene biosynthesis enzymes, using the sequences obtained from *Blastocladiella emersonii*: phytoene desaturase (Uniprot: KJ468786), lycopene cyclase / phytoene synthase (Uniprot: KJ468785), and β-carotene 15,15'-monooxygenase (BCMO1) (Uniprot: KJ468787). While functional biochemical characterization of these specific *B. emersonii* enzymes has not been performed, a BLASTP search against the SwissProt database reveals expected top hits with experimental verification of biochemical activity. For the *B. emersonii* phytoene dehydrogenase (UniProt ID: A0A060GS52) is the phytoene desaturase from *Phycomyces blakesleeanus* NRRL 1555 (UniProt ID: P54982; Functional reference: [144]). For the lycopene cyclase / phytoene synthase (UniProt ID: A0A060GVE0), the top hit is the bifunctional lycopene cyclase/phytoene synthase from *Mucor circinelloides f. lusitanicus* (UniProt ID: Q9UUQ6; Functional reference: [145]). Finally for the carotenoid dioxygenase (UniProt ID: A0A060GW07) the top hit is the β,β-carotene 15,15'-monooxygenase from *Gallus gallus* (UniProt ID: Q91993; Functional reference: [146]). Additionally, HMM profiles were generated from sequences available from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [147, 148]: phytoene desaturase, K10027 (442 Bacterial proteins, 49 Archaeal proteins); phytoene synthase, K02291 (6 Eukaryote proteins, 45

Plant proteins); β-carotene 15,15'-monooxygenase, K00515 (64 Metazoan proteins).

Due to the lack of hits in *C. lativittatus* for the *B. emersonii* phytoene desaturase sequence, I made additional queries using the phytoene desaturase from *Giberella fujikuroi* (CarB; UniProt accession: Q8X0Z0) and *Neurospora crassa* (NCU00552).

A maximum likelihood tree was constructed using RAxML (v7.5.4) with the identified BCMO1 fungal sequences and those from KEGG families using the GTR+Γ+LGF model and 100 bootstrap replicates.

Photosensory survey

Putative photosensory proteins were identified using known fungal photobiology proteins WC-1 ,WC-2, FRQ, VIVID, and FWD-1 from *N. crassa* [127]. Cryptochrome homologs were searched using HMMER (v3.0) *hmmsearch* with an HMM generated from the 20 seed sequences in the Pfam domain PF12546 (Cryptochrome_C). Hits were retained above a threshold of 1e-20. Similarly, phytochrome homologs were searched using *hmmsearch* with an HMM generated from the 80 seed sequences in PF00360 (PHY). Hits were retained above a threshold of 1e-20.

Opsin proteins were identified as containing the PF00001 (7tm_1) or PF01036 (Bac_rhodopsin) domains after prediction from Trinotate as described above. Additional support was provided by TMHMM (v2.0) as containing either 6 or 7 transmembrane domains. BacOpsin-GC fusion proteins were identified as posessing both the PF01036 and PF00211 (Guanylate_cyc) domains as predicted by Trinotate.

4.3 Results

Transcriptome Characterization

After quality trimming, obtained a total of 28,698,279 reads with an average length of 196 nt. De novo assembly of reads using Trinity [139] yielded 77,597 transcripts with an average length of 386 bp. Within these transcripts, 21,486 open reading frames (ORFs) were predicted using Transdecoder [141]. Annotation with Trinotate predicted 12,156 transcripts with a BLASTp hit, 11,040 with predicted Pfam domain(s), and 29,076 with associated GO terms.

The top 20 Pfam domains identified in the *C. lativittatus* transcriptome and their respective counts in other chytrids are provided in Table 4.1. The most striking examples of domain families which are underrepresented among other chytrids are trypsin (PF00089), glycoside hydrolase family 47 (PF01532), and papain family cysteine protease (PF00112), all three of which have some manner of protease or carbohydrate degrading functionality. An additional family which appears to be overrepresented in *C. lativittatus* is the Myosin tail family (PF01576), although the related Blastocladiomycete *C. anguillalae* also has a higher number of these proteins relative to other Blastocladiomycete and Chytridiomycete speceis. Corresponding Gene Ontology (GO) Slim classifications for recovered transcripts are shown in Figure 4.1.

Insect virulence

To test for the presence of and possible expansions in gene families that may be related to insect virulence, I scanned the *C. lativittatus* transcriptome for specific protein domains which have been previously implicated in fungal associated insect vir-

ulence, or which may be otherwise related to fungal-insect association.

Proteases- The C1 cysteine proteases are commonly found in fruit (eg. papaya) and often used as meat tenderizers. The enzymes from fig, pineapple, and papaya plants have been studied as antihelmintics and found to have high proteolytic activity against nematode cuticles [149]. The family is characterized by the Peptidase C1 (PF00112) and C1-like (PF03051) Pfam domains. The *C. lativittatus* transcriptome contains at least 56 transcripts containing peptidase C1 domains ($< 98\%$ identity). Searches of Blastocladiomycete and Chytridiomycetes genomes found no proteins containing these domains, although proteins with this domain are present in the Dikarya lineages. Phylogenetic analysis of the Pfam seed sequences and a reduced set of the fungal copies revealed a number of observations (Figure 4.4). First, the *C. lativittatus* transcripts are broadly distributed, with very few tight clusters. None of the transcripts cluster within the other Dikarya sequences; instead they fall sporadically among the other metazoan sequences, predominantly with other lower eukaryotes. One group of transcripts cluster as more recent divergences closer to the arthropod sequences.

Trypsins are serine proteases found in the digestive systems of many vertebrates [150]. These enzymes are characterized by the PF00089 Pfam domain, and 43 transcripts in *C. lativittatus* were identified as having this domain. Searches of other non-insect associated Blastocladiomycte and Chytridiomycete proteomes revealed an order of magnitude fewer proteins containing these domains. Based on the presence of the catalytic triad, a collection of three residues (His-57, Asp-102, Ser-195) which are critical to active site function, 20 *C. lativittatus* sequences were filtered and used in a phylogenetic analysis along with Pfam seed sequences. Most of the Chytrid sequences cluster with other fungi and away from the metazoan sequences.

Destruxins- The destruxins are a class of insecticidal cyclic hexadepsipep-

tides produced by some entomopathogenic fungi, most notably by species of Metarhizium [151, 135]. Based on chemical differences in the hydroxy acid, R group, and N-methylation characteristics, these compounds can be divided into a total of 12 chemically distinct classes [152, 135]. The biosynthesis of these compounds is presumed to be mediated by an NRPS gene cluster in *Metarhizium robertsii* [135]. A FASTA search with the destruxin synthase (dtxS1) protein in the *M. robertsii* gene cluster did not identify any putative homologs in our *C. lativittatus* transcriptome. No putative NRPS or PKS-related proteins searching transcriptome using antiSmash [153], though m.15019 (described in β-carotene results as a phytoene synthase) was recovered as a putative terpene synthase. Additionally, no hits for THIOL or CON using HMM searches were recovered [154]. Some hits from AMP HMM, but counts are on the order of other chytrids (15-20).

Chitin related domains- Chitin binding domains are a broad class of domains found in carbohydrate-active proteins. Overall, there are 71 different subfamilies within this broad class defined by sequence similarity in the Carbohydrate Active Enzymes database [155]. Five predicted ORFs were identified by InterPro as having a CBM18 domain and six ORFs identified with a CBM33 domain. One transcript (m.4968) was identified as posessing a chitin synthase domain (PF03142), and one transcript (m.4725) with a NADH-Ubiquinone domain (PF00361).

Adhesion-related proteins- In the infection process, when biflagellate zygotes encounter mosquito larvae, the spore is observed to settle on and attach to the host cuticle. This process is hypothesized to be facilitated by the secretion of so-called "adhesion vesicles" which contain a glue-like substance [131]. These vesicles have been observed developing prior to the attachment of the spore, localizing to points of contact between the spore and cuticle, and disappearing after host penetration [131]. While the chemical

nature of these "adhesion vesicles" remains unclear, a number of candidates exist. Fungal adhesins, for example, are membrane proteins which allow certain fungi to attach to surfaces and are usually involved in microbial community biofilm formation. One well studied example is the hyphal wall protein (Hwp1) implicated in *Candida albicans* pathogenesis [156]. However a FASTA search using *ssearch*36 did not recover any homologs of this protein in *C. lativittatus* or other Blastocladiomycete or Chytridiomycetes surveyed.

In an additional attempt to ascertain the nature of spore-cuticle attachment, I probed the FaaDB for putative adhesins in *C. lativittatus*. This method identified 16 sequences as putative adhesins. In the other Blastocladiomycte and Chytridiomycetes surveyed, 10, 5, 10, 4, and 4 proteins were predicted as such in *A. macrogynus, C. anguillulae, B. dendrobatidis, H. polyrhiza*, and *S. punctatus*, respectively.

Ecdysone receptors- The naturally occurring ecdysteroid hormone 20-hydroxyecdysone (20HE) controls moulting in arthropods [157]. There is evidence to suggest that 20HE plays a role in attracting *Coelomomyces stegomyiae* to the ovaries of adult female *Aedes aegypti* [136]. A FASTA search using *ssearch*36 with the known ecdysone receptor protein from *D. melanogaster* EcR [158] identified a single *C. lativittatus* transcript. This finding is surprising not only as it provides a straightforward answer to how *C. lativittatus* could sense its host, but also given the presumption that nuclear receptors are only limited to the metazoan lineages and not found in fungi [159]. An HMM profile constructed from arthropod EcR receptor sequences and human nuclear receptors, when searched against the *C. lativittatus* transcriptome, identified an additional three transcripts, though the originally identified transcript (m.9546) remained the highest scoring. An alignment is provided in Figure 4.6.

This 298 aa transcript is likely not full length, and only aligns to the DNA

binding region of the *D. melanogaster* receptor (approximate residues 239 to 401). The top blast hit for this transcript is the *Caenorhabditis elegans* nuclear hormone receptor family member nhr-35 (SwissProt accession: Q17771, e-val 2e-20). InterProScan [160] predicts the PF00105 domain covering positions 24-92. This domain is a Zinc Finger C4-type and is associated with nuclear receptors. No orthologs of the *C. lativittatus* transcript were detectable in any other chytrids searching with an e-value threshold of at least 1e-05.

Structurally, this transcript is most similar to the DNA-binding region of the *D. melanogaster* ecdysone receptor (PDB ID: 2HAN, chain B) [161]. These two regions have 42% sequence identity. A homology-based structure model of the *C. lativittatus* transcript using SwissModel [162] has an RMSD of 0.2 (Dali Server prediction [163]) when compared to 2HAN, chain B.

The PF00104 ligand binding domain, associated with this and other nuclear receptors in the arthropod receptors, was not predicted to be associated with this transcript. However one other *C. lativittatus* ORF is predicted to contain the PF00104 domain but has insignificant similarity to the *D. melanogaster* EcR protein (m.10080, 21.5% identity, e-val: 0.077), and as such is not a likely homolog. Nonetheless, Table 4.5 lists BLASTP results after searching the SwissProt database with m.10080. The top 5 hits are all to mammalian liver X receptors (LXRs). A total of six hits using a cutoff threshold of 1e-06 are to 20HE receptors from insects. All hits have approximately 40% coverage and approximately 25% identity.

A maximum likelihood tree (Figure 4.7) constructed from arthropod 20HE sequences, as well as human nuclear receptor sequences from all nuclear receptor families, shows the *C. lativittatus* putative DNA-binding homolog sequence falling outside of the metazoan nuclear receptor sequences.

Finally, to determine if any other, non-hormone based receptors are uniquely found in *C. lativittatus* relative to the other, non-insect associated chytrids, I searched for possible receptor candidate genes based on transmembrane domain architecture. In *C. lativittatus*, 131 transcripts were predicted to have between 6 and 9 transmembrane domains. Of these, 29 are specific and not found in the other Chytridiomycota or Blastocladiomycota genomes surveyed based on ortholog clusters generated with OrthoMCL [143]. These 29 transcripts form 12 unique paralog clusters. HMMER (v3.0) searches of the Pfam database identified domains in 8 of these clusters, while the other 4 remained unclassified (Table 4.2).

β-carotene

C. lativittatus likely has a typical β-carotene biosynthesis pathway, despite missing an enzyme in the transcriptome (Figure 4.3). To determine the molecular characteristics of the β-carotene biosynthesis and metabolism pathways in *C. lativittatus*, I queried the predicted ORFs from the transcriptome with three key enzymes from the biosynthesis pathway described in *Blastocladiella emersonii* [126]. While functional biochemical characterization of these specific *B. emersonii* enzymes has not been performed here or otherwise, a BLASTP search against the SwissProt database reveals expected top hits with experimental verification of biochemical activity.

The biosynthesis of β-carotene utilizes three enzymes and acts upon geranylgeranyl pyrophosphate. Phytoene synthase converts two molecules of geranylgeranyl pyrophosphate to one molecule of phytoene. Phytoene desaturase then works in a five-step pathway to convert phytoene into lycopene [60]. Lycopene cyclase finally acts to convert the lycopene to β-carotene [61]. The lycopene cyclase and phytoene synthase

enzymes (fulfilling the first and last steps) are encoded as a single polypeptide.

No candidate *C. lativittatus* homolog was found with the putative *B. emersonii* phytoene dehydrogenase sequence (KJ468786) in either the set of predicted ORFs (using the protein sequence in a direct search) nor in the set of assembled transcripts (using the protein sequence in a translated search). Additional queries using phytoene desaturase from *Giberella fujikuroi* (CarB; UniProt accession: Q8X0Z0) and *N. crassa* (NCU00552) were similarly unsuccessful.

One transcript, m.15019, contained a 599-aa long predicted ORF, and was identified as a putative homolog to *B. emersonii* bifunctional lycopene cyclase / phytoene synthase (KJ468785) at 38.4% identity. The best BLASTX hit of the *C. lativittatus* transcript for this ORF against the SwissProt database was a "bifunctional lycopene cyclase/phytoene synthase" from *Phycomyces blakesleeanus* (UniProt accession Q9P854; e-val 3e-95). The m.15019 transcript has an FPKM value of 3.42.

The conversion of β-carotene to retinal is facilitated by β-carotene 15,15'-monooxygenase (BCMO1). A FASTA search using the *B. emersonii* putative carotenoid dioxygenase sequence (KJ468787) identified two transcripts contained ORFs which were identified as putative homologs, m.16827 (670-aa, 44.2% identity) and m.4639 (156-aa, 26.6% identity). The top BLASTP hit against SwissProt for m.16827 was BCMO1 from *Homo sapiens* (UniProt accession Q9HAY6; e-val: 1e-44), and that for m.4639 was BCMO1 from *Mus musculus* (UniProt accession Q9JJS6; e-val: 3e-09). These transcripts had FPKM values of 2.57 and 0.997, respectively.

To provide additional support for the candidate transcripts identified above, HMM profiles were generated from sequences available from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [147, 148]. The candidate *C. lativittatus* transcripts described above were also recovered using *hmmsearch* with these HMM profiles.

Sequence searches identified all three of these β-carotene metabolism genes in the genomes of two Blastocladiomycota fungi, *Allomyces macrogynus* and *Catenaria anguillulae*. Similar searches of the genomes of the Chytridiomycota fungi *Batrachochytrium dendrobatidis*, *Homolaphlyctis polyrhiza*, and *Spizellomyces punctatus*, found an incomplete complement of these genes of this pathway. The *B. dendrobatidis* genome contains no homologs for any of these genes, while the *H. polyrhiza* genome contains a candidate phytoene desaturase homolog (top BLASTP hit against SwissProt: phytoene desaturase from *P. blakesleeanus* [P54982.1], e-val: 2e-68, 49% identity), and *S. punctatus* possesses a candidate β-carotene oxygenase homolog (top BLASTP hit against SwissProt:β,β-carotene 9',10'-oxygenase from *Macaca fascicularis* [Q8HXG8.2], e-val: 1e-37, 26% identity).

To assess the phylogenetic history of BCMO1, a maximum likelihood tree was constructed from homologs found in Ascomycete, Chytridiomycete, Blastocladiomycete, and Zygomycete species (Figure 4.8). Examination of the resulting gene tree topology provides strong support for the early-diverging fungal genes to cluster distinctly outside the metazoan gene lineages, and suggests at least 3 major duplication events. At least one duplication occurred exclusively in the metazoan lineages to give rise to BCMO1 and BCDO2. One duplication likely occurred prior to the fungal/metazoan divergence, resulting in the copies seen in the Chlorophyta, Dikarya, Zygomycota. A second duplication likely occurred after the divergence of the fungi and prior to the divergence of the Cryptomycota, resulting in two subtypes of fungal β-carotene oxygenase. Interestingly, while copies of each subtype can be found in members of the Zygomycota, only one or the other is found in the Chytridiomycota and Blastocladiomycota.

Photosensing capacity

To determine the possible nature of observed photosensory capacity in *C. lativittatus*, I searched the transcriptome for ORFs predicted to be associated with photobiology (Table 4.3) using known fungal photobiology proteins, including opsins and opsin-like proteins, circadian rhythm proteins (WC-1 and 2, FRQ, and FWD-1), cryptochromes, phytochromes, and the photoreceptor protein VIVID.

A search for homologs to *N. crassa* White Collar-1 (WC-1, NCU02356) and White Collar-2 (WC-2, NCU00902) proteins identified a 368 aa ORF (m.12730) as a potential WC-1 homolog. The putative homologs are reciprocal best-blast hits as NCU02356 is the top BLASTP hit against SwissProt using this m.12730 as a query (UniProt id: Q01371, e-val: 4e-43, 35.2% identity). Additionally, m.12730 is predicted by InterPro to contain two PAS domains (IPR000014), similar to NCU02356. However, it is much shorter: only 367 aa compared to 1167 aa for NCU02356. Other WC-1 homologs were recovered from *H. polyrhiza, S. punctatus, C. anguillalae*, and *A. macrogynus*. While all of the recovered chytrid proteins are shorter than that of *N. crassa*, the lengths range from 367 to 857aa. An alignment of the conserved PAS domains from these proteins is provided in Figure 4.9 along with the full lengths of the predicted proteins. No ORFs were predicted as WC-2 homologs, nor were there any other potential PAS-domain containing transcripts in the *C. lativitattus* transcriptome.

In addition to the white collar complex proteins WC-1 and WC-2, the blue-light sensitive photoreceptor protein VIVID (VVD) identified in *N. crassa* and other filamentous fungi is a small (186 aa), cytoplasmic flavoprotein that responds to increasing light intensity [164]. No putative homologs of the *N. crassa* VVD protein (NCU03967) were recovered in a BLASTP search against the *C. lativittatus* transcriptome. Additionally,

no homologs were observed in the other Blastocladiomycete and Chytridiomycete species surveyed. This absence is consistent with an observed absence of VVD homologs outside of the Sordariomycete lineages.

Phytochromes are, among other things, circadian rhythm regulators in plants [51], and Velvet A homologs are demonstrated to be regulators for secondary metabolism and sporulation in several fungi [165]. There were no putative phytochrome homologs identified in *C. lativittatus* using an HMM generated from seed sequences in Pfam family PF00360, nor any homologs of the phytochrome-associated Velvet protein using either the *N. crassa* Velvet A-like protein NCU01731 or an HMM generated from *N. crassa* and additional *Aspergillus* and *Fusarium* sequences. While phytochromes are known to be present in the Chytridiomycete *Spizellomyces punctatus* (see [50]), there were no homologs for Velvet or phytochrome observed in other members of the Chytridiomycota and Blastocladiomycota.

A total of 6 ORFs were predicted to be opsin-related proteins based on predicted Pfam domains and predicted seven transmembrane helical domain architecture (Table 4.3). Of particular note is a predicted 537 aa ORF (m.7819), which has two identifiable Pfam domains: a 213 aa region with similarity to bacterial rhodopsin (PF01036; e-val: 4.6e-22), and a 178 aa region with similarity to guanylate cyclase (PF00211; e-val: 3.6e-51). This architecture is similar to that found in *Allomyces macrogynus* and *Catenaria anguillalae*, and described more fully in *Blastocladiella emersonii* [126]. A homolog was also identified in *Homolaphlyctis polyrhiza*. When compared with the other examples of this protein architecture found in the Blastocladiomycota and Chytridiomycota, this *C. lativittatus* transcript shares 61.91%, 72.98%, 71.19%, 64.20%, 63.36%, and 54.55% identity with the *B. emersonii*, each of the four *A. macrogynus*, and the *H. polyrhiza* proteins, respectively.

To ascertain the putative placement of this *C. lativittatus* Opsin-GC fusion protein among other opsin proteins, a maximum likelihood tree was generated using opsin sequences from Avalar et al., with additional inclusion of the opsin-GC fusion sequences recovered from *H. polyrhiza, C. anguillalae, and C. lativittatus* (Figure 4.2). The *C. lativittatus* and *C. anguillalae* sequences cluster expectedly with the other Blastocladiomycete sequences (*A. macrogynus* and *B. emersonii*) in a well supported early-diverging fungal group. Perhaps unexpectedly, the sequence from the Chytridiomycete *H. polyrhiza* falls within, rather than outside of, the Blastocladiomycete sequences, albeit with a relatively long branch.

4.4 Discussion

The research presented in this chapter dealt with one member of only known group of insect pathogens in the basal fungal lineages, *Coelomomyces lativittatus*. The initial transcriptome generation and analysis represents the first generation of bioinformatic resources for this fungus and sought to support the phylogenetic placement within the Blastocladiomycota through three traits of *C. lativittatus* biology: I) insect pathogenicity, II) β-carotene biosynthesis, and III) photosensing capacity.

Coelomomyces lativittatus is a member of the only known genus of insect pathogens among the basal fungal lineages and has been well-studied as a potential mosquito control agent. This transcriptome study represents the first attempt at developing available genomic and proteomic resources for this and other *Coelomomyces* species. Future work will most assuredly expand on the results demonstrated here, including whole genome sequencing, developmental and life stage-specific RNA sequencing,

and proteomic extraction and characterization. Nonetheless, some observations from the analyses performed here are useful in the comparative genomics of non-insect associated early-diverging fungi, and can also provide a focus for future work in *C. lativittatus*.

While there are several examples of entomopathogenic fungi, *C. lativittatus* and other members of *Coelomomyces* are the only known members of the Blastocladiomycota which demonstrate this association, and the biological mechanism by which *Coelomomyces psorophorae* infects mosquito larvae has been documented previously [131, 16]. In general, infection is initiated by the settling of the spore onto the host cuticle, followed by encystment and secretion of thin cell wall. The appearance of an appressorium and subsequent development of a penetration tube which pierces the integument of the host then allows the fungus to enter the host hemocoel [16].

As noted previously in *C. psorophorae* [131], there is a correlation between disruption of the outermost layer of the cuticle and accumulation of an amorphous, electron-dense material at the cuticle-contacting tips of penetration tubes. As the appressorium tip is the site of actual penetration through the cuticle and into the mosquito larvae, a speculative explanation of the observable electron-dense material would be the proteases and other degradation-related proteins unique to *C. lativittatus* recovered in this study. Indeed, a hypothesis postulated at the time suggested that this material may be enzymatic in nature [131].

A comparison of counts of the top 20 Pfam domains (Table 4.1) suggests four protein families which appear to be uniqely expanded in *C. lativittatus* relative to the other non-insect associated Blastocladiomycete and Chytridiomycete species. These include "myosin tail" (PF01576), "glyco_hydro_47" (PF01532), "trypsin" (PF00089), and "C1 peptidase" (PF00112). Of these four, the latter two have clear protease and degradation functions. While the PF00112 phylogenetic history is a little unclear, at least

two *C. lativittatus* specific groups can be identified, one of which appears to cluster with known arthropod sequences. An additional *C. lativittatus* sequence clusters with the plant sequences from papaya and pineapple, known to have culticle degrading activity in nematodes. Further work, especially life-stage dependent RNAseq experiments are completely necessary to confirm this hypothesis and would be critical in order to map expression levels of these and other protease genes before, during, and after infection. Aspects of infection which cannot be described further in this study are the adhesion vesicles, which are hypothesized to secrete the "glue" that attaches the spore to the host, and the observed pseudopodia structures, which appear after the settling of the spore.

Searches for a 20-hydroxyecdysone receptor based on similarity to known arthropod receptors identified at least one candidate transcript with similarity to the DNA binding domain of the ecdysone receptor in *D. melanogaster*. This DNA binding domain profile, Pfam id: PF00105, is specifically associated with nuclear receptors. Sequence, structure, and phylogenetic analysis suggests that while it is significantly diverged in sequence, it may be a hormone receptor and is unlikely to be the result of a horizontal transfer event or sequence contamination. Furthermore, homologous sequences are not found in other Blastocladiomycete and Chytridiomycete species surveyed, providing a tantalyzing explanation of *Coelomomyces* observed affinity for 20-hydroxyecdysone from mosquitoes. However, I am hesitant to hold this as undeniable evidence of presence of this receptor in *C. lativittatus*; rather it is submitted as a starting point for future analyses. Further work to evaluate gene expression changes in this and other transcripts when C. lativittatus is exposed to the anophelid larvae or the 20-hydroxyecdysone hormone are absolutely crucial and will provide better insight into these candidate genes.

Sensory results are consistent with previous hypotheses, but undetermined if this represents a specific insect association aspect.

The presence of a moderately complex sensory network governing the full *C. lativittatus* life cycle can be inferred from experimental research on other *Coelomomyces* species. For example, *C. psorophorae* zygotes need to seek out *Culiseta inornata* larvae [130]. Once infected, the zygotes must develop into sporangia, regulate the timing of meiospore release, and these meiospores need to find the crustacean host: the copepod *Cyclops vernalis* in the case of *C. psorophorae* [166]. Once inside the crustacean host, similar regulation of sporangial development and spore release must also take place, but in a much different environment. Reared under identical conditions, dehiscence of *C. dodgei* and *C. punctatus* occurs at significantly different times [137] suggesting the presence of a photoperiod dependent regulatory mechanism. These spores must then seek out members of the opposite mating type [137], fuse to form zygotes, and exhibit phototactic capabilities to swim upwards to the water surface [137].

