Genomic resources enable insight into the developmental transcriptome of the blastoclad fungus, *Coelomomyces lativittatus*, an obligate parasite of mosquitoes and microcrustaceans

Cassandra L. Ettinger[a,#], Talieh Ostovar[a,b,$], Mark Yacoub[a,$], Steven Ahrendt[a,*], Robert H. Hice[c], Brian A. Federici[c,d], Jason E. Stajich[a,d,#]

[a]Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, California, USA.

[b]UCR/SDSU Joint Doctoral Program in Evolutionary Biology, San Diego State University, San Diego, CA, USA.

[c]Department of Entomology, University of California, Riverside, Riverside, California, USA.

[d]Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, USA

[*]**Present address** U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[$]**These authors contributed equally to this work.**

[#]**Address correspondence to**: Cassandra L. Ettinger, cassande@ucr.edu; Jason E. Stajich, jason.stajich@ucr.edu

Keywords: *Coelomomyces lativittatus*, Blastocladiomycota, transcriptome, genome assembly, genome annotation, host-associated, alternations of generations

Abstract:

Species of the phylum Blastocladiomycota, early diverging zoosporic (flagellated) lineages of fungi, are vastly understudied. This phylum includes the genus *Coelomomyces* which consists of more than eighty fungal species that are obligate parasites of arthropods. Known *Coelomomyces* species lack a complete asexual life cycle, instead surviving through an obligate heteroecious alternation of generations life cycle. Despite their global distribution and interesting life cycle, little is known about the genomics of any *Coelomomyces* species. To address this, we generated three draft-level genomes and annotations for *C. lativittatus* representing its haploid meiospore, orange (male) gamete, and amber (female) gamete life stages. These draft genome assemblies ranged in size from 5002 to 5799 contigs with a total length of 19.8-22.8 Mb and a mean of 7416 protein-coding genes. We then demonstrated the utility of these genomic resources by combining the draft annotations as a reference for analysis of *C. lativittatus* transcriptomes. We analyzed transcriptomes from across host-associated life stages including infection of larva and excised mature sporangia from the mosquito, *Anopheles quadrimaculatus*. We identified differentially expressed genes and over-enriched GO terms both across and within life stages and used these to make hypotheses about *C. lativittatus* biology. Generally, we found the *C. lativittatus* transcriptome to be a complex and dynamic expression landscape; GO terms related to metabolism and transport processes were over-enriched during infection and terms related to dispersal were over-enriched during sporulation. We further leveraged these genomic resources to identify five HMG box genes in *C. lativittatus*, three belonging to clades with mating type (MAT) loci from other fungi, as well as four ortholog expansions in *C. lativittatus* compared to other fungi. The *C. lativittatus* genomes and transcriptomes reported here are a valuable resource and may be leveraged toward furthering understanding of the biology of these and other early diverging fungal lineages.

Introduction:

Fungi contribute to critical roles in the global ecosystem, yet knowledge of their biology, genetics and biochemistry largely stems from observations of only two phyla, the Ascomycota and Basidiomycota (i.e., the Dikarya). Zoosporic (flagellated) lineages of fungi make up additional fungal phyla (including the Blastocladiomycota and Chytridiomycota), but are generally understudied [1]. The phylogenetic placement of these early-diverging zoosporic lineages is controversial and under constant revision as new genomic data becomes available [1,2]. For example, fungi belonging to the Blastocladiomycota were originally placed together with lineages in the Chytridiomycota, but now Blastocladiomycota is its own phylum [3,4].

Within the Blastocladiomycota, the genus *Coelomomyces* (Blastocladiales; Coelomomycetaceae) consists of more than eighty highly fastidious fungal species that are obligate fatal parasites, primarily of mosquitoes and microcrustaceans [5–8]. These fungi have a worldwide distribution and over the last hundred years have been reported infecting all major genera of mosquitoes, i.e., *Aedes*, *Culex*, and *Anopheles*, each of which contain many species that transmit pathogens that cause medically important diseases such as malaria, filariasis, and various viral encephalitides. Moreover, because it is difficult to detect *Coelomomyces* infections in larval and adult mosquitoes, it is estimated that there are more than several hundred species worldwide yet to be described [6], making these fungi a very large group for which we know virtually nothing about their genomics and biochemistry. This lack of knowledge is due to the failure, despite numerous attempts, to culture any species of *Coelomomyces in vitro*, the primary difficulty being the lack of a cell wall on hyphae growing in the vegetative stages of their mosquito and copepod hosts. As far as is known, *Coelomomyces* species lack a complete asexual life cycle, instead surviving through an obligate alternation of generations in which a sporophytic phase parasitizes mosquitoes (e.g., larvae) and a gametopytic phase parasitizes

microcrustaceans (e.g., copepods) (Figure 1) [5,6,9]. This type of lifecycle is uncommon in fungi, though a similar heteroecious life cycle is observed in the rust fungi [10].

The *Coelomomyces* life cycle begins when a biflagellate zygote encounters a mosquito larva. The motile spore encysts on the intersegmental membrane of the mosquito cuticle, a process facilitated by the secretion of adhesion vesicles [11]. The encysted spore develops an appressorium and penetration tube, which penetrates through the host cuticle [12]. Once inside the mosquito larva, the so-called hyphagens grow into coenocytic hyphae that ramify the hemocoel over a period of seven to ten days, which then form sporangia at their tips [6,9]. The mosquito larvae subsequently die and putrefy, liberating the sporangia. Meiosis then occurs in the sporangia resulting in haploid uniflagellate meiospores, which after sporangial dehiscence seek out and infect a crustacean host (typically copepods, though ostracods can serve as hosts as in some *Coelomomyces* species [13]).

The penetration of copepods is thought to occur in a manner similar to that of the mosquito larvae [9,12], after which hyphae of the gametophyte form a holocarpic gametangium that cleaves into gametes. The meiospores that infect the copepods are of opposite mating types, thus forming gametangia in the copepod host that can generate gametes of opposite mating types. When gametogenesis is complete, the gametangium bursts, killing the copepod host and allowing the gametes to escape through fissures in the intersegmental membranes. If a copepod is infected by meiospores of each mating type, the gametangia burst simultaneously, and gametes of opposite mating types mate forming biflagellate zygotes within the dead copepod, which then seek out a mosquito larval host after release, thereby completing the alternations of generations life cycle [5]. If only a single mating type gametangium develops within a copepod, the gametes swim to the surface where they seek a mate in the water in which the mosquito larvae are breeding. In some species, such as *C. punctatus* and *C. dodgei*,

the gametangia and gametes of different mating types, much like those of *Blastocladiella emersonii*, are different colors, apparently due to different levels of β-carotene, with the so-called male isoforms being bright orange, and the females light amber [14,15].

As noted above, despite their worldwide distribution and relatively unique life cycle, little is known about the biology, biochemistry, or genomics of *Coelomomyces* species. Modern molecular and genomic techniques allow us to circumvent the need for *in vitro* culturing and to expand foundational knowledge of this important fungal genus. Toward this goal, we have established an *in vivo* culture of *C. lativittatus*, a close relative of *C. dodgei* and *C. punctatus* [6,16], which we maintain using the mosquito, *Anopheles quadrimaculatus*, and the copepod, *Acanthocyclops vernalis*. The research presented here represents the first exploratory investigation of *Coelomomyces* genomics and the *C. lativittatus* transcriptome.

To begin to answer questions related to *Coelomomyces* biology, we generated draft genomes and annotations for *C. lativittatus* from three life stages: (i) meiospores, (ii) orange (male) gametes, and (iii) amber (female) gametes. We sought to demonstrate the value of these genomic resources by using them as reference for transcriptome analysis. We generated transcriptomes from across life stages including infection of A. *quadrimaculatus* hosts and sporangia excised from *A. quadrimaculatus*, to elucidate genes involved in the unique biology and alternation of generations life cycle of this fungus. We then further used the genomic resources to search for mating type loci in *C. lativittatus*, as well as look at expansions of orthologous genes compared to close relatives in the Blastocladiomycota and Chytridiomycota. The *C. lativittatus* genomes and transcriptomes reported here provide an invaluable foundational resource for understanding the biology of this and other unique and important understudied fungal lineages in various worldwide aquatic ecosystems.

Methods:

Study system

Larvae and copepods used for maintenance of the *in vivo* culture of *C. lativittatus* were, respectively, *A. quadrimaculatus* and *A. vernalis*. These were maintained in culture as described previously [17].

DNA extraction methods & sequencing

We sought to generate genomes from three haploid life stages (i) meiospores, (ii) orange (male) gametes, and (iii) amber (female) gametes.

To obtain *C. lativittatus* meiospores, mosquito larvae with advanced infections were collected when full of sporangia, within a day or two of death. To induce germination of sporangia, each larva was surface-sterilized by rinsing it in 70% ethanol for 20 seconds, after which each was placed in 1 ml of double-distilled water in a 22 mm plastic Petri dish at room temperature. The larvae were dissected with jeweler's forceps and most of the cuticle and midgut were removed from the water. Typically, the sporangia dehisced releasing meiospores 48 – 72 hours after incubation at room temperature. Meiospore samples were first collected using a 1 ml pipette and centrifuged using a table-top mini-fuge for 3 seconds to sediment any sporangia in the sample. The meiospores were then pelleted by centrifugation at 16,000xg for 2 minutes.

To obtain *C. lativittatus* gametes, infected copepods containing the orange and amber mating types were separated prior to copepod lysis. Liquid was removed and then copepods were surface-sterilized by rinsing in 70% ethanol and then washed with double-distilled water to reduce contaminants. After the gametes were released from the copepods, the copepod carcasses were removed by allowing them to settle in the microfuge tube. Samples were spun for 3 minutes at 6000xg, and the supernatant was removed.

