

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

September 2015

Dissertation Committee:

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

Copyright by  
Steven Robert Ahrendt  
2015

The Dissertation of Steven Robert Ahrendt is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgments

There were many people who contributed immensely to the publication of this dissertation, and to whom I am deeply indebted. First and foremost, my research mentor Dr. Jason Stajich, a mentor in every sense of the word, who allowed me the freedom to pursue my own directions while simultaneously providing the guidance to see this dissertation to completion. My co-advisor, Dr. Angelina Chang, whose input gave a fresh perspective on aspects of this research. My undergraduate advisor Dr. Jamie Foster at the Kennedy Space Center, without whose inspiration I would never have pursued graduate research at all. My good friend Jen Mobbes, who taught me all about being a graduate student. I would like to extend a huge thank you to the members of the Stajich lab, specifically former postdoc John Abramayan, and the ever-capable undergraduate students Na Jeong, Sapphire Ear, Dylan McVay, and Spencer Swanson; also to the members of the Borkovich lab, whose equipment and expertise I borrowed early and often. All of my committee members, Dr. Jason Stajich, Dr. Angelina Chang, and Dr. Katherine Borkovich, who guided me through this entire process. My parents, without whose constant support I would not be here today. And finally Kat, who stuck it out with me until the end, for which I am eternally grateful.

## 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, September 2015  
Dr. Jason E. Stajich , Chairperson

Species belonging to the early diverging fungal lineages (Blastocladiomycota, Chytridiomycota, Cryptomycota, and Neocallimastigomycota) reproduce via motile uniflagellated zoospores. These organisms are traditionally understudied, despite being active decomposers, parasites, and symbionts with other organisms in the ecosystem. This dissertation research uses a comparative approach to answer questions about how these organisms interact with their environment, regarding putative photoreception (Chapters 3 and 4), molecular aspects of the evolutionary transition from aquatic motile single cells to terrestrial multicellular organisms (Appendix A), potential anti-fungal and competitive behavior and its mechanisms (Chapter ??).

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>x</b>
<b>1 Introduction to the Basal Fungal lineages</b>	<b>1</b>
1.1 Overview of fungal phylogenetics and the importance of the basal lineages	1
1.2 Development of bioinformatic resources for basal fungi . . . . .	5
1.3 Goals of this body of research . . . . .	7
1.3.1 Competition-based secondary metabolism . . . . .	8
1.3.2 Entomopathogenic fungi . . . . .	9
1.3.3 Mechanics and evolution of Fungal photobiology . . . . .	10
1.3.4 Eukaryotic Flagellar motility . . . . .	12
<b>2 Growth suppression properties of the Chytridiomycete <i>Homolaphlyctis polyrhiza</i> JEL142</b>	<b>14</b>
2.1 Introduction . . . . .	14
2.1.1 Background of chytrids and <i>Hp</i> . . . . .	15
2.2 Methods . . . . .	17
2.2.1 Fungal strains and maintenance . . . . .	17
2.2.2 Bioactivity on solid media . . . . .	18
2.2.3 Bioactivity in liquid media . . . . .	18
2.2.4 Reassembly and annotation . . . . .	20
2.2.5 Secretome and small-metabolite screen . . . . .	20
2.2.6 Temperature and pH growth assays . . . . .	20
2.3 Results . . . . .	21
2.3.1 The physiology of <i>Hp</i> suggests that it is more tolerant of environmental stresses than <i>Bd</i> . . . . .	21
2.3.2 <i>Hp</i> inhibition is unique among the chytrids and broadly active against other fungal species . . . . .	22
2.3.3 Liquid filtrate screening suggests a non-protein compound is responsible for bioactivity . . . . .	23
2.3.4 Secretome + antiSMASH analyses, coupled with improved genome assembly and annotation predicts a few unique secondary metabolite gene clusters with potential relevance . . . . .	24
2.3.5 Genome reassembly and reannotation . . . . .	25
2.3.6 Miscellaneous . . . . .	25

2.3.6.1	N. crassa hyphal tip morphology . . . . .	25
2.3.6.2	Kinase mutant screen . . . . .	25
2.4	Discussion . . . . .	25
<b>3</b>	<b>Structural characteristics of opsin-like proteins found in basal fungal lineages</b>	<b>31</b>
3.1	Introduction . . . . .	31
3.2	Methods . . . . .	34
3.2.1	Homology modeling and Docking . . . . .	34
3.2.2	RMSD calculation . . . . .	35
3.3	Results . . . . .	36
3.3.1	Homology models are of reasonable quality . . . . .	36
3.3.2	Structural conservation reveals <i>S. punctatus</i> structure to be most likely functional as photoreceptor . . . . .	36
3.3.3	<i>in silico</i> chemical screen . . . . .	39
3.3.4	Molecular Dynamics simulations . . . . .	40
3.4	Conclusions . . . . .	40
<b>4</b>	<b>A comparative analysis of auxillary components of the rhodopsin and opsin-like signalling pathways in fungi</b>	<b>45</b>
4.1	Introduction . . . . .	45
4.2	Methods . . . . .	47
4.2.1	General Photosensory overview . . . . .	47
4.2.2	G-protein Analysis . . . . .	47
4.2.3	PLC . . . . .	47
4.2.4	PDE . . . . .	47
4.2.5	RGS . . . . .	47
4.3	Results . . . . .	47
4.3.1	Photosensory . . . . .	47
4.3.2	G-protein Analysis . . . . .	47
4.3.2.1	G $\alpha$ . . . . .	47
4.3.2.2	G $\beta$ . . . . .	49
4.3.2.3	G $\gamma$ . . . . .	49
4.3.3	Phototaxis and Heterologous Protein expression . . . . .	49
4.4	Conclusions . . . . .	49
<b>5</b>	<b>Transcriptome analysis in <i>Coelomomyces lativittatus</i></b>	<b>53</b>
5.1	Introduction . . . . .	53
5.2	Results . . . . .	56
5.2.1	Transcriptome Characterization . . . . .	56
5.2.2	Insect Virulence . . . . .	56
5.2.2.1	C1 cysteine proteases . . . . .	57
5.2.2.2	Trypsin proteases . . . . .	58
5.2.2.3	Destruxins . . . . .	58
5.2.2.4	Chitin related domains . . . . .	58
5.2.2.5	Adhesion-related proteins . . . . .	59
5.2.2.6	Ecdysone receptors . . . . .	60
5.2.3	$\hat{\text{I}}\ddot{\text{s}}$ -carotene . . . . .	62

5.2.4	Sensing: Photosensing capacity is consistent with ideas about C.lat abilities . . . . .	65
5.2.5	"interesting" animal homologs . . . . .	67
5.3	Discussion . . . . .	67
5.3.1	Sensory results are consistent with previous hypotheses, but undetermined if this represents a specific insect association aspect.	70
5.3.2	$\beta$ -carotene biosynthesis and metabolism pathways are present and nearly complete in the <i>C. lativittatus</i> transcriptome. . . . .	72
<b>6</b>	<b>Conclusions</b>	<b>79</b>
6.1	Inhibition . . . . .	80
6.2	Rhodopsin . . . . .	81
6.3	<i>Coelomomyces</i> . . . . .	81
<b>Bibliography</b>		<b>83</b>
<b>A Comparative genomics analysis of Flagellar motility</b>		<b>89</b>
A.1	Introduction . . . . .	89
A.2	Results and Discussion . . . . .	89
A.3	Methods . . . . .	90

# List of Figures

2.1	Inhibition of <i>N. crassa</i> is observed with <i>H. polychiza</i> and not other chytrids.	27
2.2	Inhibitory properties of <i>H. polychiza</i> are not specific to <i>N. crassa</i> . . . . .	29
2.3	Filter-sterilized <i>Hp</i> -conditioned media effectively inhibits <i>N. crassa</i> growth.	30
2.4	<i>N. crassa</i> displays no avoidance behavior of solid agar blocks. . . . .	30
3.1	Structural details of the <i>S. punctatus</i> chytriopsin homology model. . . . .	43
5.1	Distribution of <i>Aspergillus</i> GO-Slim terms associated with <i>C. lat</i> transcriptome. . . . .	77
5.2	Opsin-GC fusion proteins; Chytridiomycota and Blastocladiomycota cluster	78
5.3	$\beta$ -carotene enzyme presence / absence . . . . .	78
A.1	Heatmap cluster analysis of flagellar proteins from <i>Naegleria gruberi</i> . .	91

# List of Tables

2.1	Kinase and phosphatase mutants used in screening against <i>H. polyrhiza</i> sporangia. . . . .	28
3.1	PROCHECK Ramachandran plot results for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures. . . . .	42
3.2	Pairwise backbone RMSD measurements for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures, calculated using the Bio3D package. . . . .	42
3.3	Conserved Rhodopsin motifs in X-ray crystal structures and chytropsin homology models. Abbreviations: Bt1U19= <i>Bos taurus</i> X-ray crystal structure (PDBID 1U19); Tp2Z73= <i>Todares pacificus</i> X-ray crystal structure (PDBID 2Z73); . . . . .	44
4.1	G-protein subunits identified through sequence similarity . . . . .	50
4.2	G-protein subunit comparison . . . . .	51
4.3	PDE . . . . .	52
5.1	PFAM . . . . .	74
5.2	FAADB . . . . .	75
5.3	OrthMCL . . . . .	76

# **Chapter 1**

## **Introduction to the Basal Fungal lineages**

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

Various studies have attempted to estimate the date of the divergence of fungi from animals [1]. These studies place this evolutionary split approximately 1 billion years ago, with a range of around  $\pm 500$  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 contains seven phyla: Microsporidia, Chytridiomycota, Blastocladiomycota, Neocallimastigomycota, Glomeromycota, Basidiomycota, and Ascomycota, with the most recent inclusion of the phylum Cryptomycota [7].

The most recently diverged 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"). Collectively, it is the most widely studied group and is home to several model and non-model organisms of research interest, including *Neurospora crassa*, *Saccharomyces cerevisiae*, and *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 previously attributed to the phylum Zygomycota [8]: Mucormycotina, Entomophthoromycotina, Zoopagomycotina, and Kickxellomycotina. These subphyla contain primarily terrestrial fungi of various medical and industrial research interest.

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. These organisms exist as either motile, flagellated zoospores, or as sessile, non-flagellated sporangia. The asexual life cycle displayed by these lineages involves oscillation between both stages, wherein the zoospore seeks an appropriate environmental substrate, encysts upon it and retracts its flagellum, develops a cell wall and undergoes several rounds of mitotic cell division, and ultimately produces and releases hundreds of new zoospores [citation]. Zoospore-producing sporangia (zoosporangia) are thin-walled, while resting spores are thick-walled, developing into sporangia only after a period of dormancy [9]. Eucarpic chytrids have a sporangium

and filamentous rhizoids, while holocarpic chytrids produce thalli which are entirely converted to sporangia during reproduction. monocentric = one spore for one sporangium, polycentric = multiple sporangia on a rhizomycelial network. Certain species within the Blastocladiomycota have demonstrated sexual reproduction with zoospores of different mating types [citation needed].

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 (Glomeromycota, Ascomycota, and Basidiomycota). While these lineages only represent less than 2% of all described fungi [10], 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 [?]. Members of these lineages fulfill nearly all varieties of ecological niches, from saprotrophic feeders of forest detritus (*Spizellomyces punctatus* (Chytridiomycota)), to pathogenic associations with insects (*Coelomomyces lativittatus* (Blastocladiomycota)), plants (*Olpidium brassicae* (Chytridiomycota)), vertebrates (*Batrachochytrium dendrobatidis* (Chytridiomycota) [11]), nematodes (*Catenaria anguillulae* (Blastocladiomycota)) and algae (*Zygorhizidium pltonicum* [12]), and even to intracellular symbioses with other chytrids (*Rozella allomyces* (Cryptomycota)). 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 [?]. 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.

Revisions of fungal phylogenetics continued through the 20th century and included Alexopolous's series on Introductory Mycology in the 1950s.

Significant microscopy work was carried out on chytrid species as early as 1953 by Koch and others, which allowed for the discussion of zoosporic ultrastructure characters. In the 1970s, the then recent technology of electron microscopy allowed for finer grain examinations of chytrid ultrastructure.

In 1993, Mims and Blackwell for fourth edition (1996) using DNA-based phylogenetics (specifically SSU rDNA). Importantly, established Fungi as monophyletic, established chytrids as Fungi, and excluded other heterokont flagellates (eg oomycetes).