Given this evidence in other *Coelomomyces* species, *C. lativittatus* likely also possesses a complex sensory network relative to chytrids which display no insect association. The demonstrated ability for photoperiod regulation prompted our search for transcripts predicted to be involved in photosensing. From this search, one putative homolog of the *N. crassa* White collar-1 protein was identified. The remaining components of the white collar / circadian rhythm process (White collar-2, FRQ, FWD-1), however, were not recovered. The White collar-2 protein is present, however, in the chytridiomycete *S. punctatus* and the Blastocladiomycete *C. anguillulae*, but is absent in the other Chytridiomycete and Blastocladiomycete species surveyed. Therefore it is not necessarily unusual to find its incomplete presence in *C. lativittatus*, especially given

the limitations of the current transcriptome study.

Several proteins predicted to be opsins or opsin-related were identified based on transmembrane domain architecture and Pfam domain identification. Notably, one transcript is predicted to have a type 1 microbial rhodopsin domain fused with a guanylate cyclase domain. This structure is similar to a novel fusion protein recently described in the Blastocladiomycete *B. emersonii*, and additional homologs can be identified in the genome assemblies of other Blastocladiomycetes *A. macrogynus* and *C. anguillulae* [126]. The mechanism of activity of the fusion protein described for *B. emersonii* is that light activates the type 1 rhodopsin domain, which in turn activates the coupled GC domain. This facilitates synthesis of cGMP, which activates K^+-selective cyclic nucleotide gated channels. Voltage-activated Ca^{2+} channels, activated by the resulting hyperpolarization of the plasma membrane, would elevate Ca^{2+} levels, prompting interaction with the flagellum and ultimately mediating phototaxis [126].

The presence of this fusion protein in *C. lativittatus*, in addition to its previously described presence in *B. emersonii, A. macrogynus*, and *C. anguillulae*, supports the hypothesis that the novel fusion gene appeared prior to the divergence of the Blastocladiomycota lineage as it can now be said to be present in all three of the Blastocladiaceae, Catenariaceae, and Coelomomycetaceae families. Furthermore, its presence in the Chytridiomycete *H. polyrhiza* suggests that the fusion may have appeared earlier. However the fusion architecture does not appear in any of the other Chytridiomycete genomes surveyed, suggesting that its presence in *H. polyrhiza* is the result of either a recent fusion event, duplication and losses in the other Chytridiomycete lineages, or HGT event.

β-carotene biosynthesis and metabolism pathways are present and nearly complete in the *C. lativittatus* transcriptome.

Endogenous β-carotene production is biologically important for many reasons, one of which is that it functions as precursor to retinal, the critical component of rhodopsin-mediated photoreception. Carotenoid production in the Blastocladiomycota is well known, and while γ-carotene is the predominant molecule in *B. emersonii* and several Allomyces species, β-carotene specifically is known to be produced by *C. dodgei*. In this and other *Coelomomyces* species, the relative levels of β-carotene are indicative of mating type, implying that either the production and/or regulation of β-carotene is at some level influenced by or related to the same mechanics which govern sexual reproduction. The extent of this relationship remains to be explored.

The presence of nearly all critical enzymes in the retinal biosynthesis pathway in *C. lativittatus* is consistent with these previous observations about, and suggests a fairly straightforward biological mechanism for, β-carotene production in *Coelomomyces*. Additionally, the presence of this pathway, coupled with the identification of multiple opsin-related transcripts, suggests that *Coelomomyces* has the biochemical capacity for rhodopsin-mediated photoreception.

However, the lack of a phytoene dehydrogenase homolog, the first step in β-carotene biosynthesis, is unusual given its presence in related Blastocladiomycetes A. macrogynus, *B. emersonii*, and *C. anguillulae*. This absence suggests that either this gene is not transcriptionally active during the life stage sampled, mRNA transcripts from this gene were not recovered at detectable levels during RNA extraction, or *C. lativittatus* uses a novel mechanism for conversion of phytoene to lycopene to produce β-carotene.

The enzyme responsible for conversion of β-carotene to retinal is β-carotene monoxygenase. This is one of two enzymes capable of cleaving β-carotene, the other being β-carotene dioxygenase. Two transcripts were recovered bearing similarities to the predicted BCMO1 protein from *B. emersonii*, and strong similarity to BCMO1 profiles generated from known metazoan sequences. A phylogenetic reconstruction positions these transcripts expectedly within a Blastocladiomycete specific group of monooxygenase homologs, itself within a fungal-specific group. This suggests at least one duplication event occurred after the divergence of Fungi from the metazoan lineages.

These findings presented in this chapter are descriptive and represent the first insights into the deeper molecular biology of this insect pathogen. The extent of the proteins, pathways, and networks studied in this work can be elucidated more completely once an annotated genome and life-stage specific transcriptomes are generated. For example, the diversity and copy number of sensory proteins actually present in the genome is likely to be higher than those captured by this transcriptome study, and a clearer picture of β-carotene biosynthesis will almost certainly be observed.

Nonetheless, this initial work still provides a perspective on the underlying biology in the single chytrid pathogen of insects. Near-term future work build upon these findings and deal with obtaining a draft reference genome from *C. lativittatus*, enhanced transcriptome sequencing accounting for important developmental timepoints, and proteomics studies dealing with surface receptors necessary for environment sensing. Potential long term applications of this and other work include exploitation as a means of mosquito population control.

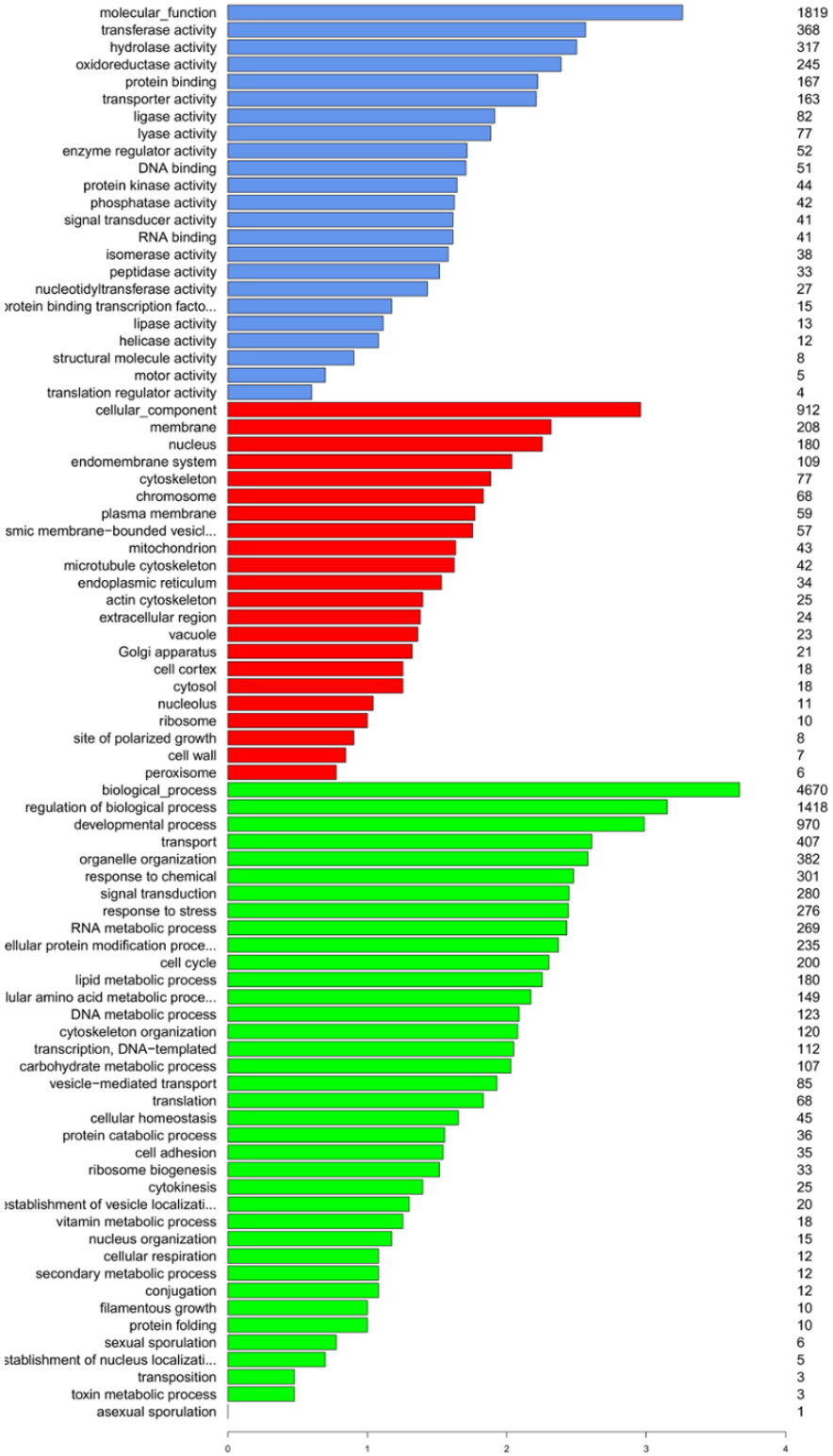


Figure 4.1: Distribution of *Aspergillus* GO-Slim terms in the *C. lativitattus* transcriptome.

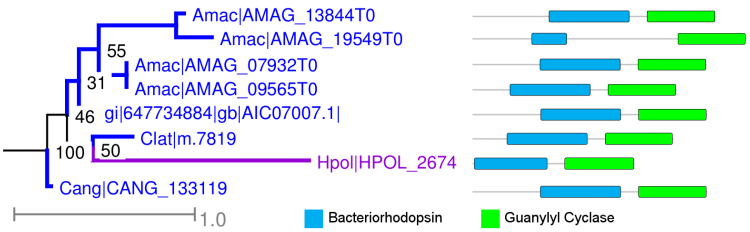
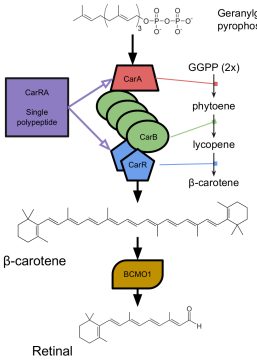


Figure 4.2: A subset of bacterial opsin-guanylate cyclase fusion proteins. Proteins with similar architecture as described for *B. emersonii* identified in Blastocladiomycete and Chytridiomycete proteomes, and *C. lativitattus* transcriptome.



Organism	phytoene dehydrogenase: CarB	lycopene cyclase / phytoene synthase: CarRA	β-carotene monooxygenase: BCMO1
A. macrogynus	✔	✔	✔
B. emersonii	✔	✔	✔
C. anguillulae	✔	✔	✔
C. lattivitatus		✔	✔
B. dendrobatidis			
H. polyrhiza	✔		
S. punctatus			✔

Figure 4.3: Presence and absence of β-carotene related proteins in species belonging to the Chytridiomycota and Blastocladiomycota.

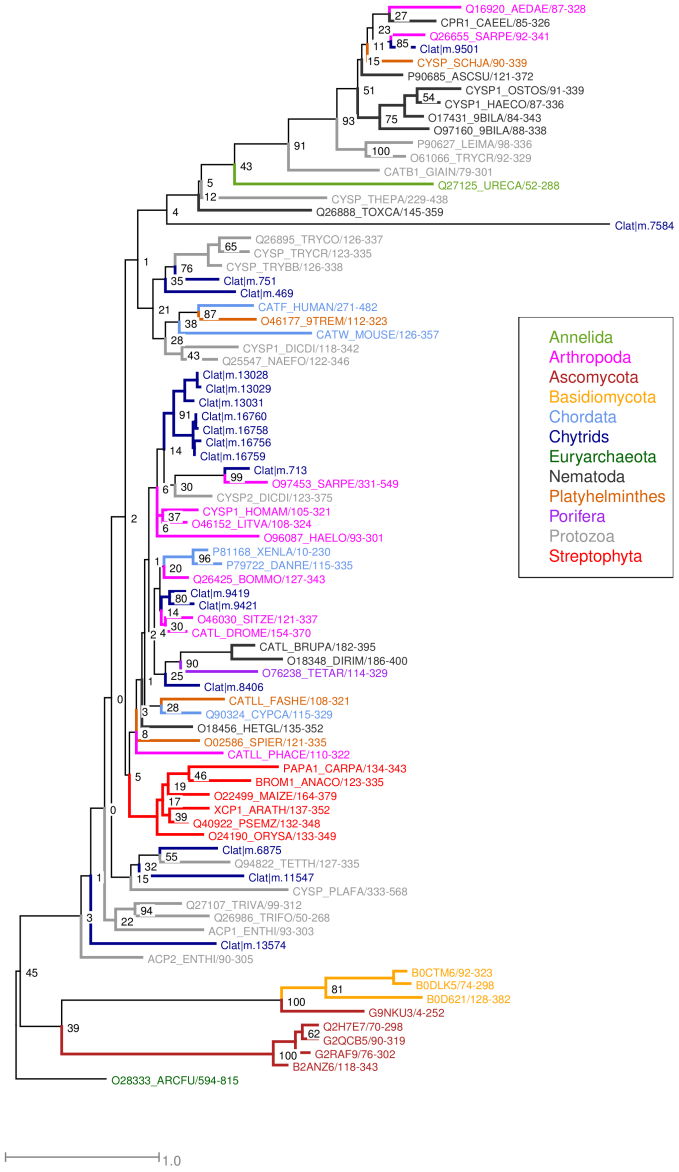


Figure 4.4: RAxML tree of select PF00112 seed sequences, unique *C. lativitattus* protein sequences, and sequences from Dikarya species.

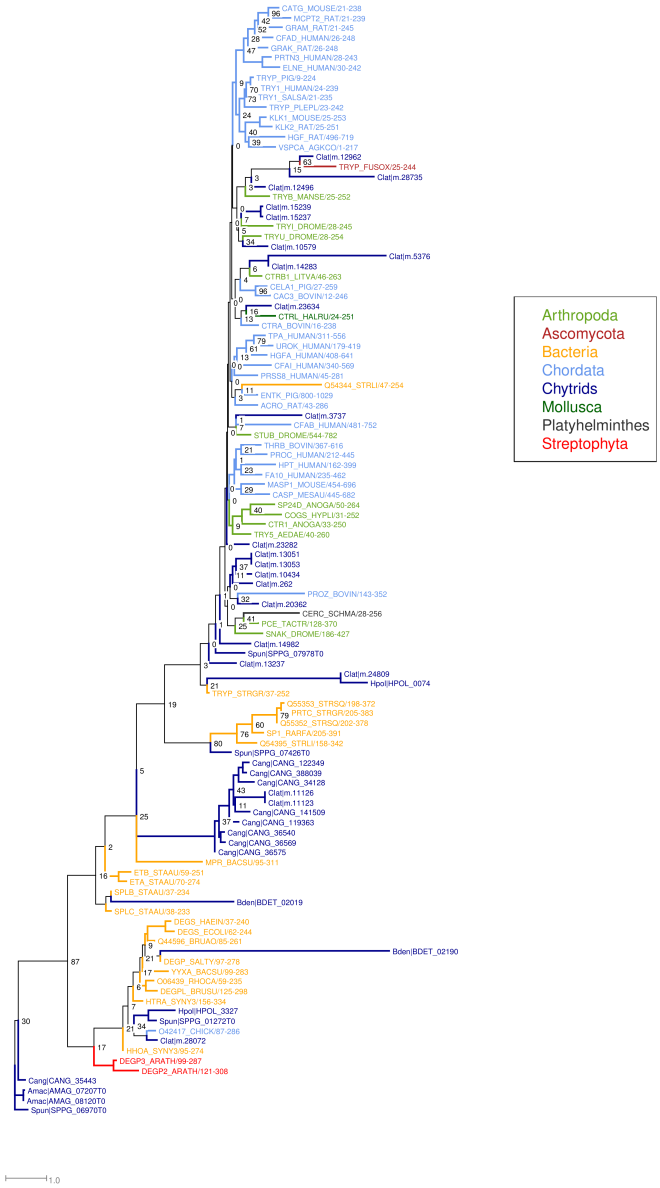


Figure 4.5: RAxML tree of select PF0089 seed sequences and unique *C. lativitattus* protein sequences

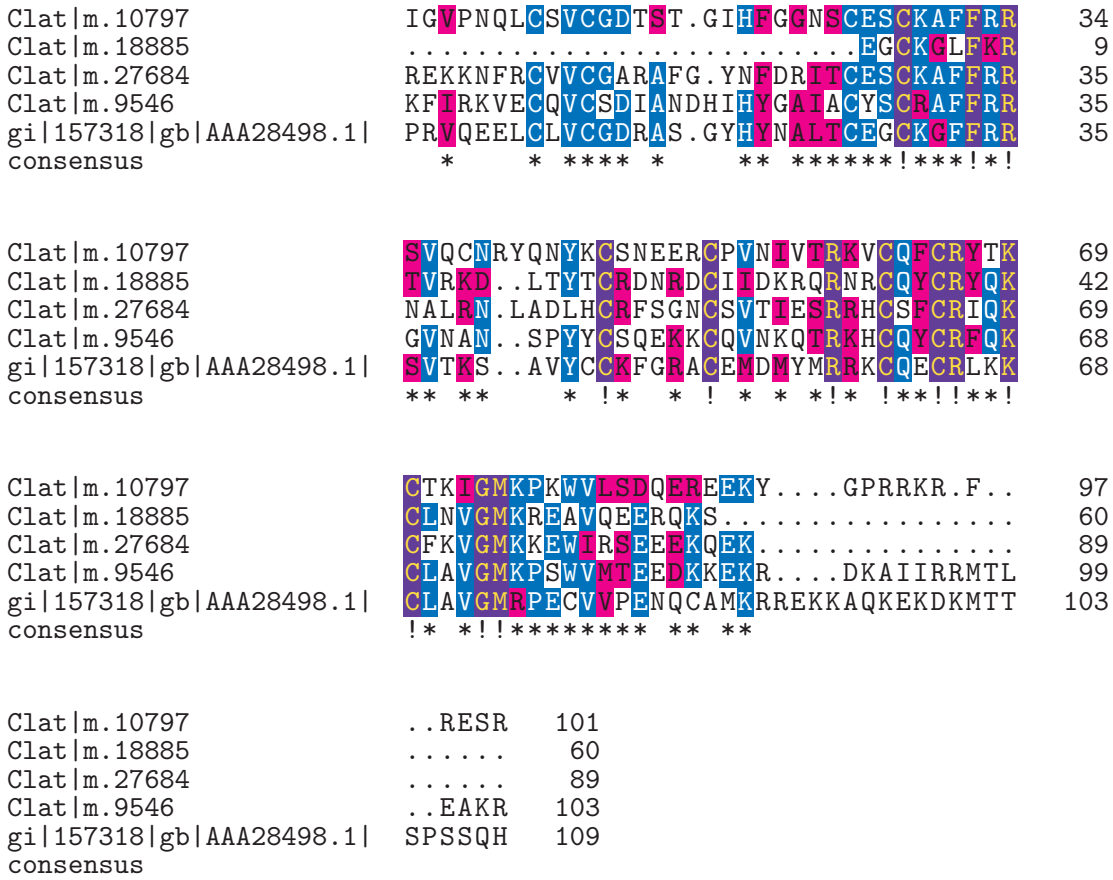


Figure 4.6: Alignment of *D. melanogaster* EcR and putative 20HE receptors identified in *C. lativittatus*

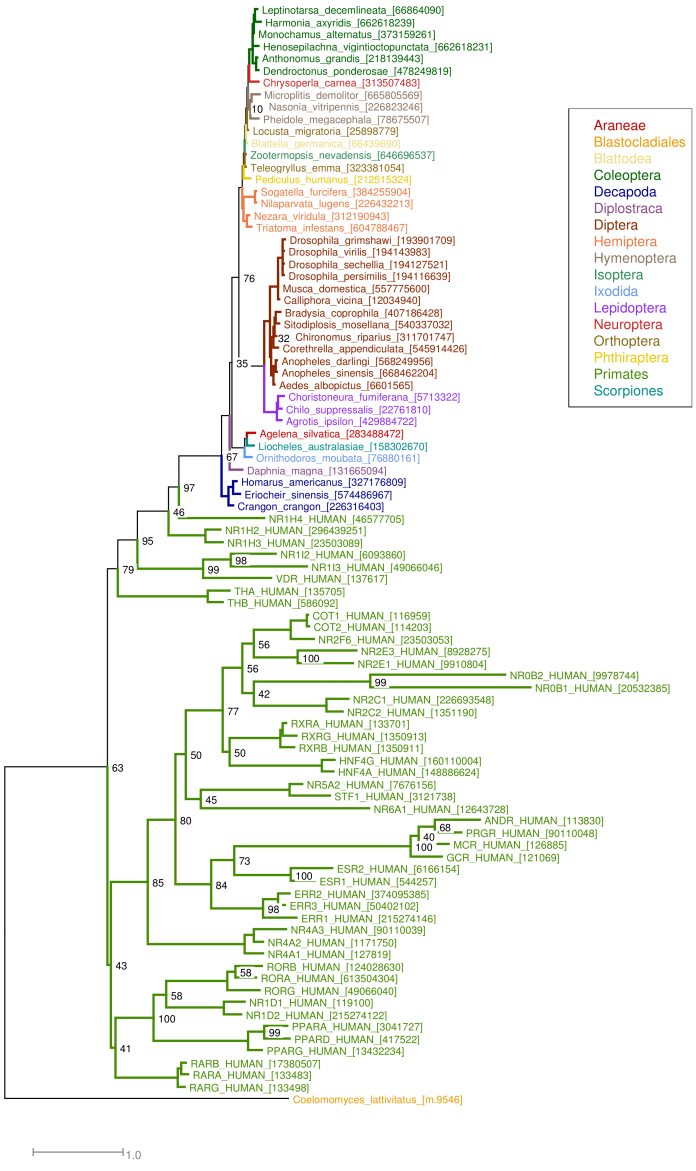


Figure 4.7: RAxML tree of select 20-hydroxyecdysone receptors from insects and Human nuclear receptors, and unique *C. lativitattus* protein sequences

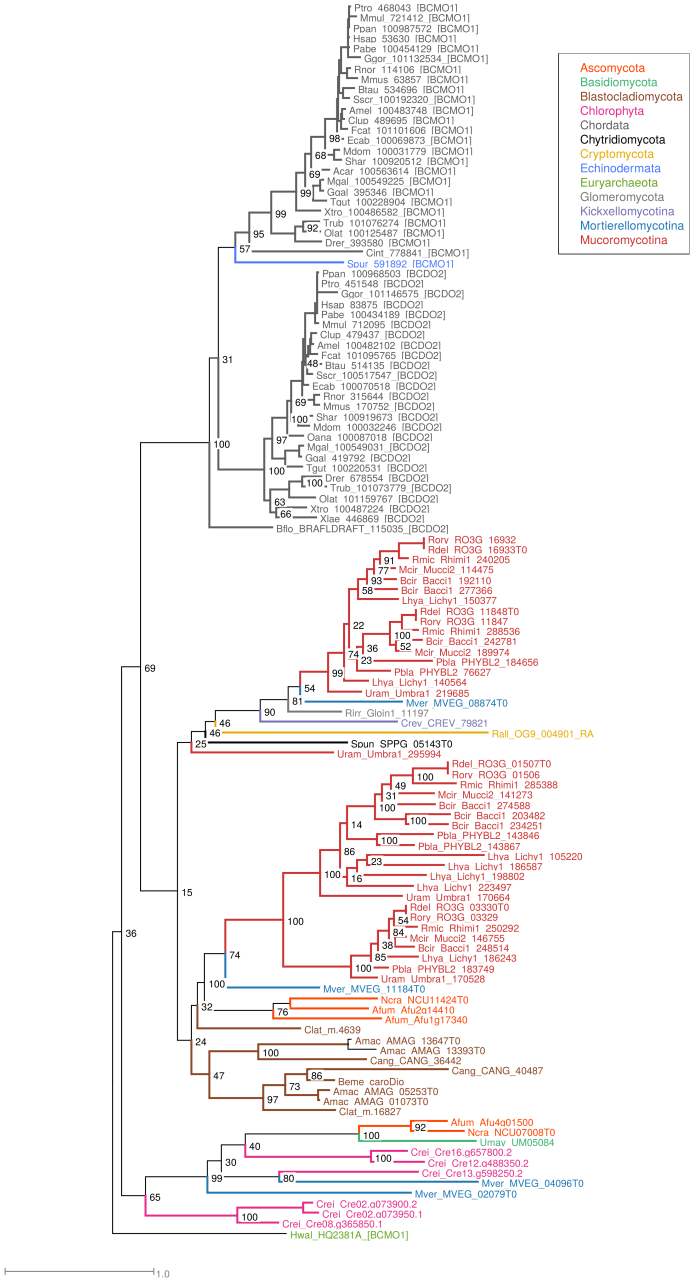


Figure 4.8: Phylogenetic history of BCMO1 homologs found in Fungi and Metazoan lineages using a maximum likelihood tree constructed with RAxML from identified BCMO1 fungal sequences and those from KEGG families.