DNA was extracted from the resulting meiospore and gamete samples using a Qiagen genomic DNA purification kit with Qiagen genomic-20/G tips following the standard manufacturer protocol. DNA was then amplified using the Qiagen repli-G Whole Genome Amplification Kit according to the standard manufacturer protocol. Illumina Libraries were prepared with the NEB DNA Library Prep Kit. Libraries were sequenced on an Illumina HiSeq 2500 with 100bp PE sequencing by the UC Riverside Genomics Core Facilities.

RNA extraction and library preparation

For transcriptome analysis of the sporophyte, hyphae were excised from infected fourth instar larvae of *A. quadrimaculatus* during either early, middle, or late stages of fungal development. For the purpose of this study, we define early, middle, and late infection stages as follows. Typically, the early stage of obvious infection appears as a few unpigmented, i.e., white, sporangia at the tips of hyphae about six days after molting to the fourth instar. The fat body in these larvae is quite well developed, and you can see hyphae adhering to this tissue in each of the larval abdominal segments and in the thorax. The middle stage occurs over days 7 and 8, during which the number of sporangia increases significantly, with most being mature, meaning rusty brown in color. The late stage occurs during days 9 and 10, by which time many larvae are full of sporangia and die, although many other larvae survive another 4-5 days before dying. These larvae continue to grow, being at least twice the size at which healthy larvae pupate. Two replicate samples were collected from each time point.

For the sporangia transcriptomes, mosquito larvae with advanced infections were collected when full of sporangia, within a day or two of death. To induce sporangia to undergo meiosis and germinate, each larva was surface-sterilized by rinsing it in 70% ethanol for 20 seconds,

after which each was placed in 1 ml of double-distilled water in a 22 mm plastic Petri dish at room temperature. The sporangia were excised from the larvae using jeweler's forceps, afterwhich the cuticle and midgut were removed from the water. Typically, the sporangia dehisced releasing meiospores 48 – 72 hours after incubation at room temperature. Sporangia were collected at 0, 24, 36, and 48 hour time points starting from the period the sporangia were excised from the mosquito larva (0 hour) through dehiscence (48 hour), when the uniflagellate meiospores burst out of the sporangia. Samples were collected using a 1 ml pipette and centrifuged using a table-top mini-fuge for 3 seconds to obtain the sporangia in the sample. Two replicate samples were collected from each time point.

RNA from all samples was extracted with Trizol (Life Technologies, Grand Island, NY) as per the manufacturer's protocol. 1.2 µg of RNA was used as the starting material for the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, MA). Poly-A RNA was purified as per instructions and converted to adapter-ligated, size-selected cDNA. An aliquot of the library was cloned into pJet1.2 (Thermo Fisher Scientific, Waltham, MA) and clones sequenced with standard methods to check library quality. An aliquot was also run on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). The resulting sequencing libraries were sequenced by the Institute for Integrative Genome Biology Core facility at the University of California at Riverside using the MiSeq instrument with 100 bp paired-end reads (Illumina, San Diego, CA).

Genome assembly

Genomes for the meiospore (MEIOSPORE), orange gamete (ORANGE) and amber gamete (AMBER) libraries were assembled using the Automatic Assembly of the Fungi (AAFTF) pipeline v. 0.2.3 [18]. Briefly, this involved first trimming and filtering reads using bbduk.sh from BBTools v. 37.76 [19]. Next, assemblies were produced using AAFTF command assemble

which relies on SPAdes v. 3.12.0 [20] run with default parameters to select optimal K-mer size and screened for contaminant vectors with AAFTF vecscreen step using NCBI BLAST [21]. The AAFTF sourpurge step was run which utilizes sourmash v. 3.5.0 [22] to further purge any remaining contaminant contigs and AAFTF rmdup step using Minimap2 v. 2.17 [23] was run to identify duplicate contigs for removal. Finally, AAFTF polish step ran Pilon v. 1.22 [24] to polish the resulting contigs in each assembly to remove potentially mis-called consensus nucleotides or INDELs by SPAdes.

Assembly evaluation for each genome was performed using QUAST v. 5.0.0 [25] and BUSCO v. 5.0.0 [26] against both the eukaryote_odb10 and fungi_odb10 gene sets. BUSCO assessment was also performed for reference genomes from other fungal lineages for comparison (Table S1). We performed telomere searches against the *Coelomomyces* assemblies using find_telomeres.py (https://github.com/markhilt/genome_analysis_tools) to test for telomeric repeats at the ends of the scaffolds and determine chromosome completeness.

Contamination screen and removal

Microbial contamination was assessed in each of the assemblies using the BlobTools2 pipeline [27]. This involved first assigning taxonomy against the UniProt Reference Proteomes database (v. 2020_10) to each contig using diamond (v. 2.0.4) and command-line BLAST v. 2.2.30+ [21,28]. Next, read coverage was calculated by aligning the reads from the MEIOSPORE, AMBER and ORANGE libraries to their respective genome assemblies with BWA v.0.7.17 [29] and sorted using samtools v. 1.11 [30]. Finally, the BlobToolKit Viewer was used to visualize the GC content, read coverage and predicted taxonomies of contigs to identify contaminants.

Microbial contamination was further assessed with the anvi'o v.7 pipeline [31,32], a complementary method, for the AMBER assembly. This involved first obtaining read coverage

from each of the three genomic samples (AMBER, ORANGE, and MEIOSPORE) to the AMBER assembly with bowtie2 v. 2.4.2 [33] and samtools v. 1.11 [30]. A contig database for the AMBER assembly was then generated using ‘anvi-gen-contigs-database’ which uses Prodigal v. 2.6.3 [34] to predict open-reading frames. This command also identifies single-copy bacterial [35], archaeal [35] and protista [36] genes using HMMER v. 3.2.1 [37] and ribosomal RNA genes using barrnap [38]. We predicted taxonomy for each predicted ORF using Kaiju v. 1.7.2 [39] with the NCBI BLAST non-redundant protein database nr including fungi and microbial eukaryotes v. 2020-05-25. We then constructed anvi’o profiles for each sample (AMBER, ORANGE, and MEIOSPORE) using ‘anvi-profile’ with the ‘--cluster-contigs’ option and a contig length cut-off of >2.5 kbp. These profiles were then merged using 'anvi-merge'. Contaminant contigs in the AMBER assembly were then identified through ‘anvi-interactive’ using a combination of hierarchical clustering based on coverage and tetranucleotide frequency, taxonomic identity, and GC content.

Contaminant contigs (e.g., any contig identified by the BlobTools2 pipeline as assigned to bacteria, archaea or viruses and any contig identified using the anvi’o pipeline) were subsequently removed from the draft assemblies.

Genome annotation

Genome annotation was performed using the Funannotate pipeline v.1.7.4 [40]. This first involved using RepeatModeler v.1.0.11 [41] and RepeatMasker v.4.0.7 [42] to generate a library of predicted repetitive elements and then soft mask these elements in the draft genomes. Next the RNASeq data was assembled using Trinity v. 2.10.0 in Genome-Guided mode and aligned with PASA v.2.3.3 to train the *ab initio* gene prediction algorithms augustus and SNAP [43,44]. Consensus gene models were generated using EVidenceModeler v.1.1.1 [45] on predicted protein-coding gene models from a combination of algorithms including CodingQuarry v. 2.0,

Augustus v. 3.3.3, GeneMark-ETS v. 4.38, GlimmerHMM, and SNAP v 2013_11_29 [46–50]. Transfer rRNA genes were predicted using tRNAscan-SE v. 1.3.1 [51]. Protein annotations were predicted for the consensus gene models based on similarity to Pfam [52], CAZyme domains [53,54], MEROPS [55], eggNOG v. 1.0.3 [56], InterProScan v. 5 [57], and Swiss-Prot [58] using HMMER v.3 [37] or diamond BLASTP [59]. Phobius [60] and SignalP v. 4.1c [61] were also run to predict transmembrane proteins and secreted proteins respectively. Any problematic gene models that were flagged by Funannotate were manually curated as needed. The annotation results were summarized in custom code written in R v. 4.0.3 using the tidyverse v. 1.3.0 package [62,63]. The annotated genomes of the MEIOSPORE, ORANGE, and AMBER assemblies were aligned and mapped to the RNA sequencing reads using HISAT2 [64].

Transcriptome analysis

Given the obligate relationship of *Coelomomyces* with its hosts, we chose a reference-based transcriptome approach as initial results from *de-novo* approaches revealed host contamination even after removal using a reference host transcriptome. To provide a more comprehensive gene set to use as a reference for transcriptome analysis, we combined the predicted transcript sets from all three genome annotations. We then used CD-HIT-EST to cluster transcripts at 90% sequence identity and evaluated this combined set (AOM90) with BUSCO v. 5.0.0 [26,65]. For comparative purposes, a protein alignment of *C. lativittatus* with other fungal lineages (Table S1), was constructed using the PHYling_unified (https://github.com/stajichlab/PHYling_unified) pipeline, which uses HMMER v.3 [37] and ClipKIT [66] to search for markers in the protein sequences, build, and trim an alignment based on BUSCO fungi_odb10 HMMs. A maximum likelihood phylogeny was built from this alignment using IQ-TREE2 v.2.2.0 [67].

Mosquito host transcripts were removed from the transcriptome data using BBMap against an *A. quadrimaculatus* (Accession: GBTE00000000) reference transcriptome prior to read

quantification [19,68]. Host-filtered transcriptome read counts were quantified against the AOM90 transcript set using kallisto [69]. The count data was then imported into R for analysis with the DESeq2, ggplot2 and GSEABase packages [63,70–72].

General expression trends across all samples were visualized using variance stabilized count data. We then used DESeq2 to identify differentially expressed transcripts between life stages (e.g., sporangial vs. infection). Significant genes were defined as transcripts with a false discovery rate adjusted *p*-value of > 0.01 and a |log2 fold change| > 2. Gene Ontology (GO) enrichment analysis was performed to assess if the differentially expressed transcripts were significantly over-enriched in any particular functions ($p < 0.05$). This analysis was performed at each of three GO classes: biological processes (BP), molecular functions (MF), and cellular components (CC). Significantly over-enriched GO terms were simplified using Revigo with the default settings [73].