In 1999, James et al attempted to reconcile the previous ultrastructure-based phylogeny (using characters such as zoospore discharge, thallus development, ultrastructural features) with modern molecular techniques (eg ssu rDNA) used in assessing phylogenetic relationships in other fungal groups.

By 2006, at the time of the establishment of the Deep Hypha coordination network, the Chytridiomycota was accepted as one of the four major fungal phyla (alongside the Zygomycota, Basidiomycota, and Ascomycota; hereafter, the phylogenetic details of these latter three lineages will be discussed in much less detail, if at all). Additionally, the Chytridiomycota was agreed to be the most basal lineage primarily due to the presence of a flagellated life stage. The typical life cycle for these basal lineages begins with motile, posteriorly uniflagellate zoospores. After locating a suitable substrate, these zoospores encyst and develop into sporangia. Within the sporangia, asexual reproduction results in the production of more zoospores. The cycle is renewed when these

zoospores are discharged from the sporangia.

Concerning chytrids as a system for study, interest waned in the latter 20th century, driven in part because few species are of substantial economic importance [?] as well as moderate difficulty in collection and culturing methods. More widespread interest renewed in the early 2000s due to the acceptance of *Batrachochytrium dendrobatidis* as a global pathogen and the causative agent of worldwide amphibian decline (first recognized as concerning in 1989, though data indicates that widespread decline began in 1970 in US, Puerto Rico, and northeastern Australia [?]).

In 1998, *Bd*-caused chytridiomycosis emerged through experimental research as leading front-runner for cause of widespread decline [?], and in 1999, *B. dendrobatidis* isolated from a blue poison dart frog which had died in Washington, DC, was formally described [11].

## 1.2 Development of bioinformatic resources for basal fungi

As genomic resources became more widely available, chytrid genomes became more easily developed. The first available resource was an expressed sequence tag (EST) dataset published in 2005 for the Blastocladiomycete *Blastocladiella emersonii* [13]. 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/annotation/genome/batrachochytrium\\_dendrobatidis/Multilocus](http://www.broadinstitute.org/annotation/genome/batrachochytrium_dendrobatidis/Multilocus)). The resulting assembly is 23.72 Mb, represents 7.4X coverage of this diploid strain, and was the first whole genome assembly for any chytrid. In 2008, a second draft genome

was released for *B. dendrobatidis* strain JAM81 through the Joint Genome Institute (<http://genomeportal.jgi-psf.org/Batde5/Batde5.home.html>). This assembly is 24.3 Mb and represents 8.74X coverage.

As part of a push to understand the Origins of Multicellularity, the genomes for the *Allomyces macrogynus* (Blastocladiomycete) and the exclusively terrestrial *Spizellomyces punctatus* (Chytridiomycete) (24.1 Mb) were sequenced by the Broad Institute in 2009.

In 2011, the non-pathogenic *Homolaphlyctis polyrhiza* JEL142, the closest relative to *Bd* was sequenced for comparison to try to identify aspects of pathogenicity in *Bd* [14]. The resulting assembly size was inferred at 26.7 Mb (haploid) and represented 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 helped assess the assembly and annotation with Dr. Stajich; the results of which are described in Chapter ??.

*Gonapodya prolifera* (Monoblepharidomycota) is an aquatic fungus with both sexual and asexual reproductive schemes, and encompassing both hyphal and zoosporic growth stages. A genome assembly was published in 2011 through JGI. *Catenaria anguillulae* (Blastocladiomycota) is ,

Piromyces and Orpinomyces are members of the Neocallimastigomycota  
Rozella allomycis, the intracellular parasite of Allomyces sp. was sequenced in 2013.

Also 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 ??.

Rapid advances in the feasibility 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: the amphibian pathogen *Batrachochytrium dendrobatidis* and its closest, non-pathogenic relative, *Homolaphlyctis polyrhiza*; the saprobic Blastocladiomycete *Spizellomyces punctatus*, and related *Catenaria anguillulae*; the Monoblepharidomycete *Gonapodya prolifera*; the Neocallimastigomycetes *Piromyces sp.* and *Orpinomyces sp.*; and the aquatic Blastocladiomycete *Allomyces macrogynus* and its intracellular, Microsporidia parasite, *Rozella allomycis*.

### 1.3 Goals of this body of research

Due to the continued development and availability of bioinformatic resources for fungi in general, and the basal lineages specifically, this thesis research focuses on comparative genomics of these systems. It is presented in four chapters comprising three aspects of biology focused on sensing and interpretation of biotic and environmental signals.

### 1.3.1 Competition-based secondary metabolism

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 [15]. 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 ?? using the non-pathogenic Chytridiomycete *Homolaphlyctis polyrhiza* JEL142. In this chapter, I describe the initial observation of *Hp* inhibition of the vegetative hyphal growth of *N. crassa* via an unknown secreted compound. Then, I go on to describe the observational assays using the sporangia of related Chytridiomycetes, and probing the breadth of non-Chytridiomycete fungi whose growth is susceptible to *Hp*, encompassing the 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.

### 1.3.2 Entomopathogenic fungi

Adhesion: Intricate details of fungal adhesion (eg kinetics of macromolecular attachment, nature of reactions, and strucre/chemicsty of fungal and insect surfaces) has gone almost completely unknown for three decades. [?]. Various hypotheses have been suggested and examined, including eletrocstatic forces (see Cherbit+Delmas1979 and Coucias+Latge1986), surface receptors, glycoproteins, and various haemagglutinins (see grula1984 and latge1984b) (also boucias1988), and various carbohydrate compounds (see kerwin+washino1986b).

Germination: Appressoirium development an secretion of adhesive (see zaccharuck1981). spore germination, while recognized as critical, is poorly understood process. Disconnect between in vitro and in vivo tests/observations

- What types of fungi?
- What means of infection/pathogenesis?
- General focus on mosquito control. A subset of fungi are entomopathogenic with respect to mosquitos. Of these, only *Coelomomyces lativittatus* (*Cl*) (Blastocladiomycota) is the only known chytrid. While initially studied as a promising avenue for mosquito control, and an alternative to traditional pesticides, difficulties in culturing *Cl* 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.

### 1.3.3 Mechanics and evolution of Fungal photobiology

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 phototoxic (see [16] and [17]), however the full extend 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 (ie Glomeromycota and Dikarya), with little emphasis placed on the basal lineages [18]. 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, and opsins [18]. 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 [19]. 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 [18]. Cryptochromes, found predominantly in plants, animals, and insects, are blue-light sensitive proteins involved in circadian rhythm regulation and light activated DNA damage repair [?]. Additional evidence suggests cryptochrome proteins play a role in mediating

the phototactic behavior of sponge larvae [20].

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 [21], [22], [23].

The largest family of membrane receptors by far is that of the seven-transmembrane (7TM) receptors, comprising upwards of 800 genes [24]. 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 retinylidene chromophore, typically 11-*cis*-retinal [25].

The retinal chromophore utilized in Type 2 rhodopsin-mediated photoreception is biosynthesized from  $\beta$ -carotene [26]. Photoisomerization of this chromophore results in a conformational shift to the all-*trans* isomer [27] 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 4, 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 3.

Biosynthesis of  $\beta$ -carotene begins with phytoene cyclase converting two molecules of geranylgeranyl pyrophosphate to one molecule of phytoene. Phytoene desaturase then works in a five-step pathway to convert phytoene into lycopene [28]. Lycopene

cyclase finally acts to convert lycopene to  $\beta$ -carotene [29]. Subsequently, two different cleavage enzymes can potentially act on  $\beta$ -carotene. The enzyme  $\beta,\beta$ -carotene 15,15'-monooxygenase 1 (BCMO1) cleaves  $\beta$ -carotene into two all-*trans*-retinal molecules, and is considered a key enzyme for retinoid metabolism [30]. The structurally related enzyme  $\beta,\beta$ -carotene 9',10'-dioxygenase (BCDO2) also acts on  $\beta$ -carotene to produce  $\beta$ -apo-10'-carotenal and  $\beta$ -ionone, however its physiological role is less well-characterized [31]. Comparative analysis and discussion of retinal biosynthesis enzymes is provided in Chapter ??.

#### 1.3.4 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. The chytrid flagellar apparatus is composed of the flagellar stalk (axoneme), the kinetosome (basal body), and the rootlet system [32]. Additionally the characteristic 9+2 arrangement characteristic in other eukaryotic flagellated organisms persists.

During the course of fungal evolution, there was a transition from flagellated motile aquatic single celled organisms to terrestrial multicellular organisms. Prior to the divergence of the Glomeromycota, there was at least one instance of the loss of the flagellum []. Though there exists support for up to X different loss events [], including A [] and B [].

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 []. In the Blastocladiomycota, *Coelomomyces* species are biflagellate

during a part of their life cycle. A more complete description of the life cycle is provided in Chapter ???. Briefly, however, prior to infection of the mosquito, the uniflagellate gametes of opposing mating types fuse to form a biflagellate zygote. []

A short treatment of comparative genomics work on the chytrid flagellar apparatus is given 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.

iiżf

## Chapter 2

# Growth suppression properties of the Chytridiomycete

*Homolaphlyctis polyrhiza* JEL142

### 2.1 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.

Resource competition likely plays a role in the evolution of natural antifungal production [15]. 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.

There are many fungal pathogens of economically important crops (eg. *Fusarium graminearum*, *Magnaporthe grisea*, *Ashbya gossypii*) and humans (eg. *Aspergillus fumigatus*). Research into new natural antifungal treatment is increasingly important as certain pathogenic fungi become progressively resistant to current, conventionally synthesized antifungal drugs (see reviews by Bossche et al. in 1998, Kontoyiannis and Lewis in 2002, and Pfaller in 2012).

There are numerous examples of fungi as potent producers of these types of compounds, including toxins [36], antimicrobials [37], and plant hormone mimics [38]. The thrust of research into fungal-derived secondary metabolites focuses on Ascomycete and Basidiomycete fungi due to their diverse product types [39].

### 2.1.1 Background of chytrids and Hp

However, flagellated fungal lineages, commonly known as "chytrids", are a relatively understudied group, both in general and in this regard specifically. These lineages, closest to the Fungal-Animal divergence and sister to other fungal groups, are known for having a uniflagellate, zoosporic stage in their life cycles and fulfil crucial ecological roles such as terrestrial decomposition. Additionally, several species are pathogenic (eg *Batrachochytrium dendrobatidis* [11], *Coelomomyces* spp. [40], and *Rozella allomycis* [41]), with intracellular and/or extracellular lifestages. Comparative genomic analyses have identified a variety of degradation enzymes in their genomes, but the full extent of competition-based secondary metabolite production 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 polryhiza* (*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 [42, 9, 43] and was provided a formal name in 2011 [44]. A draft 454 genome assembly was produced in that same year for comparative insights into pathogenicity of *Bd* [14].

This isolate was collected in Maine, USA from a 1.7 ha oligotrophic and fishless lake with a pH of 4.6 [45, 46]. 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, but phylogenetic reanalysis prompted the description of a novel genus and species [44].

While working with *Hp* in the Stajich lab for the phototaxis and protein expression experiments described in Chapter 3, I made the discovery of accidental contamination of *Hp* plates by *N. crassa*. Specifically, I first noticed the distinctive zone created by *N. crassa* hyphae growing near *Hp* sporangia (illustrated in Figure ??A). I also had previously observed this type of contamination on plates of related Chytridiomycetes *Bd* and *Sp*, which did not demonstrate this inhibitory effect when co-cultured with *N. crassa*.

This property of *Hp* is reliable and reproducible. This chapter, therefore, describes the experimental and computational work towards identification and characterization of a compound responsible for hyphal growth suppression of *N. crassa* by *Hp*. This represents the first study of *Hp* in this context and presents novel findings about microbial interactions within early branching fungal lineages. Significant help, specifically concerning media prep and general strain maintenance, was provided by undergraduate students in the Stajich lab Na Jeong and Sapphire Ear, and visiting student in the Re-

search Experience for Undergraduates (REU) program Spencer Swansen.

## 2.2 Methods

### 2.2.1 Fungal strains and maintenance

Cultures of *Homolaphlyctis polyrhiza* JEL142, *Operculomyces laminatus* JEL223, *Rhizoclostratium 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). Unless 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<sub>2</sub>O. Sporangia samples for inoculation were obtained by removing a block (approx. 1 cm<sup>2</sup>) of agar containing actively-growing chytrid sporangia.