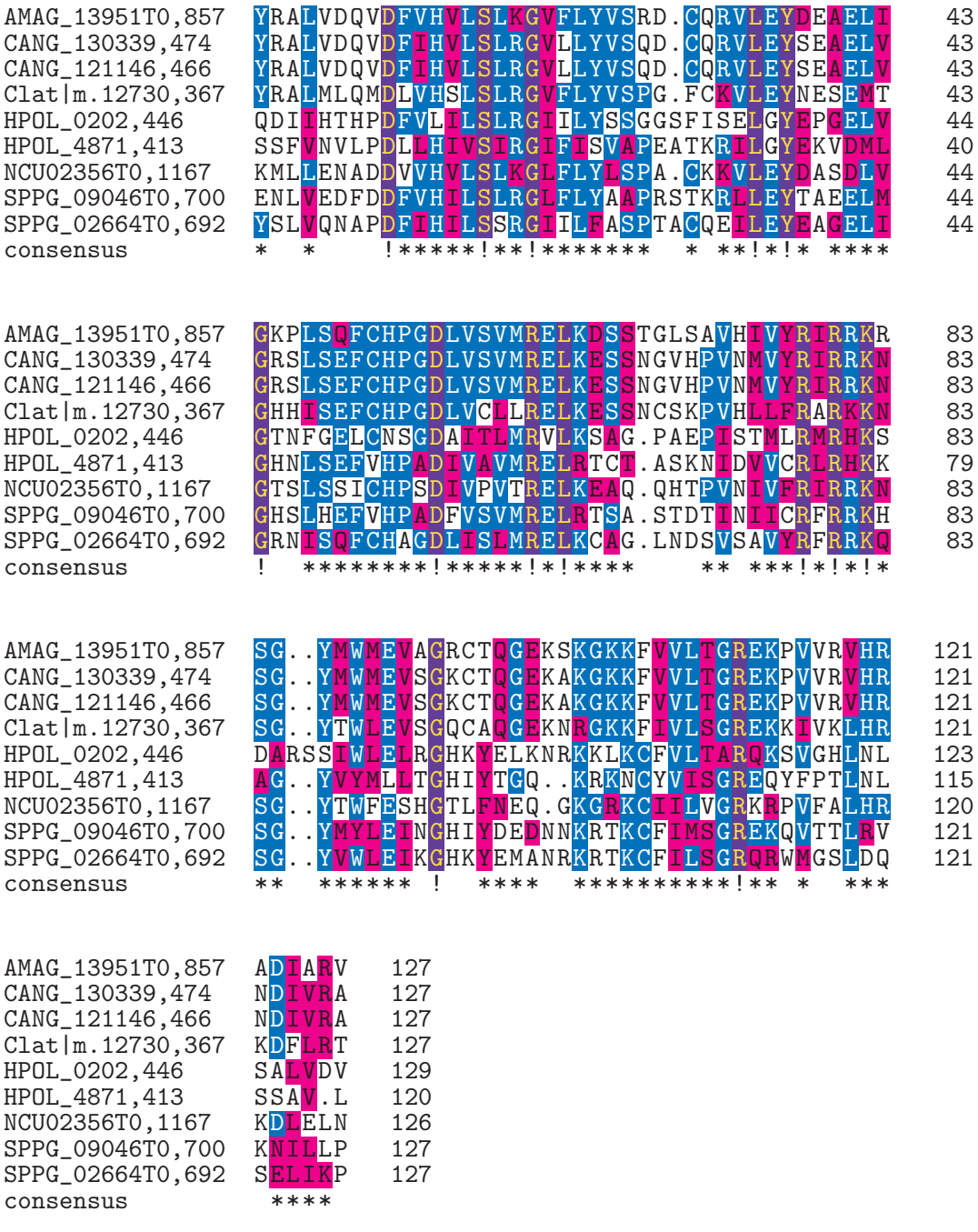


Figure 4.9: Multiple sequence alignment of PAS-domain in fungal WC1 homologs

Table 4.1: Comparisions of counts of top 20 PFAM domains from *C. lativittatus* as identified in other chytrids. Details for organism abbreviations can be found in appendix.

	Description	*Clat*	*Amac*	*Cang*	*Spun*	*Bden*	*Hpol*
PF00069	Pkinase	297	409	168	354	404	149
PF00400	WD40	211	1287	1164	1710	1075	170
PF00076	RRM 1	143	201	69	278	240	53
PF00153	Mito carr	119	222	40	294	231	49
PF00012	HSP70	112	42	8	18	16	9
PF00118	Cpn60 TCP1	93	22	11	20	31	10
PF00071	Ras	86	80	29	108	96	47
PF00270	DEAD	70	115	51	136	126	81
PF00036	efhand	67	27	34	60	68	30
PF00112	Peptidase C1	65	0	1	0	0	0
PF01576	Myosin tail 1	59	1	2	4	3	10
PF00004	AAA	58	128	50	134	135	93
PF00271	Helicase C	52	140	77	184	199	71
PF00227	Proteasome	52	26	16	28	43	19
PF01532	Glyco hydro 47	52	2	3	8	12	8
PF07690	MFS 1	50	144	40	120	69	36
PF02985	HEAT	50	34	38	56	136	58
PF00005	ABC tran	50	217	91	168	231	96
PF00009	GTP EFTU	48	79	19	92	48	54
PF00089	Trypsin	47	2	2	8	4	2

Table 4.2: Unique clusters of *C. lativitattus* proteins with 7-9 predicted membrane-spanning domains, grouped using OrthoMCL

Cluster	Protein	Description
1127	Clat|m.15402	7tm_1;PF00001
1127	Clat|m.16182	7tm_1;PF00001
1127	Clat|m.18794	7tm_1;PF00001
1127	Clat|m.9338	7tm_1;PF00001
1175	Clat|m.12314	K_trans;PF02705
1175	Clat|m.12319	K_trans;PF02705
1175	Clat|m.12322	K_trans;PF02705
1176	Clat|m.15582	Grp1_Fun34_YaaH;PF01184
1176	Clat|m.15583	Grp1_Fun34_YaaH;PF01184
1176	Clat|m.15584	Grp1_Fun34_YaaH;PF01184
1177	Clat|m.16070	DUF3533;PF12051
1177	Clat|m.16072	DUF3533;PF12051
1177	Clat|m.16075	DUF3533;PF12051
1269	Clat|m.11119	UAA transporter;PF08449
1269	Clat|m.11121	UAA transporter;PF08449
1270	Clat|m.12151	–
1270	Clat|m.12173	–
1271	Clat|m.8230	–
1271	Clat|m.8232	–
1272	Clat|m.12638	–
1272	Clat|m.12641	–
1273	Clat|m.14468	Sodium:sulfate symporter;PF00939
1273	Clat|m.14469	Sodium:sulfate symporter;PF00939
1274	Clat|m.14625	–
1274	Clat|m.14626	–
1275	Clat|m.4725	NADH-Ubiquinone/plastoquinone;PF00361
1275	Clat|m.8681	NADH-Ubiquinone/plastoquinone;PF00361
1276	Clat|m.4968	Chitin synthase;PF03142
1276	Clat|m.4970	Chitin synthase;PF03142

Table 4.3: Predicted *C. lativittatus* proteins associated with photosensing. PFAM Family definitions: PF13426, "PAS domain" [PAS_9]; PF08447, "PAS fold" [PAS_3]; PF00001, "7 transmembrane receptor (rhodopsin family)" [7tm_1]; PF01036, "bacteriorhodopsin-like protein" [Bac_rhodopsin]; PF10317, "Serpentine type 7TM GPCR chemoreceptor Srd" [7TM_GPCR_Srd]; PF00211, "Adenylate and Guanylate cyclase catalytic domain" [Guanylate_cyc]

OrfID	Type	UniProtKB BLAST match	InterPro
m.12730	WC-1	Neurospora crassa [Q01371], 4e-43	–
m.7819	OpGC	*Halobacterium sp.* NRC-1 [P71411], 7.4e-07 *Drosophila melanogaster* [Q8INF0], 8e-44	PF01036 PF00211
m.9338	Op	*Danio rerio* [Q2KNE5], 3e-08	PF10317 PF00001
m.15402	Op	*Apis mellifera* [Q17053], 8e-10	PF00001
m.16182	Op	*Cambarus hubrichti* [O18312], 1e-04	PF00001
m.18794	Op	*Limulus polyphemus* [P35360], 3e-10	PF00001
m.11198	Op	*Halobacterium sp.* AUS-2 [P29563], 3e-30	PF01036

Table 4.4: Fungal Adhesin and Adhesin-like Database (FAADB) predictions for chytrid proteins. Descriptions for chytrid proteins identified using positive dataset from predicted adhesin and adhesin-like proteins.

	SVM.Score	UniprotKB.ID	Identity	E.value
Amac|AMAG_17113T0	-0.74	Q9DBG3.1	55.26	0.0
Amac|AMAG_05862T0	-0.61	P00940.2	58.06	2e-99
Amac|AMAG_12167T0	-0.54	P00940.2	59.27	8e-101
Amac|AMAG_08758T0	-0.19	P46226.3	59.84	1e-104
Amac|AMAG_10182T0	-0.17	P46226.3	60.24	6e-105
Amac|AMAG_18542T0	-0.16	P48491.2	58.19	2e-64
Amac|AMAG_16371T0	0.15	–	–	–
Amac|AMAG_08832T0	0.17	Q966L9.1	54.63	8e-06
Amac|AMAG_17977T0	0.21	–	–	–
Amac|AMAG_16084T0	0.32	P29141.1	42.00	6e-40
Amac|AMAG_04496T0	0.34	Q6H236.1	45.71	1e-08
Amac|AMAG_19066T0	0.76	–	–	–
Amac|AMAG_20477T0	0.83	–	–	–
Amac|AMAG_04749T0	1.33	–	–	–
Amac|AMAG_03430T0	1.43	–	–	–
Cang|CANG_48379	-0.78	Q06852.2	29.47	7e-09
Cang|CANG_125451	-0.68	P58559.1	58.02	2e-125
Cang|CANG_33361	-0.09	P48494.3	59.04	7e-102
Cang|CANG_38430	0.60	P22105.3	39.67	1e-20
Cang|CANG_69396	0.94	–	–	–
Clat|m.10957	-0.78	P27393.1	63.61	4e-65
Clat|m.22466	-0.75	Q6CJG5.2	48.60	6e-48
Clat|m.2165	-0.68	A7S7F2.1	43.48	2e-47
Clat|m.18806	-0.68	P48501.1	56.19	1e-86
Clat|m.13634	-0.63	Q90XG0.1	71.02	2e-122
Clat|m.1572	-0.58	O09452.1	84.01	2e-166
Clat|m.15929	-0.52	P07487.2	80.28	2e-77
Clat|m.10361	-0.46	P20445.2	61.34	9e-99
Clat|m.18916	-0.42	P48501.1	60.87	4e-102
Clat|m.11233	-0.30	Q96UF2.1	68.82	1e-132
Clat|m.4480	-0.07	O77458.1	64.24	5e-61
Clat|m.13062	0.52	Q6BMK0.1	74.52	6e-83
Clat|m.745	0.90	Q92824.4	36.65	7e-26
Clat|m.16183	1.05	P22105.3	51.01	4e-18
Clat|m.13342	1.13	Q92824.4	43.22	8e-43
Clat|m.11209	2.24	–	–	–
Bden|BDET_06684	-0.79	Q5R2J2.1	71.56	1e-177
Bden|BDET_05736	-0.17	P00939.1	57.66	2e-88
Bden|BDET_05372	-0.03	Q95P23.1	28.02	9e-21
Bden|BDET_01761	0.45	Q9AVB0.1	64.57	3e-20
Bden|BDET_00626	0.49	–	–	–
Bden|BDET_00287	0.67	Q9AVB0.1	58.01	8e-22
Bden|BDET_02239	1.45	P23253.1	54.09	2e-22
Bden|BDET_00436	1.48	–	–	–
Bden|BDET_03668	1.51	Q63425.2	39.51	6e-06
Bden|BDET_06100	1.70	–	–	–
Hpol|HPOL_4940	-0.79	P00940.2	56.63	1e-95
Hpol|HPOL_4790	-0.46	O57479.3	75.61	8e-148
Hpol|HPOL_4769	-0.24	P52041.2	40.00	3e-09
Hpol|HPOL_2513	-0.14	Q1ZXD6.1	32.78	6e-13
Spun|SPPG_01012T0	-0.68	P30741.2	61.60	1e-102
Spun|SPPG_08522T0	-0.47	Q6ZRI0.3	31.26	9e-10
Spun|SPPG_00685T0	-0.15	Q5H8C1.3	33.98	2e-14
Spun|SPPG_01988T0	0.05	–	–	–

Table 4.5: BLASTP results using *C. lativittatus* ORF m.10080 as a query against SwissProt database. Top five hits are to liver X receptors (LXRs) from various mammalian species. The next siz hits are only those which were to 20-hydroxy-ecdysone receptors, all of which are from insects

	Coverage	E val	% Id	Description
Q5E9B6.1	40%	1e-15	28%	Liver X receptor alpha [*Bos taurus*]
Q13133.2	40%	1e-15	28%	Liver X receptor alpha [*Homo sapiens*]
Q62685.1	40%	3e-15	28%	Liver X receptor alpha [*Rattus norvegicus*]
Q9Z0Y9.3	40%	9e-15	28%	Liver X receptor alpha [*Mus musculus*]
P55055.2	40%	6e-10	28%	Liver X receptor beta [*Homo sapiens*]
P49880.2	36%	1e-07	28%	20-hydroxy-ecdysone receptor [*Aedes aegypti*]
P49883.1	42%	6e-06	24%	20-hydroxy-ecdysone receptor [*Manduca sexta*]
P49881.1	42%	8e-06	24%	20-hydroxy-ecdysone receptor [*Bombyx mori*]
P34021.1	36%	1e-05	26%	20-hydroxy-ecdysone receptor [*Drosophila melongaster*]
O18531.2	36%	1e-05	27%	20-hydroxy-ecdysone receptor [*Lucilia cuprina*]
P49882.1	36%	4e-05	25%	20-hydroxy-ecdysone receptor [*Chironomus tentans*]

Chapter 5

Growth suppression properties of the Chytridiomycete *Homolaphlyctis polyrhiza* JEL142

5.1 Introduction

Interactions between microorganisms are facilitated by biological signals. These include proteins, small molecules, and various chemical compounds, either bound to the cell surface or secreted into the environment. Many of these compounds can be classified as secondary metabolites: chemicals not required for growth or development of the organism. A number of well-known naturally-derived antifungal products, including toxins [167], antimicrobials [168], and plant hormone mimics [169], can be classified as secondary metabolites. Secretion by an organism in a resource-limited environment of secondary metabolites which also happen to negatively impact neighboring organisms would confer a selective advantage upon the producer. Thus, resource competition likely

plays a role in the evolution of natural antifungal production [66].

In humans, the most common fungal pathogen is *Candida albicans* (Ascomycota) [170]. It is a commensal of healthy gut flora, but overgrowth results in opportunisit infection and the disease candidiasis. *Cryptococcus neoformans* (Basidiomycota) is another common human pathogen. It is typically found in soil and associated with pigeon guano but causes severe meningitis in patients with AIDS or others with similarly compromised immune systems [171].

After surveying the international fungal pathology community regarding plant pathogens, the filamentous Ascomycete *Magnaporthe oryzae* was deemed to be the most important in terms of scientific/economic merit [172]. *M. oryzae* is the causative agent of rice blast disease, a devastating disease given that roughly one-half of the global population is reliant on rice. As such, this fungus has developed into model system for studying plant-pathogen interactions. The Ascomycete *Botrytis cinerea*, also known as grey mould, was deemed to be the second most important fungal plant pathogen. Due to its broad host range which includes over 200 species, it is the most extensively studied necrotrophic fungal pathogen [172].

Other important fungal plant pathogens infect wheat (*P. graminis* f. sp. *tritici*, causes black rust; *P. striiformis* f. sp. *tritici*, causes yellow rust; *P. triticina*, causes brown rust), cereals (*Fusarium graminearum*, *Blumeria graminis*, *Mycosphaerella graminicola*), tomatoes (*Fusarium oxysporum*), and corn (*Ustilago maydis*).

Therefore, research into new natural antifungal treatment is increasingly important as these and other pathogenic fungi become progressively resistant to current, conventionally synthesized fungicides and antifungal drugs (see: [173], [174], [175], [176], and [177]). The bulk of research into fungal-derived secondary metabolites has thus far focused primarily on Ascomycete and Basidiomycete fungi due to their diverse prod-

uct types [178]. Chytrids are a relatively understudied group for secondary metabolites and represent unexplored diversity which may harbor novel compounds and molecules. Comparative genomic analyses have identified a variety of degradation enzymes, but a detailed examination of secondary metabolites involved in competition zoosporic fungi has yet to be explored. It is likely that these organisms have evolved mechanisms for mediating chemical interactions with other organisms, and therefore could present a potentially valuable source of novel antifungal compounds.

Homolaphlyctis polyrhiza (*Hp*) is a non-pathogenic member of the Chytridiomycota and is most closely related to the amphibian pathogen *B. dendrobatidis* (*Bd*). The specific isolate, *Hp* JEL 142, has been used previously in phylogenetic studies of chytrids [31, 29, 39] and was provided a formal name in 2011 [40]. A draft 454 genome assembly was produced in that same year in order to gain comparative insights into pathogenicity of *Bd* [38].

This isolate of *Hp* was collected by Dr. Joyce Longcore's research group in Maine, USA from a 1.7 ha oligotrophic and fishless lake with a pH of 4.6 [179, 180]. It was cultured using onionskin bait and isolated into pure culture on mPmTG nutrient agar. Its overall morphology is typical of members of the Rhizophydiales [181] and includes a single lobed mitochondrion, a single libpid globule, microbody and rumposome placement, and Kinetosome / nonflagellated centriole ultrastruture. Phylogenetic reanalysis based on maximum likelihood using rRNA subunits established the placement of *Hp* as sister to *Bd*, and prompted the description of a novel genus (*Homolaphlyctis*, in recognition of the researchers who helped with collection and analysis) and species (*H. polyrhiza*, in reference to its multiple rhizoidal axes) [40].

While working with *Hp* in the Stajich lab for the phototaxis and protein expression experiments described in Chapter 3, unintentional contamination of *Hp* plates

by *Neurospora crassa* led to the unique observation of growth inhibition by a chytrid on the filamentous ascomycete. Specifically, I observed a distinctive zone created by *N. crassa* hyphae growing near *Hp* sporangia, while *N. crassa* contamination on plates of related Chytridiomycetes *Batrachochytrium dendrobatidis* (*Bd*) and *Spizellomyces punctatus* (*Sp*) did not result in the same zone of inhibition.

This property of *Hp* is reliable and reproducible. This chapter describes the experimental and computational work towards identification and characterization of a compound responsible for hyphal growth suppression of *N. crassa* by *Hp*. This study is the first of *Hp* in the context and presents novel findings about microbial interactions within early branching fungal lineages. Assistance in preparing media and general strain maintenance was provided by undergraduate students in the Stajich lab, Na Jeong and Sapphire Ear, and visiting student in the Research Experience for Undergraduates (REU) program Spencer Swansen.

5.2 Methods

Fungal strains and maintenance

Cultures of *Homolaphlyctis polyrhiza* JEL142, *Operculomyces laminatus* JEL223, *Rhizoclosmatium hyalinus* JEL800, and *Obelidium mucronatum* JEL802 were individually maintained on mPmTG [peptonized milk (0.4 g/L), tryptone (0.4 g/L), dextrose (2 g/L)], *Batrachochytrium dendrobatidis* JEL423 cultures were maintained on 1% Tryptone [tryptone (10 g/L), dextrose (3.2 g/L)], and *Spizellomyces punctatus* SW-1 cultures were maintained on PmTG [peptonized milk (0.5 g/L), tryptone (1 g/L), dextrose (5 g/L)]. All chytrid cultures were maintained at room temperature (approx. 23°C). Un-

less otherwise specified, all experiments using chytrids were carried out using the specific media described above for each species. Motile chytrid zoospores, when required, were obtained from actively growing (ie 2-4 day old) plates by flooding and subsequently (after 30-45 minutes at room temperature) collecting 2-4 mls of sterile di H_2O. Sporangia samples for inoculation were obtained by removing a block (approx. 1 cm^2) of agar containing actively-growing chytrid sporangia.

Vogel's minimal medium (VM) [182] was used for vegetative growth of *Neurospora crassa* FGSC 2489, *Neurospora discreta* FGSC 8579 and *Neurospora tetrasperma* FGSC 2508. *N. crassa* kinase/phosphotase mutants (Table 5.1; [183]) were maintained on VM + Hygromycin and grown similarly to *N. crassa* WT strains. *Trichoderma reesei* FGSC 10290, *Phycomyces blakesleeanus*, and *Ashbya gossypii* cultures were maintained on PDA media [potato dextrose agar (39 g/L)] at 28°C, 20°C, and 30°C, respectively. *Aspergillus nidulans* FGSC A4 cultures were maintained on minimal media [dextrose (10 g/L), nitrate salts (50 mL/L), Trace elements (1 ml/L)], at 28°C. *Saccharomyces cerevisiae* strains MAU99 and AH109 were maintained on YPDA [yeast extract (10 g/L), peptone (20 g/L), dextrose (20 g/L), adenine sulfate (0.003%)] media at 35°C. *Coprinopsis cinerea* FGSC 9003 cultures were maintained on YPD media at 37°C. Plates and slants of 1% tryptone, PmTG, mPmTG, VM, YPD, and PDA media included 1% agar. Plates and slants of MM included 1.8% agar. Plates and slants of YPDA included 2% agar.

Bioactivity on solid media and liquid media

To assess the breadth of fungal species susceptible to *Hp* sporangia, *Hp* was individually co-cultured on mPmTG media with the selection of species described above.

A block of agar containing active *Hp* sporangia was added to mPmTG plate, flushed with sterile diH_2O, and incubated for 48h, at which point they were inoculated at a single point with 1x10^6 *Neurospora* conidia in diH_2O. For *Neurospora* experiments, plates were left to grow for an additional 24 - 48 hrs.

Similarly, for other fungi (which are much slower growing than *Neurospora*), *Hp* and the target fungus were inoculated at the same time on an mPmTG solid media plate. Growth was monitored every other day until the other fungi reached *Hp* sporangia, approximately 8-10 days.

To obtain aliquots of filter-sterilized*Hp*-conditioned media ("filtrate"), initial preparations (in triplicate) were constructed by adding eight blocks of agar containing actively-growing *Hp* sporangia, each approximately 1cm^2 to 10 ml of liquid mPmTG. Initially, to determine if the compound was being constitutively produced or is a response mechanism to the presence of another organism, a second preparation was made identically to the one just described, to which an additional 0.5 μl of a 100 μl *N. crassa* conidial suspension was added. To control for possibility that growth rate changes could be caused by nutrient depletion by *N. crassa* in this second trial, a third preparation was made with eight mPmTG agar blocks with no *Hp* sporangial growth in 10 ml mPmTG, to which 0.5 μl of a 100 μl *N. crassa* conidial suspension was added. These cultures preparations were left to incubate at room temperature for 72hrs, at which point the total filtrate was obtained by passing each individual replicate through a 0.22 μm syringe filter into a 50 ml conical tube. All replicates for a given treatment were pooled after filtration.

Large scale preparations were constructed in a similar manner by adding twenty blocks of agar containing *Hp* sporangia to 50 ml of mPmTG media, allowed to grow for 72hrs and filtered using a 0.22 μm syringe filter. A dilution series was prepared using

twenty-five blocks of agar in 75 ml of mPmTG media and filtered using a 0.22 μm Stericup filter (Millipore). Conditioned *Hp* media filtrate was used to dilute VM media at values of 0%, 10%, 25%, 50%, 75%, 90%, and 100% filtrate, achieving a total volume of 3 ml in which 1x10^5 *N. crassa* conidia were inoculated. Similarly, additional dilutions were prepared using fresh mPmTG media, *Bd* filtrate (prepared similar to *Hp* described above), and fresh 1% Tryptone media.

For temperature assays, the filtrate from each small scale preparation was subsequently divided into eight 2 ml aliquots (in duplicate) to test for bioactivity over three separate experiments. The control sample was left undisturbed and uninoculated throughout the experiment. Another sample was inoculated with 2x10^6 conidia but otherwise untreated. The filtrate was subjected to six different temperature treatments to test its thermostability: -80°C (30 min), -20°C (1 hr), 4°C (1 hr), 28°C (1 hr), 65°C (1 hr), and 90°C (30 min). Each experimental treatment was inoculated with 2x10^6 conidia. These temperature treatments were allowed to return to room temperature prior to inoculation.

To test the sensitivity of *Bd* to *Hp*, 3 ml of the filtrate from the large scale preparation was used as an incubation medium for one blcok of 1% Tryptone agar containing actively growing *Bd* sporangia. Similarly, 3 ml of mPmTG media was used as a positive control. These preparations were incubated for 96h at room temperature to allow for potential zoospore production and release. After 96h, the suspension was mixed briefly and 1 ml was added to mPmTG plates, while another 1 ml was added to 1% Tryptone plates. The plates were incubated at room temperature for a maximum of 14d and photographed periodically.

Reassembly and annotation

Hp sporangia material was collected by scraping actively growing *Hp* plates with sterile spatulas. Genomic DNA was extracted using a modified bead-beating procedure. Briefly, approximately 100 mg of material containing both zoospores and sporangia was scraped from actively-growing, zoospore-rich *B. dendrobatidis* and *S. punctatus* plates. The material was added to approximately 100 mg of silicon beads (0.5mm dia.) and mixed with 600 μl of Cell lysis solution (Qiagen, Germantown, MD) and 3 μl of proteinase K. The solution was homegenized with a bead beater using a 30s pulse at 4°C and subsequently incubated for 1.5h at 55°C. 200 μl of protein precipitation solution (Qiagen, Germantown, MD) was added to the mixture and iced for 15 min. After centrifugation at 14000xg for 3 min at room temperature, the supernatant was collected, mixed with 600 μl isopropanol, and spun at 1400xg for 1 min at room temperature. Pellet was washed with 600 μl ice cold 70% EtOH and spun at 1400xg for 1 min at room temperature. Finally the pellet was air dried for 15 min at room temperature, resuspended in 50 μl H_2O, incubated at 65°C for 1 hr, and stored at -20°C. A DNAseq Illumina library was prepared using the NEBNext Ultra DNA library kit and submitted to the University of California, Riverside Genomics Institute for Integrateve Genomic Biology (IIGB) core facility for MiSeq Illumina HT sequencing.

Whole genome assembly was performed using the Celera Assembler [184]. HT sequencing yielded 27,541,671 reads. The final assembly is approximately 27.3 Mb and consists of 2720 contigs (N50=22,735). The GC content is 48.2%. For comparision, the previous 454 assembly utilized 922,085 reads producing 16,311 contigs (N50=36,162) and an assembly size of 26.7 Mb.

Genome annotation was performed using MAKER (v2.31.8) [185], incorporat-

ing EST evidence from *Batrachochytrium dendrobatidis*, and including RepeatMasker (v4.0.5), SNAP (v0.15.4), Augustus (v2.7), and Fgenesh. A total of 8155 proteins were predicted.

Secretome and small-metabolite prediction

Putative secondary metabolite clusters were predicted from the annotated proteomes of *Bd, Sp*, *A. macrogynus*, and *C. anguillulae* using antiSMASH [153] and SMURF [186].