In order to identify transcripts that show change in expression across the development time course conditions within each sporangium (e.g., 0 hr vs. 24 hr vs. 32 hr vs. 48 hr) and infection stages (e.g., early vs. middle vs. late infection), we performed a Likelihood Ratio Test (LRT). Significant transcripts from LRT were filtered with a false discovery rate adjusted *p*-value of > 0.01 and a |log2 fold change| > 2. The “DEGreport” package was used to cluster genes with similar expression profiles based on the LRT results across different time series conditions [74]. A GO enrichment analysis was performed to identify over-enriched GO terms in each of the clusters with similar expression patterns ($p < 0.05$). GO terms were simplified using Revigo with default settings [73].

Identification of mating type (MAT) loci in C. lativattus

To identify high mobility group box (HMG-box) genes putatively involved in mating in the *C. lativittatus* genomes, we used HMMsearch v. 3.3.2 for PFAM PF00505 with an e-value of 1e-15 [37]. Given that genes neighboring HMG-boxes are thought to be involved in mating in other fungi [75], we used Clinker [76] and Cblaster [77] to assess the syntenic regions surrounding the HMG-boxes in the *Coelomomyces* assemblies to identify conserved regions neighboring HMG-boxes. To confirm phylogenetic placement of the identified HMG-box genes, we aligned the candidate genes from *C. lativattatus* with those of other fungi (see Table S1) using HMMalign [37]. We then constructed a maximum likelihood phylogenetic tree of the HMG-box genes using IQ-TREE [67]. Finally, to compare expression of HMG-box genes across *C. lativittatus* life stages, the variance stabilized expression levels of the HMG-box genes were plotted using ggplot2 [71]. Pairwise t-tests were performed to assess differential expression between life stages.

Identification of orthologous gene expansions relative to other fungal lineages

Orthofinder v. 2.5.4 was used to identify whether any differentially expressed transcripts represented genes expanded in copy number in *C. lativittatus* compared to other fungi (Table S1) [78]. We filtered the Orthofinder results to orthogroups containing genes with differentially expressed transcripts in the RNAseq data. These results were subsetted by orthogroups that were at least |log2 fold change| > 2 higher in copy number in *C. lativittatus* compared to the other fungi. Orthogroup expansions were confirmed through phylogenetic analyses. Briefly, a nucleotide alignment of all genes in an orthogroup of interest was produced using MUSCLE v. 5.1 and then a phylogenetic tree was built with FastTree using -gamma and -lg [79,80].

Results

Bacterial contaminants were identified and removed from draft assemblies

Given the obligate nature of *C. lativittaus* with its hosts, microbial contamination was assessed in each assembly using BlobTools2 (Figure S1). This approach flagged 1969, 11, and 24 contigs as putative contaminants in the AMBER, MEIOSPORE and ORANGE assemblies respectively. All prokaryotic contigs identified using BlobTools2 were removed from the ORANGE and MEIOSPORE assemblies. However, given the large number of contaminants predicted to be present in the AMBER assembly, we decided to also use a complimentary method, anvi'o, to visualize and confirm contaminant contigs. This second method enabled us to identify 1127 contaminant contigs. Altogether, both methods identified a shared 1005 contigs, while BlobTools2 identified an additional 964 unique contigs and anvi'o identified an additional 122 unique contigs. Ultimately, a total of 2091 contaminant contigs were identified and removed from the AMBER assembly.

C. lativittatus genome assemblies are fragmented and incomplete

After contaminant removal, the draft genome assemblies ranged in size from 5002 to 5799 contigs with a total length between 19.8 Mb and 22.8 Mb (Table 1). The assemblies were considerably fragmented with an average N50 of 6128 bp with the longest contigs ranging from 30,249 to 56,469 bp in length.

Assessment with BUSCO further confirmed that the draft assemblies were partially complete (Table S1). Mean completeness when run in 'genome' mode was only a moderate 43.1% using the fungi set, and a slightly higher 56.3% using the eukaryota set. These values were lower than the mean completeness of other blastoclads (fungi: 75.5%, eukaryota: 83.8%), including a recent long-read genome from *B. emersonii* (fungi: 81.8%) [81]. Additionally, fragmented BUSCOs accounted for an average 15.5% and 22.3% in the *C. lativittatus* draft assemblies for the fungi and eukaryota sets respectively and many of the BUSCO marker genes were not found (fungi: 41.4%, eukaryota: 21.3%).

C. lativittatus genome annotations provide sufficient reference for transcriptome work

The total number of gene models predicted for each assembly ranged from 7344 to 7677 (Table 1). On average, there were 7416 protein-coding genes and 59 tRNA genes, with 63.2% of these having a hit to at least one functional database. Using HISAT2, the overall alignment of RNAseq reads to the AMBER, ORANGE, and MEIOSPORE assemblies was 55.4%, 55.17%, and 59.01% respectively.

To provide a more comprehensive gene set for transcriptome analysis, we combined the three predicted transcript sets together. We clustered the combined set at 90% identity using CD-HIT-EST to remove redundancy and to generate a final set (AOM90) of 8645 transcripts. Combining the three transcript sets led to improved BUSCO scores (Table S1); using BUSCO 'protein' mode, completeness increased from an average of 42.4% to 62.5% for the fungi set, and from 61.8 to 82.8% for the eukaryota set (Figure S2). While these improved scores were still lower than the average BUSCO protein scores for other blastoclads (Figure 2; fungi: 84.1, eukaryota: 90.1%), we believe that the AOM90 combined transcript set provides a sufficiently robust reference for transcriptome analysis.

Differentially expressed transcripts and GO Enrichment analysis across life stages

Initial analysis of the transcriptome profiles supported a distinct divide between infection and sporangial life stages with each stage clustering separately (Figure 3A) and clear differences in expression in the most abundant differentially expressed transcripts (Figure 3B). We followed this analysis with differential transcript expression analysis using DESeq2 to identify specific transcripts responsible for these patterns.

We found 1262 differentially expressed transcripts between life stages, with 395 transcripts enriched during infection compared to 867 enriched during sporangial life stages ($p < 0.01$, log2foldchange > 2). Of these, 575 (45.6%) had no matches to any of the databases used for annotation. Interestingly, while more transcripts were enriched in sporangial stages, many of the most abundant transcripts were representative of transcripts enriched in the infection stages and many of these transcripts were unannotated. For example, of the top 25 most abundant differentially expressed transcripts, 22 were upregulated during infection relative to sporangial stages (Figure 3B). Further, 18 of these 25 transcripts had no homology to any of the annotation databases, and further only one transcript was fully annotated, HMI54_014395 (*ERG10*), an acetyl-CoA C-acetyltransferase.

We performed GO enrichment analysis on the differentially expressed transcripts to identify over-enriched GO terms of interest (Table S2, $p < 0.05$). For the infection stages, GO terms from 38 biological processes (BP), 9 cellular compartments (CC), and 46 molecular functions (MF) were identified. Of these, seven of the top ten significantly over-enriched BP were metabolic processes (e.g., GO:0006082: organic acid metabolic process, GO:0046394: carboxylic acid biosynthetic process) and three were transport-related (e.g., GO:0006848: pyruvate transport, GO:1905039: carboxylic acid transmembrane transport). For the sporangial stages, GO terms from 35 BP, 26 CC, and 41 MF were identified. Of these, the most significantly over-enriched BP was related to reproduction (GO:0000003: reproduction), with four of the top ten significantly over-enriched biological processes were related to metabolic and biosynthetic processes (e.g., GO:0005975: carbohydrate metabolic process, GO:0006183: GTP biosynthetic process, GO:0006228: UTP biosynthetic process) and three were related to movement or organization of cellular machinery (e.g., GO:0006928: movement of cell or subcellular component, GO:0007010: cytoskeleton organization, GO:0007017: microtubule-based process).

Differentially expressed transcripts and GO Enrichment analysis within life stages

Identification of differentially expressed transcripts across the development time course within sporangial and infection stages was done using a LRT. We found 380 significant transcripts which were differentially expressed within infection stages ($p < 0.01$). Of these, we found two clusters of differentially expressed transcripts with similar expression patterns in each, group 1 and group 2 with 167 and 213 transcripts, respectively (Figure 4A). The same analysis with the sporangial stages indicated 3701 differentially expressed transcripts ($p < 0.01$) and out of which, we identified 7 clusters, each including differentially expressed transcripts with similar patterns of expression. There were 1083, 965, 859, 219, 355, 156 and 64 genes in groups 1 to 7, respectively (Figure 4B).

In order to identify over-enriched GO terms of interest, we performed GO enrichment analysis on the differentially expressed transcripts across the developmental conditions within infection and sporangial stages (Figure S3, $p < 0.05$). Within the infection stages, we found 30 BP, 12 CC, and 33 MF over-enriched GO terms. Of these, the top significantly over-enriched GO terms were in the CC category (GO:0000786: nucleosome, GO:0043228: non-membrane-bounded organelle, GO:0043232: intracellular non-membrane-bounded organelle, GO:0044815: DNA packaging complex, GO:0032993: protein-DNA complex), and were also seen in group 2 with decreasing expression across infection stages (Figure 4A).

Furthermore, we found 19 BP, 10 CC and 28 MF over-enriched GO terms represented within the sporangial stages and of them, the most significantly over-enriched GO terms were related to structural molecule activity and binding (GO:0005488: binding, GO:0003735: structural constituent of ribosome, GO:0004100: chitin synthase activity, GO:0005198: structural molecule activity, GO:0003779: actin binding) as well as the GO terms related to peptide biological

processes (GO:0006412: translation, GO:0043043: peptide biosynthetic process, GO:0006518: peptide metabolic process). Terms related to structural molecule activity and binding were generally seen in expression pattern groups 1 and 5 both of which generally increased in expression over time (Figure 4B).