Vogel's minimal medium (VM) (Davis1970) was used for vegetative growth of *Neurospora crassa* FGSC 2489, *Neurospora discreta* FGSC 8579 and *Neurospora tetrasperma* FGSC 2508. *Neurospora crassa* kinase/phosphotase mutants (Table ??; [47]) 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. *Corpinopsis 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.

### 2.2.2 Bioactivity on solid media

To assess the breadth of fungal species susceptible to *Hp* sporangia, *Hp* was individually co-cultured on mPmTG media with the variety of fungi described above. A block of agar containing active *Hp* sporangia was added to mPmTG plate, flushed with sterile diH<sub>2</sub>O, and incubated for 48h, at which point they were inoculated at a single point with 1x10<sup>6</sup> *Neurospora* conidia in diH<sub>2</sub>O. 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.

### 2.2.3 Bioactivity in liquid media

Figure ?? illustrates the general protocol for obtaining aliquots of filter-sterilized *Hp*-conditioned media ("filtrate") to test for bioactivity. For small quantities, initial preparations (in triplicate) were constructed by adding eight blocks of agar containing actively-

growing *Hp* sporangia, each approximately 1cm<sup>2</sup> to 10 ml of liquid mPmTG. To initially help determine if the compound was being constitutively-produced or a response mechanism, a second preparation was made identically to the one just described, to which an additional 0.5  $\mu$ l of a 100  $\mu$ l *N. crassa* conidial suspension was added. To help control for potential nutrient depletion by *N. crassa* in this second preparation, a third preparation contained eight mPmTG agar blocks with no *Hp* sporangial growth in 10 ml mPmTG, and 0.5  $\mu$ l of a 100  $\mu$ l *N. crassa* conidial suspension. These 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  $\mu$ m syringe filter into a 50 ml conical tube. All replicates for a given treatment were pooled after filtration.

Scaled up experiments were performed in a similar manner: twenty blocks of agar containing *Hp* sporangia were added to 50 ml of mPmTG media, and filtered using a 0.22  $\mu$ m syringe filter.

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<sup>6</sup> 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<sup>6</sup> 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 block 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.

#### **2.2.4 Reassembly and annotation**

*Hp* sporangia material was collected by scraping plates with sterile spatulas. Material was ground using bead-beating methods, and DNA obtained using extraction methods. 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.

#### **2.2.5 Secretome and small-metabolite screen**

A comparative analysis of putative small metabolite and secretome composition in chytrids was performed using antiSMASH [48], and a predictive workflow optimized for fungi described by [49].

Scripts used in these analyses are located at <https://github.com/sahrendt0/Scripts/tree/master>

#### **2.2.6 Temperature and pH growth assays**

To compare to the assays described in [50], *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 different pH levels: 4.0, 5.0, 6.0, 6.8, 8.0, and 9.0.

## 2.3 Results

### 2.3.1 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 get a sense of 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<sub>450</sub>, a measure of cell density, of *Hp* at low pH (4 and 5) 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 ??A). Similarly, *Hp* and *Bd* were grown at temperatures ranging from 18°C to 30°C, and the OD<sub>450</sub> assessed

during a 21-day period (Figure ??B). Taken together, these results suggest that *Hp* has a higher tolerance for environmental changes than *Bd*. This finding is not terribly surprising given that this isolate was obtained from an acidic lake [44].

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

Our initial observations were repeatable, and under controlled circumstances, we were able to demonstrate that *Hp* sporangia are reliably and repeatedly capable of inhibition of *N. crassa* vegetative growth on solid mPmTG media (Figure ??A). The related species, *Bd* and *Sp*, do not demonstrate any inhibitory activity (Figure ??B and C). Additional Chytridiomycota species (*Operculomyces laminatus* JEL 223, *Rhizoclosterium hyalinus* JEL 800, and *Obelidium mucronatum* JEL 802) do not inhibit hyphal growth of *N. crassa* (Figure ??D-F), providing additional support that this behavior 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 2.4).

To determine if the sensitivity to *Hp* sporangia was unique to *N. crassa*, we screened a panel of fungi from among the Ascomycota, Basidiomycota, and Zygomycota. Figure 2.2A recapitulates our previous observations about *N. crassa*, while Figure 2.2B 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, *Trichoderma reesei* (Sordariomycetes; Hypocreales; Hypocreaceae), *Aspergillus nidulans* (Eurotiomycetes; Euro-

tiales; Trichocomaceae), and *Ashbya gossypii* (Saccharomycetes; Saccharomycetales; Saccharomycetaceae) are all completely sensitive to *Hp* (Figure 2.2D-F, respectively).

Within the Basidiomycota, *Coprinopsis cinerea* (Agaricomycetes; Agaricales; Psathyrellaceae) is completely sensitive to *Hp* sporangia (Figure 2.2G). Growth of members of the order Mucorales had mixed sensitivities. *Phycomyces blakesleeanus* (Mucorales; Phycomycetaceae) was completely sensitive to *Hp* sporangia, yet *Rhizopus oryzae* (Mucorales; Mucoraceae) appeared to be insensitive (Figure 2.2H & I, respectively). Despite a possibly limited 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 (*Saccharomyces cerevisiae* MAU99 and AH109) and *E. coli* DH5 $\alpha$  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 ??).

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

We were interested in obtaining a sample of bioactive liquid, on which we 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  $\ddot{\text{A}}$ IJconditioned media $\ddot{\text{A}}$ I 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 ??A). Additionally, conditioned media derived from "Hp alone" and "Hp+Nc" preparations (Prep 1 and 2, Figure S3) were indistinguishable from one an-

other in their effects on *N. crassa* growth.

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 2.3D-E), but was ineffective after treatment of -80°C and 90°C (Figure 2.3F-G). *N. crassa* has no problem growing in fresh mPmTG media (Figure 2.3H). Additionally, the conditioned media was ineffective if left sterile at room temperature for a period of 96h, 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 (not shown or pull up figure / possibly take a week to repeat this).

### **2.3.4 Secretome + antiSMASH analyses, coupled with improved genome assembly and annotation predicts a few unique secondary metabolite gene clusters with potential relevance**

Secretome prediction made using workflow from [49] (Figure 8).

Secondary metabolite clusters were predicted using antiSMASH [48] (Figure 9).

Several unique proteins were found to be related to terpene synthases, and these were unique to Hp when compared to other Chytridiomycota and Blastocladiomycota species.

Make a tree from these enzymes.

### **2.3.5 Genome reassembly and reannotation**

### **2.3.6 Miscellaneous**

#### **2.3.6.1 N. crassa hyphal tip morphology**

<Figure 3>

#### **2.3.6.2 Kinase mutant screen**

In order to address potential mechanisms of action for this behavior, we randomly assayed 2 *N. crassa* phosphatase and 17 kinase mutants from [47]. All were found to be inhibited by growing sporangia of *H. polyrhiza* (Table 2.1).

## **2.4 Discussion**

This study was an attempt to characterize a startling and unexpected observation: actively growing *H. polyrhiza* sporangia inhibit the vegetative, hyphal growth of *N. crassa*. The findings demonstrate that i) *H. polyrhiza* 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.

*H. polyrhiza*, 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 the broad fungal species screen support this presumption, and it appears that *H. polychiza* is active against a multitude of fungal species.

Compared with other chytrid isolates we 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 assay 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 *H. polychiza* 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. Filter-sterilized conditioned media was subjected to temperatures of -80°C, -20°C, 4°C, 23°C, 28°C, 65°C, and 90°C, which are temperatures which would be expected during routine storage and/or shipping processes. With the exception of -80°C and 90°C, treatment at these temperatures did not negatively impact the growth suppression properties of the filtered conditioned media.

We performed both in silico secretome and secondary metabolite screens using the previously published 454-based *H. polychiza* genome. The secondary metabolite screen in particular provided a rough idea of candidate enzymes which were expanded

in *H. polyrhiza*.

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.

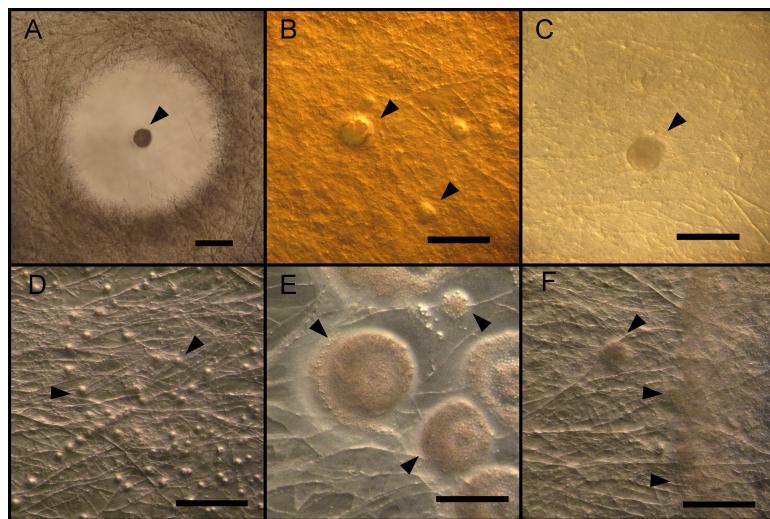


Figure 2.1: *Batrachochytrium dendrobatidis* JEL 423 (*Bd*), the closest relative of *Hp*, is well known as a global emerging pathogen of amphibians and is the causative agent for recent worldwide amphibian decline. *Spizellomyces punctatus* SW-1 (*Sp*) is a soil saprobe and is not known to be pathogenic. 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.

Mutant	Group	Family	NCU	Neurospora_gene
x1	pp2A	-	-	-
x2	ppt-1	-	-	-
x3	CMGC	MAPK	NCU07024	os-2
x4	STE	STE11	NCU03071	os-4
x5	STE	STE7	NCU00587	os-5
1846	CAMK	CAMKL	NCU00914	stk-16
1849	Other	PEK	NCU01187	cpc-3
1885	STE	STE20	NCU03894	stk-4
1888	Other	WEE	NCU04326	stk-29
1916	Unclassified	-	NCU06421	stk-41
1917	Unclassified	-	NCU06422	stk-42
1919	Unclassified	-	NCU06583	stk-44
1920	Other	VPS15	NCU06626	stk-45
1925	AGC	YANK	NCU07062	stk-49
1936	CGMC	CLK/SRPK	NCU10004	stk-56
1937	Unclassified	-	NCU05638	stk-34
1944	CMGC	CDK	NCU07880	prk-6
1946	Other	IKS	NCU08177	stk-51
2294	CAMK	CAMKL	NCU04747	stk-31

Table 2.1: Kinase and phosphatase mutants used in screening against *H. polyrhiza* sporangia.

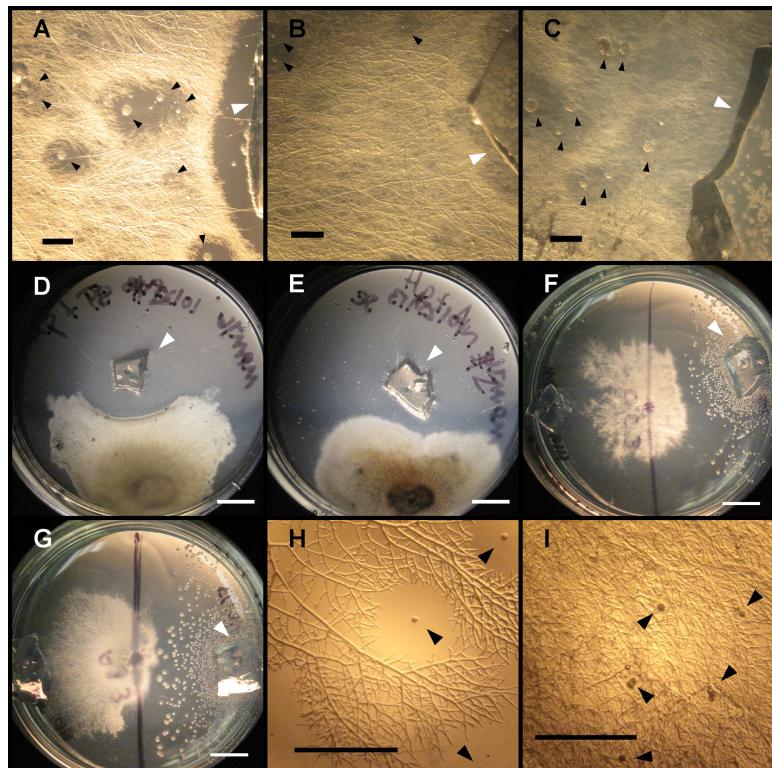


Figure 2.2: A selection of fungal species from the Ascomycota, Basidiomycota, and Mucorales were assayed for sensitivity against *Hp* sporangia. A) *Neurospora crassa* (Sordariomycetes; Sordariales; Sordariaceae), B) *Neurospora tetrasperma*, C) *Neurospora discreta*, D) *Trichoderma reesei* (Sordariomycetes; 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.