To identify putative secretome proteins, these proteomes (as well as the transcriptome from *C. lativitattus*) were also subjected to a predictive workflow optimized for fungi [187]. The workflow described therein evaluated the prediction accuracies for a number of programs for eukaryotic secretome prediction: SignalP [188], Phobius [189, 190], TargetP [191, 192], and WolfPSort [193, 194]. These programs were evaluated indiviually and in combination with TMHMM [142] and PS-Scan [195] and ranked by Mathews' Correlation Coefficient (MCC) [196, 197, 198] which is a measure of Sensitivity and Specificity. The fungal prediction workflow with the highest accuracy (MCC: 84%) combined SignalP/TMHMM/WolfPsort/Phobius/PS-Scan, in that order. Here, I used SignalP (v4.1), WolfpSort (v0.2), and Phobius (v1.01) in the specific order given above to predict secreted proteins from the proteomes for chytrids listed in Table B.1. TMHMM (v2.0) and PS-Scan (v1.79) were used to filter out transmembrane and ER targeting proteins, respectively.

Temperature and pH growth assays

To compare to the assays described in [199], *Hp* was grown at a range of temperatures: 18°C, 23°C, 25°C, 28°C, and 30°C. Briefly, a 15 ml mPmTG inoculum was incubated at 23°C for 1 week. 1 ml of culture was used to inoculate 30 ml of mPmTG in 50 ml conical tubes.

Similarly, *Hp* and *Bd* were grown at a range of pHs by preparing mPmTG and 1% Tryptone media at pH levels: 4.0, 5.0, 6.0, 6.8, 8.0, and 9.0. Low-pH media were prepared with the addition of 1M HCl, while high-pH media were prepared with addition of 1M NaCl. pH levels were checked with a pH meter prior to autoclaving.

5.3 Results

Hp inhibition is unique among the chytrids and broadly active against other fungal species

My initial observation was repeatable, and under controlled circumstances, I was able to demonstrate that *Hp* sporangia are reliably and repeatedly capable of inhibition of *N. crassa* vegetative growth on solid mPmTG media (Figure 5.1A). The other tested Chytridiomycota species, *Bd* and *Sp*, do not demonstrate any inhibitory activity (Figure 5.1B and C). Additional Chytridiomycota species (*Operculomyces laminatus* JEL 223, *Rhizoclosmatium hyalinus* JEL 800, and *Obelidium mucronatum* JEL 802) do not inhibit hyphal growth of *N. crassa* (Figure 5.1D-F), supporting the conclusion that this property is unique to *Hp*.

This phenomenon is not the result of any thigmotropism-related response or object avoidance type behavior in *N. crassa*, as hyphae of *N. crassa* are not inhibited

by an agar block lacking *Hp* sporangia (Figure 5.5).

To determine if the sensitivity to *Hp* sporangia was unique to *N. crassa*, I screened a panel of fungi from among the Ascomycota, Basidiomycota, and Zygomycota. Figure 5.2A recapitulates the previous observations about *N. crassa*, while Figure 5.2B and C demonstrate, respectively, that vegetative hyphal growth of *N. discreta* and *N. tetrasperma* can also be suppressed by the presence of *Hp* sporangia.

Within the Ascomycota, but outside of the genus Neurospora, *T. reesei* (Sordariomycetes; Hypocreales; Hypocreaceae), *A. nidulans* (Eurotiomycetes; Eurotiales; Trichocomaceae), and *A. gossypii* (Saccharomycetes; Saccharomycetales; Saccharomycetaceae) are all completely sensitive to *Hp* (Figure 5.2D-F, respectively).

Within the Basidiomycota, *C. cinerea* (Agaricomycetes; Agaricales; Psathyrellaceae) is completely sensitive to *Hp* sporangia (Figure 5.2G). Growth of members of the order Mucorales had mixed sensitivities. *P. blakesleeanus* (Mucorales; Phycomycetaceae) was completely sensitive to *Hp* sporangia, yet *R. oryzae* (Mucorales; Mucoraceae) appeared to be insensitive (Figure 5.2H & I, respectively). With a limited, but phylogenetically diverse panel of ten fungi, the broadly observed pattern of sensitivity suggests that the responsible compound has a generalized mechanism of action.

Additionally, the growth in *Hp* conditioned media filtrate of two strains of yeast (*S. cerevisiae* MAU99 and AH109) and *E. coli* DH5α suggested that this filtrate inhibited liquid growth of these organisms, and also provides evidence that the antimicrobial properties of *Hp* extend to bacteria as well (Figure 5.7).

The physiology of *Hp* suggests that it is more tolerant of environmental stresses than *Bd*

To assess whether or not the observed growth suppression effect was the result of acidification of the local environment, *Hp* was cultured on plates augmented with bromophenol blue or phenol red. After 96h growth, neither plates changed color, indicating that *Hp* sporangia do not decrease the pH of mPmTG media below 6.8. However, phenol red plates which were additionally inoculated with *N. crassa* did change from red to yellow in the areas around *N. crassa* hyphae, suggesting that *N. crassa* acidifies mPmTG media to below pH 6.8. This effect is not observed in bromophenol blue plates, suggesting that the final pH lies between 6.8 and 4.6.

To assess the environmental tolerances of *Hp* compared to *Bd*, these two species were individually cultured in liquid media (mPmTG and 1% Tryptone, respectively) prepared at different pHs, ranging from 4.0 to 9.0. The average OD_{450}, a measure of cell density, of *Hp* at low pH (4.0 and 5.0) was 0.037 and 0.068 respectively, which was higher than that for *Bd* (0.001 and 0.004 respectively). This suggests that *Hp* is more tolerant of acidic environments than *Bd* (Figure 5.4A). Similarly, *Hp* and *Bd* were grown at temperatures ranging from 18°C to 30°C, and the OD_{450} assessed during a 21-day period (Figure 5.4B). Taken together, these results suggest that *Hp* has a higher tolerance for environmental changes than *Bd*. This finding is consistent with the fact that this isolate was obtained from an acidic lake [40].

Liquid filtrate screening suggests a non-protein compound is responsible for bioactivity

I was interested in obtaining a sample of bioactive liquid, on which I could perform chemical profiling to isolate a responsible compound. To test for bioactivity, *Hp* cultured in liquid mPmTG media for a period of 72 hrs and subsequently filter-sterilized, producing "conditioned media" to which *N. crassa* conidia was reintroduced.

This conditioned media, obtained from *Hp* and absent of all sporangia, retains the same inhibitory properties against *N. crassa* observed with *Hp* sporangia on solid agar plates (Figure 5.3A). Similarly, *N. crassa* was grown in mPmTG media, and after filter-sterilization, conditioned media was demonstrated to still provide sufficient sugar content to support growth of reintroduced *N. crassa* conidia (Figure 5.3B), supporting the idea of the presence of an active inhibitory compound rather than an effect of nutrient limitation. Additionally, conditioned media derived from "*Hp* alone" and "*Hp*+*Nc*" preparations (Figure 5.3A & C) were indistinguishable from one another in their effects on *N. crassa* growth, providing evidence to suggest that the compound is constitutively produced by *Hp* and not as a response to a foreign competitor.

The stability of the conditioned media was tested at seven different temperatures, ranging from -80°C to 90°C. The conditioned media retained bioactivity between -20°C and 60°C (Figure 5.3D-E), but was ineffective after treatment of -80°C and 90°C (Figure 5.3F-G). *N. crassa* has no problem growing in fresh mPmTG media (Figure 5.3H). Additionally, the conditioned media was ineffective if left sterile at room temperature for a period of 96h in the light, and subsequently inoculated with *N. crassa* conidia.

After treatment of 100 mg/ml of Proteinase K, the conditioned media was no

less effective at inhibition of *N. crassa* hyphal growth than untreated media. Taken together, these results suggest that a protein is not responsible for the observed bioactivity.

A dilution series was prepared using VM media suggested that *Hp* filtrate loses efficacy at 75% of growth media. Figure 5.6 shows VM media diluted with either *Hp* filtrate, fresh mPmTG, *Bd* filtrate, or fresh 1% Tryptone. Overall, mPmTG media appears to be a poorer nutrient source than 1% Tryptone. However, the dilution with *Hp* filtrate is the only one in which a clear decrease in *N. crassa* aerial hyphal growth can be observed as the concentration of additive increases. With all other additives, hyphal growth remains abundant.

in silico analyses

Secondary metabolite clusters were predicted from the revised Illumina genome assembly using antiSMASH [153] and SMURF [186]. Using antiSMASH, several proteins in the *Hp* genome were predicted to be likely homologs of terpene synthases. These genes were unique to *Hp* when compared with other Chytridiomycota and Blastocladiomycota species. A list of the putative terpene synthase proteins from the Illumina assembly, along with their top BlastP results against the SwissProt database, is provided in Table 5.3. The screen using antiSMASH also predicted 26 putative NRPS genes.

Using SMURF, there were no putative secondary metabolite clusters predicted for *Hp*. For comparison, 2 clusters were predicted in *B. dendrobatidis*, 9 in *S. punctatus*, and 2 in *A. macrogynus*.

To corroborate the experimental results which suggested the active molecule is not a protein, I performed *in silico* secretome predictions using the available chytrid

proteomes with a workflow described previously for assessing secretome prediction programs [187]. Table 5.2 lists the number of proteins in the examined chytrid genomes which appear to be secreted. In *Hp*, 58 proteins were predicted to be secretome proteins. Using the secretome protein predictions from all surveyed chytrid genomes, OrthoMCL produced 350 ortholog clusters. However, none of these were unique to *Hp*. Furthermore, none of the candidates were predicted to synthesize known antimicrobial or toxic proteins, which is supported by the temperature and proteinase treatments and further suggests that the compound is not enzymatic in nature.

Kinase mutant screen

Finally, in order to address potential mechanisms of action for this behavior, an undergraduate student Na Jeong randomly assayed 2 *N. crassa* phosphotase and 17 kinase mutants from [183]. All were found to be inhibited by growing sporangia of *Hp* (Table 5.1).

5.4 Discussion

This study was an attempt to characterize a startling and unexpected observation: actively growing *Hp* sporangia inhibit the vegetative, hyphal growth of *N. crassa*. The findings demonstrate that i) *Hp* is unique among surveyed chytrids in its ability to inhibit filamentous growth of *N. crassa*, ii) the interaction is not limited to *N. crassa* but is broadly active against a number of Ascomycete, Basidiomycete, and Zygomycete species, and iii) in silico secondary metabolite analysis suggests a terpene synthase-

related enzyme as a possible candidate based on observed unique expansion of genes.

Hp, an aquatic, non-parasitic member of the Chytridiomycota isolated from a lake in the northeastern United States, and *N. crassa*, a filamentous, multicellular, non-parasitic fungus primarily observed on decaying plant matter, are unlikely to come in specific contact with one another in the environment, so this response was presumed to be a non-specific interaction. The results from screening a phylogentically broad set of fungal species support this hypothesis, and it appears that compound(s) produced by *Hp* are active against a multitude of fungal species. One logical explanation for such a broadly active inhibitory compound is that it is in fact a competitive advantage. A diffusible compound with anti-microbial properties would allow *Hp*, in its natural aquatic environment, to clear a localized niche in which it would have unobstructed access to nutrients.

Compared with other chytrid isolates which were assayed, *Hp* stands alone in its antifungal properties. Although the described basal lineages represent only 2% of all described fungi, and the estimated diversity of these lineages is presumed high, this is the first time an interaction of this type has been described in any of these lineages. While it is possible that this behavior is not unique to this specific *Hp* isolate, a more exhaustive assessment of chytrid isolates is necessary to elaborate on this hypothesis.

The observation that conditioned media derived from "*Hp* alone" and "*Hp*+*Nc*" preparations were indistinguishable from one another in their effects on *N. crassa* growth suggests that the compound is being constitutively produced from *Hp* and is not a response to a fungal competitor. Analysis of the transcriptome would be beneficial and would in theory easily identify any constitutively active genes in the sporangia life stage, from which this compound is expected to originate.

The liquid compound profiling suggests that the compound is not protein-

based, given that filter-sterilized conditioned media was a) effective after 65°C treatment, a temperature at which many proteins denature, and b) effective after treatment of Proteinase K. For additional support for this hypthesis, I performed both *in silico* secretome and secondary metabolite predictions. While the former did not predict any potential secreted proteins with known antimicrobial or antifungal properties, the latter methods predicted several terpene synthase related enzymes which appear to be uniquely present in *Hp* compared to the other basal fungi proteomes searched.

Taken together, these data tentatively suggest that the compound could be a constitutively-produced compound with broadly specific activity. While its mechanism of action is unknown, as are its chemical structure and related biosynthetic pathway(s), it is unlikely to be a protein or protein complex and is more likely to be a secondary metabolite. Future work will necessarily focus on chemical profiling of bioactive spent media filtrate to generate a working hypothesis for the chemical nature of the product. A starting point for this research is provided in the form of in silico genomic and transcriptomic research. Finally, it is worth noting that the relative ease with which this observation was made speaks to the necessity for further research into these basal lineages, of which intimate genomic and biochemical knowledge is lacking.

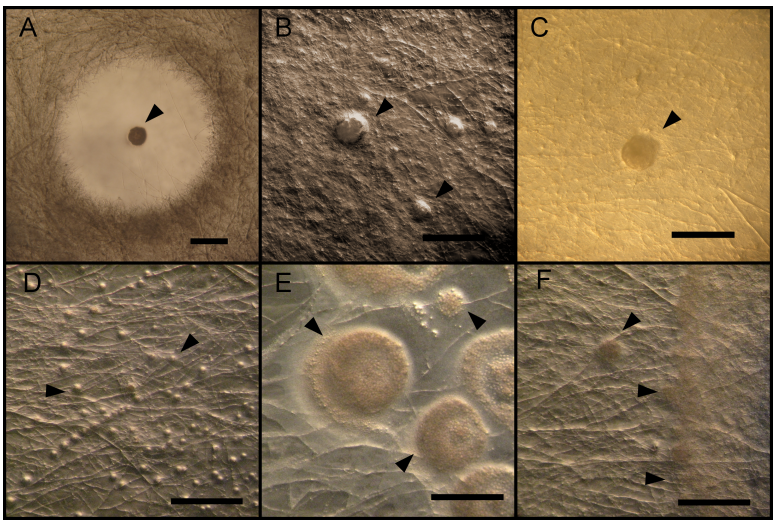


Figure 5.1: *Batrachochytrium dendrobatidis* JEL 423 (*Bd*), the closest relative of *Hp*, is well known as a global emerging pathogen of amphibians and is the causative agent for recent worldwide amphibian decline. *Spizellomyces punctatus* SW-1 (*Sp*) is a soil saprobe and is not known to be pathogenic. A) *Hp* cultured with *N. crassa* produces a cleared zone. B) *Bd*, C) *Sp*, D) *Operculomyces laminatus* JEL 223, E) *Rhizoclosmatium hyalinus* JEL 800, F) *Obelidium mucronatum* JEL 802 do not display this property. Black scale bars = 1mm. Black arrowheads illustrate location of chytrid sporangia.

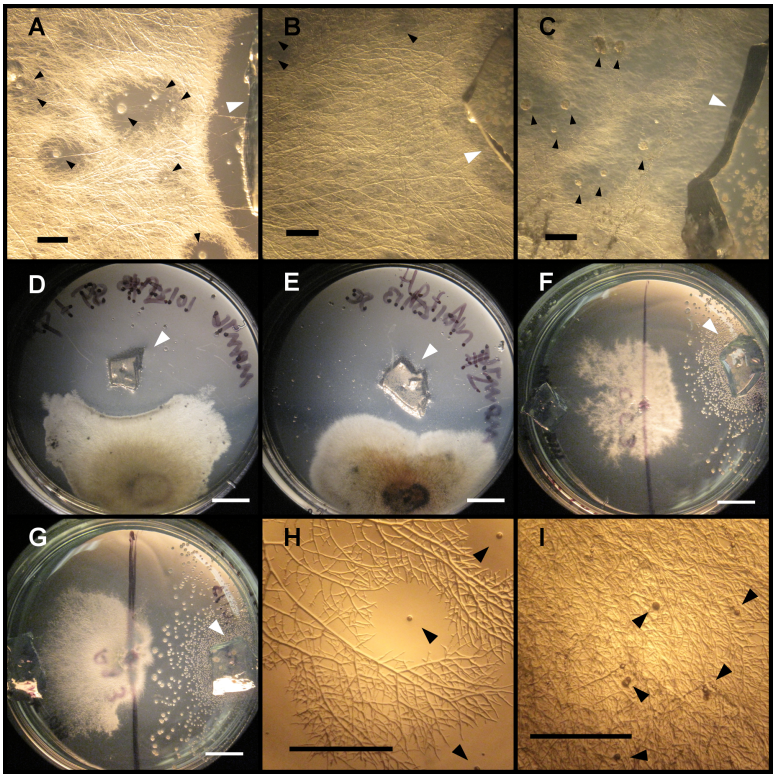


Figure 5.2: A selection of fungal species from the Ascomycota, Basidiomycota, and Mucorales were assayed for sensitivity against *Hp* sporangia. A) *Neurospora crassa* (Sordariomycetes; Sordariales; Sordariaceae), B) *Neurospora tetrasperma*, C) *Neurospora discreta*, D) *Trichoderma reesei* (Sordariomycetes; Hypocreales; Hypocreaceae), E) *Aspergillus nidulans* (Eurotiomycetes; Eurotiales; Trichocomaceae), F) *Coprinopsis cinerea* (Agaricomycetes; Agaricales; Psathyrellaceae), G) *Ashbya gossypii* (Saccharomycetes; Saccharomycetales; Saccharomycetaceae), H) *Phycomyces blakesleeanus* (Mucorales; Phycomycetaceae), and I) *Rhizopus oryzae* (Mucorales; Mucoraceae). Scale bars for A-C,H-I = 1mm. Scale bars for D-G = 5mm. Black arrowheads in A-C,H-I indicate *Hp* sporangia. White arrowheads in A-G indicate blocks of agar with active *Hp* sporangia.

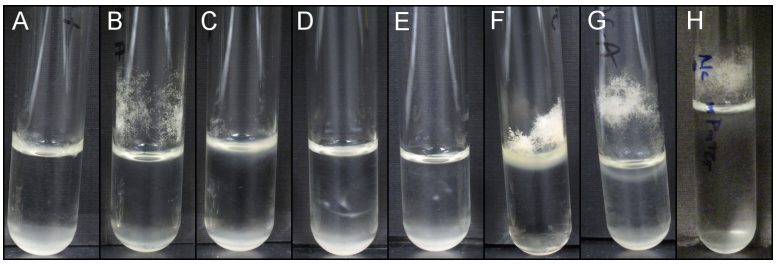


Figure 5.3: Test tubes containing filter-sterilized *Hp*-conditioned media ("filtrate") were inoculated with *N. crassa* conidia and left to incubate at room temperature (23°C) for 96h. For A), B), and C), filtrate was derived from initial preparations of *Hp* alone, *N. crassa* alone, and *Hp*+*N. crassa*, respectively. Panel B) establishes that nutrient limitation is not responsible for inhibitory observation as growth of *N. crassa* can still be supported, while A) and C) establish that *Hp* is not exhibiting this behavior as a response to the presence of another fungus. D), E), F), G) contain *Hp*-derived mPmTG filtrate after low (-20°C), high (60°C), ultra-low (-80°) and ultra-high (90°) temperature treatments, respectively. For D-G, N. crassa inoculation occurred after media was allowed to return to room temperature. H) *N. crassa* growth in fresh mPmTG media.

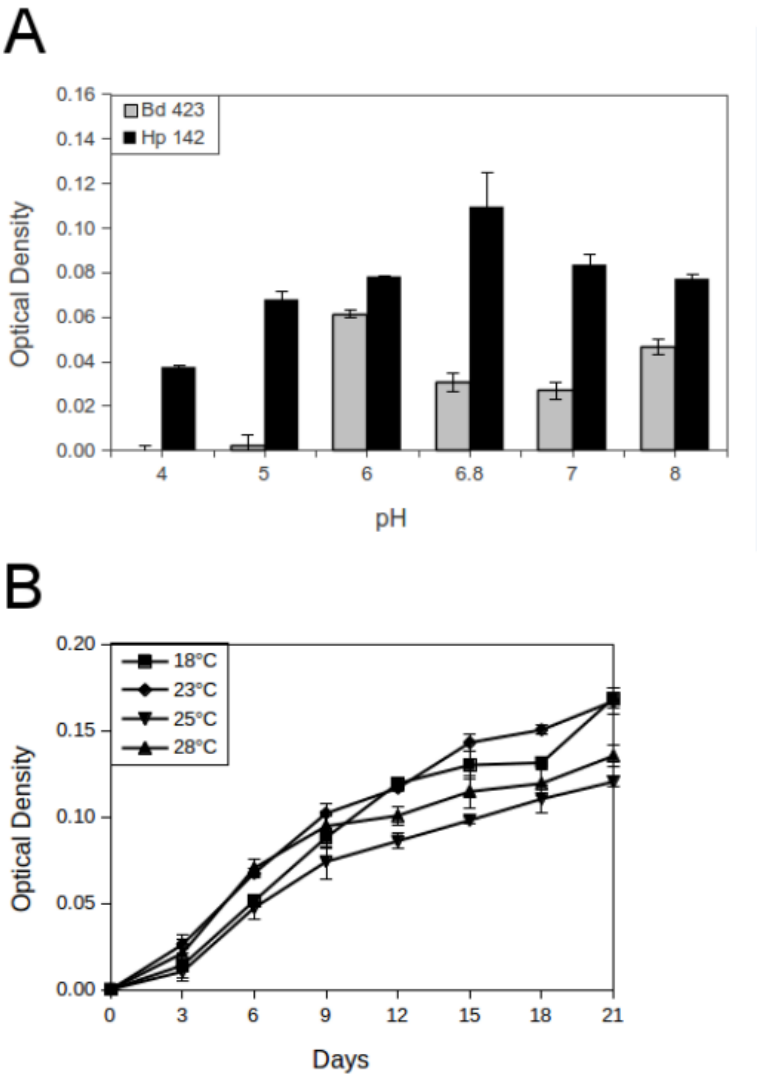


Figure 5.4: A) *Hp* and *Bd* were grown at different pH levels, ranging from 4 to 8. A pH of 6.8 represents the standard pH at which the respective optimal media is prepared. B) *Hp* growth after 21 days at various temperatures ranging from 18°C to 28°C.

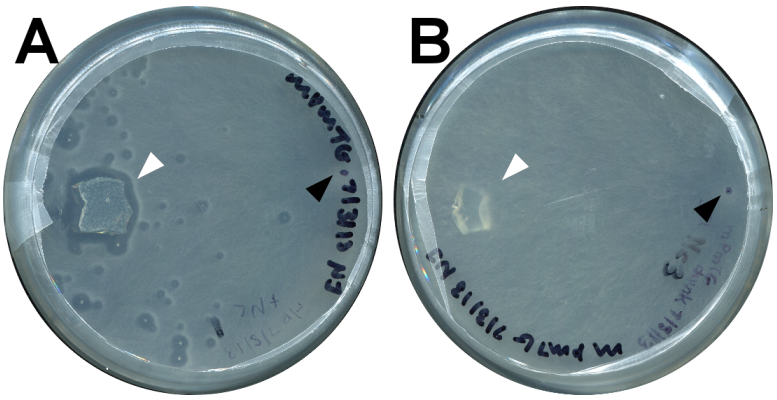


Figure 5.5: Solid agar plates of mPmTG media containing A) actively growing *H. polyrhiza* and B) no *H. polyrhiza* growth. After 72 hours post-inoculation of *N. crassa*, zones of inhibition are clearly visible surrounding *H. polyrhiza* sporangia. White arrowheads in A and B indicate agar blocks containing or lacking, respectively, *H. polyrhiza* sporangia. Black arrowheads indicate point of inoculation with *N. crassa* conidia. Plates are 100mm in diameter.

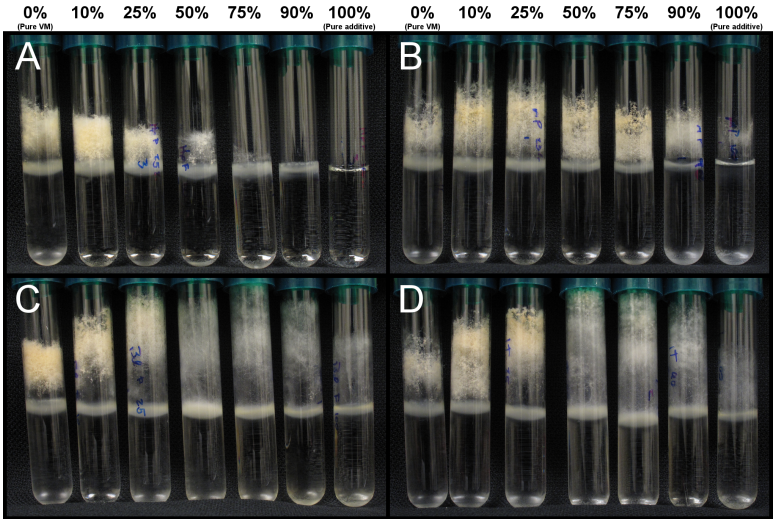


Figure 5.6: VM media diluted with various additives: A) *Hp* filtrate, B) fresh mPmTG, C) *Bd* filtrate, D) fresh 1% Tryptone. Dilution series in all panels (left to right): 0% (pure VM), 10%, 25%, 50%, 75%, 90%, 100% (pure additive).

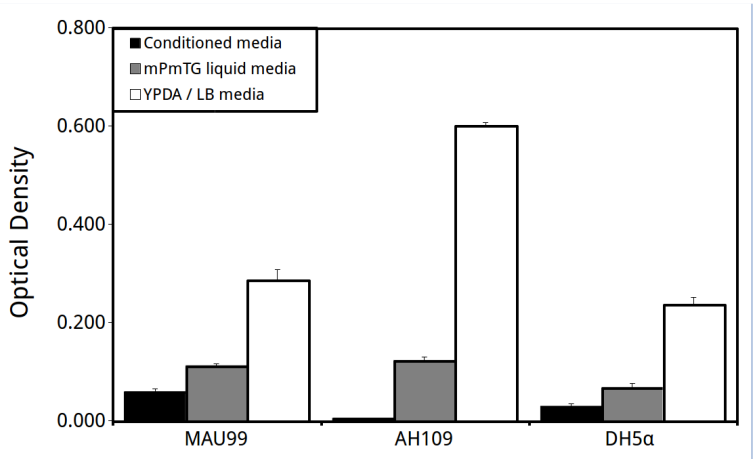


Figure 5.7: OD_{600} measurements of liquid cultures of *S. cerevisiae* MAU99 and AH109, and *E. coli* DH5α after 72h stationary incubation in different media preparations at room temperature: filter-sterilized *Hp*-conditioned media (black), *Hp* liquid mPmTG media (grey), and either liquid YPDA or LB for yeast or *E. coli*, respectively (white).