HMG-box loci were identified with differential expression across life stages

A total of five unique HMG-box genes were identified, with all five HMG-box genes present in the MEIOSPORE assembly and three in each of the AMBER and ORANGE assemblies. The identified HMG-box genes were found on small fragmented contigs (average: 8,300 bp) which contained only a few neighboring genes (average: 3 genes). Despite their small size, synteny analysis across the three assemblies found that the AMBER and ORANGE assemblies share two HMG-box loci with each other. The third HMG-box genes in the ORANGE and AMBER assemblies were only shared with the MEIOSPORE assembly, which has four syntenic orthologous HMG-boxes (Figure S4). We tested the five HMG-box loci in *C. lativittatus* for synteny against *Allomyces macrogynus* to determine whether neighboring genes around HMG-boxes are conserved in sexually reproducing chytrids [82]. We were unable to determine synteny of neighboring genes around HMG-box loci, possibly due to the fragmented scaffolds where these genes are found in our *C. lativittatus* assemblies. Phylogenetic analyses of the HMG-box genes showed that four of the HMG-box genes generally fell in clades with other blastoclads or chytrids, while HMI54_015288 fell into a clade with Dikarya. Further, three of the five HMG-box orthologs from *C. lativitattus* (HMI56_006544, HMI55_007199, and HMI54_004920) were present in a clade containing known Mating-related HMG-boxes (Figure S5). HMI56_006544 and HMI54_004920 were present in all three *C. lativitattus* assemblies while HMI55_007199 was only present in the ORANGE assembly. Interestingly HMI56_006544 is closely related to SexM in *Phycomyces blakesleeanus,* however we were unable to identify a SexP homolog in *C. lativitattus.* The number of identified HMG-box genes here is in line with

that of other blastoclads (i.e., *Blastocladiella britannica* and *Catenaria anguillulae*; Figure 5). Of the five identified HMG-box genes, HMI56_007461 was significantly over-expressed in the infection stages compared to the sporangial stages (Figure S6; $p < 0.01$).

C. lativattatus may have expanded gene families of orthologs related to its unique life history

In order to determine if gene families expanded in *C. lativattatus* we performed Orthofinder analysis comparing *C. lativattatus* with other fungal taxa (Table S1). In total we identified 37,398 orthogroups among the fungal taxa. Of these, 182 orthogroups were exclusively present in *C. lativattatus* and absent in all other fungal taxa.

We identified 10 orthogroups that both contained genes with differentially expressed transcripts and were expanded in copy number compared to other fungi. Of these 10 orthogroups, only four had predicted functional annotations (Figure 5). We tested the four orthogroups for duplication errors and removed any sequences that appeared truncated or had 100% sequence similarity to an ortholog from a different *C. lativittatus* assembly. Most of the copy number expansions for these four orthogroups appeared to occur on clades exclusive to *C. lativittatus.* The putative functions for the four validated orthogroups were replication timing regulatory factor 1 (*RIF1*), chitin deacetylase, adenosine monophosphate (AMP)-activated protein kinase 1 (*AMPK1*), and *Egh16*-like virulence factor. Within these orthgroups, only the *RIF1* orthogroup contained a gene that had higher expression in infection stages; the other three orthogroups had genes with higher expression during sporangial stages.

Discussion

C. lativittatus annotated genomes are an important community resource

The genomes assembled and annotated here, while incomplete, are a critical community resource as little genomic data exists for members of the Blastocladiomycota. The smallest

public genome is *B. britannica* with a genome size of 19 Mb with 9,431 predicted gene models and the largest is *Allomyces macrogynus* with a genome size of 47 Mb and 19,446 predicted genes [83]. *C. lativittatus* falls on the shorter end of this range with an average genome length of 21.5 Mb and average 7475 predicted gene models, possibly due to the incomplete nature of the draft genomes reported here. While incomplete, we think that the three *C. lativattaus* genomes assembled and annotated in this study provide much needed genomic resources for these obligate fungi as demonstrated by their utility when combined as a reference for the transcriptome work herein. The phylogenetic placement of Blastocladiomycota has been disputed, but inclusion of additional genomes like those reported here can help elucidate these ancient phylogenetic relationships [84].

Transcriptomic landscape of C. lativittatus life stages provides insight into Blastocladiomycota biology

The transcriptome of *C. lativittatus* is a complex, dynamic, and underexplored landscape. The results of this study highlight a need for future refinement and exploration of gene annotation in this species, as evidenced by the 45.6% of differentially expressed transcripts with no annotation, and the majority of the top 25 abundant differentially expressed transcripts lacking functional annotation. In spite of these shortcomings, we were able to make generalizable insights about *C. lativittatus* biology from the subset of transcripts that were annotated with GO terms. Overall, during infection, GO terms were over-enriched related to metabolism and transport processes. While during sporangial stages, GO terms were enriched related to dispersal (i.e., cell signaling, locomotion and transport machinery). Looking at the expression patterns within life stages, we can begin to see more complicated trends emerge.

Within infection stages, we identified two different patterns of gene expression. In the first, gene expression increased over the course of development time (early, middle and late infection),

with over-enriched GO terms related to membrane transport and metabolic processes. In the second, gene expression decreased over time with GO terms related to DNA replication, nucleic acid and amino acid biosynthetic processes.

The enrichment of metabolism and membrane transport processes in the transcriptomes during infection stages is similar to reports from other early diverging fungal lineages. Upregulation of transport-related pathways has been reported in chytrid infection of frog hosts, which the authors suggest might be related to nutrient availability and proliferation related to host association [85]. Further, in the Microsporidia, *Vavraia culicis*, over-enriched GO terms for growth, metabolism and replication were identified and posited to be related to its generalist lifestyle and ability to infect multiple types of hosts [68]. Here, we observed an over-enrichment of metabolism terms as part of a pattern of increasing expression (group 1, Figure 4A), and an over-enrichment of replication-related terms as part of a pattern of decreasing expression (group 2, Figure 4A). Thus, we may be observing a shift in priorities during infection, with early infection stages marked by increased replication as hyphagens grow into coenocytic hyphae inside the host and later infection stages marked by increased metabolism as the fungus proliferates and begins preparing to make sporangia.

Within sporangial stages, we identified seven expression patterns with two patterns displaying higher expression over these developmental stages, two patterns displaying lower expression, and three patterns with a relatively higher gene expression at the second time point (24 hrs), followed by stable or decreasing expression. In general, expression patterns with higher expression were enriched in GO terms related to chitin activity as well as terms related to dispersal and microtubule-based processes. While, in decreasing expression patterns the over-enriched GO terms were mostly related to metabolism and transcription. The other three expression patterns, which displayed the highest expression at the second developmental time

point, were functionally different from each other. One group was over-enriched in GO terms related to dispersal and structural machinery, another in terms related to oxidative stress responses, and the third in terms related to lipid metabolism. Similarly, time series clustering of the transcriptome profiles of differentially expressed genes during the sporulation of the blastoclad, *Blastocladiella emersonii,* showed eight different patterns [86].

The enrichment of reproduction and dispersal machinery-related mechanisms during the sporangial stages likely relates to the production of meiospores. For example, signal transduction and microtubule and cytoskeleton biogenesis were similarly reported to be over-enriched during sporulation in *B. emersonii* [86]. These authors also observed a decrease in transcription and metabolism during sporulation which they attributed to the nutritional starvation required in order to sporulate. Additionally, previous investigations into protein synthesis in chytrids [87] and blastoclads [88,89] suggest that translation does not occur until zoospore germination and that zoospores are likely partially dependent on maternal mRNA and ribosomes for initial protein production. Launden *et. al.* posited that the chytrid zoospore life stage is optimized for dispersal to new hosts rather than general metabolism [90]. The authors also reported complex lipid dynamics throughout the lifecycle of the chytrid, *Rhizoclosmatium globosum.* Of particular note, they observed increased expression of genes related to lipid transport and metabolism in zoospores, which often have large amounts of intracellular storage lipids. In this study across sporangial stages, we observed an over-enrichment of dispersal and microtubule-based terms as part of patterns with increasing expression, and transcription terms as part of patterns with decreasing expression (Figure 4B). We also observed one pattern of increasing and then decreasing expression related to lipid metabolism (group 7, Figure 4B). Therefore, here we may be discerning the metabolic preparation and production of meiospores for optimal dispersal, survival and host identification.

Later stages of host association are likely characterized by increased immune response in the host and countered by increased stress response by the fungus in order to continue to evade the host immune system. Upregulation of stress response genes has been reported in the plant pathogen *Zymoseptoria tritici* during late stages of sporulation in its host, which the authors posited might be protective [91]. Similarly, we observed an enrichment of terms related to oxidative stress responses in the later stages of sporulation (group 6, Figure 4B), which we speculate may assist *Coelomomyces* evade host defenses during meiospore production.

Mating-type genes may be useful for future work on evolution of sex in fungi

Unlike in animals and plants which have sex-specific chromosomes, sex determination in fungi is regulated by only a handful of genes. These mating-type (*MAT*) loci include HMG-box genes [92]. While *MAT* loci in Dikarya have been widely studied [93], the *MAT* loci of early diverging lineages of fungi have received relatively less attention [94]. Given its obligate sexual two-host alternation of generations life cycle and the ability to separate sexed haploid gametes by color (orange = male; amber = female), *C. lativattatus* provides an intriguing system for investigating the evolution of sex in early diverging fungi. Using the genomes generated here, we identified five putative HMG-box genes, including one gene, HMI56_007461, which was differentially expressed between life stages, and three genes which were in a clade with known Mating-related genes. Additionally, HMI56_006544, which was highly expressed during the sporangial stage, is homologous to the SexM gene of *Phycomyces blakesleeanus.* Interestingly, the three *C. lativitattus* HMG-boxes within the clade containing Mating-related genes were up-regulated in the sporangial stage. Future work should tease apart the role in mating of these putative HMG-box genes in *C. lativattatus* and also place these genes in a comparative framework in order to further investigate the evolution of sex determination in fungi.