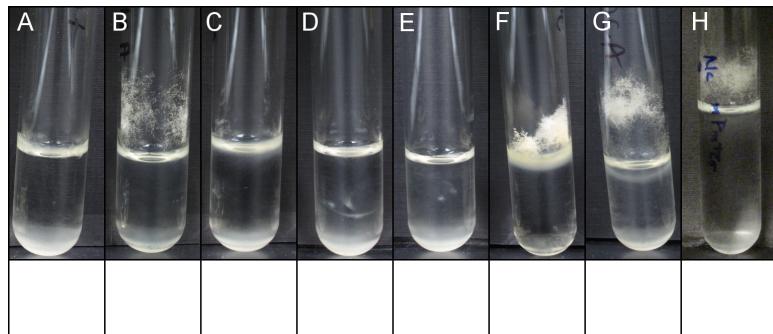


Figure 2.3: Test tubes containing filter-sterilized *Hp*-conditioned media ("filtrate") were inoculated with *N. crassa* conidia and left to incubate at room temperature for 96h (see methods for more detail). For A), B), and C), filtrate was derived from initial preparations of *Hp* alone, *N. crassa* 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.

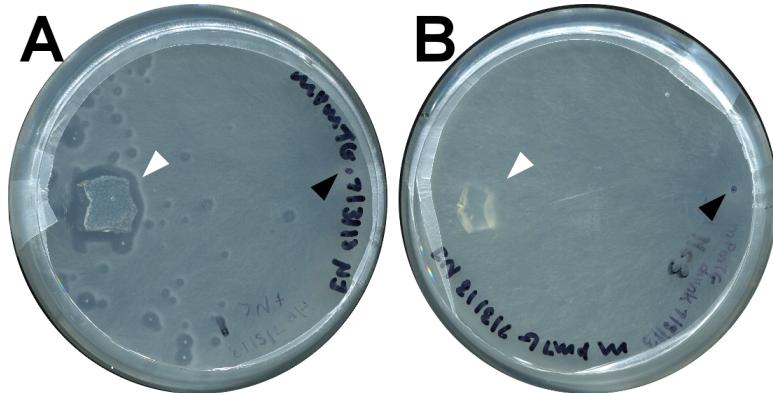


Figure 2.4: Solid agar plates of mPmTG media containing A) actively growing *H. polryrhiza* and B) no *H. polryrhiza* growth. After 72 hours post-inoculation of *N. crassa*, zones of inhibition are clearly visible surrounding *H. polryrhiza* sporangia. White arrowheads in A and B indicate agar blocks containing or lacking, respectively, *H. polryrhiza* sporangia. Black arrowheads indicate point of inoculation with *N. crassa* conidia. Plates are 100mm in diameter.

## **Chapter 3**

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

### **3.1 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 (citation needed). 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 [18]. Identified types of fungal photosensory proteins include phytochrome, cryptochrome, white-collar, and opsin [18] (Figure ??). 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 [19]. 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 [18].

The white-collar complex, on the other hand, is very well characterized in Fungi. First studied in the model filamentous Ascomycete fungus *Neurospora crassa*, this complex 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 [21, 22, 23].

Cryptochromes, found in plants, animals, and insects, are blue-light sensitive proteins involved in circadian rhythm regulation and light activated DNA damage repair [18].

Opsins are a broad class of photosensitive seven-transmembrane proteins that respond to light through photoisomerization of a retinaldehyde chromophore, typically 11-*cis*-retinal. Despite similar overall structure and mechanism of activation, this group can be classified into two types based on sequence similarity and function [51].

Rhodopsin is formed by the combination of a retinaldehyde chromophore and an opsin apoprotein. This construct is generated when the chromophore is joined to opsin via a covalent Schiff-base to a conserved lysine residue. While retinal (A1) is the most common chromophore in vertebrates and invertebrates, others are observed elsewhere in nature. For example, 3,4-dehydroretinal (A2) is observed in fish, amphibia, and reptiles. Switching between the A1 and A2 can be employed as a light adaptation strategy in certain freshwater fish [52]. 3-hydroxyretinal (A3) is found in insects, while 4-hydroxyretinal (A4) is observed in the firefly squid. The A1 retinal can adopt different isomer conformations, including all-*trans*, 13-*cis*, 11-*cis*, or 9-*cis* [52]. Evidence suggests that the 11-*cis*-retinal isomer has been selected for as the optimal chromophore due to the energetic stability of the resulting chromophore-opsin construct [53].

Type 1 rhodopsins are activated by the photoisomerization of all-*trans*-retinal to 13-*cis*-retinal, and function predominantly as ion transporters. Type 2 rhodopsins are G-protein coupled receptors (GPCRs), respond to 11-*cis*-retinal to all-*trans*-retinal photoisomerization, and function as visual receptors. These latter types have previously only been found in animals [51]. The exact nature of the evolutionary relationship between the two types has not been clearly established and is currently the subject of discussion [54, 52].

Previous work has demonstrated that certain species of early-diverging fungi are phototoxic. The phototactic capabilities of the marine fungus *Rhizophyllum litorium* were quantified [17]. This fungus demonstrated the most rapid response to light at a wavelength of 400 nm. Regardless of intensity, no response was seen at wavelengths greater than 600 nm. While the evidence strongly suggests a blue-light sensitive photoreceptor, the researchers were unable to specifically characterize the pigment.

An independent study looked at the phototactic capabilities of the Blastocla-

diomycete *Allomyces reticulatus* [16]. This fungus has a visible, red-pigmented eyespot which screens light for the membrane-bound photosensitive protein. Careful analysis determined that action spectrum of the phototactic zoospores peaks at  $536 \pm 4$  nm, similar to that of the human green-sensitive cone, strongly suggesting a Type 2 rhodopsin.

The increasing availability of genomes from the traditionally understudied early-diverging fungal lineages coupled with a fairly well understood and important environmental sensing system results in an opportunity to expand on known information about the photosensory response in fungi.

In this chapter, I describe a computational and exploratory approach toward understanding the complete view of rhodopsin-specific photoreception in early-diverging fungi, beginning with the creation of the sensory protein and concluding with the cellular response.

## 3.2 Methods

### 3.2.1 Homology modeling and Docking

Based on the T-Coffee multiple sequence alignment, 3D models of the *B. dendrobatidis*, *S. punctatus*, and *A. macrogynus* proteins were built using Modeller [55] with explicit loop refinement and refined with OPUSRota [56]. Models were scored using PROCHECK [57, ?] and Verify3D [?].

The melatonin model was constructed using the human melatonin sequence (UniProt ID: P48039) and subjected to homology modeling with Modeller (v9.9) [doi:10.1002/0471140866] using the *T. pacificus* rhodopsin crystal structure (2Z73) as a template. The Modeller-generated homology model was of better stereochemical quality than the iTasser gener-

ated Melatonin model using the GPCR database. As such, the highest quality Modeller-generated model was selected for further side chain refinements with OPUSRota [doi:10.1110/ps.035022.1]. Similar to the Bd model generation.

Automated protein-ligand docking was accomplished using Autodock 4 [58] 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 [59] 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).

Structure figures were created using the PyMOL Molecular Graphics System, Version 1.7.4 [?]. Membrane topology figure panels were drawn using the TExtopo package [60].

### 3.2.2 RMSD calculation

The loop regions in all models were removed such that the models contained only the seven transmembrane helix regions (scripts on github). The helix-only structures were then aligned using the STAMP Structural alignment method [?] of VMD (v 1.9.1) [61], and the RMSD values of the backbones of the aligned helix-only structures were computed using the *rmsd()* function in the Bio3D R package [62].

### 3.3 Results

#### 3.3.1 Homology models are of reasonable quality

Ramachandran plots were generated for all structure models using PROCHECK [57, ?]. These plots graphically display the backbone dihedral angles  $\psi$  and  $\phi$  of each amino acid residue in a protein and are indicative of model quality, and are summarized in Table ???. For *B. dendrobatis*, *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 [?], are provided in Table ???. The *B. dendrobatis* 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.

#### 3.3.2 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 configurations. A number of structural features provide support for the relationship between chytropsins and other members of the opsin family. Based on sequence similarity, the chytropsin sequences are expected to have seven transmembrane architecture. Table 3.2 displays the pairwise backbone RMSD calculations which display high degree of agree-

ment with rhodopsin crystal structures. Coupled with the overall structure alignment presented in Figure ??, this expected architecture is confirmed.

These values are from a Ramachandran plot [63] generated using PROCHECK [57], 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 phi and psi 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  $\text{â}^{\circ}\text{IJ}$ most favored $\text{â}^{\circ}\text{I}$  regions. Since our models have reasonably high percentages in the  $\text{â}^{\circ}\text{IJ}$ most favored $\text{â}^{\circ}\text{I}$  regions, and reasonably low percentages in the  $\text{â}^{\circ}\text{IJ}$ disallowed $\text{â}^{\circ}\text{I}$  regions, this table suggests that our chytriopsin and melatonin homology models are of reasonable quality.[Lower resolution of the structure for cow vs squid]

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, and that melatonin receptors are less similar to animal rhodopsins than are our chytriopsins.

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-stabilization 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 [64, 27]. 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 possesses both the conserved lysine (K320) and a suitable counterion (D94) in positions favorable for proper function. The structures of *B. dendrobatis* and *A. macrogynus*, on the other hand, lack the conserved lysine and counterion residues in analogous positions. Binding pocket residues, and lysine, counterion, and H-bond network residues are compared in Figure ??.

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 [27]. 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. den-*

*dendrobatis* 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 ?? and ??.

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 [27]. The residues that correspond to the disulfide bond are conserved in the three chytrid structures: Cys80-Cys155 in *B. dendrobatis*, Cys91-Cys166 in *S. punctatus*, and Cys131-Cys220 in *A. macrogynus*. The salt bridge residues are relatively conserved in *B. dendrobatis*, 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 ??.

### 3.3.3 *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. 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 (Figure ??).

### 3.3.4 Molecular Dynamics simulations

## 3.4 Conclusions

The opsin class of visual receptors can be divided into two subtypes based on function. While both types have similar tertiary structure (eg seven transmembrane helices), the Type 2 rhodopsins have thus far only been identified in metazoan lineages. 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.

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  $\beta$ -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. dendrobatis* 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 [53], is not necessary for activation of the light-driven cascades in bacteriorhodopsin [65] and rhodopsin [66]. In the case of bacteriorhodopsin specifically, a K216A mutant was generated and homologously expressed in *H. salinarium*

L33 and provided with retinylidene-n-alkylamines to achieve a Schiff-base construct without the covalent linkage. As this mutation mirrors the *B. dendrobatis* protein (alanine present at the critical position), this suggests that perhaps a non-traditional chromophore is required for rhodopsin function in *B. dendrobatis*.

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 [?] compared to incorrectly modeled structures. As indicated in Table 4, the models for *B. dendrobatis* 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 [51], this finding supports the hypothesis that the chytrid sequences are Type 2 and not Type 1 rhodopsins.

	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 3.1: PROCHECK Ramachandran plot results for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures.

	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 3.2: Pairwise backbone RMSD measurements for chytriopsin and melatonin homology models, and animal rhodopsin crystal structures, calculated using the Bio3D package.