Table 5.1: List of *N. crassa* kinase and phosphatase mutants used in screening against *H. polyrhiza* sporangia. Group/Family abbreviations: S/T (serine/threonine), CMGC (cyclin-dependent, mitogen-activated,glycogen synthase, and cyclin-dependent protein kinase-like kinases), STE (sterile), CAMK/L (Ca^{2+}/calmodulin-dependent protein kinase, "CAMK"-like kinase), AGC, PPP (phosphoprotein phosphatase), MAPK (mitogen-activated protein kinase), PEK (pancreatic alpha-subunit of eukaryotic initiation factor kinase), WEE (small), YANK (yet another novel kinase), CLK/SRPK (cyclin-dependent protein kinase-like kinase / serine-rich protein kinase), VPS (vacuole protein sorting), CDK (cyclin-dependent protein kinase), IKS (Ira1 kinase suppressor).

Group	Family	NCU	*N. crassa* gene
S/T	PPP	NCU06563	*pp2A*
S/T	PPP	NCU01433	*ppt-1*
CMGC	MAPK	NCU07024	*os-2*
STE	STE11	NCU03071	*os-4*
STE	STE7	NCU00587	*os-5*
CAMK	CAMKL	NCU00914	*stk-16*
Other	PEK	NCU01187	*cpc-3*
STE	STE20	NCU03894	*stk-4*
Other	WEE	NCU04326	*stk-29*
Unclassified	–	NCU06421	*stk-41*
Unclassified	–	NCU06422	*stk-42*
Unclassified	–	NCU06583	*stk-44*
Other	VPS15	NCU06626	*stk-45*
AGC	YANK	NCU07062	*stk-49*
CGMC	CLK/SRPK	NCU10004	*stk-56*
Unclassified	-	NCU05638	*stk-34*
CMGC	CDK	NCU07880	*prk-6*
Other	IKS	NCU08177	*stk-51*
CAMK	CAMKL	NCU04747	*stk-31*

Table 5.2: Number of proteins predicted to be in the secretomes of basal fungi, using secretome predictive workflow optimized for Fungi and incorporating SignalP, Phobius, TargetP, WolfPSort, TMHMM, and PS-Scan (see [187]).

Species	Num of proteins
R. allomycis	32
G. prolifera	101
Piromyces sp. E2	269
Orpinomyces sp. C1	280
H. polyrhiza	58
S. punctatus	66
B. dendrobatidis	133
A. macrogynus	57
C. anguillulae	101
C. lativittatus	98

Table 5.3: Top-scoring BLAST hits in SwissProt for putative terpene synthase proteins identified in *Hp* using antiSMASH.

H. polyrhiza Query	SwissProt top hit	%ID	AlnLength	Eval
Hpol|HPOL_06475	Q6GLK6.1; Acyl-CoA synthetase fam.	44.89	528	3e-139
Hpol|HPOL_00800	P53798.2; Squalene synthase	44.85	379	3e-102
Hpol|HPOL_00236	Q8QZR1.1; Tyrosine aminotransferase	52.25	423	5e-151
Hpol|HPOL_00238	D3ZN43.1; NADH dehydrogenase	43.38	219	2e-63

Chapter 6

Conclusions

The objective of this dissertation research was to enhance the current knowledge about basal fungal groups. Genomic resources and methodologies are becoming increasingly more accessible, and so the ability to incorporate this data into existing studies is becoming more widespread. This objective was addressed with a particular focus on the mechanisms by which basal fungi interact with other organisms and with their environment.

Chapter 2 and Chapter 3 investigated the structural mechanics and components of basal fungal rhodopsin photosensory pathways. Sunlight is one of the most obvious environmental sources of information, and the most easily studied. It should come as no surprise that photoreception exists in some form in all three domains of life, however varied in its implementation. Found in the metazoan lineages, perhaps the most well known photoreceptor proteins is the Type 2 rhodopsin, a photoreceptive GPCR proteins which accomplishes its signaling via trandsucin and cGMP phosphodiesterase signaling cascade. The research in Chapters 2 and 3 address four key questions:

"How similar, structurally, are the opsin-like proteins from chytrids to each other and to other experimentally verified photoreceptors?", "How does this structural similarity (or dissimilarity) impact functional photoreception?", "What is the complement of downstream associated compenents (e.g., heterotrimeric G-protein subunits)?", and "Do the auxilary protein presence/absence correlate with known aspects of the evolution of photobiology in fungi?"

Sequence homology predicts a number of seven-transmembrane domain proteins in various chytrid species, with particular similarity to the members of the Type 2 GPCR rhodopsin family. I used sequences from *B. dendrobatidis*, *S. punctatus*, and *A. macrogynus* in homology modeling against known rhodopsin crystal structures to produce 3D structural models of these "chytriopsins". Quality metrics indicate that these computational models are reasonably sound predictions of the protein structures as they exist in the cell.

Therefore they are good candidates for analysis in ligand docking and molecular dynamics simulations to infer biological function. The structural analysis suggests that the rhodopsin-like protein identified in *S. punctatus* is most likely to be functional given the conservation of important structural features, most importantly the conserved lysine residue to which the chromophore ligand is bound. Automated non-covalent docking screens of retinal-like ligands against other chytropsin proteins from *B. dendrobatidis* and *A. macrogynus* suggest that even though these proteins lack the conserved lysine in a proper position, the binding pockets are spatially and chemically able to accomodate the 11-*cis*-retinal chromophore.

Covalent docking prediction of the *Sp* protein using 11-*cis*-retinal suggests that it is the most likely to be functionally active. This prediction was further refined with MD simulations using AMBER. The results were compared to the experimentally ver-

ified interaction between the *Todares pacificus* rhodopsin and 11-*cis*-retinal (PDB ID: 2Z73).

With respect to the proteins associated with rhodopsin signaling in animals, namely the heterotrimeric G proteins and effectors like phosphodiesterase, sequence similarity searcher uncovered homologs in the basal fungal lineages.

Chapter 4 dealt with the transcriptome analysis of the entomopathogenic Blastocladiomycete *Coelomomyces lativittatus*. Species in the basal fungal lineages occupy a diverse collection of environmental niches, including symbionts, pathogens, and saprotrophs. However, species in the genus *Coelomomyces*, itself a member of the Blastocladiomycota, are the only known basal fungi which are pathogenic in arthropods. Specifically, their development requires oscillation between two hosts: mosquito larvae and microcrustaceans. This lifecycle has made them both attractive targets for research into non-pesticide-based mosquito control, yet also difficult systems in which to pursue this research. The transcriptome analysis presented in Chapter 4 was addressing two major points about entompathogenic chytrids, using *Coelomomyces lativittatus* as a model: "How do aspects of the transcriptome regarding already known aspects of *C. lativittatus* biology, specifically regarding β-carotene biosynthesis, environment sensing, and insect-association?", and "How does the protein complement of *C. lativittatus* compare related Chytridiomycete and Blastocladiomycete species which are not insect-associated?"

PFAM prediction suggests a number of candidate enzymes related to insect virulence, the most prominent being members of the C1 peptidase family. These have documented antihelmintic effects and cuticle degrading activity in nematodes. Furthermore, this family seems expanded in *C. lativittatus* relative to other basal fungi, which are not insect-associated organisms.

The transcriptome analysis demonstrated that *C. lativittatus* likely has a complete complement of β-carotene processing enzymes, despite the apparent lack of one enzyme in the pathway: phytoene desaturase, the first enzyme in the three step cascade and responsible for converting phytoene to lycopene. Because the other two enzymes are accounted for, and because *Coelomomyces* are experimentally verified producers of β-carotene, it is unlikely that there is a novel biosynthetic pathway at work and instead the phyotene desaturase mRNA was not expressed at a high enough level to be recovered in either RNA extraction or during the HT sequencing.

The descriptive results from this investigation serve as starting points for future work involving genomic, transcriptomic, and proteomic analyses from multiple developmental stages and timepoints.

The research presented in this dissertation will inform future work on chytrids, and expand the knowledgebase for basal fungal lineages.

In Chapter 5 I explored the inhibitory properties of the chytridiomycte *Homolaphlyctis polyrhiza*. My surprising and unexpected observation in the Stajich lab of the non-pathogenic Chytridiomycte isolate *Homolaphlyctis polyrhiza* JEL142 inhibiting vegetative hyphal growth of a *Neurospora crassa* contaminant prompted an investigation into its biological nature. My work for this investigation pursued three major questions: "Is this a unique property of *Hp*?", "Is this a specific interaction with *N. crassa*?", and "What is the chemical nature of the responsible compound?"

By examining five other chytrid species in culture, I have evidence supporting the idea that this is a specific behavior for *Hp*. No other chytrid surveyed displayed the appropriate inhibitory phenotype, including the most closely related chytrid to *Hp*, the amphibian pathogen *Batrachochytrium dendrobatidis*.

Next, I expanded the potential targets of *Hp* to include members of the Ascomycota, Basidiomycota, and Zygomycota. All of these targets were susceptible to *Hp*, with the exception of *Rhizopus oryzae*.

Finally, I looked for the chemical nature for the mechanism of action by liquid assay and computational screening. I determined that the active compound is soluble and stable for at least 96h absent any *Hp* sporangia. Additionally, liquid obtained from preparations featuring both *Hp*+*N. crassa* and *Hp* alone were indistinguishably effective against reintroduction of *N. crassa* conidia.

Taken together, these data suggest that the compound is a constitutively-produced secondary metabolite compound with broadly specific activity. Its mechanism of action is unknown, as are its chemical structure and related biosynthetic pathway(s). Near term future work will necessarily focus on chemical profiling of bioactive spent media filtrate to generate a working hypothesis for the chemical nature of the product. A starting point for this research is provided in the form of *in silico* genomic and transcriptomic research. Finally, it is worth noting that the relative ease with which this discovery was made speaks to the necessity for further research into these basal lineages, of which intimate genomic and biochemical knowledge is lacking.

Bibliography

[1] J. W. Taylor and M. L. Berbee, "Dating divergences in the Fungal Tree of Life: review and new analyses," *Mycologia*, vol. 98, pp. 838–849, Nov. 2006.

[2] M. L. Berbee and J. W. Taylor, "Dating the evolutionary radiations of the true fungi," *Can. J. Bot.*, vol. 71, no. 8, pp. 1114–1127, 1993.

[3] R. F. Doolittle, D. F. Feng, S. Tsang, G. Cho, and E. Little, "Determining divergence times of the major kingdoms of living organisms with a protein clock," *Science*, vol. 271, pp. 470–477, 26 Jan. 1996.

[4] D. Y. Wang, S. Kumar, and S. B. Hedges, "Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi," *Proc. Biol. Sci.*, vol. 266, pp. 163–171, 22 Jan. 1999.

[5] M. L. Berbee and J. W. Taylor, "Dating the molecular clock in fungi – how close are we?," *Fungal Biol. Rev.*, vol. 24, pp. 1–16, Feb. 2010.

[6] D. S. Hibbett, M. Binder, J. F. Bischoff, M. Blackwell, P. F. Cannon, O. E. Eriksson, S. Huhndorf, T. Y. James, P. M. Kirk, R. Lücking, H. T. Lumbsch, F. Lutzoni, P. B. Matheny, D. J. McLaughlin, M. J. Powell, S. Redhead, C. L. Schoch, J. W. Spatafora, J. A. Stalpers, R. Vilgalys, M. C. Aime, A. Aptroot, R. Bauer, D. Begerow, G. L. Benny, L. A. Castlebury, P. W. Crous, Y.-C. Dai, W. Gams, D. M. Geiser, G. W. Griffith, C. Gueidan, D. L. Hawksworth, G. Hestmark, K. Hosaka, R. A. Humber, K. D. Hyde, J. E. Ironside, U. Kõljalg, C. P. Kurtzman, K.-H. Larsson, R. Lichtwardt, J. E. Longcore, J. Miądlikowska, A. Miller, J.-M. Moncalvo, S. E. Mozley-Standridge, F. Oberwinkler, E. Parmasto, V. Reeb, J. D. Rogers, C. Roux, L. Ryvarden, J. P. Sampaio, A. Schüßler, J. Sugiyama, R. G. Thorn, L. Tibell, W. A. Untereiner, C. Walker, Z. Wang, A. Weir, M. Weiss, M. M. White, K. Winka, Y.-J. Yao, and N. Zhang, "A higher-level phylogenetic classification of the Fungi," *Mycol. Res.*, vol. 111, pp. 509–547, May 2007.

[7] M. D. M. Jones, T. A. Richards, D. L. Hawksworth, and D. Bass, "Validation and justification of the phylum name *Cryptomycota* phyl. nov," *IMA Fungus*, vol. 2, pp. 173–175, Dec. 2011.

[8] M. M. White, T. Y. James, K. O'Donnell, M. J. Cafaro, Y. Tanabe, J. Sugiyama, N. Carolina, and K. O. Donnell, "Phylogeny of the Zygomycota based on nuclear ribosomal sequence data," *Mycologia*, vol. 98, no. 6, pp. 872–884, 2006.

[9] T. Y. James, P. M. Letcher, J. E. Longcore, S. E. Mozley-Standridge, D. Porter, M. J. Powell, G. W. Griffith, and R. Vilgalys, "A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota)," *Mycologia*, vol. 98, no. 6, pp. 860–871, 2006.

[10] C. J. Alexopoulos, C. W. Mims, and M. Blackwell, *Introductory Mycology.* John Wiley and Sons, Inc., fourth ed., 1996.

[11] J. E. Stajich, M. L. Berbee, M. Blackwell, D. S. Hibbett, T. Y. James, J. W. Spatafora, and J. W. Taylor, "The Fungi," *Curr. Biol.*, vol. 19, pp. R840–5, Sept. 2009.

[12] M. J. Powell, "Looking at mycology with a janus face: A glimpse at Chytridiomycetes active in the environment," *Mycologia*, vol. 85, no. 1, pp. 1–20, 1993.

[13] L. Tedersoo, M. Bahram, S. Põlme, U. Kõljalg, N. S. Yorou, R. Wijesundera, L. Villarreal Ruiz, A. M. Vasco-Palacios, P. Q. Thu, A. Suija, M. E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Põldmaa, M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Pärtel, E. Otsing, E. Nouhra, A. L. Njouonkou, R. H. Nilsson, L. N. Morgado, J. Mayor, T. W. May, L. Majuakim, D. J. Lodge, S. S. Lee, K.-H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T. W. Henkel, H. Harend, L.-D. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. De Kesel, T. Dang, X. Chen, F. Buegger, F. Q. Brearley, G. Bonito, S. Anslan, S. Abell, and K. Abarenkov, "Global diversity and geography of soil fungi," *Science*, vol. 346, p. 1256688, 28 Nov. 2014.

[14] K. R. Freeman, A. P. Martin, D. Karki, R. C. Lynch, M. S. Mitter, A. F. Meyer, J. E. Longcore, D. R. Simmons, and S. K. Schmidt, "Evidence that chytrids dominate fungal communities in high-elevation soils," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 18315–18320, 27 Oct. 2009.

[15] A. T. C. Paulitz, J. A. Menge, and T. C. Paulitz, "Is *Spizellomyces punctatum* a parasite or saprophyte of Vesicular-Arbuscular mycorrhizal fungi?," *Mycologia*, vol. 76, no. 1, pp. 99–107, 1984.

[16] S. L. Zebold, H. C. Whisler, J. A. Shemanchuk, and L. B. Travland, "Host specificity and penetration in the mosquito pathogen *Coelomomyces psorophorae*," *Can. J. Bot.*, vol. 57, no. 24, pp. 2766–2770, 1979.

[17] J. P. Tewari and P. Bains, "Fungi associated with the roots of clover in Alberta. i. *Olpidium brassicae* and *Ligniera* sp," *Can. Plant Dis. Surv.*, vol. 63, p. 2, 1983.

[18] J. E. Longcore, A. P. Pessier, and D. K. Nichols, "*Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians," *Mycologia*, vol. 91, no. 2, pp. 219–227, 1999.

[19] J. W. Deacon and G. Saxena, "Orientated zoospore attachment and cyst germination in *Catenaria anguillulae*, a facultative endoparasite of nematodes," *Mycol. Res.*, vol. 101, pp. 513–522, May 1997.

[20] H. M. Canter, "Studies on British chytrids: XXVI. A critical examination of *Zygorhizidium melosirae* Canter and *Z. planktonicum* Canter," *Journal of the Linnean Society of London, Botany*, vol. 60, pp. 85–97, 1 Feb. 1967.

[21] A. A. Held, "Encystment and germination of the parasitic chytrid *Rozella allomycis* on host hyphae," *Can. J. Bot.*, vol. 51, no. 10, pp. 1825–1835, 1973.

[22] H. M. Fitzpatrick, *The Lower Fungi - Phycomycetes*. first ed., 1930.

[23] W. J. Koch, "Studies of the motile cells of chytrids. I. Electron microscope observations of the flagellum, blepharoplast, and rhizoplast," *Am. J. Bot.*, vol. 45, no. 10, pp. 811–819, 1958.

[24] W. J. Koch, "Studies of the motile cells of chytrids. II. Internal structure of the body observed with light microscopy," *Am. J. Bot.*, vol. 45, no. 1, pp. 60–72, 1958.

[25] W. J. Koch, "Studies of the motile cells of chytrids. III. Major types," *Am. J. Bot.*, vol. 48, no. 9, pp. 786–788, 1961.

[26] T. D. Bruns, T. J. White, and J. W. Taylor, "Fungal molecular systematics," *Annu. Rev. Ecol. Syst.*, vol. 22, pp. 525–564, 1 Jan. 1991.

[27] T. D. Bruns, R. Vilgalys, S. M. Barns, D. Gonzalez, D. S. Hibbett, D. J. Lane, L. Simon, S. Stickel, T. M. Szaro, and W. G. Weisburg, "Evolutionary relationships within the Fungi: analyses of nuclear small subunit rRNA sequences," *Mol. Phylogenet. Evol.*, vol. 1, pp. 231–241, Sept. 1992.

[28] P. Wainright, G. Hinkle, M. Sogin, and S. Stickel, "Monophyletic origins of the Metazoa: an evolutionary link with Fungi," *Science*, vol. 260, pp. 340–342, Apr. 1993.

[29] T. Y. James, F. Kauff, C. L. Schoch, P. B. Matheny, V. Hofstetter, C. J. Cox, G. Celio, C. Gueidan, E. Fraker, J. Miądlikowska, H. T. Lumbsch, A. Rauhut, V. Reeb, A. E. Arnold, A. Amtoft, J. E. Stajich, K. Hosaka, G.-H. Sung, D. Johnson, B. O'Rourke, M. Crockett, M. Binder, J. M. Curtis, J. C. Slot, Z. Wang, A. W. Wilson, A. Schüssler, J. E. Longcore, K. O'Donnell, S. E. Mozley-Standridge, D. Porter, P. M. Letcher, M. J. Powell, J. W. Taylor, M. M. White, G. W. Griffith, D. R. Davies, R. A. Humber, J. B. Morton, J. Sugiyama, A. Y. Rossman, J. D. Rogers, D. H. Pfister, D. Hewitt, K. Hansen, S. Hambleton, R. A. Shoemaker, J. Kohlmeyer, B. Volkmann-Kohlmeyer, R. A. Spotts, M. Serdani, P. W. Crous, K. W. Hughes, K. Matsuura, E. Langer, G. Langer, W. A. Untereiner, R. Lücking, B. Büdel, D. M. Geiser, A. Aptroot, P. Diederich, I. Schmitt, M. Schultz, R. Yahr, D. S. Hibbett, F. Lutzoni, D. J. McLaughlin, J. W. Spatafora, and R. Vilgalys, "Reconstructing the early evolution of Fungi using a six-gene phylogeny," *Nature*, vol. 443, pp. 818–822, Oct. 2006.

[30] L. Berger, R. Speare, P. Daszak, D. E. Green, A. A. Cunningham, C. L. Goggin, R. Slocombe, M. A. Ragan, A. D. Hyatt, K. R. McDonald, H. B. Hines, K. R. Lips, G. Marantelli, and H. Parkes, "Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, pp. 9031–9036, July 1998.

[31] T. Y. James, D. Porter, C. A. Leander, R. Vilgalys, and J. E. Longcore, "Molecular phylogenetics of the *Chytridiomycota* supports the utility of ultrastructural data in chytrid systematics," *Can. J. Bot.*, vol. 78, no. 3, pp. 336–350, 2000.

[32] P. M. Letcher, J. E. Longcore, and M. J. Powell, "*Irineochytrium*, a new genus in Chytridiales having zoospores and aplanospores," *Mycologia*, vol. 106, pp. 1188–1198, Nov. 2014.

[33] W. J. Davis, P. M. Letcher, J. E. Longcore, and M. J. Powell, "*Fayochytriomyces*, a new genus within Chytridiales," *Mycologia*, vol. 107, pp. 432–439, Mar. 2015.

[34] D. R. Simmons, T. Y. James, A. F. Meyer, and J. E. Longcore, "*Lobulomycetales*, a new order in the Chytridiomycota," *Mycol. Res.*, vol. 113, pp. 450–460, Apr. 2009.

[35] K. F. Ribichich, S. M. Salem-Izacc, R. C. Georg, R. Z. N. Vêncio, L. D. Navarro, and S. L. Gomes, "Gene discovery and expression profile analysis through sequencing of expressed sequence tags from different developmental stages of the chytridiomycete *Blastocladiella emersonii*," *Eukaryot. Cell*, vol. 4, pp. 455–464, Feb. 2005.

[36] "*Batrachochytrium dendrobatidis* JAM81." `http://genomeportal.jgi-psf.org/Batde5/Batde5.home.html`.

[37] I. n. Ruiz-Trillo, G. Burger, P. W. H. Holland, N. King, B. F. Lang, A. J. Roger, and M. W. Gray, "The origins of multicellularity: a multi-taxon genome initiative," *Trends Genet.*, vol. 23, pp. 113–118, Mar. 2007.

[38] S. Joneson, J. E. Stajich, S.-H. Shiu, and E. B. Rosenblum, "Genomic transition to pathogenicity in chytrid fungi," *PLoS Pathog.*, vol. 7, p. e1002338, Nov. 2011.

[39] P. M. Letcher, M. J. Powell, and M. C. Viusent, "Rediscovery of an unusual chytridiaceous fungus new to the order Rhizophydiales," *Mycologia*, vol. 100, pp. 325–334, Mar. 2008.

[40] J. E. Longcore, P. M. Letcher, and T. Y. James, "*Homolaphlyctis polyrhiza* gen. et sp. nov., a species in the *Rhizophydiales* (*Chytridiomycetes*) with multiple rhizoidal axes," *Mycotaxon*, vol. 118, pp. 433–440, 2011.

[41] J. S. Karling, *Chytridiomycetarum Iconographia*. Lubrect and Cramer, 1977.

[42] Y. Chang, S. Wang, S. Sekimoto, A. L. Aerts, C. Choi, A. Clum, K. M. LaButti, E. A. Lindquist, C. Yee Ngan, R. A. Ohm, A. A. Salamov, I. V. Grigoriev, J. W. Spatafora, and M. L. Berbee, "Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants," *Genome Biol. Evol.*, vol. 7, pp. 1590–1601, 14 May 2015.

[43] C. G. Orpin, "Studies on the rumen flagellate *Neocallimastix frontalis*," *J. Gen. Microbiol.*, vol. 91, pp. 249–262, Dec. 1975.

[44] Y. W. Ho and D. J. S. Barr, "Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from Malaysia," *Mycologia*, vol. 87, pp. 655–677, 1 Sept. 1995.

[45] N. H. Youssef, M. Couger, C. G. Struchtemeyer, A. S. Liggenstoffer, R. A. Prade, F. Z. Najar, H. K. Atiyeh, M. R. Wilkins, and M. S. Elshahed, "Genome of the anaerobic fungus *Orpinomyces* sp. C1A reveals the unique evolutionary history of a

remarkable plant biomass degrader," *Appl. Environ. Microbiol.*, vol. 79, pp. 4620–4634, May 2013.

[46] R. J. Gruninger, A. K. Puniya, T. M. Callaghan, J. E. Edwards, N. Youssef, S. S. Dagar, K. Fliegerova, G. W. Griffith, R. Forster, A. Tsang, T. McAllister, and M. S. Elshahed, "Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential," *FEMS Microbiol. Ecol.*, vol. 90, pp. 1–17, Oct. 2014.

[47] T. Y. James, A. Pelin, L. Bonen, S. Ahrendt, D. Sain, N. Corradi, and J. E. Stajich, "Shared signatures of parasitism and phylogenomics unite Cryptomycota and Microsporidia," *Curr. Biol.*, vol. 23, pp. 1548–1553, Aug. 2013.

[48] J. Saranak and K. W. Foster, "Rhodopsin guides fungal phototaxis," *Nature*, vol. 387, pp. 465–466, May 1997.

[49] L. K. Muehlstein, J. P. Amon, and D. L. Leffler, "Phototaxis in the marine fungus *Rhizophydium littoreum*," *Appl. Environ. Microbiol.*, vol. 53, pp. 1668–1672, July 1987.

[50] A. Idnurm, S. Verma, and L. M. Corrochano, "A glimpse into the basis of vision in the kingdom Mycota," *Fungal Genet. Biol.*, vol. 47, pp. 881–892, Nov. 2010.

[51] N. C. Rockwell, Y.-S. Su, and J. C. Lagarias, "Phytochrome structure and signaling mechanisms," *Annu. Rev. Plant Biol.*, vol. 57, pp. 837–858, Jan. 2006.