Ortholog expansions in C. lativittatus may relate to host-association

Compared to other fungal lineages, *C. lativattatus* genomes had an enriched number of gene copies in four orthologs with functional annotations representing chitin deacetylase, *AMPK1*, *Egh16*-like virulence factor, and *RIF1* (Figure 5). We believe these expanded orthogroups may be good candidates for future investigations elucidating mechanisms behind *C. lativattatus*-host interactions.

In *C. lativattatus,* we found that members of the chitin deacetylase orthogroup were upregulated during sporangial stages and that “chitin synthase activity” was also an over-enriched GO term in group 1, which is a group represented by increasing expression (Figure 4B). Chitin deacetylases catalyze the deacetylation of chitin, an important structural component of fungal cell walls and insect cuticles, and have been previously reported in many fungal species [95]. Chitin-binding proteins and chitin deacetylation are thought to protect fungal pathogens against plant chitinases during infection [96,97], and have also been shown to be upregulated during infection of amphibian hosts by the chytrid pathogen, *Batrachochytrium dendrobatidis* [85]. Further, the *B. dendrobatidis* genome has gene expansions of chitin-binding proteins which can confer protection against chintanse activity when expressed in the fungus, *Trichoderma reeseii [98,99]*. Thus, it is possible that the expansion in the chitin deacetylase orthogroup here may be related to *C. lativattatus* defense against its two hosts.

AMPK1 genes are sensors that modulate energy metabolism and homeostasis, and can be important for regulating stress responses [100]. These genes can also be used to alter host energy metabolism by microbial pathogens [101]. Increased counts of *AMPK1* orthologs and higher expression in sporangial stages in *C. lativattatus* may be related to regulation of increased stress during its two-host life cycle or to evasion of host immune detection.

In *C. lativattatus,* we found that members of the *Egh16*-like expanded orthogroup displayed higher expression during sporangial stages. *Egh16*-like virulence factors are proteins related to appressorium formation and pathogenesis which are present in pathogenic fungi including insect pathogens such as *Metarhizium acridum* [102,103]. *Egh16*-like factors have been postulated to aid in the penetration of insect cuticles [102]. Thus, it is possible that the expanded *Egh16*-like virulence factor orthogroup is contributing to *C. lativattatus* virulence.

RIF1 is important for telomere length control and subtelomeric silencing in fungi and other eukaryotes [104]. Subtelomeric regions have increased variability, caused by duplications and rearrangements that can result in functional novelty, including secondary metabolites related to pathogenicity and virulence. Silencing of subtelomeric regions is one way pathogens can control secondary metabolite expression [105,106]. Given the increased expression of members of the expanded *RIF1* ortholog during infection, these genes may have roles in silencing subtelomeric regions and may be another tool in the *C. lativattatus* toolbox for interacting with its hosts.

Conclusion

We generated three draft genomes and annotations for *C. lativittatus* and characterized the *C. lativittatus* transcriptome landscape across infection and sporangial life stages. Little is known about blastoclads or zoosporic fungi and thus, the genomic and transcriptomic data described here represent a valuable foundational resource. In the transcriptome investigation, we identified differentially expressed transcripts and over-enriched GO terms that provide insight into the blastoclad life cycle, with infection stages marked by an enrichment of metabolism and transport processes and sporangial stages by dispersal-related processes. Further, *C. lativittatus* has an interesting obligate alternation of generations life cycle with two hosts and here we found several ortholog expansions in virulence related genes which may have roles in modulating its host-associated lifecycle. As additional genomic data from other blastoclads and

zoosporic fungi is generated, more powerful comparative approaches can be used to assess the evolutionary relationships of these lineages to other fungi, as well as better understand their complex life histories. We hope the genomic and transcriptomic resources here set the stage for this future work by providing some foundational knowledge of these unique fungi.

Data Availability

The decontaminated *C. lativittatus* genome assemblies and annotation are deposited at DDBJ/ENA/GenBank under the accessions JADGJU000000000, JADGJV000000000 and JADHYY000000000. The versions described in this paper are JADGJU010000000, JADGJV010000000 and JADHYY010000000. The raw sequence reads for the genome assemblies are available through BioProjects PRJNA631428, PRJNA631429, and PRJNA631430. The raw sequence reads for the transcriptome work are available under GEO accession GSE222209. Analysis scripts for this work are available on GitHub (https://github.com/stajichlab/Chytrid_Coelomomyces_Project) and archived in Zenodo (https://doi.org/10.5281/zenodo.7435008).

Acknowledgments

Funding to support sequencing work was provided by a seed grant from the UC Riverside Office of Research and Chancellor's Strategic Investment Funds project "*Coelomomyces Genomics for Mosquito Vector Control*". CLE was supported by funds from California Department of Food and Agriculture agreement # 20–0267. MY was supported by funds from Gordon and Betty Moore Foundation (award #9337). J.E.S. is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and partially supported by USDA Agriculture Experimental Station at the University of California, Riverside and NIFA Hatch projects CA-R-PPA-5062-H. Data analyses performed at the High-Performance Computing Cluster at the University of California Riverside in the Institute of Integrative Genome Biology were supported by NSF grant DBI-

1429826 and NIH grant S10-OD016290. We thank Jericho Ortañez for assistance in experiments for generating genomic DNA and RNA.

References

1. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a Fully Resolved Fungal Tree of Life. Annual Review of Microbiology. 2020. pp. 291–313. doi:10.1146/annurev-micro-022020-051835

2. Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, et al. A genome-scale phylogeny of the kingdom Fungi. Curr Biol. 2021;31: 1653–1665.e5. doi:10.1016/j.cub.2021.01.074

3. James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, et al. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). Mycologia. 2006;98: 860–871. doi:10.3852/mycologia.98.6.860

4. Porter TM, Martin W, James TY, Longcore JE, Gleason FH, Adler PH, et al. Molecular phylogeny of the Blastocladiomycota (Fungi) based on nuclear ribosomal DNA. Fungal Biol. 2011;115: 381–392. doi:10.1016/j.funbio.2011.02.004

5. Whisler HC, Zebold SL, Shemanchuk JA. Life history of *Coelomomyces psorophorae*. Proc Natl Acad Sci U S A. 1975;72: 693–696. doi:10.1073/pnas.72.2.693

6. Couch JN, Bland CE. The Genus *Coelomomyces*. Orlando, FL 32887: Academic Press, Inc; 1985.

7. Shen D, Nyawira KT, Xia A. New discoveries and applications of mosquito fungal pathogens. Current Opinion in Insect Science. 2020. pp. 111–116. doi:10.1016/j.cois.2020.05.003

8. Powell MJ. Blastocladiomycota. Handbook of the Protists. 2017. pp. 1497–1521. doi:10.1007/978-3-319-28149-0_17

9. Federici BA, Chapman HC. *Coelomomyces dodgei*: establishment of an in vivo laboratory culture. J Invertebr Pathol. 1977;30: 288–297. doi:10.1016/0022-2011(77)90136-7

10. Duplessis S, Lorrain C, Petre B, Figueroa M, Dodds PN, Aime MC. Host Adaptation and Virulence in Heteroecious Rust Fungi. Annu Rev Phytopathol. 2021;59: 403–422. doi:10.1146/annurev-phyto-020620-121149

11. Travland LB. Initiation of infection of mosquito larvae (*Culiseta inornata*) by *Coelomomyces psorophorae*. J Invertebr Pathol. 1979;33: 95–105. doi:10.1016/0022-2011(79)90136-8

12. Zebold SL, Whisler HC, Shemanchuk JA, Travland LB. Host specificity and penetration in the mosquito pathogen *Coelomomyces psorophorae*. Can J Bot. 1979;57: 2766–2770. doi:10.1139/b79-327

13. Whisler HC, Karanja DMS, Shemanchuk JA, Zebold SL, Romney SV, Nielsen LT. The life

history and *in vivo* culture of *Coelomomyces utahensis* (Blastocladiomycetes). J Invertebr Pathol. 2009;100: 40–43. doi:10.1016/j.jip.2008.09.009

14. Federici BA. Differential pigmentation in the sexual phase of *Coelomomyces*. 1977;267: 514–515. doi:10.1038/267514a0

15. Federici BA, Thompson SN. β-Carotene in the Gametophytic Phase of *Coelomomyces dodgei*. Experimental Mycology. 1979;3: 281–284. doi:10.1016/S0147-5975(79)80053-5

16. Federici BA. Experimental hybridization of *Coelomomyces dodgei* and *Coelomomyces punctatus*. Proc Natl Acad Sci U S A. 1979;76: 4425–4428. doi:10.1073/pnas.76.9.4425

17. Federici BA. Species-specific gating of gametangial dehiscence as a temporal reproductive isolating mechanism in *Coelomomyces*. Proc Natl Acad Sci U S A. 1983;80: 604–607. doi:10.1073/pnas.80.2.604

18. Stajich J, Palmer J. stajichlab/AAFTF: v0.2.3 release with. 2019 [cited 10 Jun 2022]. doi:10.5281/zenodo.3437300

19. Bushnell B. BBMap: Short read aligner for DNA and RNA-seq data. 2014. Available: https://sourceforge.net/projects/bbmap/

20. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020;70: e102. doi:10.1002/cpbi.102

21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST : architecture and applications. BMC Bioinformatics. 2009. doi:10.1186/1471-2105-10-421

22. Brown CT, Titus Brown C, Irber L. sourmash: a library for MinHash sketching of DNA. The Journal of Open Source Software. 2016. p. 27. doi:10.21105/joss.00027

23. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191

24. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9: e112963. doi:10.1371/journal.pone.0112963

25. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29: 1072–1075. doi:10.1093/bioinformatics/btt086

26. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31: 3210–3212. doi:10.1093/bioinformatics/btv351

27. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. G3 Genes|Genomes|Genetics. 2020. pp. 1361–1374. doi:10.1534/g3.119.400908

28. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18: 366–368. doi:10.1038/s41592-021-01101-x

29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.

Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352

31. Eren AM, Murat Eren A, Kiefl E, Shaiber A, Veseli I, Miller SE, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology. 2021. pp. 3–6. doi:10.1038/s41564-020-00834-3

32. Eren AM, Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015. p. e1319. doi:10.7717/peerj.1319

33. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009. p. R25. doi:10.1186/gb-2009-10-3-r25

34. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11: 119. doi:10.1186/1471-2105-11-119

35. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019;35: 4162–4164. doi:10.1093/bioinformatics/btz188

36. Delmont T. Assessing the completion of eukaryotic bins with anvi'o. 2018. Available: http://merenlab.org/delmont-euk-scgs

37. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195. doi:10.1371/journal.pcbi.1002195

38. Seemann T. barrnap: BAsic Rapid Ribosomal RNA Predictor. Available: https://github.com/tseemann/barrnap

39. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7: 11257. doi:10.1038/ncomms11257

40. Palmer JM, Stajich J. Funannotate v1.8.1: Eukaryotic genome annotation. 2020. Available: http://doi.org/10.5281/zenodo.4054262

41. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117: 9451–9457. doi:10.1073/pnas.1921046117

42. Smit AFA, Hubley R, Green P. RepeatMasker. 2013-2015. Available: http://www.repeatmasker.org

43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29: 644–652. doi:10.1038/nbt.1883

44. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic

Acids Res. 2003;31: 5654–5666. doi:10.1093/nar/gkg770

45. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9: R7. doi:10.1186/gb-2008-9-1-r7

46. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34: W435–9. doi:10.1093/nar/gkl200

47. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Research. 2008. pp. 1979–1990. doi:10.1101/gr.081612.108

48. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20: 2878–2879. doi:10.1093/bioinformatics/bth315

49. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5: 59. doi:10.1186/1471-2105-5-59

50. Testa AC, Hane JK, Ellwood SR, Oliver RP. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics. 2015;16: 170. doi:10.1186/s12864-015-1344-4

51. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25: 955–964. doi:10.1093/nar/25.5.955

52. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42: D222–30. doi:10.1093/nar/gkt1223

53. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42: D490–5. doi:10.1093/nar/gkt1178

54. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 2018;46: D516–D521. doi:10.1093/nar/gkx894

55. Rawlings ND, Barrett AJ, Bateman A. Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. Curr Protoc Bioinformatics. 2014;48: 1.25.1–33. doi:10.1002/0471250953.bi0125s48

56. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44: D286–93. doi:10.1093/nar/gkv1248

57. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031

58. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. Methods Mol Biol. 2016;1374: 23–54. doi:10.1007/978-1-4939-3167-5_2

59. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12: 59–60. doi:10.1038/nmeth.3176

60. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338: 1027–1036. doi:10.1016/j.jmb.2004.03.016

61. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-0036-z

62. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019. p. 1686. doi:10.21105/joss.01686

63. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: https://www.R-project.org/

64. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology. 2019. pp. 907–915. doi:10.1038/s41587-019-0201-4

65. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28: 3150–3152. doi:10.1093/bioinformatics/bts565

66. Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020;18: e3001007. doi:10.1371/journal.pbio.3001007

67. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534. doi:10.1093/molbev/msaa015

68. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nature Communications. 2015. doi:10.1038/ncomms8121

69. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8

71. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available: https://ggplot2.tidyverse.org

72. Morgan M, Falcon S, Gentleman R. GSEABase: Gene set enrichment data structures and methods. 2022. Available: http://bioconductor.org/packages/GSEABase/

73. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6: e21800. doi:10.1371/journal.pone.0021800

74. Pantano L. DEGreport: Report of DEG analysis. 2022. Available: http://lpantano.github.io/DEGreport/

75. Vossenberg BTLH van de, van de Vossenberg BTLH, Warris S, Nguyen HDT, van Gent-Pelzer MPE, Joly DL, et al. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. Scientific Reports. 2019. doi:10.1038/s41598-019-45128-9

76. Gilchrist CLM, Chooi Y-H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics. 2021. doi:10.1093/bioinformatics/btab007

77. Gilchrist CLM, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. Bioinformatics Advances. 2021. doi:10.1093/bioadv/vbab016

78. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

79. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010. p. e9490. doi:10.1371/journal.pone.0009490

80. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

81. Leonard G, Galindo LJ, Milner DS, Avelar GM, Gomes-Vieira AL, Gomes SL, et al. A Genome Sequence Assembly of the Phototactic and Optogenetic Model Fungus Blastocladiella emersonii Reveals a Diversified Nucleotide-Cyclase Repertoire. Genome Biol Evol. 2022;14. doi:10.1093/gbe/evac157

82. Lee SC, Ni M, Li W, Shertz C, Heitman J. The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev. 2010;74: 298–340. doi:10.1128/MMBR.00005-10

83. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014;42: D699–704. doi:10.1093/nar/gkt1183

84. Amses KR, Rabern Simmons D, Longcore JE, Mondo SJ, Seto K, Jerônimo GH, et al. Diploid-dominant life cycles characterize the early evolution of Fungi. Proceedings of the National Academy of Sciences. 2022. doi:10.1073/pnas.2116841119

85. Ellison AR, DiRenzo GV, McDonald CA, Lips KR, Zamudio KR. First in Vivo *Batrachochytrium dendrobatidis* Transcriptomes Reveal Mechanisms of Host Exploitation, Host-Specific Gene Expression, and Expressed Genotype Shifts. G3. 2017;7: 269–278. doi:10.1534/g3.116.035873

86. Vieira ALG, Gomes SL. Global gene expression analysis during sporulation of the aquatic fungus *Blastocladiella emersonii*. Eukaryot Cell. 2010;9: 415–423. doi:10.1128/EC.00312-09

87. Léjohn HB, Lovett JS. Ribonucleic acid and protein synthesis in *Rhizophlyctis rosea*

zoospores. J Bacteriol. 1966;91: 709–717. doi:10.1128/jb.91.2.709-717.1966

88. Lovett JS. Reactivation of ribonucleic acid and protein synthesis during germination of *Blastocladiella* zoospores and the role of the ribosomal nuclear cap. J Bacteriol. 1968;96: 962–969. doi:10.1128/jb.96.4.962-969.1968

89. Schmoyer IR, Lovett JS. Regulation of protein synthesis in zoospores of *Blastocladiella*. J Bacteriol. 1969;100: 854–864. doi:10.1128/jb.100.2.854-864.1969

90. Laundon D, Chrismas N, Bird K, Thomas S, Mock T, Cunliffe M. A cellular and molecular atlas reveals the basis of chytrid development. Elife. 2022;11. doi:10.7554/eLife.73933

91. Keon J, Rudd JJ, Antoniw J, Skinner W, Hargreaves J, Hammond-Kosack K. Metabolic and stress adaptation by *Mycosphaerella graminicola* during sporulation in its host revealed through microarray transcription profiling. Mol Plant Pathol. 2005;6: 527–540. doi:10.1111/j.1364-3703.2005.00304.x

92. Benkhali JA, Coppin E, Brun S, Peraza-Reyes L, Martin T, Dixelius C, et al. A Network of HMG-box Transcription Factors Regulates Sexual Cycle in the Fungus *Podospora anserina*. PLoS Genetics. 2013. p. e1003642. doi:10.1371/journal.pgen.1003642

93. Wallen RM, Perlin MH. An Overview of the Function and Maintenance of Sexual Reproduction in Dikaryotic Fungi. Front Microbiol. 2018;9: 503. doi:10.3389/fmicb.2018.00503

94. Idnurm A, Walton FJ, Floyd A, Heitman J. Identification of the sex genes in an early diverged fungus. Nature. 2008. pp. 193–196. doi:10.1038/nature06453

95. Zhao Y, Park R-D, Muzzarelli RAA. Chitin Deacetylases: Properties and Applications. Marine Drugs. 2010. pp. 24–46. doi:10.3390/md8010024

96. Gueddari NEE, El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. New Phytologist. 2002. pp. 103–112. doi:10.1046/j.1469-8137.2002.00487.x

97. van den Burg HA, Harrison SJ, Joosten MHAJ, Vervoort J, de Wit PJGM. *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Mol Plant Microbe Interact. 2006;19: 1420–1430. doi:10.1094/MPMI-19-1420

98. Liu P, Stajich JE. Characterization of the Carbohydrate Binding Module 18 gene family in the amphibian pathogen *Batrachochytrium dendrobatidis*. Fungal Genet Biol. 2015;77: 31–39. doi:10.1016/j.fgb.2015.03.003

99. Abramyan J, Stajich JE. Species-specific chitin-binding module 18 expansion in the amphibian pathogen *Batrachochytrium dendrobatidis*. MBio. 2012;3: e00150–12. doi:10.1128/mBio.00150-12

100. Hardie DG, Ross FA, Hawley SA. AMPK: a nutrient and energy sensor that maintains energy homeostasis. Nat Rev Mol Cell Biol. 2012;13: 251–262. doi:10.1038/nrm3311

101. Mesquita I, Moreira D, Sampaio-Marques B, Laforge M, Cordeiro-da-Silva A, Ludovico

P, et al. AMPK in Pathogens. Experientia Supplementum. 2016. pp. 287–323. doi:10.1007/978-3-319-43589-3_12

102. Cao Y, Zhu X, Jiao R, Xia Y. The Magas1 gene is involved in pathogenesis by affecting penetration in *Metarhizium acridum*. J Microbiol Biotechnol. 2012;22: 889–893. doi:10.4014/jmb.1111.11055

103. Grell MN, Mouritzen P, Giese H. A *Blumeria graminis* gene family encoding proteins with a C-terminal variable region with homologues in pathogenic fungi. Gene. 2003;311: 181–192. doi:10.1016/s0378-1119(03)00610-3

104. Sreesankar E, Senthilkumar R, Bharathi V, Mishra RK, Mishra K. Functional diversification of yeast telomere associated protein, Rif1, in higher eukaryotes. BMC Genomics. 2012;13: 255. doi:10.1186/1471-2164-13-255

105. Wyatt NA, Richards JK, Brueggeman RS, Friesen TL. A Comparative Genomic Analysis of the Barley Pathogen f. Identifies Subtelomeric Regions as Drivers of Virulence. Mol Plant Microbe Interact. 2020;33: 173–188. doi:10.1094/MPMI-05-19-0128-R

106. Diotti R, Esposito M, Shen CH. Telomeric and Sub-Telomeric Structure and Implications in Fungal Opportunistic Pathogens. Microorganisms. 2021;9. doi:10.3390/microorganisms9071405

Figures/Tables:

Table 1. Genome assembly and annotation statistics. Various statistics calculated by QUAST for each of the assemblies are provided here including the total number of contigs in the assembly, the total assembly length, percent GC content, the N50 and the L50. All statistics from QUAST are based on contigs of size >= 500 bp, unless specifically noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs in each assembly). We also report the percentage of RNAseq reads that were aligned to each assembly using HISAT, the percentage of each assembly masked by RepeatMasker (this value includes total interspersed repeats, simple repeats and low complexity regions), and the number of predicted telomeres. A summary of genome annotation results is reported here including the total number of gene models, mRNAs, tRNAs, the mean exon length (bp), tRNA length (bp), mRNA length (bp), and the percentage of proteins with a PFAM domain, InterProScan, EggNog, COG, or GO term matches.