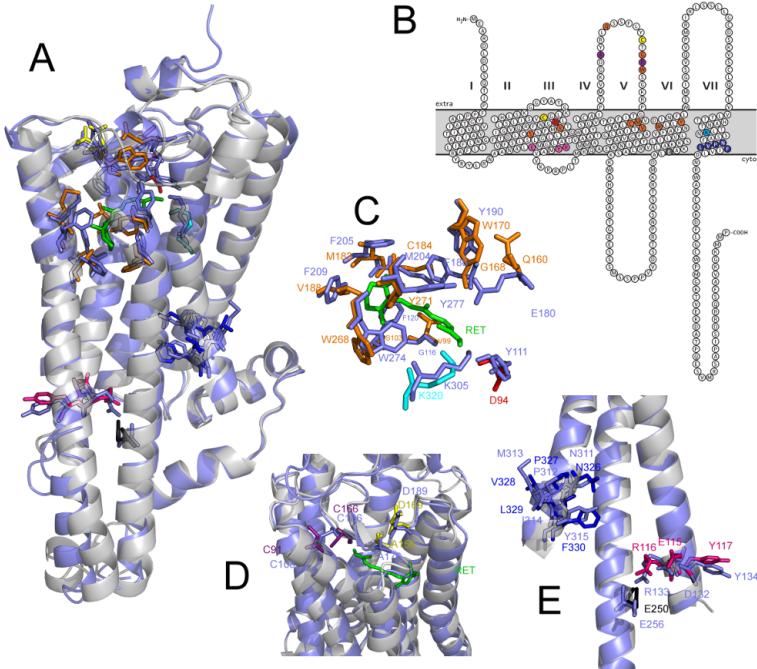


Figure 3.1: Key *S. punctatus* residues are colored according to function: orange ( $\alpha$ AIJbinding pocket $\alpha$ I residues), red (putative counterion), purple (disulfide bond), yellow (salt bridge), dark blue (NPxxY motif), and pink & black (ion lock). Light purple functional and backbone residues 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)  $\text{Ca}$  backbone structural alignment of *S. punctatus* homology model and *T. pacificus* x-ray crystal structure, displayed in a cartoon representation, with key residues displayed as sticks. 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*. The lysine residue from *S. punctatus* (cyan) is present in an optimal position to interact with the 11-cis-retinal ligand (green), an aspect unique among the chytriodopsins studied. The putative *S. punctatus* counterion D94 (red) is also situated in an ideal position for photoisomerization. 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*. In *T. pacificus*, residues R133 and E256 (light purple) form a salt bridge in the inactive state to hold the structure in place, which is broken upon receptor activation. The corresponding *S. punctatus* residues R116 (pink) and E250 (black) are in equivalent positions and presumably carry out the same function.

	Description	Bt1U19	Tp2Z73	Bd	Sp	Am	Cl1540
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	—	E122	F120	A92	S103	V143	—
5	Binding pocket	W265	W274	W317	W268	W468	—
6	H-bond core/Photoisomerase counterion	E181	E180	R149	Q160	—	—
7	—	I189	F188	Y157	G168	—	—
8	Binding pocket/H-bond network w/Y268	Y191	Y190	Y159	W170	—	—
9	—	M207	M204	L173	M183	L237	—
10	—	F208	F205	I174	C184	A238	—
11	—	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	L336
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 3.3: Conserved Rhodopsin motifs in X-ray crystal structures and chytropsin homology models. Abbreviations: Bt1U19=*Bos taurus* X-ray crystal structure (PDBID 1U19); Tp2Z73=*Todares pacificus* X-ray crystal structure (PDBID 2Z73);

## **Chapter 4**

# **A comparative analysis of auxillary components of the rhodopsin and opsin-like signalling pathways in fungi**

### **4.1 Introduction**

Opsins are a broad class of photosensitive seven-transmembrane proteins that respond to light through photoisomerization of a retinaldehyde chromophore, typically 11-*cis*-retinal. Despite similar overall structure and mechanism of activation, this group can be classified into two types based on sequence similarity and function [51].

Four major G $\alpha$  protein subfamilies have been identified: G<sub>s</sub>, G<sub>i</sub>, G<sub>q</sub>, and G<sub>12</sub> [?]. Additional fungal [?] and plant [?] subfamilies exist as well. The type of G-protein to which the receptor is coupled can help classify the Type 2 rhodopsins. The G<sub>s</sub> 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  $\text{Na}^+$  and  $\text{Ca}^{2+}$  channels [?]. 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 (Borkovich MycotaIII).  $G_o$  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 (Borkovich MycotaIII). 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- $\beta$ 1. (Borkovich MycotaIII). Finally, the  $G_{12}$  group contains the  $G_{12}$  and  $G_{13}$  subunits. The function of these subunits is currently unknown, however mutations of a gene with high sequence similarity in drosophila causes a maternal effect (Borkovich MycotaIII).

Motifs that describe the fungal group come from seven identified fungal  $G\alpha$  proteins with no similarity to  $G\alpha$  proteins of the previously identified groups. In *S. cerevisiae*, the GP1 protein functions as a mating factor, while the GP2 protein functions in intracellular cAMP regulation. Members of the plant group may play a role in signal transduction from hormone receptors.

Additionally, at least four  $\beta$ -subunits and six  $\gamma$ -subunits have been identified [?] in Fungi, and it has been demonstrated previously that the  $G\beta/\gamma$  subunit is responsible for signal transduction in yeast [?].

## 4.2 Methods

### 4.2.1 General Photosensory overview

### 4.2.2 G-protein Analysis

### 4.2.3 PLC

### 4.2.4 PDE

### 4.2.5 RGS

## 4.3 Results

### 4.3.1 Photosensory

### 4.3.2 G-protein Analysis

#### 4.3.2.1 $G\alpha$

Multiple  $G\alpha$  proteins were identified at an e-val  $< 1e^{-20}$  in all surveyed chytrid genomes. Counts of predicted proteins are presented in Table 4.1.

One of the *S. punctatus*  $G\alpha$  proteins, SPPG05404, 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  $G\alpha$  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, how-

ever, also possesses 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*, BDEG06989 and BDEG06990, have high similarity to GNA-1 at 66.3% and 73.7% identity, respectively. However, only BDEG06989 contains an N-terminal myristoylation motif. Neither of them contain the C-terminal pertussis motif.

SPPG05884 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).

*G. prolifera* possesses 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 G $\alpha$  proteins to identified *N. crassa* G $\alpha$  proteins are described in Table ??.

A multiple sequence alignment, highlighting the pertussis and myristoylation motifs shared among all fungal G $\alpha$  proteins containing such motifs is presented in Figures ?? and ??.

A phylogenetic analysis was performed on all identified fungal G $\alpha$  proteins using RaxML. The majority of identified but uncharacterized chytrid G $\alpha$  proteins were most closely related to the Group IV family. This group remains largely uncharacterized, though the *Ustilago maydis* homolog is induced during pathogenic development [?]. All other characterized G $\alpha$  groups (I, II, and III) contain chytrid members, unsurprisingly

suggesting an ancient origin for these families. Of note, certain duplication events seem to have occurred after species divergence, and the fungal and mammalian G $\alpha$  proteins represent distinct groups.

#### 4.3.2.2 G $\beta$

All surveyed chytrids were predicted to possess one or more G $\beta$  proteins. Counts and percent identity with *N. crassa* GNB-1 are presented in Tables ?? and ??.

#### 4.3.2.3 G $\gamma$

*A. macrogynus*, *B. dendrobatis*, *H. polyrhiza*, and *Piromyces* were each predicted to possess a single G $\gamma$  protein (e-val < 1e $-20$ ). Of these identified G $\gamma$  proteins, only that from *B. dendrobatis* contained a pertussis toxin sensitivity motif, with 51.4% identity to NCU00041 (*N. crassa* GNG-1). Counts and percent identity with *N. crassa* GNG-1 are presented in Tables ?? and ??.

### 4.3.3 Phototaxis and Heterologous Protein expression

In order to determine the phototactic abilities of *B. dendrobatis* and *S. punctatus*, I followed the procedure outlined in [?] for *Rhizophyllum littoreum*. Our setup consisted of a slide projector for a light source, and a mirror to redirect the light source.

## 4.4 Conclusions

Species	G $\alpha$	G $\beta$	G $\gamma$
<i>Amac</i>	19	4	1
<i>Cang</i>	13	1	1
<i>Clat</i>	-	-	-
<i>Bden</i>	10	2	1
<i>Hpol</i>	6	1	1
<i>Spun</i>	5	3	0
<i>Ncra</i>	3	1	1

Table 4.1: G-protein subunits identified through sequence similarity

Species	Class	Name	Broad	NCBI	Identity <sup>a</sup>
<i>Neurospora crassa</i>	I	GNA-1	NCU06493 <sup>pm</sup>	-	100
-	II	GNA-2	NCU06729	-	100
-	III	GNA-3	NCU05206 <sup>m</sup>	-	100
<i>Allomyces macrogynus</i>	I	-	AMAG_03583 <sup>pm</sup>	-	71
-	I	-	AMAG_04903 <sup>pm</sup>	-	69
-	I	-	AMAG_15273	-	88
-	I	-	AMAG_04635 <sup>m</sup>	-	68
-	I	-	AMAG_17306 <sup>m</sup>	-	65
-	I	-	AMAG_13117 <sup>m</sup>	-	65
-	I	-	AMAG_15402 <sup>m</sup>	-	61
-	I	-	AMAG_06540 <sup>m</sup>	-	61
-	I	-	AMAG_16300 <sup>m</sup>	-	57
-	I	-	AMAG_08894 <sup>m</sup>	-	57
-	I	-	AMAG_17089	-	57
-	III	-	AMAG_03685 <sup>m</sup>	-	64
-	III	-	AMAG_09154 <sup>m</sup>	-	64
-	III	-	AMAG_03372 <sup>m</sup>	-	62
-	III	-	AMAG_04691	-	62
<i>Batrachochytrium dendrobatidis</i>	I	-	BDEG_06990	-	74
-	I	-	BDEG_06989 <sup>m</sup>	-	66
-	III	-	BDEG_03796	-	37
-	I	-	BDEG_02250	-	34
-	I	-	BDEG_00355	-	42
-	III	-	BDEG_00354	-	45
-	I	-	BDEG_00353	-	40
-	I	-	BDEG_00317	-	38
-	I	-	BDEG_00316	-	37
-	I	-	BDEG_00315	-	38
<i>Rozella allomycis</i>	I	-	OG9_002085-RA <sup>pm</sup>	-	70
-	-	-	OG9_000744-RA	-	37/36/36
-	-	-	OG9_003339-RA <sup>m</sup>	-	33/33/32
-	-	-	OG9_004960-RA	-	37/37/39
<i>Spizellomyces punctatus</i>	I	-	SPPG_05404 <sup>pm</sup>	-	77
-	I	-	SPPG_05884 <sup>m</sup>	-	76
-	III	-	SPPG_01130 <sup>m</sup>	-	70
-	-	-	SPPG_02793 <sup>m</sup>	-	52/42/46
-	-	-	SPPG_08686	-	43/40/37
<i>Piromyces sp.</i>	I	-	PirE2_1_18092 <sup>pm</sup>	-	74
-	III	-	PirE2_1_48456 <sup>p</sup>	-	66
<i>Gonapodya prolifera</i>	I	-	jgi 101126 <sup>pm</sup>	-	74
-	I	-	jgi 136309	-	57
-	III	-	jgi 137180 <sup>m</sup>	-	63
-	I	-	jgi 269854 <sup>m</sup>	-	74

Table 4.2: G-protein subunit comparison

Org	PDEA	PDEB	PDEG	Species
9	9	0		
6	6	0		
7	7	0		
3	2	0		
9	10	0		
10	10	0		
10	9	0		
23	24	1		
12	12	1		
11	11	2		
2	2	1		
2	2	0		
6	7	0		
9	12	0		
3	2	0		
3	0	0		
1	2	0		
0	1	0		
1	1	1		
5	5	0		

Table 4.3: PDE

# Chapter 5

## Transcriptome analysis in *Coelomomyces lativittatus*

### 5.1 Introduction

Species of Coelomomyces (Blastocladiomycota; Blastocladiales; Coelomomyctaceae) belong to the basal fungal lineages along with the Chytridiomycota, Neocalimastigomycota, and Cryptomycota. These species in general are obligate parasites which cycle between insect and crustacean hosts [67]. This process 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 [68]. After secretion of a thin cell wall, the encysted spore develops an appressorium and penetration tube which breaks through the host cuticle [69]. 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 [70]). The

penetration of copepods is thought to occur in a manner similar to that of the mosquito larvae [69]. 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 [67].

This work described in this chapter is motivated by a desire to ultimately understand the entomopathogenic nature of *Coelomomyces lativittatus* specifically, while also adding to the growing body of knowledge regarding chytrid biology generally.

*Coelomomyces* species have been studied previously in the context of mosquito control [71]. 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.