[52] A. S. Rivera, N. Ozturk, B. Fahey, D. C. Plachetzki, B. M. Degnan, A. Sancar, and T. H. Oakley, "Blue-light-receptive cryptochrome is expressed in a sponge eye lacking neurons and opsin," *J. Exp. Biol.*, vol. 215, pp. 1278–1286, 15 Apr. 2012.

[53] P. Ballario and G. Macino, "White collar proteins: PASsing the light signal in *Neurospora crassa*," *Trends Microbiol.*, vol. 5, pp. 458–462, Nov. 1997.

[54] J. Purschwitz, S. Müller, C. Kastner, and R. Fischer, "Seeing the rainbow: light sensing in fungi," *Curr. Opin. Microbiol.*, vol. 9, pp. 566–571, Dec. 2006.

[55] L. M. Corrochano, "Fungal photoreceptors: sensory molecules for fungal development and behaviour," *Photochem. Photobiol. Sci.*, vol. 6, pp. 725–736, July 2007.

[56] K. L. Pierce, R. T. Premont, and R. J. Lefkowitz, "Seven-transmembrane receptors," *Nat. Rev. Mol. Cell Biol.*, vol. 3, pp. 639–650, Sept. 2002.

[57] G. Wald, "The molecular basis of visual excitation," *Nature*, vol. 219, pp. 800–807, 1968.

[58] J. von Lintig and K. Vogt, "Filling the gap in vitamin a research: Molecular identification of an enzyme cleaving β-carotene to retinal," *J. Biol. Chem.*, vol. 275, pp. 11915–11920, 21 Apr. 2000.

[59] S. O. Smith, "Structure and activation of the visual pigment rhodopsin," *Annu. Rev. Biophys.*, vol. 39, pp. 309–328, June 2010.

[60] A. Hausmann and G. Sandmann, "A single five-step desaturase is involved in the carotenoid biosynthesis pathway to β-carotene and torulene in *Neurospora crassa*," *Fungal Genet. Biol.*, vol. 30, pp. 147–153, July 2000.

[61] F. X. Cunningham, Jr., Z. Sun, D. Chamovitz, J. Hirschberg, and E. Gantt, "Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942," *Plant Cell*, vol. 6, pp. 1107–1121, 1 Aug. 1994.

[62] G. Lietz, A. Oxley, C. Boesch-Saadatmandi, and D. Kobayashi, "Importance of β,β-carotene 15,15'-monooxygenase 1 (BCMO1) and β,β-carotene 9',10'-dioxygenase 2 (BCDO2) in nutrition and health," *Mol. Nutr. Food Res.*, vol. 56, no. 2, pp. 241–250, 2012.

[63] G. P. Lobo, A. Isken, S. Hoff, D. Babino, and J. von Lintig, "BCDO2 acts as a carotenoid scavenger and gatekeeper for the mitochondrial apoptotic pathway," *Development*, vol. 139, pp. 2966–2977, Aug. 2012.

[64] Robert A. Samson, Harry C. Evans, and J.-P. Latgé, *Atlas of Entomopathogenic Fungi*. Springer Berlin Heidelberg, 1988.

[65] J. N. Couch and C. E. Bland, *The Genus Coelomomyces*. Orlando, FL 32887: Academic Press, Inc, 1985.

[66] M. F. Vicente, A. Basilio, A. Cabello, and F. Peláez, "Microbial natural products as a source of antifungals," *Clin. Microbiol. Infect.*, vol. 9, pp. 15–32, Jan. 2003.

[67] D. J. Barr, "The phylogenetic and taxonomic implications of flagellar rootlet morphology among zoosporic fungi," *Biosystems.*, vol. 14, no. 3-4, pp. 359–370, 1981.

[68] K. Inaba, "Molecular architecture of the sperm flagella: Molecules for motility and signaling," *Zoolog. Sci.*, vol. 20, pp. 1043–1056, Sept. 2003.

[69] C. D. Silflow and P. A. Lefebvre, "Assembly and motility of eukaryotic cilia and flagella. Lessons from *Chlamydomonas reinhardtii*," *Plant Physiology*, vol. 127, pp. 1500–1507, Dec. 2001.

[70] Y. J. Liu, M. C. Hodson, and B. D. Hall, "Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes," *BMC Evol. Biol.*, vol. 6, p. 74, 29 Sept. 2006.

[71] L. E. Padua, H. C. Whisler, B. P. Gabriel, and S. L. Zebold, "*In vivo* culture and life cycle of *Coelomomyces stegomyiae*," *J. Invertebr. Pathol.*, vol. 48, pp. 284–288, Nov. 1986.

[72] A. R. Cashmore, J. A. Jarillo, Y. J. Wu, and D. Liu, "Cryptochromes: blue light receptors for plants and animals," *Science*, vol. 284, pp. 760–765, 30 Apr. 1999.

[73] J. L. Spudich, C. S. Yang, K. H. Jung, and E. N. Spudich, "Retinylidene proteins: structures and functions from archaea to humans," *Annu. Rev. Cell Dev. Biol.*, vol. 16, pp. 365–392, Jan. 2000.

[74] A. Terakita, "The opsins," *Genome Biol.*, vol. 6, no. 3, p. 213, 2005.

[75] Y. Shichida and T. Matsuyama, "Evolution of opsins and phototransduction," *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 364, pp. 2881–2895, Oct. 2009.

[76] S. Sekharan and K. Morokuma, "Why 11-*cis*-retinal? why not 7-*cis*, 9-*cis* or 13-*cis*-retinal in the eye?," *J. Am. Chem. Soc.*, vol. 133, pp. 19052–19055, 30 Nov. 2011.

[77] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, "Pfam: the protein families database," *Nucleic Acids Res.*, vol. 42, pp. D222–30, Jan. 2014.

[78] F. Armougom, S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas, and C. Notredame, "Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee," *Nucleic Acids Res.*, vol. 34, pp. W604–8, July 2006.

[79] C. Notredame, D. G. Higgins, and J. Heringa, "T-Coffee: A novel method for fast and accurate multiple sequence alignment," *J. Mol. Biol.*, vol. 302, pp. 205–217, Sept. 2000.

[80] K. Katoh, K. Misawa, K.-I. Kuma, and T. Miyata, "MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform," *Nucleic Acids Res.*, vol. 30, pp. 3059–3066, July 2002.

[81] K. Katoh, K.-I. Kuma, H. Toh, and T. Miyata, "MAFFT version 5: improvement in accuracy of multiple sequence alignment," *Nucleic Acids Res.*, vol. 33, pp. 511–518, 20 Jan. 2005.

[82] R. C. Edgar, R. M. Drive, and M. Valley, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res.*, vol. 32, pp. 1792–1797, Jan. 2004.

[83] N. Eswar, B. Webb, M. a. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using MODELLER," *Curr. Protoc. Protein Sci.*, vol. Chapter 2, p. Unit 2.9, Nov. 2007.

[84] M. Lu, A. D. Dousis, and J. Ma, "OPUS-Rota: a fast and accurate method for side-chain modeling," *Protein Sci.*, vol. 17, pp. 1576–1585, Sept. 2008.

[85] Y. Zhang, "I-TASSER server for protein 3D structure prediction," *BMC Bioinformatics*, vol. 9, p. 40, 23 Jan. 2008.

[86] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, pp. 1157–1174, 15 July 2004.

[87] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," *J. Cheminform.*, vol. 4, p. 17, 13 Aug. 2012.

[88] M.-Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Sci.*, vol. 15, pp. 2507–2524, Nov. 2006.

[89] R. a. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, pp. 283–291, Apr. 1993.

[90] M. Wiederstein and M. J. Sippl, "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins," *Nucleic Acids Res.*, vol. 35, pp. W407–10, July 2007.

[91] R. Lüthy, J. U. Bowie, and D. Eisenberg, "Assessment of protein models with three-dimensional profiles," *Nature*, vol. 356, pp. 83–85, 5 Mar. 1992.

[92] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.3r1." August 2010.

[93] E. Beitz, "TeXtopo: shaded membrane protein topology plots in LaTeX2ÉŻ," *Bioinformatics*, vol. 16, no. 11, pp. 1050–1051, 2000.

[94] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *Journal of Computational Chemistry*, vol. 30, pp. 2785–2791, 30 Dec. 2009.

[95] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, "Chapter 12 - PubChem: Integrated platform of small molecules and biological activities," in *Annual Reports in Computational Chemistry* (Ralph A. Wheeler and David C. Spellmeyer, ed.), vol. Volume 4, pp. 217–241, Elsevier, 2008.

[96] J. J. Irwin and B. K. Shoichet, "ZINC–a free database of commercially available compounds for virtual screening," *J. Chem. Inf. Model.*, vol. 45, pp. 177–182, Jan. 2005.

[97] R. B. Russell and G. J. Barton, "Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels," *Proteins*, vol. 14, pp. 309–323, Oct. 1992.

[98] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *J. Mol. Graph.*, vol. 14, pp. 33–8, 27–8, Feb. 1996.

[99] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, "Bio3d: an R package for the comparative analysis of protein structures," *Bioinformatics*, vol. 22, pp. 2695–2696, Nov. 2006.

[100] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York, , and P. A. Kollman, *AMBER 2015*, 2015.

[101] A. Onufriev, D. Bashford, and A. David, "Modification of the generalized Born model suitable for macromolecules," *J. Phys. Chem. B*, vol. 104, no. 15, pp. 3712–3720, 2000.

[102] D. Bashford and D. A. Case, "Generalized Born models of macromolecular solvation effects," *Annu. Rev. Phys. Chem.*, vol. 51, pp. 129–152, 2000.

[103] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *J. Mol. Biol.*, vol. 7, pp. 95–99, July 1963.

[104] R. R. Birge, "Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin," *Biochim. Biophys. Acta*, vol. 1016, pp. 293–327, Apr. 1990.

[105] U. Schweiger, J. Tittor, and D. Oesterhelt, "Bacteriorhodopsin can function without a covalent linkage between retinal and protein," *Biochemistry*, vol. 33, pp. 535–541, Jan. 1994.

[106] E. a. Zhukovsky, P. R. Robinson, and D. D. Oprian, "Changing the location of the schiff base counterion in rhodopsin," *Biochemistry*, vol. 31, pp. 10400–10405, Oct. 1992.

[107] M. C. Lagerström and H. B. Schiöth, "Structural diversity of G protein-coupled receptors and significance for drug discovery," *Nat. Rev. Drug Discov.*, vol. 7, pp. 339–357, Apr. 2008.

[108] J. R. Hepler and A. G. Gilman, "G proteins," *TIBS*, vol. 17, pp. 383–387, Oct. 1992.

[109] A. G. Gilman, "G proteins: transducers of receptor-generated signals," *Annu. Rev. Biochem.*, vol. 56, pp. 615–649, Jan. 1987.

[110] D. E. Clapham and E. J. Neer, "G protein $\beta\gamma$ subunits," *Annu. Rev. Pharmacol. Toxicol.*, vol. 37, pp. 167–203, 1997.

[111] K. L. Clark, D. Dignard, D. Y. Thomas, and M. Whiteway, "Interactions among the subunits of the G protein involved in *Saccharomyces cerevisiae* mating," *Mol. Cell. Biol.*, vol. 13, pp. 1–8, Jan. 1993.

[112] S. R. Neves, P. T. Ram, and R. Iyengar, "G protein pathways," *Science*, vol. 296, pp. 1636–1639, 31 May 2002.

[113] K. A. Borkovich, "Signal transduction pathways and heterotrimeric G proteins," in *The Mycota III* (K. Esser and P. A. Lemke, eds.), pp. 211–233, 1996.

[114] M. Bölker, "Sex and crime: heterotrimeric G proteins in fungal mating and pathogenesis," *Fungal Genet. Biol.*, vol. 25, pp. 143–156, Dec. 1998.

[115] L. Li, S. J. Wright, S. Krystofova, G. Park, and K. A. Borkovich, "Heterotrimeric G protein signaling in filamentous fungi," *Annu. Rev. Microbiol.*, vol. 61, pp. 423–452, Jan. 2007.

[116] S. Liu and R. A. Dean, "G protein alpha subunit genes control growth, development, and pathogenicity of *Magnaporthe grisea*," *Mol. Plant. Microbe. Interact.*, vol. 10, pp. 1075–1086, Dec. 1997.

[117] L. De Vries, B. Zheng, T. Fischer, E. Elenko, and M. G. Farquhar, "The regulator of G protein signaling family," *Annu. Rev. Pharmacol. Toxicol.*, vol. 40, pp. 235–271, 2000.

[118] M. Conti, "Phosphodiesterases and cyclic nucleotide signaling in endocrine cells," *Mol. Endocrinol.*, vol. 14, pp. 1317–1327, Sept. 2000.

[119] P. Deterre, J. Bigay, F. Forquet, M. Robert, and M. Chabre, "cGMP phosphodiesterase of retinal rods is regulated by two inhibitory subunits," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, pp. 2424–2428, Apr. 1988.

[120] S. Krystofova and K. A. Borkovich, "The heterotrimeric g-protein subunits GNG-1 and GNB-1 form a G$\beta\gamma$ dimer required for normal female fertility, asexual development, and Gα protein levels in *Neurospora crassa*," *Eukaryot. Cell*, vol. 4, pp. 365–378, Feb. 2005.

[121] W. R. Pearson and D. J. Lipman, "Improved tools for biological sequence comparison," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, pp. 2444–2448, Apr. 1988.

[122] S. R. Eddy, "Accelerated profile HMM searches," *PLoS Comput. Biol.*, vol. 7, p. e1002195, Oct. 2011.

[123] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, pp. 1312–1313, 1 May 2014.

[124] E. Beitz, "TeXshade: shading and labeling of multiple sequence alignments using LaTeX2ÉŻ," *Bioinformatics*, vol. 16, pp. 135–139, 1 Feb. 2000.

[125] J. A. Bieszke, E. N. Spudich, K. L. Scott, K. A. Borkovich, and J. L. Spudich, "A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment," *Biochemistry*, vol. 38, pp. 14138–14145, Oct. 1999.

[126] G. M. Avelar, R. I. Schumacher, P. A. Zaini, G. Leonard, T. A. Richards, and S. L. Gomes, "A Rhodopsin-Guanylyl cyclase gene fusion functions in visual perception in a fungus," *Curr. Biol.*, 14 May 2014.

[127] K. A. Borkovich, L. A. Alex, O. Yarden, M. Freitag, G. E. Turner, N. D. Read, S. Seiler, D. Bell-Pedersen, J. Paietta, N. Plesofsky, M. Plamann, M. Goodrich-Tanrikulu, U. Schulte, G. Mannhaupt, F. E. Nargang, A. Radford, C. Selitrennikoff, J. E. Galagan, J. C. Dunlap, J. J. Loros, D. Catcheside, H. Inoue, R. Aramayo, M. Polymenis, E. U. Selker, M. S. Sachs, G. A. Marzluf, I. Paulsen, R. Davis, D. J. Ebbole, A. Zelter, E. R. Kalkman, R. O'Rourke, F. Bowring, J. Yeadon, C. Ishii, K. Suzuki, W. Sakai, and R. Pratt, "Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism," *Microbiol. Mol. Biol. Rev.*, vol. 68, pp. 1–108, Mar. 2004.

[128] F. D. Ivey, P. N. Hodge, G. E. Turner, and K. A. Borkovich, "The gαi homologue *gna-1* controls multiple differentiation pathways in *Neurospora crassa*," *Mol. Biol. Cell*, vol. 7, pp. 1283–1297, Aug. 1996.

[129] N. G. Abdulaev, M. P. Popp, W. C. Smith, and K. D. Ridge, "Functional expression of bovine opsin in the methylotrophic yeast *Pichia pastoris*," *Protein Expr. Purif.*, vol. 10, pp. 61–69, June 1997.

[130] H. C. Whisler, S. L. Zebold, and J. A. Shemanchuk, "Life history of *Coelomomyces psorophorae*," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 72, pp. 693–696, Feb. 1975.

[131] L. B. Travland, "Initiation of infection of mosquito larvae (*Culiseta inornata*) by *Coelomomyces psorophorae*," *J. Invertebr. Pathol.*, vol. 33, pp. 95–105, Jan. 1979.

[132] H. C. Whisler, D. M. S. Karanja, J. A. Shemanchuk, S. L. Zebold, S. V. Romney, and L. T. Nielsen, "The life history and *in vivo* culture of *Coelomomyces utahensis* (blastocladiomycetes)," *J. Invertebr. Pathol.*, vol. 100, pp. 40–43, Jan. 2009.

[133] E.-J. Scholte, B. G. J. Knols, R. A. Samson, and W. Takken, "Entomopathogenic fungi for mosquito control: a review," *J. Insect Sci.*, vol. 4, p. 19, 23 June 2004.

[134] R. J. St. Leger, A. K. Charnley, and R. M. Cooper, "Cuticle-degrading enzymes of entomopathogenic fungi: Synthesis in culture on cuticle," *J. Invertebr. Pathol.*, vol. 48, pp. 85–95, July 1986.

[135] B. Wang, Q. Kang, Y. Lu, L. Bai, and C. Wang, "Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, pp. 1287–1292, 24 Jan. 2012.

[136] C. J. Lucarotti, "Invasion of *Aedes aegypti* ovaries by *Coelomomyces stegomyiae*," *J. Invertebr. Pathol.*, vol. 60, pp. 176–184, Sept. 1992.

[137] B. A. Federici, "Species-specific gating of gametangial dehiscence as a temporal reproductive isolating mechanism in *Coelomomyces*," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 80, pp. 604–607, Jan. 1983.

[138] B. A. Federici and S. N. Thompson, "β-carotene in the gametophytic phase of *Coelomomyces dodgei*," *Experimental Mycology*, vol. 3, pp. 281–284, Sept. 1979.

[139] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev, "Full-length transcriptome assembly from RNA-Seq data without a reference genome," *Nat. Biotechnol.*, vol. 29, pp. 644–652, July 2011.

[140] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A flexible trimmer for Illumina sequence data," *Bioinformatics*, 1 Apr. 2014.

[141] B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev, "*De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis," *Nat. Protoc.*, vol. 8, pp. 1494–1512, Aug. 2013.

[142] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *J. Mol. Biol.*, vol. 305, pp. 567–580, Jan. 2001.

[143] L. Li, C. J. Stoeckert, Jr, and D. S. Roos, "OrthoMCL: identification of ortholog groups for eukaryotic genomes," *Genome Res.*, vol. 13, pp. 2178–2189, Sept. 2003.

[144] M. J. Ruiz-Hidalgo, E. P. Benito, G. Sandmann, and A. P. Eslava, "The phytoene dehydrogenase gene of *Phycomyces*: regulation of its expression by blue light and vitamin a," *Mol. Gen. Genet.*, vol. 253, pp. 734–744, 27 Feb. 1997.

[145] A. Velayos, A. P. Eslava, and E. A. Iturriaga, "A bifunctional enzyme with lycopene cyclase and phytoene synthase activities is encoded by the *carrp* gene of *Mucor circinelloides*," *Eur. J. Biochem.*, vol. 267, pp. 5509–5519, Sept. 2000.

[146] A. Wyss, G. Wirtz, W.-D. Woggon, R. Brugger, M. Wyss, A. Friedlein, H. Bachmann, and W. Hunziker, "Cloning and expression of β,β-carotene 15,15'-dioxygenase," *Biochem. Biophys. Res. Commun.*, vol. 271, pp. 334–336, 10 May 2000.

[147] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, pp. 27–30, 1 Jan. 2000.

[148] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Data, information, knowledge and principle: back to metabolism in KEGG," *Nucleic Acids Res.*, vol. 42, pp. D199–205, Jan. 2014.

[149] G. Stepek, J. M. Behnke, D. J. Buttle, and I. R. Duce, "Natural plant cysteine proteinases as anthelmintics?," *Trends Parasitol.*, vol. 20, pp. 322–327, July 2004.

[150] N. D. Rawlings and A. J. Barrett, "[2] families of serine peptidases," in *Proteolytic Enzymes: Serine and Cysteine Peptidases* (Alan J. Barrett, ed.), vol. Volume 244, pp. 19–61, Academic Press, 1994.

[151] B. G. G. Donzelli, S. B. Krasnoff, Y. Sun-Moon, A. C. L. Churchill, and D. M. Gibson, "Genetic basis of destruxin production in the entomopathogen *Metarhizium robertsii*," *Curr. Genet.*, vol. 58, pp. 105–116, Apr. 2012.

[152] M. S. C. Pedras, L. Irina Zaharia, and D. E. Ward, "The destruxins: synthesis, biosynthesis, biotransformation, and biological activity," *Phytochemistry*, vol. 59, pp. 579–596, Mar. 2002.

[153] K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, and T. Weber, "antiSMASH 2.0–a versatile platform for genome mining of secondary metabolite producers," *Nucleic Acids Res.*, vol. 41, pp. W204–12, 3 June 2013.

[154] K. E. Bushley and B. G. Turgeon, "Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships," *BMC Evol. Biol.*, vol. 10, p. 26, 26 Jan. 2010.

[155] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, "The carbohydrate-active enzymes database (CAZy) in 2013," *Nucleic Acids Res.*, vol. 42, pp. D490–5, Jan. 2014.

[156] J. F. Staab, S. D. Bradway, P. L. Fidel, and P. Sundstrom, "Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1," *Science*, vol. 283, pp. 1535–1538, 5 Mar. 1999.

[157] C. S. Thummel and J. Chory, "Steroid signaling in plants and insects—common themes, different pathways," *Genes Dev.*, vol. 16, pp. 3113–3129, 15 Dec. 2002.

[158] M. R. Koelle, W. S. Talbot, W. A. Segraves, M. T. Bender, P. Cherbas, and D. S. Hogness, "The *Drosophila ecr* gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily," *Cell*, vol. 67, pp. 59–77, 4 Oct. 1991.

[159] H. Escriva, M.-C. Langlois, R. L. Mendonça, R. Pierce, and V. Laudet, "Evolution and diversification of the nuclear receptor superfamily," *Ann. N. Y. Acad. Sci.*, vol. 839, no. 1, pp. 143–146, 1998.

[160] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter, "InterProScan 5: genome-scale protein function classification," *Bioinformatics*, vol. 30, pp. 1236–1240, 1 May 2014.

[161] M. Jakób, R. Kołodziejczyk, M. Orłowski, S. Krzywda, A. Kowalska, J. Dutko-Gwóźdź, T. Gwóźdź, M. Kochman, M. Jaskólski, and A. Ozyhar, "Novel DNA-binding element within the C-terminal extension of the nuclear receptor DNA-binding domain," *Nucleic Acids Res.*, vol. 35, pp. 2705–2718, 10 Apr. 2007.

[162] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, pp. 195–201, 15 Jan. 2006.

[163] L. Holm and P. Rosenström, "Dali server: conservation mapping in 3D," *Nucleic Acids Res.*, vol. 38, pp. W545–9, July 2010.

[164] C. Schwerdtfeger and H. Linden, "VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation," *EMBO J.*, vol. 22, pp. 4846–4855, 15 Sept. 2003.

[165] A. M. Calvo, "The VeA regulatory system and its role in morphological and chemical development in fungi," *Fungal Genet. Biol.*, vol. 45, pp. 1053–1061, July 2008.

[166] H. C. Whisler, S. L. Zebold, and J. A. Shemanchuk, "Alternate host for mosquito parasite *Coelomomyces*," *Nature*, vol. 251, pp. 715–716, 1974.

[167] M. Kokkonen, L. Ojala, P. Parikka, and M. Jestoi, "Mycotoxin production of selected *Fusarium* species at different culture conditions," *Int. J. Food Microbiol.*, vol. 143, pp. 17–25, 30 Sept. 2010.

[168] P. Wiemann and N. P. Keller, "Strategies for mining fungal natural products," *J. Ind. Microbiol. Biotechnol.*, vol. 41, pp. 301–313, Feb. 2014.

[169] B. J. Howlett, "Secondary metabolite toxins and nutrition of plant pathogenic fungi," *Curr. Opin. Plant Biol.*, vol. 9, pp. 371–375, Aug. 2006.

[170] J. C. Sherris, *Sherris medical microbiology: an introduction to infectious diseases.* McGraw-Hill Medical Publishing, 1994.

[171] A. Casadevall, "Progress in understanding fungal pathogenesis," *Curr. Opin. Microbiol.*, vol. 16, pp. 375–376, Aug. 2013.

[172] R. Dean, J. A. L. Van Kan, Z. A. Pretorius, K. E. Hammond-Kosack, A. Di Pietro, P. D. Spanu, J. J. Rudd, M. Dickman, R. Kahmann, J. Ellis, and G. D. Foster, "The top 10 fungal pathogens in molecular plant pathology," *Mol. Plant Pathol.*, vol. 13, pp. 414–430, May 2012.

[173] H. Vanden Bossche, F. Dromer, I. Improvisi, M. Lozano-Chiu, J. H. Rex, and D. Sanglard, "Antifungal drug resistance in pathogenic fungi," *Med. Mycol.*, vol. 36 Suppl 1, pp. 119–128, 1998.

[174] D. P. Kontoyiannis and R. E. Lewis, "Antifungal drug resistance of pathogenic fungi," *Lancet*, vol. 359, pp. 1135–1144, 30 Mar. 2002.

[175] M. Kretschmer, M. Leroch, A. Mosbach, A.-S. Walker, S. Fillinger, D. Mernke, H.-J. Schoonbeek, J.-M. Pradier, P. Leroux, M. A. De Waard, and M. Hahn, "Fungicide-driven evolution and molecular basis of multidrug resistance in field populations of the grey mould fungus *Botrytis cinerea*," *PLoS Pathog.*, vol. 5, p. e1000696, Dec. 2009.

[176] M. Leroch, M. Kretschmer, and M. Hahn, "Fungicide resistance phenotypes of *Botrytis cinerea* isolates from commercial vineyards in South West Germany," *Journal of Phytopathology*, vol. 159, pp. 63–65, 1 Jan. 2011.

[177] M. A. Pfaller, "Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment," *Am. J. Med.*, vol. 125, pp. S3–13, Jan. 2012.

[178] J. Bérdy, "Thoughts and facts about antibiotics: Where we are now and where we are heading," *J. Antibiot.*, vol. 65, pp. 385–395, 18 Apr. 2012.