Software	Statistic	Amber	Orange	Meiospore
QUAST	# contigs (>= 0 bp)	5799	5413	5002
	Largest contig	56469	30249	37025
	Total length (>= 0 bp)	22882397	19891335	21857322
	GC (%)	33.43	32.32	32.33
	N50	6067	5623	6695
	N90	1782	1603	1996
	L50	1089	1030	988
	L90	3884	3610	3254
	# N's per 100 kbp	480.4	362.27	101.38
Telomeres	Telomeres TOTAL	23	18	22
	Telomeres FRWRD	8	9	10
	Telomeres RVRS	15	9	12
HISAT2	Overall RNAseq alignment to genome	55.40%	55.17%	59.01%
RepeatMasker	Repeative regions (%)	10.77%	12.57%	14.04%
Funannotate	Total number of gene models	7404	7344	7677

	Number of mRNA genes	7352	7289	7608
	Number of tRNA genes	52	55	69
	Mean exon length	367.89	358.18	372.63
	Mean tRNA length	74.58	74.62	74.91
	Mean mRNA length	1501.03	1309.02	1465.69
	Genes with PFAM hit (%)	48.74	44.91	48.04
	Genes with InterProScan hit (%)	65.11	63.82	64.91
	Genes with EggNog hit (%)	54.16	52.78	53.82
	Genes with COG hit (%)	53.28	52.03	52.98
	Genes with GO Term (%)	46.83	45.68	46.72

Figure 1. Alternation of generations life cycle of *Coelomomyces lativittatus*. Diagram showing the general alternation of generations life cycle of *C. lativittatus* between copepod and mosquito hosts. We have circled and highlighted the different life stages that were used in this study. Genomic sequencing was performed on haploid stages: including male gametes (orange), female gametes (yellow) and meiospores (purple). RNA sequencing was performed on diploid stages including mosquito larval infection stages (green) as well as sporangial stages excised from mosquito larva (blue).

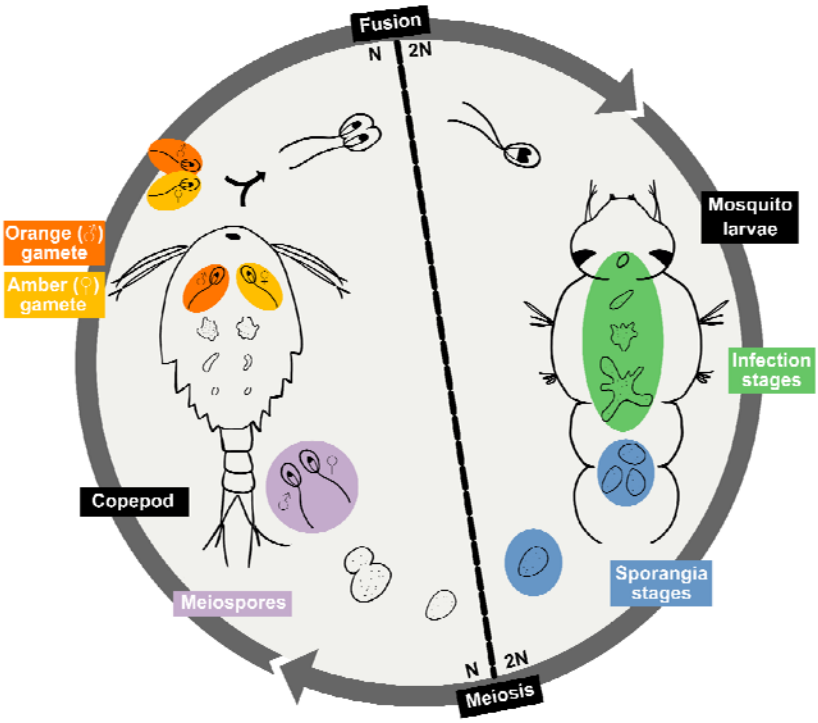


Figure 2. *C. lativittatus* protein set is comparable to those of other fungal taxa. Moving from left to right, here we show a maximum likelihood phylogeny which shows the relationship of *C. lativittatus* to other fungal lineages. This tree was generated using IQ-TREE2 on an alignment of BUSCO fungi_odb10 HMMs constructed using the PHYling_unified pipeline. Taxa labels in the phylogeny are shown colored by assigned fungal phylum. In association with this phylogeny, we then depict a barchart of predicted gene counts for each taxa with counts by fungal lifestyle (saprobe = purple, algal associated = yellow, animal associated = pink). Next, we show barcharts of BUSCO 'protein' completion status for the eukaryota_odb10 and fungal_odb10 sets. Bars show the percent of genes found in each genome annotation as a percentage of the total gene set and are colored by BUSCO status (missing = grey, fragmented = yellow, complete and duplicated = green, complete and single-copy = blue). The values depicted here for *C. lativittatus* gene counts and BUSCO scores are based on the combined clustered transcript set (AOM90). The scores for individual *C. lativittatus* assemblies can be seen in Figure S2.

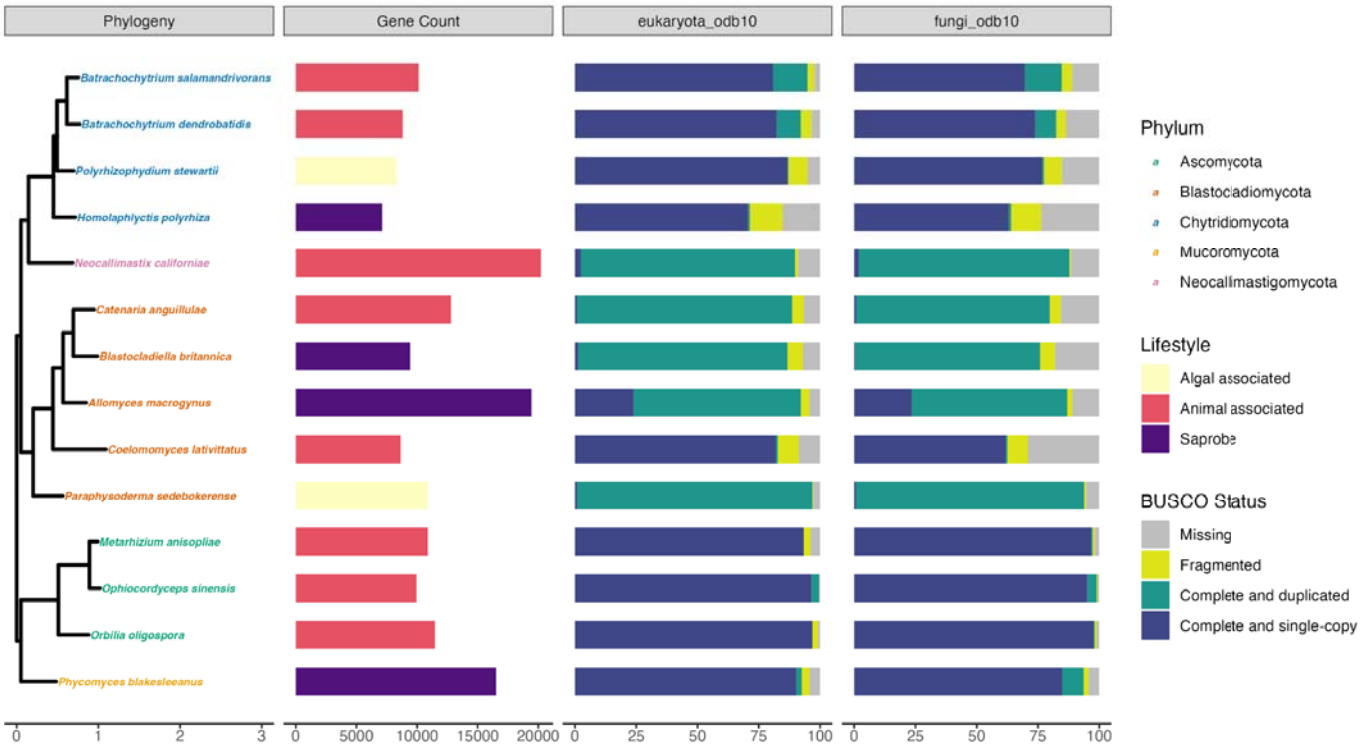


Figure 3. Transcript expression differs between *C. lativittatus* life stages. (A) Principal Component Analysis (PCA) of variance stabilized transcriptomic count data. Samples are colored by time points while shapes are used to broadly represent life stages (circle = infection, triangle = sporangial). Ellipses represent the 95% confidence interval around the centroid of each life stage. (B) Heatmap showing the variance stabilized counts of the 25 most expressed transcripts with differential expression across life stages. Replicates are indicated by ‘A’ or ‘B’.