There are many examples of entomopathogenic organisms specializing in mosquito hosts, covering 13 genera across 2 kingdoms (Fungi and Chromista) [71]. 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 [?]. Later studies added to this repertoire the production of toxic cyclic peptides known as destruxins [?].

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)

[72]. Other evidence demonstrates a species-specific, photoperiod-dependent periodicity of gamete release from the copepod host [73], strongly implying that *Coelomomyces* has the molecular capacity for some manner of circadian rhythm regulation.

*Coelomomyces* are known producers of  $\beta$ -carotene [74], 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").  $\beta$ -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, an ongoing effort toward the assembly and annotation of a *Coelomomyces* transcriptome 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 *Coelomomyces lattivitatus*. In this chapter, I compare expressed protein functions 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.

## 5.2 Results

### 5.2.1 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 [75] yielded 77,597 transcripts with an average length of 386 bp. Within these transcripts, 21,486 open reading frames (ORFs) were predicted using Transdecoder [76]. 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 5.1. 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 *Cantenaria anguillalae* also has a higher number of these proteins relative to other Blastocladiomycete and Chytridiomycete species. Corresponding Gene Ontology (GO) Slim classifications for recovered transcripts are shown in Figure 5.1.

### 5.2.2 Insect Virulence

At least one gene family appears to be expanded in *C. lativittatus*, with implications in insect pathogenicity. Also, based on the hypothesis that *C. lativittatus* senses Anopholes hormones, we predict that there are receptors in the fungus that are similar

to or can bind the hormones like 20HE.

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

#### **5.2.2.1 C1 cysteine 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 [77]. The family is characterized by the Peptidase C1 (PF00112) and C1-like (PF03051) Pfam domains. The *C. lativittatus* transcriptome contains at least 56 transcripts with 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 the fungal copies revealed a number of observations (Fig4\_PF00112). First, the *C. lat* transcripts are broadly distributed, with very few tight clusters. A group of 5 transcripts fall nicely within a “fungal-specific” lineage containing the only other fungal sequences (derived from the Ascomycota and Basidiomycota). Another cluster of *C. lat* transcripts cluster as more recent divergences closer to the arthropod sequences.

### **5.2.2.2 Trypsin proteases**

Trypsin are serine proteases found in the digestive systems of many vertebrates. These enzymes are characterized by the PF00089 Pfam domain, and 43 transcripts in *C. lat* were identified. (more here; trees are running)

### **5.2.2.3 Destruxins**

The destruxins are a class of insecticidal cyclic hexadepsipeptides produced by some entomopathogenic fungi, most notably by species of *Metarhizium* (Donzelli et al. 2012; Wang et al. 2012). 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 (Pedras et al. 2002; Wang et al. 2012). The biosynthesis of these compounds is presumed to be mediated by an NRPS gene cluster in *Metarhizium robertsii* (Wang et al. 2012). 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, though m.15019 (described in  $\beta$ -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 (Bushley and Turgeon 2010). Some hits from AMP HMM, but counts are on the order of other chytrids ( 15-20).

### **5.2.2.4 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 (<http://www.cazy.org/>; accessed Oct 17, 2014) (Lombard et al. 2014). (presence/absence among chytrids and

why that would be important to note here) Five predicted ORFs were identified by InterPro as having a CBM18 domain and six ORFs identified with a CBM33 domain. All of the respective transcripts have minimal gene expression.

There was one transcript (m.4968) with a chitin synthase domain annotation (PF03142), and one transcript (m.4725) with a NADH-Ubiquinone domain (PF00361). The FPKM values of these transcripts were 637, and 130, respectively.

#### **5.2.2.5 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 (Travland 1979). 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 (Travland 1979b). 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 *C. albicans* pathogenesis (Staab et al. 1999). However there are no examples of this protein in *C. lativittatus* or other Blastocladiomycete or Chytridiomycetes surveyed. In an additional attempt to ascertain the nature of spore-cuticle attachment, we probed the Fungal Adhesin and Adhesin-like Database (FaaDB, <http://bioinfo.icgeb.res.in/faap/faap.html>), a set of experimentally verified and well-annotated fungal adhesins from several different fungi, predominantly Dikarya. The positive dataset was searched against the *C. lativittatus* transcriptome and well-scoring hits were recovered and submitted to the FAApred

SVM-based prediction method trained on both positive and negative adhesin datasets. In *C. lativittatus*, this method identified 16 sequences were as putative adhesins. In the other Blastocladiomycete 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.

#### 5.2.2.6 Ecdysone receptors

The naturally occurring ecdysteroid hormone 20-hydroxyecdysone (20HE) controls moulting in arthropods (Thummel and Chory 2002). There is evidence to suggest that 20HE plays a role in attracting *C. stegomyiae* to the ovaries of adult female *Aedes aegypti* (Lucarotti 1992). A FASTA search with the known ecdysone receptor protein from *D. melongaster* EcR (Koelle et al. 1991) 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 fact that nuclear receptors are not known to be in fungi and are presumed to be only limited to the metazoan lineages (Escriva et al. 1998). 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) was still the highest scoring. An alignment is provided in (20HE\_alignment.pdf).

This 298 aa transcript is likely not full length, and only aligns to the DNA binding region of the *D. mel* receptor (approximate residues 239 to 401). The top blast hit for this transcript is the *C. elegans* nuclear hormone receptor family member nhr-35 (SwissProt accession: Q17771, e-val 2e-20). InterProScan (Jones et al. 2014) 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. mel* ecdysone receptor (PDB ID: 2HAN, chain B) (JakÅšb et al. 2007). These two regions have 42% sequence identity. A homology-based structure model of the *C. lativittatus* transcript using SwissModel (Arnold et al. 2006) has an RMSD of 0.2 (Dali Server prediction (Holm and RosenstrÃ m 2010)) 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. mel* EcR protein (m.10080, 21.5% identity, e-val: 0.077). Nonetheless, (TabS1\_LBD\_blastHits) lists BLASTP results after searching m.10080 against SwissProt. The top 5 hits are all to mammalian liver X receptors (LXRs). A total of six hits below a threshold of 1e-06 are to 20HE receptors from insects. All hits have approximately 40% coverage and approximately 25% identity.

A maximum likelihood tree (Fig\_20HE\_NR\_RAxMLTree) 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 clustering outside of the metazoan nuclear receptor sequences.

Finally, we wished to determine if any unique receptors are found in *C. lat* relative to the other, non-insect associated chytrids. Noting that these are phylogenetically very different lineages, we 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 (A. mac, B. den, S. pun, and H. pol) based on ortholog clusters generated with OrthoMCL (Li et al. 2003). These 29 transcripts form 12 unique paralog clusters (meaning groups of only C.lat genes; 29 transcripts distributed among 12 groups). HMMER3 searches of Pfam database identified domains in 8 of these clusters, while the other 4 remained unclassified.

### 5.2.3 $\hat{\text{I}}\check{\text{s}}$ -carotene

*C. lativittatus* likely has a typical  $\hat{\text{I}}\check{\text{s}}$ -carotene biosynthesis pathway, despite missing enzyme in transcriptome. To determine the molecular characteristics of the  $\hat{\text{I}}\check{\text{s}}$ -carotene biosynthesis and metabolism pathways in *C. lativittatus*, we began by querying the predicted ORFs from the transcriptome with three key enzymes from the biosynthesis pathway described in *Blastocladiella emersonii* (Avelar et al. 2014). 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 (supTab\_Beme\_verificationOfSequences).

The pathway starts with a phytoene desaturase which is necessary for phytoene to lycopene conversion. 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 *Neurospora crassa* (NCU00552) were similarly unsuccessful.

The next enzyme in the process is a lycopene cyclase / phytoene synthase. One transcript, m.15019, contained a 599-aa long predicted ORF, 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  $\Delta$ I<sub>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.</sub>

The third key enzyme in the  $\beta$ -carotene biosynthesis and metabolism is  $\beta$ -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 (Kanehisa and Goto 2000; Kanehisa et al. 2014). When available, the Metazoan, Eukaryote, and/or Plant genes were used. Otherwise, the bacterial and archaeal protein sequences were used (supTab\_KEGGHMM\_list). The candidate *C. lativittatus* transcripts above were also recovered from these HMM searches.

Sequence searches identified all three of these  $\beta$ -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  $\hat{\beta}$ -carotene oxygenase homolog (top BLASTP hit against SwissProt:  $\hat{\beta},\hat{\beta}$ -carotene 9 $\alpha,\beta$ ,10 $\alpha,\beta$ -oxygenase from *Macaca fascicularis* [Q8HXG8.2], e-val: 1e-37, 26% identity). A comparative summary of these results is provided in (Figure\_bcaroPresenceAbsence).

To assess the phylogenetic history of BCMO1, a maximum likelihood tree was constructed from homologs found in Ascomycete, Chytridiomycete, Blastocladiomycete, and Zygomycete species (FigS1\_BCMO1DO2). 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  $\hat{\beta}$ -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.

The resulting gene tree was reconciled with a non-binary species tree generated with NCBI using Notung (v2.6) (BCMO1DO2\_reconciledTree.png <messy figure, might not be ultimately necessary>). This analysis supports the events described above and suggests that there were multiple additional duplications in specific zygomycete lineages.

### **5.2.4 Sensing: Photosensing capacity is consistent with ideas about C.lat abilities**

To determine the possible nature of observed photosensory capacity in *C. lativittatus*, we searched our transcriptome for ORFs predicted to be associated with photobiology (SupTab\_photosensing) 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 *Neurospora 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. They 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 NCU02256. However, it is much shorter: only 368 aa compared to 1167 aa for NCU02256. Other WC-1 homologs were recovered from *H. polyrhiza*, *S. punctatus*, *C. anguillalae*, and *A. macrogynus* (PAS\_alignment). No ORFs were predicted as WC-2 homologs, nor were there any other potential PAS-domain containing transcripts.

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 (Schwerdtfeger and Linden 2003). 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 (Rockwell et al. 2006), and Velvet A homologs are demonstrated to be regulators for secondary metabolism and sporulation in several fungi (Calvo 2008). 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 (Idnurm et al. 2010), 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 (SupTab\_photosensing). 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* (Avelar et al. 2014). 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 placement of this *C. lativittatus* Opsin-GC fusion protein among other opsin proteins, a maximum likelihood tree was generated using opsin se-

quences from Avalar et al., with additional inclusion of the opsin-GC fusion sequences recovered from *H. polyrhiza*, *C. anguillalae*, and *C. lativittatus* (OpsinGCFusion\_tree). 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 (OpsinGCFusion\_chytridCluster).

### 5.2.5 "interesting" animal homologs

Initial BLAST results against the nr database revealed a number of hits with putative animal homologs. An attempt to classify these hits further was made using a search against the Swissprot database. Of the 10 genes initially identified, only 8 had SwissProt hits with  $e\text{-val} < e\text{-}10$ . The number of hits to these 8 transcripts varied considerably, ranging from 2 up to 2407 (Rob\_mammalian\_hits). Nonetheless, all hits scoring better than  $e\text{-}10$  were collected, aligned to the *C. lat* transcript with Mafft, trimmed with trimal, and trees constructed with FastTree. (Trees are pretty messy)

## 5.3 Discussion

*Coelomomyces lativittatus* is the only known insect pathogen among the early-diverging 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*.

Insect Virulence discussion: Check out previously described repertoire in other fungi to see whatâŽs there. Surprising result is apparent unique expansion of C1 cysteine protease proteins unique to *C. lativittatus*, with previously inferred cuticle-degrading activity.

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 *C. psorophorae* infects mosquito larvae has been documented previously (Travland 1979a; Zebold et al. 1979). 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 (Zebold et al. 1979).

As noted by Travland (Travland 1979a) in *C. psorophorae*, 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 (Travland 1979a)

suggested that this material may be enzymatic in nature.

A comparison of counts of the top 20 PFAM domains (Table1\_Clat\_top20\_comparison) suggests four protein families which appear to be uniquely 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 cuticle 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

tantalizing explanation of Coelomomyces observed affinity for 20-hydroxyecdysone from mosquitoes. However, we are 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.

### **5.3.1 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 *Cu. inornata* larvae (Whisler et al. 1975). 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 *Cy. vernalis* in the case of *C. psorophorae* (Whisler et al. 1974). 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 (Federici 1983) suggesting the presence of a photoperiod dependent regulatory mechanism. These spores must then seek out members of the opposite mating type (Federici 1983), fuse to form zygotes, and exhibit phototactic capabilities to swim upwards to the water surface (Federici 1983).