[179] R. B. Davis, D. S. Anderson, S. A. Norton, and M. C. Whiting, "Acidity of twelve northern new england (U.S.A.) lakes in recent centuries," *J. Paleolimnol.*, vol. 12, pp. 103–154, 1 Jan. 1994.

[180] T. E. Rhodes and R. B. Davis, "Effects of late Holocene forest disturbance and vegetation change on Acidic Mud Pond, Maine, USA," *Ecology*, vol. 76, pp. 734–746, 1 Apr. 1995.

[181] D. J. S. Barr, "Taxonomy and phylogeny of chytrids," *Biosystems.*, vol. 10, pp. 153–165, 1978.

[182] D. R. H. de Serres, F. J., "Genetic and microbiological research techniques for *Neurospora crassa*," *Meth. Enzymol.*, 1970.

[183] G. Park, J. A. Servin, G. E. Turner, L. Altamirano, H. Colot, V, P. Collopy, L. Litvinkova, L. Li, C. A. Jones, F.-G. Diala, J. C. Dunlap, and K. A. Borkovich, "Global analysis of serine-threonine protein kinase genes in *Neurospora crassa*," *Eukaryot. Cell*, vol. 10, pp. 1553–1564, Nov. 2011.

[184] E. W. Myers, G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo, M. J. Flanigan, S. A. Kravitz, C. M. Mobarry, K. H. Reinert, K. A. Remington, E. L. Anson, R. A. Bolanos, H. H. Chou, C. M. Jordan, A. L. Halpern, S. Lonardi, E. M. Beasley, R. C. Brandon, L. Chen, P. J. Dunn, Z. Lai, Y. Liang, D. R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G. M. Rubin, M. D. Adams, and J. C. Venter, "A whole-genome assembly of *Drosophila*," *Science*, vol. 287, pp. 2196–2204, 24 Mar. 2000.

[185] B. L. Cantarel, I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado, and M. Yandell, "MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes," *Genome Res.*, vol. 18, pp. 188–196, Jan. 2008.

[186] N. Khaldi, F. T. Seifuddin, G. Turner, D. Haft, W. C. Nierman, K. H. Wolfe, and N. D. Fedorova, "SMURF: Genomic mapping of fungal secondary metabolite clusters," *Fungal Genet. Biol.*, vol. 47, pp. 736–741, Sept. 2010.

[187] X. J. Min, "Evaluation of computational methods for secreted protein prediction in different eukaryotes," *J. Proteomics Bioinform.*, vol. 3, pp. 143–147, 21 Apr. 2010.

[188] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: SignalP 3.0," *J. Mol. Biol.*, vol. 340, pp. 783–795, 16 July 2004.

[189] L. Käll, A. Krogh, and E. L. L. Sonnhammer, "A combined transmembrane topology and signal peptide prediction method," *J. Mol. Biol.*, vol. 338, pp. 1027–1036, 14 May 2004.

[190] L. Käll, A. Krogh, and E. L. L. Sonnhammer, "Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server," *Nucleic Acids Res.*, vol. 35, pp. W429–W432, 1 July 2007.

[191] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence," *J. Mol. Biol.*, vol. 300, pp. 1005–1016, 21 July 2000.

[192] O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen, "Locating proteins in the cell using TargetP, SignalP and related tools," *Nat. Protoc.*, vol. 2, no. 4, pp. 953–971, 2007.

[193] J. Sprenger, J. L. Fink, and R. D. Teasdale, "Evaluation and comparison of mammalian subcellular localization prediction methods," *BMC Bioinformatics*, vol. 7 Suppl 5, p. S3, 18 Dec. 2006.

[194] P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai, "WoLF PSORT: protein localization predictor," *Nucleic Acids Res.*, vol. 35, pp. W585–7, July 2007.

[195] E. de Castro, C. J. A. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo, "ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins," *Nucleic Acids Res.*, vol. 34, pp. W362–5, 1 July 2006.

[196] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta*, vol. 405, pp. 442–451, 20 Oct. 1975.

[197] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, pp. 412–424, May 2000.

[198] K. M. Menne, H. Hermjakob, and R. Apweiler, "A comparison of signal sequence prediction methods using a test set of signal peptides," *Bioinformatics*, vol. 16, pp. 741–742, Aug. 2000.

[199] J. S. Piotrowski, S. L. Annis, and J. E. Longcore, "Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians," *Mycologia*, vol. 96, no. 1, pp. 9–15, 2004.

[200] L. T. Haimo and J. L. Rosenbaum, "Cilia, flagella, and microtubules," *J. Cell Biol.*, vol. 91, pp. 125s–130s, Dec. 1981.

[201] S. Sekimoto, D. Rochon, J. E. Long, J. M. Dee, and M. L. Berbee, "A multigene phylogeny of *Olpidium* and its implications for early fungal evolution," *BMC Evol. Biol.*, vol. 11, p. 331, Jan. 2011.

[202] L. K. Fritz-Laylin, M. L. Ginger, C. Walsh, S. C. Dawson, and C. Fulton, "The *Naegleria* genome: a free-living microbial eukaryote lends unique insights into core eukaryotic cell biology," *Res. Microbiol.*, vol. 162, pp. 607–618, July 2011.

[203] A. Burmester, E. Shelest, G. Glöckner, C. Heddergott, S. Schindler, P. Staib, A. Heidel, M. Felder, A. Petzold, K. Szafranski, M. Feuermann, I. Pedruzzi, S. Priebe, M. Groth, R. Winkler, W. Li, O. Kniemeyer, V. Schroeckh, C. Hertweck, B. Hube, T. C. White, M. Platzer, R. Guthke, J. Heitman, J. Wöstemeyer, P. F. Zipfel, M. Monod, and A. A. Brakhage, "Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi," *Genome Biol.*, vol. 12, p. R7, 19 Jan. 2011.

[204] W. C. Nierman, A. Pain, M. J. Anderson, J. R. Wortman, H. S. Kim, J. Arroyo, M. Berriman, K. Abe, D. B. Archer, C. Bermejo, J. Bennett, P. Bowyer, D. Chen, M. Collins, R. Coulsen, R. Davies, P. S. Dyer, M. Farman, N. Fedorova, N. Fedorova, T. V. Feldblyum, R. Fischer, N. Fosker, A. Fraser, J. L. García, M. J. García, A. Goble, G. H. Goldman, K. Gomi, S. Griffith-Jones, R. Gwilliam, B. Haas, H. Haas, D. Harris, H. Horiuchi, J. Huang, S. Humphray, J. Jiménez, N. Keller, H. Khouri, K. Kitamoto, T. Kobayashi, S. Konzack, R. Kulkarni, T. Kumagai, A. Lafon, A. Lafton, J.-P. Latgé, W. Li, A. Lord, C. Lu, W. H. Majoros, G. S. May, B. L. Miller, Y. Mohamoud, M. Molina, M. Monod, I. Mouyna, S. Mulligan, L. Murphy, S. O'Neil, I. Paulsen, M. A. Peñalva, M. Pertea, C. Price, B. L. Pritchard, M. A. Quail, E. Rabbinowitsch, N. Rawlins, M.-A. Rajandream, U. Reichard, H. Renauld, G. D. Robson, S. Rodriguez de Córdoba, J. M. Rodríguez-Peña, C. M. Ronning, S. Rutter, S. L. Salzberg, M. Sanchez, J. C. Sánchez-Ferrero, D. Saunders, K. Seeger, R. Squares, S. Squares, M. Takeuchi, F. Tekaia, G. Turner, C. R. Vazquez de Aldana, J. Weidman, O. White, J. Woodward, J.-H. Yu, C. Fraser, J. E. Galagan, K. Asai, M. Machida, N. Hall, B. Barrell, and D. W. Denning, "Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*," *Nature*, vol. 438, pp. 1151–1156, 22 Dec. 2005.

[205] F. S. Dietrich, S. Voegeli, S. Brachat, A. Lerch, K. Gates, S. Steiner, C. Mohr, R. Pöhlmann, P. Luedi, S. Choi, R. A. Wing, A. Flavier, T. D. Gaffney, and P. Philippsen, "The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome," *Science*, vol. 304, pp. 304–307, 9 Apr. 2004.

[206] C. H. Slamovits, N. M. Fast, J. S. Law, and P. J. Keeling, "Genome compaction and stability in microsporidian intracellular parasites," *Curr. Biol.*, vol. 14, pp. 891–896, 25 May 2004.

[207] N. Corradi, D. E. Akiyoshi, H. G. Morrison, X. Feng, L. M. Weiss, S. Tzipori, and P. J. Keeling, "Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antonospora locustae* and *Enterocytozoon bieneusi*," *PLoS One*, vol. 2, p. e1277, 5 Dec. 2007.

[208] J. E. Galagan, S. E. Calvo, C. Cuomo, L.-J. Ma, J. R. Wortman, S. Batzoglou, S.-I. Lee, M. Baştürkmen, C. C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scazzocchio, M. Farman, J. Butler, S. Purcell, S. Harris, G. H. Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G. H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J. H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E. U. Selker, D. B. Archer, M. A. Peñalva, B. R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W. C. Nierman, D. W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M. S. Sachs, S. A. Osmani, and B. W. Birren, "Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*," *Nature*, vol. 438, pp. 1105–1115, 22 Dec. 2005.

[209] J. Amselem, C. A. Cuomo, J. A. L. van Kan, M. Viaud, E. P. Benito, A. Couloux, P. M. Coutinho, R. P. de Vries, P. S. Dyer, S. Fillinger, E. Fournier, L. Gout, M. Hahn, L. Kohn, N. Lapalu, K. M. Plummer, J.-M. Pradier, E. Quévillon, A. Sharon, A. Simon, A. ten Have, B. Tudzynski, P. Tudzynski, P. Wincker, M. Andrew, V. Anthouard, R. E. Beever, R. Beffa, I. Benoit, O. Bouzid, B. Brault, Z. Chen, M. Choquer, J. Collémare, P. Cotton, E. G. Danchin, C. Da Silva, A. Gautier, C. Giraud, T. Giraud, C. Gonzalez, S. Grossetete, U. Güldener, B. Henrissat, B. J. Howlett, C. Kodira, M. Kretschmer, A. Lappartient, M. Leroch, C. Levis, E. Mauceli, C. Neuvéglise, B. Oeser, M. Pearson, J. Poulain, N. Poussereau, H. Quesneville, C. Rascle, J. Schumacher, B. Ségurens, A. Sexton, E. Silva, C. Sirven, D. M. Soanes, N. J. Talbot, M. Templeton, C. Yandava, O. Yarden, Q. Zeng, J. A. Rollins, M.-H. Lebrun, and M. Dickman, "Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*," *PLoS Genet.*, vol. 7, p. e1002230, Aug. 2011.

[210] "*Backusella circina* FSU 941." `http://genome.jgi.doe.gov/Bacci1/Bacci1.home.html`.

[211] J. E. Stajich, T. Harris, B. P. Brunk, J. Brestelli, S. Fischer, O. S. Harb, J. C. Kissinger, W. Li, V. Nayak, D. F. Pinney, C. J. Stoeckert, Jr, and D. S. Roos, "FungiDB: an integrated functional genomics database for fungi," *Nucleic Acids Res.*, vol. 40, pp. D675–81, Jan. 2012.

[212] M. J. McCullough, A. F. DiSalvo, K. V. Clemons, P. Park, and D. A. Stevens, "Molecular epidemiology of *Blastomyces dermatitidis*," *Clin. Infect. Dis.*, vol. 30, pp. 328–335, Feb. 2000.

[213] "*Catenaria anguillulae* PPL171." `http://genome.jgi.doe.gov/Catan1/Catan1.home.html`.

[214] J. E. Stajich, S. K. Wilke, D. Ahrén, C. H. Au, B. W. Birren, M. Borodovsky, C. Burns, B. Canbäck, L. A. Casselton, C. K. Cheng, J. Deng, F. S. Dietrich, D. C. Fargo, M. L. Farman, A. C. Gathman, J. Goldberg, R. Guigó, P. J. Hoegger, J. B. Hooker, A. Huggins, T. Y. James, T. Kamada, S. Kilaru, C. Kodira, U. Kües, D. Kupfer, H. S. Kwan, A. Lomsadze, W. Li, W. W. Lilly, L.-J. Ma, A. J.

Mackey, G. Manning, F. Martin, H. Muraguchi, D. O. Natvig, H. Palmerini, M. A. Ramesh, C. J. Rehmeyer, B. A. Roe, N. Shenoy, M. Stanke, V. Ter-Hovhannisyan, A. Tunlid, R. Velagapudi, T. J. Vision, Q. Zeng, M. E. Zolan, and P. J. Pukkila, "Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*)," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, pp. 11889–11894, 29 June 2010.

[215] T. J. Sharpton, J. E. Stajich, S. D. Rounsley, M. J. Gardner, J. R. Wortman, V. S. Jordar, R. Maiti, C. D. Kodira, D. E. Neafsey, Q. Zeng, C.-Y. Hung, C. McMahan, A. Muszewska, M. Grynberg, M. A. Mandel, E. M. Kellner, B. M. Barker, J. N. Galgiani, M. J. Orbach, T. N. Kirkland, G. T. Cole, M. R. Henn, B. W. Birren, and J. W. Taylor, "Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives," *Genome Res.*, vol. 19, pp. 1722–1731, Oct. 2009.

[216] "*Ciona intestinalis*." `http://genome.jgi-psf.org/Cioin2/Cioin2.home.html`.

[217] M. Matsuzaki, O. Misumi, T. Shin-I, S. Maruyama, M. Takahara, S.-Y. Miyagishima, T. Mori, K. Nishida, F. Yagisawa, K. Nishida, Y. Yoshida, Y. Nishimura, S. Nakao, T. Kobayashi, Y. Momoyama, T. Higashiyama, A. Minoda, M. Sano, H. Nomoto, K. Oishi, H. Hayashi, F. Ohta, S. Nishizaka, S. Haga, S. Miura, T. Morishita, Y. Kabeya, K. Terasawa, Y. Suzuki, Y. Ishii, S. Asakawa, H. Takano, N. Ohta, H. Kuroiwa, K. Tanaka, N. Shimizu, S. Sugano, N. Sato, H. Nozaki, N. Ogasawara, Y. Kohara, and T. Kuroiwa, "Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D," *Nature*, vol. 428, pp. 653–657, 8 Apr. 2004.

[218] S. S. Merchant, S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Maréchal-Drouard, W. F. Marshall, L.-H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.-L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa, D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J. V. Moroney, J. Moseley, C. Napoli, A. M. Nedelcu, K. Niyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J.-P. Ral, D. M. Riaño Pachón, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.-J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. Luo, D. Martínez, W. C. A. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I. V. Grigoriev, D. S. Rokhsar, and A. R. Grossman, "The *Chlamydomonas* genome reveals the evolution of key animal and plant functions," *Science*, vol. 318, pp. 245–250, 12 Oct. 2007.

[219] "*Dissoconium aciculare*." `http://genome.jgi.doe.gov/Disac1/Disac1.home.html`.

[220] "*Dendrothele bispora* CBS 962.96." `http://genome.jgi.doe.gov/Denbi1/Denbi1.home.html`.

[221] L. Eichinger, J. A. Pachebat, G. Glöckner, M.-A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabbinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, and A. Kuspa, "The genome of the social amoeba *Dictyostelium discoideum*," *Nature*, vol. 435, pp. 43–57, 5 May 2005.

[222] G. dos Santos, A. J. Schroeder, J. L. Goodman, V. B. Strelets, M. A. Crosby, J. Thurmond, D. B. Emmert, W. M. Gelbart, and FlyBase Consortium, "FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations," *Nucleic Acids Res.*, vol. 43, pp. D690–7, Jan. 2015.

[223] "*Dothidotthia symphoricarpi* CBS 119687." `http://genome.jgi.doe.gov/Dotsy1/Dotsy1.home.html`.

[224] C. Aurrecoechea, J. Brestelli, B. P. Brunk, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. T. Kraemer, W. Li, J. A. Miller, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, G. Srinivasamoorthy, C. J. Stoeckert, Jr, R. Thibodeau, C. Treatman, and H. Wang, "EuPathDB: a portal to eukaryotic pathogen databases," *Nucleic Acids Res.*, vol. 38, pp. D415–9, Jan. 2010.

[225] J. A. Chapman, E. F. Kirkness, O. Simakov, S. E. Hampson, T. Mitros, T. Weinmaier, T. Rattei, P. G. Balasubramanian, J. Borman, D. Busam, K. Disbennett, C. Pfannkoch, N. Sumin, G. G. Sutton, L. D. Viswanathan, B. Walenz, D. M. Goodstein, U. Hellsten, T. Kawashima, S. E. Prochnik, N. H. Putnam, S. Shu, B. Blumberg, C. E. Dana, L. Gee, D. F. Kibler, L. Law, D. Lindgens, D. E. Martinez, J. Peng, P. A. Wigge, B. Bertulat, C. Guder, Y. Nakamura, S. Ozbek, H. Watanabe, K. Khalturin, G. Hemmrich, A. Franke, R. Augustin, S. Fraune, E. Hayakawa, S. Hayakawa, M. Hirose, J. S. Hwang, K. Ikeo, C. Nishimiya-Fujisawa, A. Ogura, T. Takahashi, P. R. H. Steinmetz, X. Zhang, R. Aufschnaiter, M.-K. Eder, A.-K. Gorny, W. Salvenmoser, A. M. Heimberg, B. M. Wheeler, K. J. Peterson, A. Böttger, P. Tischler, A. Wolf, T. Gojobori, K. A. Remington, R. L. Strausberg, J. C. Venter, U. Technau, B. Hobmayer, T. C. G. Bosch, T. W. Holstein, T. Fujisawa, H. R. Bode, C. N. David, D. S. Rokhsar, and R. E. Steele, "The dynamic genome of *Hydra*," *Nature*, vol. 464, pp. 592–596, 25 Mar. 2010.

[226] "*Lichtheimia hyalospora*." http://genome.jgi.doe.gov/Lichy1/Lichy1.home.html.

[227] "*Lophiostoma macrostomum* CBS 122681." http://genome.jgi.doe.gov/Lopma1/Lopma1.home.html.

[228] Q. Gao, K. Jin, S.-H. Ying, Y. Zhang, G. Xiao, Y. Shang, Z. Duan, X. Hu, X.-Q. Xie, G. Zhou, G. Peng, Z. Luo, W. Huang, B. Wang, W. Fang, S. Wang, Y. Zhong, L.-J. Ma, R. J. St Leger, G.-P. Zhao, Y. Pei, M.-G. Feng, Y. Xia, and C. Wang, "Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*," *PLoS Genet.*, vol. 7, p. e1001264, 6 Jan. 2011.

[229] D. A. Martinez, B. G. Oliver, Y. Gräser, J. M. Goldberg, W. Li, N. M. Martinez-Rossi, M. Monod, E. Shelest, R. C. Barton, E. Birch, A. A. Brakhage, Z. Chen, S. J. Gurr, D. Heiman, J. Heitman, I. Kosti, A. Rossi, S. Saif, M. Samalova, C. W. Saunders, T. Shea, R. C. Summerbell, J. Xu, S. Young, Q. Zeng, B. W. Birren, C. A. Cuomo, and T. C. White, "Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection," *MBio*, vol. 3, pp. e00259–12, 4 Sept. 2012.

[230] S. C. Lee, R. B. Billmyre, A. Li, S. Carson, S. M. Sykes, E. Y. Huh, P. Mieczkowski, D. C. Ko, C. A. Cuomo, and J. Heitman, "Analysis of a food-borne fungal pathogen outbreak: virulence and genome of a *Mucor circinelloides* isolate from yogurt," *MBio*, vol. 5, pp. e01390–14, 8 July 2014.

[231] R. S. Cornman, Y. P. Chen, M. C. Schatz, C. Street, Y. Zhao, B. Desany, M. Egholm, S. Hutchison, J. S. Pettis, W. I. Lipkin, and J. D. Evans, "Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees," *PLoS Pathog.*, vol. 5, p. e1000466, June 2009.

[232] C. A. Cuomo, C. A. Desjardins, M. A. Bakowski, J. Goldberg, A. T. Ma, J. J. Becnel, E. S. Didier, L. Fan, D. I. Heiman, J. Z. Levin, S. Young, Q. Zeng, and E. R. Troemel, "Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth," *Genome Res.*, vol. 22, pp. 2478–2488, Dec. 2012.

[233] M. Falb, F. Pfeiffer, P. Palm, K. Rodewald, V. Hickmann, J. Tittor, and D. Oesterhelt, "Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*," *Genome Res.*, vol. 15, pp. 1336–1343, Oct. 2005.

[234] "*Patellaria atrata* CBS 101060." http://genome.jgi.doe.gov/Patat1/Patat1.home.html.

[235] "*Phycomyces blakesleeanus* NRRL1555." http://genome.jgi-psf.org/Phybl2/Phybl2.home.html.

[236] "*Piromyces* sp. E2." http://genome.jgi.doe.gov/PirE2_1/PirE2_1.home.html.

[237] "*Pleomassaria siparia* CBS 279.74." http://genome.jgi.doe.gov/Plesi1/Plesi1.home.html.

[238] B. M. Tyler, S. Tripathy, X. Zhang, P. Dehal, R. H. Y. Jiang, A. Aerts, F. D. Arredondo, L. Baxter, D. Bensasson, J. L. Beynon, J. Chapman, C. M. B. Damasceno, A. E. Dorrance, D. Dou, A. W. Dickerman, I. L. Dubchak, M. Garbelotto, M. Gijzen, S. G. Gordon, F. Govers, N. J. Grunwald, W. Huang, K. L. Ivors, R. W. Jones, S. Kamoun, K. Krampis, K. H. Lamour, M.-K. Lee, W. H. McDonald, M. Medina, H. J. G. Meijer, E. K. Nordberg, D. J. Maclean, M. D. Ospina-Giraldo, P. F. Morris, V. Phuntumart, N. H. Putnam, S. Rash, J. K. C. Rose, Y. Sakihama, A. A. Salamov, A. Savidor, C. F. Scheuring, B. M. Smith, B. W. S. Sobral, A. Terry, T. A. Torto-Alalibo, J. Win, Z. Xu, H. Zhang, I. V. Grigoriev, D. S. Rokhsar, and J. L. Boore, "*Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis," *Science*, vol. 313, pp. 1261–1266, 1 Sept. 2006.

[239] V. A. Manning, I. Pandelova, B. Dhillon, L. J. Wilhelm, S. B. Goodwin, A. M. Berlin, M. Figueroa, M. Freitag, J. K. Hane, B. Henrissat, W. H. Holman, C. D. Kodira, J. Martin, R. P. Oliver, B. Robbertse, W. Schackwitz, D. C. Schwartz, J. W. Spatafora, B. G. Turgeon, C. Yandava, S. Young, S. Zhou, Q. Zeng, I. V. Grigoriev, L.-J. Ma, and L. M. Ciuffetti, "Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence," *G3*, vol. 3, pp. 41–63, Jan. 2013.

[240] C. A. Levesque, H. Brouwer, L. Cano, J. Hamilton, C. Holt, E. Huitema, S. Raffaele, G. Robideau, M. Thines, J. Win, M. Zerillo, G. Beakes, J. Boore, D. Busam, B. Dumas, S. Ferriera, S. Fuerstenberg, C. Gachon, E. Gaulin, F. Govers, L. Grenville-Briggs, N. Horner, J. Hostetler, R. Jiang, J. Johnson, T. Krajaejun, H. Lin, H. Meijer, B. Moore, and P. Morris, "Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire," *Genome Biol.*, vol. 11, no. 7, p. R73, 2010.

[241] E. Tisserant, M. Malbreil, A. Kuo, A. Kohler, A. Symeonidi, R. Balestrini, P. Charron, N. Duensing, N. Frei dit Frey, V. Gianinazzi-Pearson, L. B. Gilbert, Y. Handa, J. R. Herr, M. Hijri, R. Koul, M. Kawaguchi, F. Krajinski, P. J. Lammers, F. G. Masclaux, C. Murat, E. Morin, S. Ndikumana, M. Pagni, D. Petitpierre, N. Requena, P. Rosikiewicz, R. Riley, K. Saito, H. San Clemente, H. Shapiro, D. van Tuinen, G. Bécard, P. Bonfante, U. Paszkowski, Y. Y. Shachar-Hill, G. A. Tuskan, J. P. W. Young, P. W. Young, I. R. Sanders, B. Henrissat, S. A. Rensing, I. V. Grigoriev, N. Corradi, C. Roux, and F. Martin, "Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. 20117–20122, 10 Dec. 2013.

[242] "*Rhizopus microsporus* var. *microsporus* ATCC52813." `http://genome.jgi.doe.gov/Rhimi1_1/Rhimi1_1.home.html`.

[243] L.-J. Ma, A. S. Ibrahim, C. Skory, M. G. Grabherr, G. Burger, M. Butler, M. Elias, A. Idnurm, B. F. Lang, T. Sone, A. Abe, S. E. Calvo, L. M. Corrochano, R. Engels, J. Fu, W. Hansberg, J.-M. Kim, C. D. Kodira, M. J. Koehrsen, B. Liu, D. Miranda-Saavedra, S. O'Leary, L. Ortiz-Castellanos, R. Poulter, J. Rodriguez-Romero, J. Ruiz-Herrera, Y.-Q. Shen, Q. Zeng, J. Galagan, B. W. Birren, C. A. Cuomo, and B. L. Wickes, "Genomic analysis of the basal lineage fungus *Rhizopus*

oryzae reveals a whole-genome duplication," *PLoS Genet.*, vol. 5, p. e1000549, July 2009.

[244] "*Saccharomyces cerevisiae* S288C, v.2011-02-03." `http://yeastgenome.org`.

[245] "*Sporormia fimetaria* CBS 119925." `http://genome.jgi.doe.gov/Spofi1/Spofi1.home.html`.

[246] V. Wood, R. Gwilliam, M.-A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, A. Düsterhöft, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, P. Nurse, and L. Cerrutti, "The genome sequence of *Schizosaccharomyces pombe*," *Nature*, vol. 415, pp. 871–880, 21 Feb. 2002.

[247] "*Sporobolomyces roseus*." `http://genome.jgi-psf.org/Sporo1/Sporo1.home.html`.

[248] J. C. Kissinger, B. Gajria, L. Li, I. T. Paulsen, and D. S. Roos, "ToxoDB: accessing the *Toxoplasma gondii* genome," *Nucleic Acids Res.*, vol. 31, pp. 234–236, 1 Jan. 2003.