Given this evidence in other Coelomomyces species, *C. lativittatus* likely also possesses a complex sensory network relative to chytrids which display no insect asso-

ciation. 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 *Blastocladiella emersonii*, and additional homologs can be identified in the genome assemblies of other Blastocladiomycetes *A. macrogynus* and *C. anguillulae* (Avelar et al. 2014). 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<sup>+</sup>-selective cyclic nucleotide gated channels. Voltage-activated Ca<sup>2+</sup> channels, activated by the resulting hyperpolarization of the plasma membrane, would elevate Ca<sup>2+</sup> levels, prompting interaction with the flagellum and ultimately mediating phototaxis (Avelar et al. 2014).

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 Blas-

tocladiomycota 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 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.

### **5.3.2 $\hat{\text{I}}\ddot{\text{s}}$ -carotene biosynthesis and metabolism pathways are present and nearly complete in the *C. lativittatus* transcriptome.**

Endogenous  $\hat{\text{I}}\ddot{\text{s}}$ -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  $\hat{\text{I}}\ddot{\text{s}}$ -carotene is the predominant molecule in *B. emersonii* and several Allomyces species,  $\hat{\text{I}}\ddot{\text{s}}$ -carotene specifically is known to be produced by *C. dodgei*. In this and other Coelomomyces species, the relative levels of  $\hat{\text{I}}\ddot{\text{s}}$ -carotene are indicative of mating type, implying that either the production and/or regulation of  $\hat{\text{I}}\ddot{\text{s}}$ -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,  $\hat{\text{I}}\ddot{\text{s}}$ -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  $\beta$ -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  $\beta$ -carotene.

The enzyme responsible for conversion of  $\beta$ -carotene to retinal is  $\beta$ -carotene monooxygenase. This is one of two enzymes capable of cleaving  $\beta$ -carotene, the other being  $\beta$ -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 study 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  $\beta$ -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 will deal with obtaining a draft reference genome from *C. lativittatus*, enhanced transcriptome se-

	Description	<i>Clat</i>	<i>Amac</i>	<i>Cang</i>	<i>Spun</i>	<i>Bden</i>	<i>Hpol</i>
PF00069	Pkinase	297	409	185	354	404	149
PF00400	WD40	211	1287	185	1710	1075	170
PF00076	RRM 1	143	201	94	278	240	53
PF00153	Mito carr	119	222	42	294	231	49
PF00012	HSP70	112	42	11	18	16	9
PF00118	Cpn60 TCP1	93	22	11	20	31	10
PF00071	Ras	86	80	57	108	96	47
PF00270	DEAD	70	115	94	136	126	81
PF00036	efhand	67	27	49	60	68	30
PF00112	Peptidase C1	65	0	1	0	0	0
PF01576	Myosin tail 1	59	1	42	4	3	10
PF00004	AAA	58	128	121	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	57	120	69	36
PF02985	HEAT	50	34	64	56	136	58
PF00005	ABC tran	50	217	180	168	231	96
PF00009	GTP EFTU	48	79	80	92	48	54
PF00089	Trypsin	47	2	9	8	4	2

Table 5.1: PFAM

quencing 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.

	SVM.Score	UniprotKB.ID	Description
Amac AMAG_17113T0	-0.74	Q9DBG3.1	AP-2 complex subunit beta
Amac AMAG_05862T0	-0.61	P00940.2	Triosephosphate isomerase EC=5.3.1.1
Amac AMAG_12167T0	-0.54	P00940.2	Triosephosphate isomerase EC=5.3.1.1
Amac AMAG_08758T0	-0.19	P46226.3	Triosephosphate isomerase, cytosolic EC=5.3.
Amac AMAG_10182T0	-0.17	P46226.3	Triosephosphate isomerase, cytosolic EC=5.3.
Amac AMAG_18542T0	-0.16	P48491.2	Triosephosphate isomerase, cytosolic EC=5.3.
Amac AMAG_16371T0	0.15		
Amac AMAG_08832T0	0.17	Q966L9.1	ATP-dependent RNA helicase glh-2 EC=3.6.
Amac AMAG_17977T0	0.21		
Amac AMAG_16084T0	0.32	P29141.1	Minor extracellular protease vpr EC=3.4.21.-
Amac AMAG_04496T0	0.34	Q6H236.1	Paternally-expressed gene 3 protein
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	Cell surface glycoprotein 1
Cang CANG_125451	-0.68	P58559.1	Glyceraldehyde-3-phosphate dehydrogenase 3
Cang CANG_33361	-0.09	P48494.3	Triosephosphate isomerase, cytosolic EC=5.3.
Cang CANG_38430	0.60	P22105.3	Tenascin-X
Cang CANG_69396	0.94		
Clat m.10957	-0.78	P27393.1	Collagen alpha-2(IV) chain
Clat m.22466	-0.75	Q6CJG5.2	Triosephosphate isomerase EC=5.3.1.1
Clat m.2165	-0.68	A7S7F2.1	Bystin
Clat m.18806	-0.68	P48501.1	Triosephosphate isomerase EC=5.3.1.1
Clat m.13634	-0.63	Q90XG0.1	Triosephosphate isomerase B EC=5.3.1.1
Clat m.1572	-0.58	O09452.1	Glyceraldehyde-3-phosphate dehydrogenase,
Clat m.15929	-0.52	P07487.2	Glyceraldehyde-3-phosphate dehydrogenase 2
Clat m.10361	-0.46	P20445.2	Glyceraldehyde-3-phosphate dehydrogenase H
Clat m.18916	-0.42	P48501.1	Triosephosphate isomerase EC=5.3.1.1
Clat m.11233	-0.30	Q96UF2.1	Glyceraldehyde-3-phosphate dehydrogenase 2
Clat m.4480	-0.07	O77458.1	Triosephosphate isomerase EC=5.3.1.1
Clat m.13062	0.52	Q6BMK0.1	Glyceraldehyde-3-phosphate dehydrogenase H
Clat m.745	0.90	Q92824.4	Proprotein convertase subtilisin/kexin type 5
Clat m.16183	1.05	P22105.3	Tenascin-X
Clat m.13342	1.13	Q92824.4	Proprotein convertase subtilisin/kexin type 5
Clat m.11209	2.24		
Bden BDET_06684	-0.79	Q5R2J2.1	Glyceraldehyde-3-phosphate dehydrogenase H
Bden BDET_05736	-0.17	P00939.1	Triosephosphate isomerase EC=5.3.1.1
Bden BDET_05372	-0.03	Q95P23.1	Enterin neuropeptides
Bden BDET_01761	0.45	Q9AVB0.1	Lectin-B
Bden BDET_00626	0.49		
Bden BDET_00287	0.67	Q9AVB0.1	Lectin-B
Bden BDET_02239	1.45	P23253.1	Sialidase EC=3.2.1.18
Bden BDET_00436	1.48		
Bden BDET_03668	1.51	Q63425.2	Periaxin
Bden BDET_06100	1.70		
Hpol HPOL_4940	-0.79	P00940.2	Triosephosphate isomerase EC=5.3.1.1
Hpol HPOL_4790	-0.46	O57479.3	Glyceraldehyde-3-phosphate dehydrogenase H
Hpol HPOL_4769	-0.24	P52041.2	3-hydroxybutyryl-CoA dehydrogenase EC=1
Hpol HPOL_2513	-0.14	Q1ZXD6.1	Probable serine/threonine-protein kinase rocc
Spun SPPG_01012T0	-0.68	P30741.2	Triosephosphate isomerase EC=5.3.1.1
Spun SPPG_08522T0	-0.47	Q6ZRI0.3	Otogelin
Spun SPPG_00685T0	-0.15	Q5H8C1.3	FRAS1-related extracellular matrix protein 1

	Cluster	Protein	Description
1	1127	Clat m.15402	7tm_1;PF00001
2	1127	Clat m.16182	7tm_1;PF00001
3	1127	Clat m.18794	7tm_1;PF00001
4	1127	Clat m.9338	7tm_1;PF00001
5	1175	Clat m.12314	K_trans;PF02705
6	1175	Clat m.12319	K_trans;PF02705
7	1175	Clat m.12322	K_trans;PF02705
8	1176	Clat m.15582	Grp1_Fun34_YaaH;PF01184
9	1176	Clat m.15583	Grp1_Fun34_YaaH;PF01184
10	1176	Clat m.15584	Grp1_Fun34_YaaH;PF01184
11	1177	Clat m.16070	DUF3533;PF12051
12	1177	Clat m.16072	DUF3533;PF12051
13	1177	Clat m.16075	DUF3533;PF12051
14	1269	Clat m.11119	UAA transporter;PF08449
15	1269	Clat m.11121	UAA transporter;PF08449
16	1270	Clat m.12151	–
17	1270	Clat m.12173	–
18	1271	Clat m.8230	–
19	1271	Clat m.8232	–
20	1272	Clat m.12638	–
21	1272	Clat m.12641	–
22	1273	Clat m.14468	Sodium:sulfate symporter;PF00939
23	1273	Clat m.14469	Sodium:sulfate symporter;PF00939
24	1274	Clat m.14625	–
25	1274	Clat m.14626	–
26	1275	Clat m.4725	NADH-Ubiquinone/plastoquinone;PF00361
27	1275	Clat m.8681	NADH-Ubiquinone/plastoquinone;PF00361
28	1276	Clat m.4968	Chitin synthase;PF03142
29	1276	Clat m.4970	Chitin synthase;PF03142

Table 5.3: OrthMCL

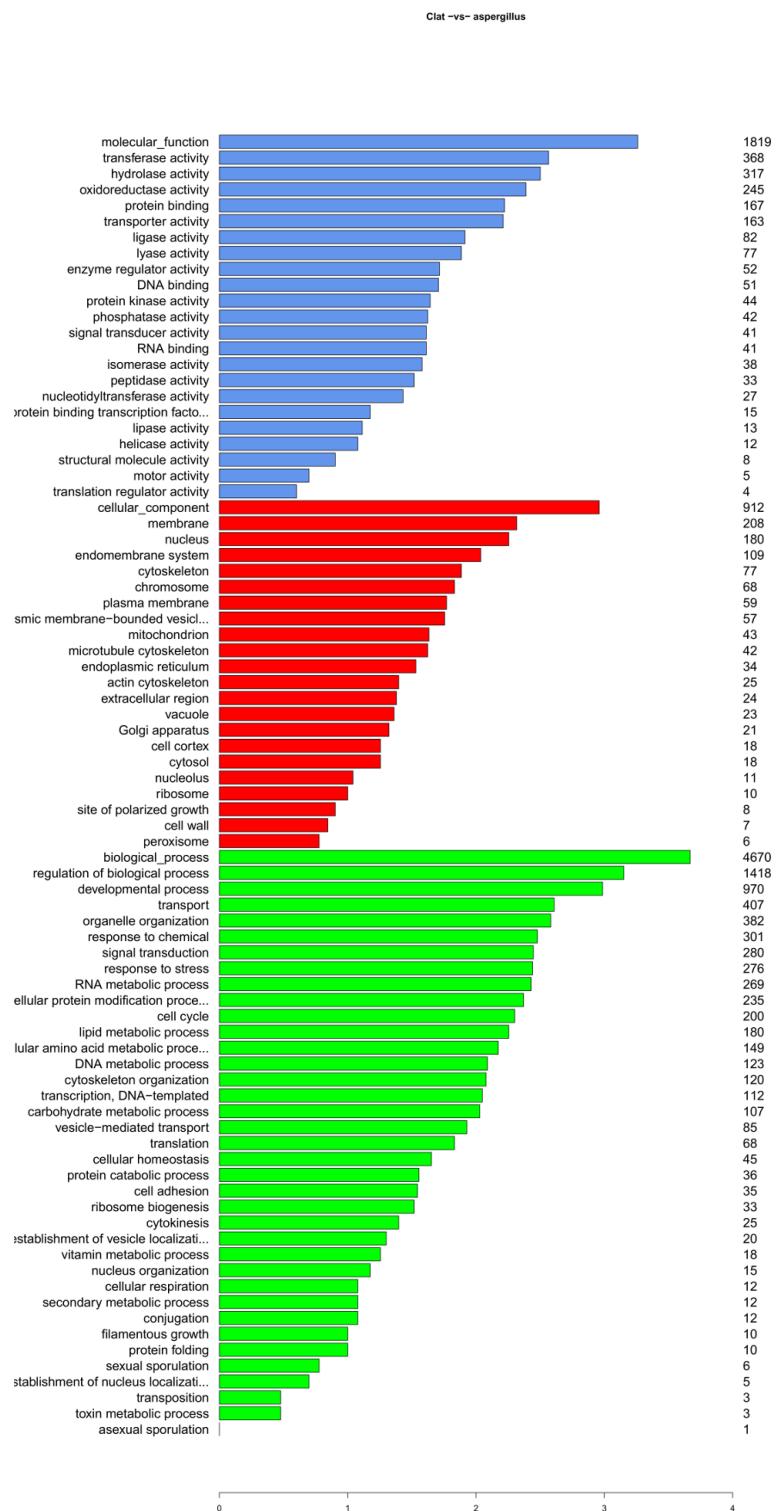


Figure 5.1: Horizontal bar chart showing the distribution of *Aspergillus* GO-Slim terms associated with the *C. lat* transcriptome. X axis is a logarithmic scale.