[249] E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar, "The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism," *Science*, vol. 306, pp. 79–86, 1 Oct. 2004.

[250] J. Kämper, R. Kahmann, M. Bölker, L.-J. Ma, T. Brefort, B. J. Saville, F. Banuett, J. W. Kronstad, S. E. Gold, O. Müller, M. H. Perlin, H. A. B. Wösten, R. de Vries,

J. Ruiz-Herrera, C. G. Reynaga-Peña, K. Snetselaar, M. McCann, J. Pérez-Martín, M. Feldbrügge, C. W. Basse, G. Steinberg, J. I. Ibeas, W. Holloman, P. Guzman, M. Farman, J. E. Stajich, R. Sentandreu, J. M. González-Prieto, J. C. Kennell, L. Molina, J. Schirawski, A. Mendoza-Mendoza, D. Greilinger, K. Münch, N. Rössel, M. Scherer, M. Vranes, O. Ladendorf, V. Vincon, U. Fuchs, B. Sandrock, S. Meng, E. C. H. Ho, M. J. Cahill, K. J. Boyce, J. Klose, S. J. Klosterman, H. J. Deelstra, L. Ortiz-Castellanos, W. Li, P. Sanchez-Alonso, P. H. Schreier, I. Häuser-Hahn, M. Vaupel, E. Koopmann, G. Friedrich, H. Voss, T. Schlüter, J. Margolis, D. Platt, C. Swimmer, A. Gnirke, F. Chen, V. Vysotskaia, G. Mannhaupt, U. Güldener, M. Münsterkötter, D. Haase, M. Oesterheld, H.-W. Mewes, E. W. Mauceli, D. DeCaprio, C. M. Wade, J. Butler, S. Young, D. B. Jaffe, S. Calvo, C. Nusbaum, J. Galagan, and B. W. Birren, "Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*," *Nature*, vol. 444, pp. 97–101, 2 Nov. 2006.

[251] "*Umbelopsis ramanniana*." `http://genome.jgi-psf.org/Umbra1/Umbra1.home.html`.

[252] M. Padamsee, T. K. A. Kumar, R. Riley, M. Binder, A. Boyd, A. M. Calvo, K. Furukawa, C. Hesse, S. Hohmann, T. Y. James, K. LaButti, A. Lapidus, E. Lindquist, S. Lucas, K. Miller, S. Shantappa, I. V. Grigoriev, D. S. Hibbett, D. J. McLaughlin, J. W. Spatafora, and M. C. Aime, "The genome of the xerotolerant mold *Wallemia sebi* reveals adaptations to osmotic stress and suggests cryptic sexual reproduction," *Fungal Genet. Biol.*, vol. 49, pp. 217–226, Mar. 2012.

[253] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuvéglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J.-M. Beckerich, E. Beyne, C. Bleykasten, A. Boisramé, J. Boyer, L. Cattolico, F. Confanioleri, A. De Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J.-M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G.-F. Richard, M.-L. Straub, A. Suleau, D. Swennen, F. Tekaia, M. Wésolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, and J.-L. Souciet, "Genome evolution in yeasts," *Nature*, vol. 430, pp. 35–44, 1 July 2004.

[254] "*Zopfia rhizophila* CBS 207.26." `http://genome.jgi.doe.gov/Zoprh1/Zoprh1.home.html`.

[255] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, and V. Buss, "The retinal conformation and its environment in rhodopsin in light of a new 2.2Å crystal structure," *J. Mol. Biol.*, vol. 342, no. 2, pp. 571–583, 2004.

[256] M. Murakami and T. Kouyama, "Crystal structure of squid rhodopsin," *Nature*, vol. 453, pp. 363–367, May 2008.

[257] A. Royant, P. Nollert, K. Edman, R. Neutze, E. M. Landau, E. Pebay-Peyroula, and J. Navarro, "X-ray structure of sensory rhodopsin II at 2.1-Å resolution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, pp. 10131–10136, Aug. 2001.

Appendix A

Comparative genomics analysis of flagellar motility

A.1 Introduction

The eukaryotic flagellum is a protenaceous structure which provides motility to certain single-celled organisms [200]. In fungi, only members of the basal lineages (specifically the Cryptomycota, Chytridiomycota, Blastocladiomycota, and Neocallimastigomycota) posess a flagellum and motile zoosporic life stage. Members of the Glomeromycota, Basidiomycota, and Ascomycota do not have such a life stage [11] and are commonly considered "terrestrial fungi". Notable exceptions to this pattern include the Microsporidian lineage, which is considered a member of the basal fungi but whose members are not flagellated, and *Olpidium brassicae*, a flagellated, single-celled organism morphologically similar to other chytrids but phylogenetically similar to terrestrial fungi [201].

The flagellum itself has a well-studied and defined substructure. Nine doublet microtubules encircle a pair of singlet microtubles. A pair of dyenin arms (one "inner"

and one "outer") connect each outer microtubule fibril to its neighbor. Flagellar motion, and thus cellular motility, is imparted by the ATP hydrolysis of these dyenin arms: the force produced by the arms causes the microtubule doublets to slide against one another, resulting in bending of the flagellum. This structural organization is referred to as a "9+2 axoneme" pattern and, though there are departures, is regarded as the common structural organization of the eukaryotic flagellum [200]. Conservation of this structure can be observed in mammalian, protozoan [68], and algal cells [69].

The axoneme is connected to the organism at its base by the basal body, a microtubule-based organelle derived from centrioles. Assembly of the flagellum is accomplishe by intraflagellar transport proteins and originate from the basal body.

A.2 Results and Discussion

A search for *Rozella* homologs of flagellar-associated proteins from the *Naeglaria* genome [202] reveal a pattern of presence/absence of proteins in Rozella which correlates with that found in the Chytridiomycota and Blastocladiomycota. This pattern, in general, differs from the Microsporidia, supporting the placement of Rozella as prior to the Chytridiomycota/Blastocladiomycota and a flagellar loss event after the divergence of the Microsporidia.

The heatmap displayed in Figure A.1 indicates presence or absence of proteins in the *Naeglaria* dataset as identified in the proteomes of other fungi searched. By clustering based on presence/absence patterns, the flagellated vs non-flagellated organisms separate into distinct groups consistent with known phylogeny (ie basal lineages separate from Dikarya and Zygomycetes). It's possible to collect a subset of genes which are

present in basal (flagellated) lineages and absent in terrestrial (non-flagellated) lineages. Presumably, these have more pronounced association with flagellar assembly and function. Furthermore, take the subset of proteins present in all lineages and do pairwise distance alignment to see if, while being from the same family / hits to same query, these genes mutated or transititioned to new structures / functionalities over evolutionary time.

A.3 Methods

The flagellar analysis was carried out using a dataset of 173 flagellar motility proteins obtained from the *Naeglaria grubeii* genome sequence [202]. These proteins were used in a FASTA search (SSEARCH v36.07) using an e-val cutoff of e-20. Heatmap generation was accomplished using *pheatmap*() in R.

Multiple sequence alignment and distance matrix generation was performed on the subset of genes present in all organisms using $tcoffee$ and *phylip*.

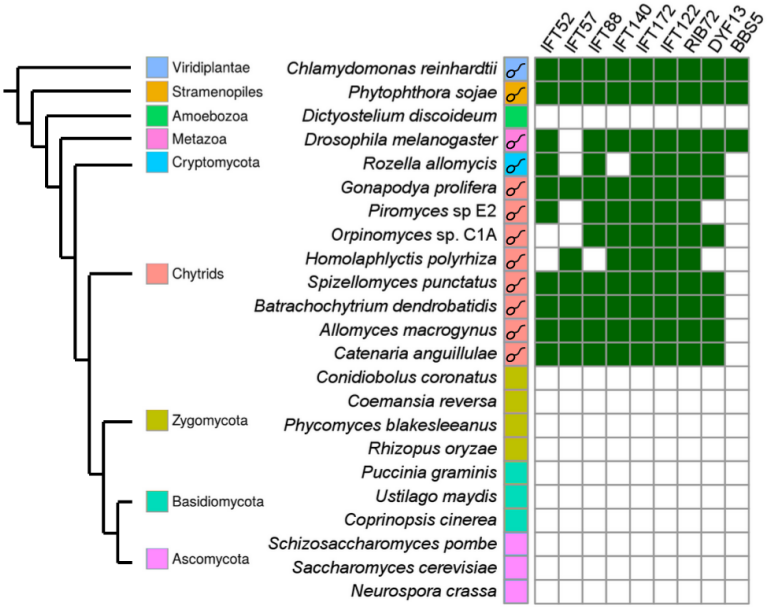


Figure A.1: A subset of proteins which cluster according to presence/absence in flagellated and non-flagellated organisms. Protein copies identified in proteomes of interest are normalized to indicate presence and absence only, where green indicates that one or more copies were found.

Appendix B

Datasets and scripts

Datasets

A list of proteomes from each organisms, the source website, and version used in these analyses is provided in Table B.1.

A table of the PDB IDs used in the structural modeling, docking, and MD simulations presented in Chapter 2 are provided in Table B.2.

Sequences used in *Pichia pastoris* cloning, described in Chapter 3, are given here:

```
>SPPG_00350|Native, putative rhodopsin in S. punctatus SW-1
CTCGCATACGGCACCGCCATTGTCCTCGTCATCATCACCGTAGCCATCATCGCAGGAAAC
CTGTTTATTCTTATCCTGTTTGCAAGATACAGATACCTCCGAAACGCCGCAAATGCCTTG
GTCTTTAGTTTGTCTTTTACAGATTTGTGTATTGGGGTCCAGAGACTTAGTTTGATATCG
ATTCACTTGTCGAAGGGGGGATATGCCACAGGCAAGCTTGGATGTGTTTTGGATGGGAAT
TTTACTGTAATGTTTTCGTCTCAATCCCTCCTAGGTCTCACGGCCCTCGGCATGGAACGG
TACATGGCAGTCGTTAAGCGCCGGCCTTTGACAATGAGACAAGCGATGATGGCCGTGGCC
GCAACATGGATATATAGTGGGGTGCTGTGCGCGATACCGTACATGGTGAAAGGGGATGCG
```

TATCGATTGCAAAGTTCATTTCTGTATTGTACTGGTGACTGGTCCGGTAAAAGGCCAGGG

ACCATCTTTTTCACCGTCATGTGTCTCGCGACCGTCACTATCGCCCTTGGAGCGTTTGGC

GTCGCATACTATTTCATTAACCGAGAAGTCCGCTTGGCCAGCAAGAAATGGGCCCACGTC

CAAGCAGGTACTGGAGAAGAGTCGTCATCAATGTCTAGTCCACCAGTGGTCGATTCCACA

CGACCCAATTCAGCATCCACACGAAAAATGGATGCATTAGAAGCTAAACTCGCTAAGAAG

AGTTTAATCCTCGTACTGGTCTTTGCGTTCAACTGGTTGCTTTACGATATCAATTTCCTT

TGGTCCATGATTGGCAGTGTCCAAGTCCCCATGCGTATTAGATTGTCTAGCCTTCTTGGG

TGCGATAGTAAAGTAAGTTGTTTGCAAACACTTGTTGTACTCACCACCACACATGCTAAT

ACCACCCTCAAAAACAGCTCAGCTTTCAACCCTGTTCTTTTTATTGTCCTCGATAACCGA

TGGGCGCGTTGCGCAAAGGATTTCCTGGGTTGCTGGCCGATGGGCAAAACGGTAGAAAAG

GATGCAACGCCGCAAGAGTCAGTCATGAGATCTGCACCCATTAGTGATCGAAGACAGTCG

TTTGCAGTGAAGGGGAAGATGATGCCTCACCACCACCACCACCACTGA

>SPPG_00350_mut|S. punctatus rhodopsin with K320A mutation

CTCGCATACGGCACCGCCATTGTCCTCGTCATCATCACCGTAGCCATCATCGCAGGAAAC

CTGTTTATTCTTATCCTGTTTGCAAGATACAGATACCTCCGAAACGCCGCAAATGCCTTG

GTCTTTAGTTTGTCTTTTACAGATTTGTGTATTGGGGTCCAGAGACTTAGTTTGATATCG

ATTCACTTGTCGAAGGGGGGATATGCCACAGGCAAGCTTGGATGTGTTTTGGATGGGAAT

TTTACTGTAATGTTTTCGTCTCAATCCCTCCTAGGTCTCACGGCCCTCGGCATGGAACGG

TACATGGCAGTCGTTAAGCGCCGGCCTTTGACAATGAGACAAGCGATGATGGCCGTGGCC

GCAACATGGATATATAGTGGGGTGCTGTGCGCGATACCGTACATGGTGAAAGGGGATGCG

TATCGATTGCAAAGTTCATTTCTGTATTGTACTGGTGACTGGTCCGGTAAAAGGCCAGGG

ACCATCTTTTTCACCGTCATGTGTCTCGCGACCGTCACTATCGCCCTTGGAGCGTTTGGC

GTCGCATACTATTTCATTAACCGAGAAGTCCGCTTGGCCAGCAAGAAATGGGCCCACGTC

CAAGCAGGTACTGGAGAAGAGTCGTCATCAATGTCTAGTCCACCAGTGGTCGATTCCACA

CGACCCAATTCAGCATCCACACGAAAAATGGATGCATTAGAAGCTAAACTCGCTAAGAAG

AGTTTAATCCTCGTACTGGTCTTTGCGTTCAACTGGTTGCTTTACGATATCAATTTCCTT

TGGTCCATGATTGGCAGTGTCCAAGTCCCCATGCGTATTAGATTGTCTAGCCTTCTTGGG

TGCGATAGTAAAGTAAGTTGTTTGCAAACACTTGTTGTACTCACCACCACACATGCTAAT

ACCACCCTCGCAAACAGCTCAGCTTTCAACCCTGTTCTTTTTATTGTCCTCGATAACCGA

TGGGCGCGTTGCGCAAAGGATTTCCTGGGTTGCTGGCCGATGGGCAAAACGGTAGAAAAG

GATGCAACGCCGCAAGAGTCAGTCATGAGATCTGCACCCATTAGTGATCGAAGACAGTCG

TTTGCAGTGAAGGGGAAGATGATGCCTCACCACCACCACCACCACTGA

>BDEG_04847_mod|Putative rhodopsin from B. dendrobatidis JEL423

ATGTCTTCCGACTATCATCCTGAATTTCTCATCACTCTGATGTTCATTTCAGTGTCAGCC

ATCATTGCCAATACCCTAATGGCCATCGCGGTAATGCGCAATCAGGCAATCCGAACAGCA

GAGTATTTTTACAATGTAAATGTTGTTTTAAGCGACTTGGTATTTGCTGTCATGGTGCTG

CTGGTCACAATTATTTCTCTAGCCAATGGTATCCCTGCCGTAACTAATTACATTGGCTGT

CAAGCCGTCGGATACGTGCTACAATCTATGGGAGCCTCGAGTGTTTTGACTATTGCCTTG

ATTACATTAAACCATTATCGGGTCATTGTTCTTGAAAAGCCTAGACTTACATACCATGAT

GTGGCAGTCCATCTTGGGGGGCTTTGGCTAATTTGTCTAACCACTTCCACTATACAGTTT

GTTTTTGGGGAAAAATTCATTCCGCGGCCTGCACATCTCCATTGCCATTACGATTACGCA

TCTAAAGATCTCCTGATTCGGATTCTGACAGGCCAGTTGATTGTGCATCTTGTTGGGTTA

CCCATTGTTGTTTGTCTGGCTTACTGGGCCATCTATAGAAAAGTGGCCGAGGTGGAGCAA

CTCGCTGTAAAAGGGAAATCGCACTATGTGACTACAGAAGTCGCATCAGAGCTTCAATCA

AAAGATCATAAGCAAACCATTGAATTGGCACCCATGCAAAACAGCTATACTATTTCTAGG

GATGTAAAAGGCAGCAACAGGATTGCCGAATCCAACGATGGAACCGCCACATTTAAACTA

ACCGTTCCCGACACCAACTCTGAAAATTTCCCAACGGCCAGCGCCGCGAGTACTAGTGCT

AGCACTCCAATCCGTCCTCATATTCCTAGCTCGCCCGTTGGTCTAAATGAGGCTTTAAGT

AATGTTGCCAAAAGAGGCTTGATGATTTCAATGTCTTTTTCCATTTGTTGGCTCCCTACG

TTGGTTGTCTGCATATTGGAAATCTTCACTGGCCAAACTATTCCCTGGCGGGTTGATGCC

TTGGTCACAATGATTGCAGCATTTAATACACTGAGTAACCCTATTGTGTTCTTTTACGTC

GATCGTCGTCTGTGGAAATCCCTACAGGAAATGGTGAAACCTCTTCATAACTACCACTTG

TAA

>Bden_EcoRI_F

CGAGAATTCCATCCTGAATTTCTCATCACTCTG

>Bden_EcoRI_His_R

CGAGAATTCTCAGTGGTGGTGGTGGTGGTGCATTTCCTGTAGGGATTTCCACAGACG

Scripts

All scripts used for analyses presented in this dissertation are available on github: https://github.com/stajichlab/ahrendt_thesis/tree/master/Scripts

Table B.1: List of all proteomes used in comparative analyses, including NCBI taxonomy database ID, phylogenetic abbreviation ("Key"), phylum designation ("Group"), isolate number/version (where applicable), and source website.

Key	Group	TaxID	Name	Isolate/Version	Reference
Aben	APF	Ascomycota	663331	*Arthroderma benhamiae*	CBS 112371	[203]
Afum	AEF	Ascomycota	746128	*Aspergillus fumigatus*	Af293	[204]
Agos	ASaF	Ascomycota	33169	*Ashbya gossypii*	ATCC 10895 v1.0	[205]
Aloc	MF	Microsporidia	278021	*Antonospora locustae*	HM-2013 v1.0	[206, 207]
Amac	CBF	Chytrids	28583	*Allomyces macrogynus*	ATCC 38327	[37]
Anid	AEF	Ascomycota	227321	*Aspergillus nidulans*	FGSC A4	[208]
Bcin	ALoF	Ascomycota	332648	*Botrytis cinerea*	B05.10	[209]
Bcir	ZMF	Zygomycota	1314798	*Backusella circina*	FSU 941	[210]
Bde2	CCF	Chytrids	684364	*Batrachochytrium dendrobatidis*	JAM 81 v1.0	[36]
Bden	CCF	Chytrids	403673	*Batrachochytrium dendrobatidis*	JEL 423	[211]
Bder	AEF	Ascomycota	559298	*Blastomyces dermatitidis*	SLH14081	[212]
Beme	CBF	Chytrids	4808	*Blastocladiella emersonii*	—	[35]
Cang	CBF	Chytrids	109876	*Catenaria anguillulae*	PL171 v1.0	[213]
Ccin	BAF	Basidiomycota	5346	*Coprinopsis cinerea*	okayama7_130	[214]
Ccor	ZEF	Zygomycota	34488	*Conidiobolus coronatus*	NRRL28638 v1.0	[42]
Cimm	AEF	Ascomycota	246410	*Coccidioides immitis*	RS	[215]
Cint	OA	Animal	7719	*Ciona intestinalis*	v2.0	[216]
Clat	CBF	Chytrids	945690	*Coelomomyces lativittatus*	—	—
Cmer	OAO	OAO	45157	*Cyanidioschyzon merolae*	10D v1.0	[217]
Cneo	BAF	Basidiomycota	5207	*Cryptococcus neoformans*	*var. grubii* H99	[211]
Cowc	OA	Animal	192875	*Capsaspora owczarzaki*	ATCC 30864	[37]
Crei	OP	Viridiplantae	3055	*Chlamydomonas reinhardtii*	v4.0	[218]
Crev	ZF	Zygomycota	61392	*Coemansia reversa*	NRRL 1564 v1.0	[42]
Daci	ADF	Ascomycota	112489	*Dissoconium aciculare*	CBS 342.82 v1.0	[219]
Dbis	BAF	Basidiomycota	1314803	*Dendrothele bispora*	CBS 962.96 v1.0	[220]
Ddis	OAM	Amoebozoa	44689	*Dictyostelium discoideum*	—	[221]
Dmel	OA	Metazoa	7227	*Drosophila melanogaster*	v5	[222]
Dsym	ADF	Ascomycota	548649	*Dothidotthia symphoricarpi*	CBS 119687 v1.0	[223]
Ebie	MF	Microsporidia	31281	*Enterocytozoon bieneusi*	H348	[207]
Ecun	MF	Microsporidia	6035	*Encephalitozoon cuniculi*	GB-M1	[224]
Eint	MF	Microsporidia	58839	*Encephalitozoon intestinalis*	ATCC 50506	[224]
Falb	OAM	Amoebozoa	691883	*Fonticula alba*	ATCC 38817 v2.0	[37]
Fgra	ASoF	Ascomycota	229533	*Fusarium graminearum*	PH-1	[211]
Gpro	CF	Chytrids	1123529	*Gonapodya prolifera*	JEL478 v1.0	[42]
Hmag	OA	Animal	6085	*Hydra magnipapillata*	strain 105	[225]
Hpol	CCF	Chytrids	166479	*Homolaphlyctis polyrhiza*	JEL 142	—
Lhya	ZF	Zygomycota	420593	*Lichtheimia hyalospora*	v1.0	[226]
Lmac	ADF	Ascomycota	372055	*Lophiostoma macrostomum*	CBS 122681 v1.0	[227]
Mani	ASoF	Ascomycota	5530	*Metarhizium anisopliae*	ARSEF 23 v1.0	[228]
Mbre	OA	Animal	81824	*Monosiga brevicollis*	MX1	[37]
Mcan	AEF	Ascomycota	554155	*Microsporum canis*	CBS 113480	[229]
Mcir	ZMF	Zygomycota	36080	*Mucor circinelloides*	f. sp. *circinelloides*	[230]
Mver	ZMF	Zygomycota	78898	*Mortierella verticillata*	NRRL 6337	[37]
Ncer	MF	Microsporidia	40302	*Nosema ceranae*	BRL01 v1.0	[231]
Ncra	ASoF	Ascomycota	5141	*Neurospora crassa*	OR74A	[211]
Npar	MF	Microsporidia	586133	*Nematocida parisii*	ERTm1	[232]
Npha	OAR	Archaea	348780	*Natromonas pharaonis*	DSM 2160	[233]
OrpC	CNF	Chytrids	1117330	*Orpinomyces* sp	C1A	[45]
Patr	ADF	Ascomycota	703506	*Patellaria atrata*	CBS 101060 v1.0	[234]
Pbla	ZMF	Zygomycota	4837	*Phycomyces blakesleeanus*	NRRL1555,v2.0.JGI	[235]
Pgra	BPF	Basidiomycota	418459	*Puccinia graminis f.sp. tritici*	CRL75-36-700-3	[211]
PirE	CNF	Chytrids	45796	*Piromyces* sp	E2 v1.0	[236]
Psip	ADF	Ascomycota	100044	*Pleomassaria siparia*	CBS 279.74	[237]
Psoj	OAO	Stramenopiles	67593	*Phytophthora sojae*	v3.0	[238]
Ptri	ADF	Ascomycota	426418	*Pyrenophora tritici-repentis*	Pt-1C-BFP	[239]
Pult	OAO	OAO	431595	*Pythium ultimum*	DAOM BR144,v.2010-07-12	[240]
Rall	MF	Cryptomycota	281847	*Rozella allomycis*	v.2012-05-03	[47]
Rirr	ZMF	Zygomycota	588596	*Rhizophagus irregularis*	DAOM 181602	[241]
Rmic	ZMF	Zygomycota	58291	*Rhizopus microsporus*	var *microsporus* ATCC52813 v1.0	[242]
Rory	ZMF	Zygomycota	64495	*Rhizopus oryzae*	99-880,v.1	[243]
Sarc	OA	Animal	72019	*Sphaeroforma arctica*	JP610	[37]
Scer	ASaF	Ascomycota	4932	*Saccharomyces cerevisiae*	S288C, v.2011-02-03	[244]
Sfim	ADF	Ascomycota	718229	*Sporormia fimetaria*	CBS 119925 v1.0	[245]
Spom	ATF	Ascomycota	4896	*Schizosaccharomyces pombe*	v.20	[246]
Spun	CCF	Chytrids	109760	*Spizellomyces punctatus*	DAOM BR117	[37]
Sros	BF	Basidiomycota	40563	*Sporobolomyces roseus*	v1.0	[247]
Srot	OA	Animal	946362	*Salpingoeca rosetta*	ATCC 50818	[37]
Sscl	ALoF	Ascomycota	665079	*Sclerotinia sclerotiorum*	1980_UF-70.30,v.2009-01.Broad	[211]

Tgon OAP	Apicomplexian	5811	*Toxoplasma gondii*	ME49,v.2008-07-23.ToxoDB-7.2	[248]
Tpse OAO	OAO	35128	*Thalassiosira pseudonana*	CCMP1335,v.JGI3	[249]
Trub AEF	Ascomycota	559305	*Trichophyton rubrum*	CBS 118892	[229]
Ttra OA	Animal	529818	*Thecamonas trahens*	ATCC 50062	[37]
Umay BUF	Basidiomycota	5270	*Ustilago maydis*	v.2011-05-24.MIPS	[250]
Uram ZMF	Zygomycota	41833	*Umbelopsis ramanniana*	AG v1.0	[251]
Uree AEF	Ascomycota	336963	*Uncinocarpus reesii*	UAMH 1704	[215]
Vcul MF	Microsporidia	103449	*Vavraia culicis* floridensis	v.1.Broad	[224]
Wseb BAF	Basidiomycota	148960	*Wallemia sebi*	CBS-633.66,v.JGI1	[252]
Ylip ASaF	Ascomycota	4952	*Yarrowia lipolytica*	CLIB122,v.14-APR-2010	[253]
Zrhi ADF	Ascomycota	1314779	*Zopfia rhizophilia*	CBS 207.26 v1.0	[254]

Table B.2: PDB IDs and descriptions of structures used in docking and Molecular dynamics simulations

PDBID	Description	Reference
1U19	Crystal structure of Bovine rhodopsin at 2.2 angstroms resolution	[255]
2Z73	Crystal structure of Squid rhodopsin	[256]
1H68	Sensory rhodopsin II	[257]