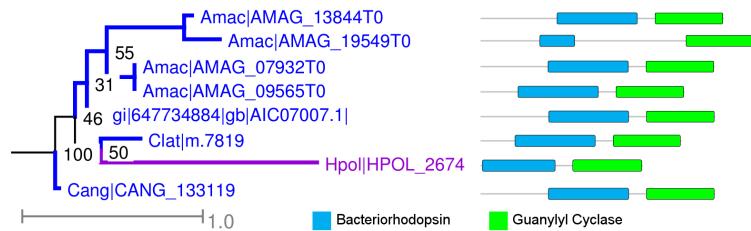


Figure 5.2: A subset of opsin

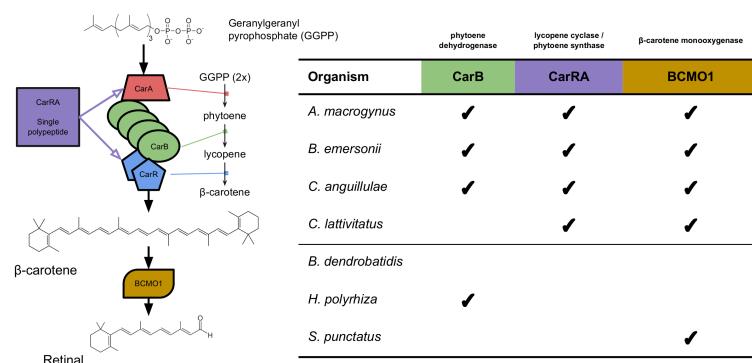


Figure 5.3: Presence and absence of  $\beta$ -carotene related proteins in species belonging to the Chytridiomycota and Blastocladiomycota.

# **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 through the following specific aims:

1. Describe and identify a compound responsible for antifungal behavior (Chapter 2)
2. Characterize the identification of an opsin-like protein, through both structural analyses (Chapter 3) as well as accessory and downstream proteins (Chapter 4)
3. Produce and interpret a transcriptome for mosquito pathogen, thereby laying groundwork for enhanced genomic resources for this organism (Chapter 5)

## 6.1 Inhibition

Attributed to Louis Pasteur in 1854: "In the fields of observation, chance favors only the prepared mind." The observation in the Stajich lab of the non-pathogenic Chytridiomycete 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:

1. Is this a unique property of *Hp*?
2. Is this a specific interaction with *N. crassa*?
3. 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*.

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.

## 6.2 Rhodopsin

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. The research in Chapters 3 and 4 address four key questions:

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

## 6.3 *Coelomomyces*

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 ?? was addressing two major points about entomopathogenic chytrids, using *Coelomomyces lativittatus* as a model:

1. What is possible to infer from the transcriptome regarding already known aspects of *C. lativittatus* biology, specifically regarding  $\beta$ -carotene biosynthesis, environment sensing, and insect-association?
2. How does the protein complement of *C. lativittatus* compare related Chytridiomycete and Blastocladiomycete species which are not insect-associated?

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.

# 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. Köljalg, 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] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] J. Saranak and K. W. Foster, “Rhodopsin guides fungal phototaxis,” *Nature*, vol. 387, pp. 465–466, May 1997.
- [17] L. K. Muehlstein, J. P. Amon, and D. L. Leffler, “Phototaxis in the marine fungus rhizophydiuum littoreum,” *Appl. Environ. Microbiol.*, vol. 53, pp. 1668–1672, July 1987.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] P. Ballario and G. Macino, “White collar proteins: PASSing the light signal in neurospora crassa,” *Trends Microbiol.*, vol. 5, pp. 458–462, Nov. 1997.
- [22] 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.
- [23] L. M. Corrochano, “Fungal photoreceptors: sensory molecules for fungal development and behaviour,” *Photochem. Photobiol. Sci.*, vol. 6, pp. 725–736, July 2007.

- [24] 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.
- [25] G. Wald, "The molecular basis of visual excitation," *Nature*, vol. 219, pp. 800–807, 1968.
- [26] J. von Lintig and K. Vogt, "Filling the gap in vitamin a research: Molecular identification of an enzyme cleaving  $\beta$ -carotene to retinal," *J. Biol. Chem.*, vol. 275, pp. 11915–11920, 21 Apr. 2000.
- [27] S. O. Smith, "Structure and activation of the visual pigment rhodopsin," *Annu. Rev. Biophys.*, vol. 39, pp. 309–328, June 2010.
- [28] A. Hausmann and G. Sandmann, "A single Five-Step desaturase is involved in the carotenoid biosynthesis pathway to  $\beta$ -carotene and torulene in *neurospora crassa*," *Fungal Genet. Biol.*, vol. 30, pp. 147–153, July 2000.
- [29] 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.
- [30] G. Lietz, A. Oxley, C. Boesch-Saadatmandi, and D. Kobayashi, "Importance of  $\beta,\beta$ -carotene 15,15 $\alpha$ -monooxygenase 1 (BCMO1) and  $\beta,\beta$ -carotene 9 $\alpha$ ,10 $\alpha$ -dioxygenase 2 (BCDO2) in nutrition and health," *Mol. Nutr. Food Res.*, vol. 56, no. 2, pp. 241–250, 2012.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] D. P. Kontoyiannis and R. E. Lewis, "Antifungal drug resistance of pathogenic fungi," *Lancet*, vol. 359, pp. 1135–1144, 30 Mar. 2002.
- [35] M. A. Pfaller, "Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment," *Am. J. Med.*, vol. 125, pp. S3–13, Jan. 2012.
- [36] 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.
- [37] P. Wiemann and N. P. Keller, "Strategies for mining fungal natural products," *J. Ind. Microbiol. Biotechnol.*, vol. 41, pp. 301–313, Feb. 2014.
- [38] B. J. Howlett, "Secondary metabolite toxins and nutrition of plant pathogenic fungi," *Curr. Opin. Plant Biol.*, vol. 9, pp. 371–375, Aug. 2006.
- [39] 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.

- [40] J. N. Couch, “Mass production of coelomomyces, a fungus that kills mosquitoes,” *Proceedings of the National Academy of Sciences*, vol. 69, pp. 2043–2047, Aug. 1972.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] 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.
- [53] 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.

- [54] A. Terakita, “The opsins,” *Genome Biol.*, vol. 6, no. 3, p. 213, 2005.
- [55] 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.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] E. Beitz, “TEXtopo: shaded membrane protein topology plots in LaTEX2ÉZ,” *Bioinformatics*, vol. 16, no. 11, pp. 1050–1051, 2000.
- [61] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, pp. 33–8, 27–8, Feb. 1996.
- [62] 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.
- [63] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, “Stereochemistry of polypeptide chain configurations,” *J. Mol. Biol.*, vol. 7, pp. 95–99, July 1963.
- [64] R. R. Birge, “Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin,” *Biochim. Biophys. Acta*, vol. 1016, pp. 293–327, Apr. 1990.
- [65] 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.
- [66] 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.
- [67] 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.
- [68] L. B. Travland, “Initiation of infection of mosquito larvae (*Culiseta inornata*) by *Coelomomyces psorophorae*,” *J. Invertebr. Pathol.*, vol. 33, pp. 95–105, Jan. 1979.

- [69] 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.
- [70] 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.
- [71] 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.
- [72] C. J. Lucarotti, “Invasion of aedes aegypti ovaries by *Coelomomyces stegomyiae*,” *J. Invertebr. Pathol.*, vol. 60, pp. 176–184, Sept. 1992.
- [73] 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.
- [74] B. A. Federici and S. N. Thompson, “Beta-Carotene in the gametophytic phase of *Coelomomyces dodgei*,” *Experimental Mycology*, vol. 3, pp. 281–284, Sept. 1979.
- [75] 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.
- [76] 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.
- [77] 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.
- [78] 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.
- [79] S. F. Altschul, T. L. Madden, A. a. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” *Nucleic Acids Res.*, vol. 25, pp. 3389–3402, Sept. 1997.

## Appendix A

# Comparative genomics analysis of Flagellar motility

### A.1 Introduction

### A.2 Results and Discussion

A search for Rozella homologs of flagellar-associated proteins from the *Naegleria* genome [78] 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.

(Figure A.1).

### A.3 Methods

The flagellar analysis was carried out using a dataset of 173 flagellar motility proteins obtained from the *Naegleria gruberii* genome sequence [78]. These proteins were used in a FASTA search (SSEARCH v36.07) using an e-val cutoff of e-20. Supplemental table X only considers proteins which are present in at least one Dikarya, Microsporidia, or Chytridiomycota proteom. PSI-BLAST [79] was used to identify homologs of the three polar-tube proteins (PTP1, PTP2, and PTP3) characterized previously [PMIDs: 11719806, 12076771].

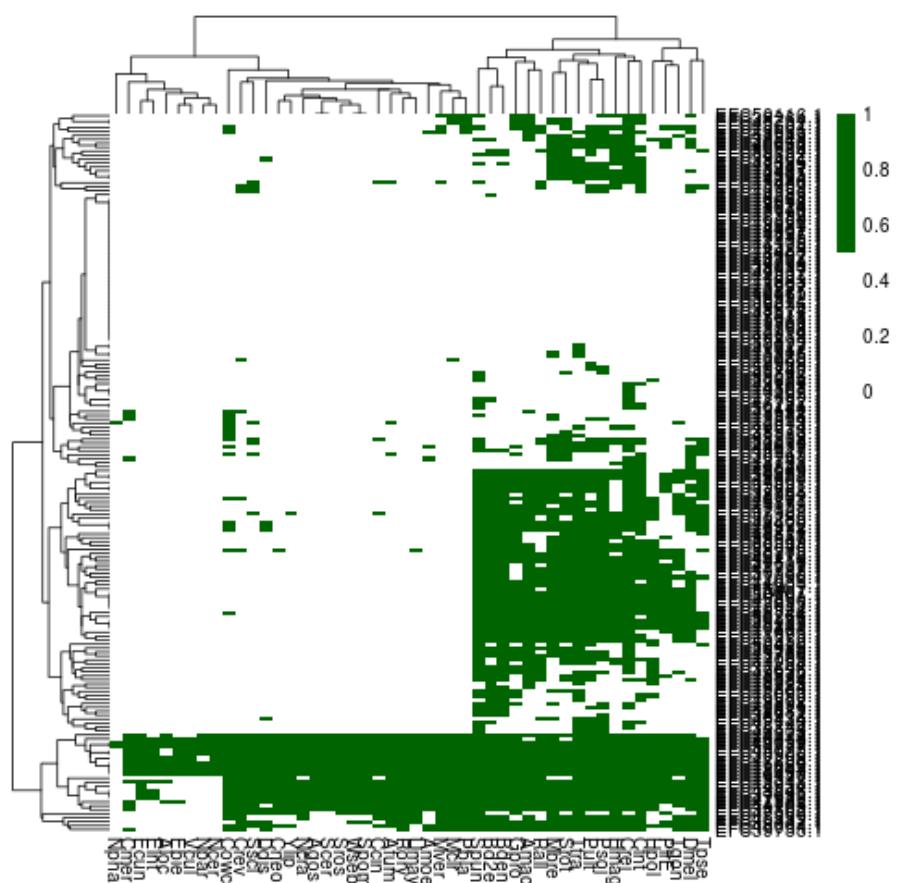


Figure A.1: Protein copies identified in proteomes of interest are normalized to indicate presence and absence only, where green indicates that one or more copies were found. Proteins are listed on the right, and proteomes given on the bottom. Both rows and columns are clustered by the "complete" method