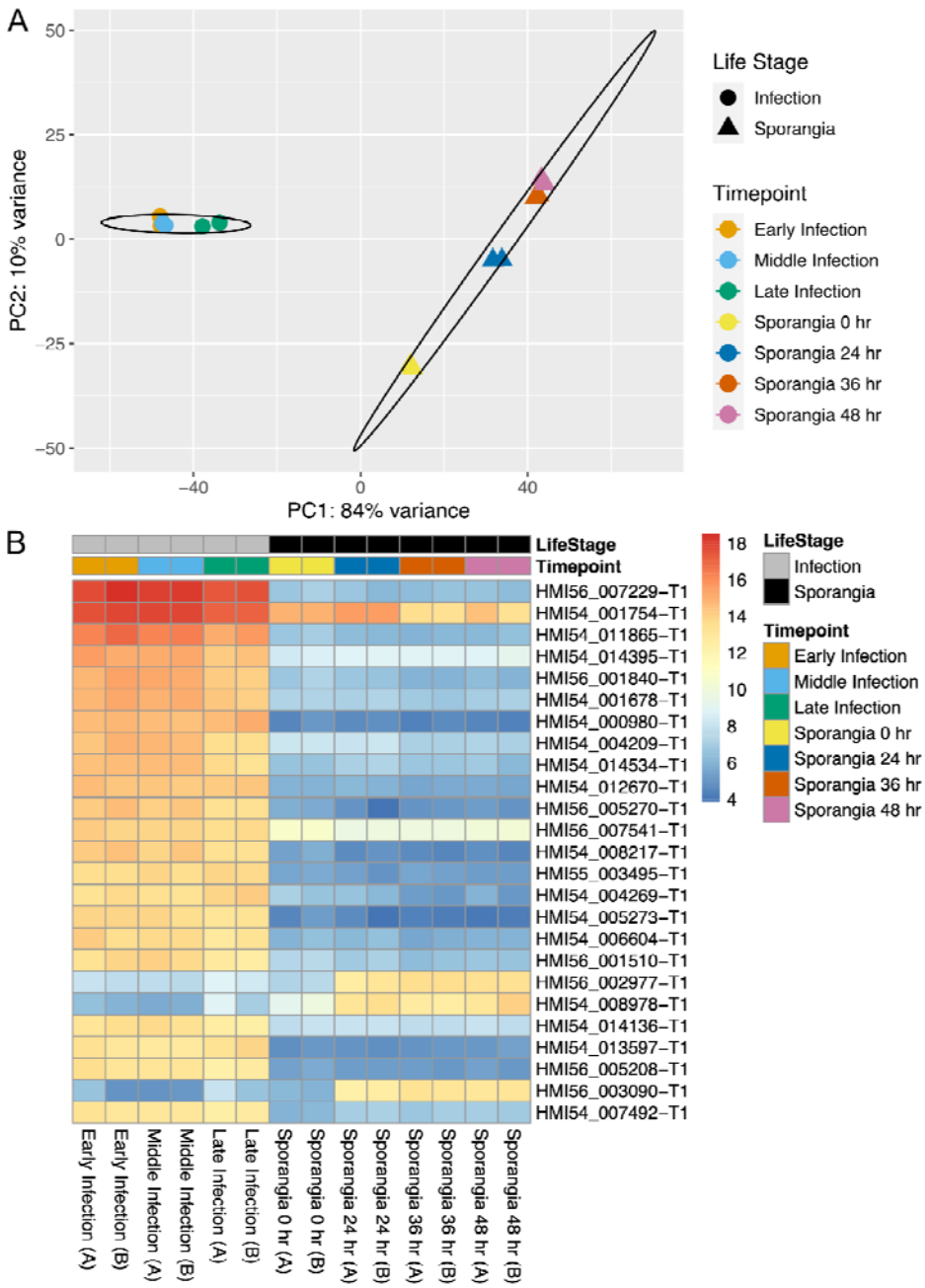
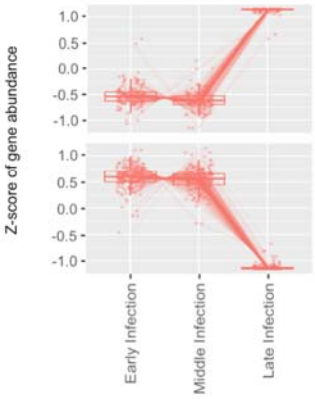

A
PC2: 10% variance
PC1: 84% variance
50
25
0
−25
−50
−40
0
40
Life Stage
Infection
Sporangia
Timepoint
Early Infection
Middle Infection
Late Infection
Sporangia 0 hr
Sporangia 24 hr
Sporangia 36 hr
Sporangia 48 hr
B
LifeStage
Timepoint
HMI56_007229−T1
HMI54_001754−T1
HMI54_011865−T1
HMI54_014395−T1
HMI56_001840−T1
HMI54_001678−T1
HMI54_000980−T1
HMI54_004209−T1
HMI54_014534−T1
HMI54_012670−T1
HMI56_005270−T1
HMI56_007541−T1
HMI54_008217−T1
HMI55_003495−T1
HMI54_004269−T1
HMI54_005273−T1
HMI54_006604−T1
HMI56_001510−T1
HMI56_002977−T1
HMI54_008978−T1
HMI54_014136−T1
HMI54_013597−T1
HMI56_005208−T1
HMI56_003090−T1
HMI54_007492−T1
Early Infection (A)
Early Infection (B)
Middle Infection (A)
Middle Infection (B)
Late Infection (A)
Late Infection (B)
Sporangia 0 hr (A)
Sporangia 0 hr (B)
Sporangia 24 hr (A)
Sporangia 24 hr (B)
Sporangia 36 hr (A)
Sporangia 36 hr (B)
Sporangia 48 hr (A)
Sporangia 48 hr (B)
18
16
14
12
10
8
6
4
LifeStage
Infection
Sporangia
Timepoint
Early Infection
Middle Infection
Late Infection
Sporangia 0 hr
Sporangia 24 hr
Sporangia 36 hr
Sporangia 48 hr

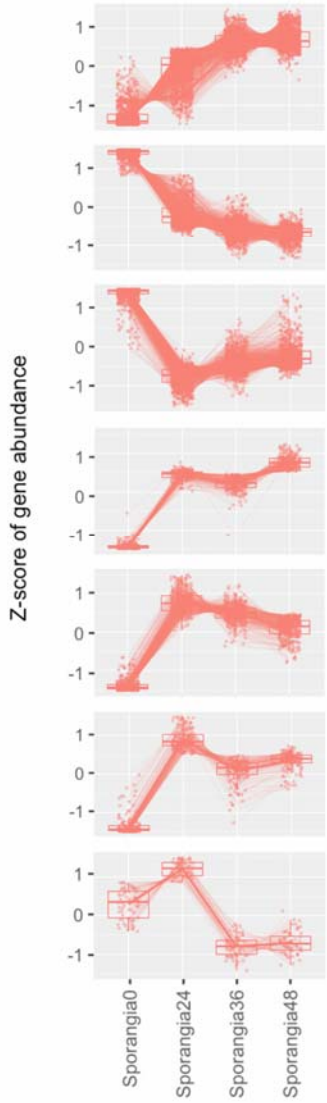
Figure 4. Transcript expression differs across the development time course conditions in each sporangium and infection stage. (A) The plots on the left are the two possible groups with specific transcript expression patterns across early, middle, and late infection timepoints and number of transcripts in each group. The table on the right shows the top five over enriched GO terms for each of the groups of transcripts. (B) The plots on the left are the seven groups with specific transcript expression patterns across 0 hr, 24 hr, 36 hr and 48 hr timepoints within the sporangial life stages. The table on the right shows the top five over enriched GO terms for each of the seven groups of transcripts.

(A)



Group	Over-represented GO term	GO ID
1 167 genes	aspartate family amino acid metabolic process	GO:0009066
	methionine metabolic process	GO:0006555
	alpha-amino acid metabolic process	GO:1901605
	active transmembrane transporter activity	GO:0022804
	sulfur amino acid metabolic process	GO:0000096
2 213 genes	cellular amide metabolic process	GO:0043603
	organonitrogen compound biosynthetic process	GO:1901566
	non-membrane-bounded organelle	GO:0043228
	nucleosome	GO:0000786
	DNA packaging complex	GO:0044815

(B)



Group	Over-represented GO term	GO ID
1 1083 genes	structural constituent of ribosome	GO:0003735
	structural molecule activity	GO:0005198
	chitin synthase activity	GO:0004100
	translation	GO:0006412
	cellular amide metabolic process	GO:0043603
2 965 genes	RNA metabolic process	GO:0016070
	nucleic acid metabolic process	GO:0090304
	nucleobase-containing compound metabolic process	GO:0006139
	gene expression	GO:0010467
	heterocycle metabolic process	GO:0046483
3 859 genes	nucleic acid binding	GO:0003676
	binding	GO:0005488
	RNA binding	GO:0003723
	ribosome biogenesis	GO:0042254
	RNA metabolic process	GO:0016070
4 219 genes	microtubule-based movement	GO:0007018
	movement of cell or subcellular component	GO:0006928
	microtubule-based process	GO:0007017
	microtubule motor activity	GO:0003777
	dynein complex	GO:0030286
5 355 genes	structural constituent of ribosome	GO:0003735
	structural molecule activity	GO:0005198
	7S RNA binding	GO:0008312
	branched-chain-amino-acid transaminase activity	GO:0004084
	glucose-6-phosphate dehydrogenase activity	GO:0004345
6 156 genes	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as	GO:0016620
	oxidoreductase activity, acting on the aldehyde or oxo group of donors	GO:0016903
	response to oxidative stress	GO:0006979
	proteasome assembly	GO:0043248
	proline metabolic process	GO:0006560
7 64 genes	reactive oxygen species metabolic process	GO:0072593
	superoxide metabolic process	GO:0006801
	lipid metabolic process	GO:0006629
	methionyl-tRNA aminoacylation	GO:0006431
	cellular lipid metabolic process	GO:0044255

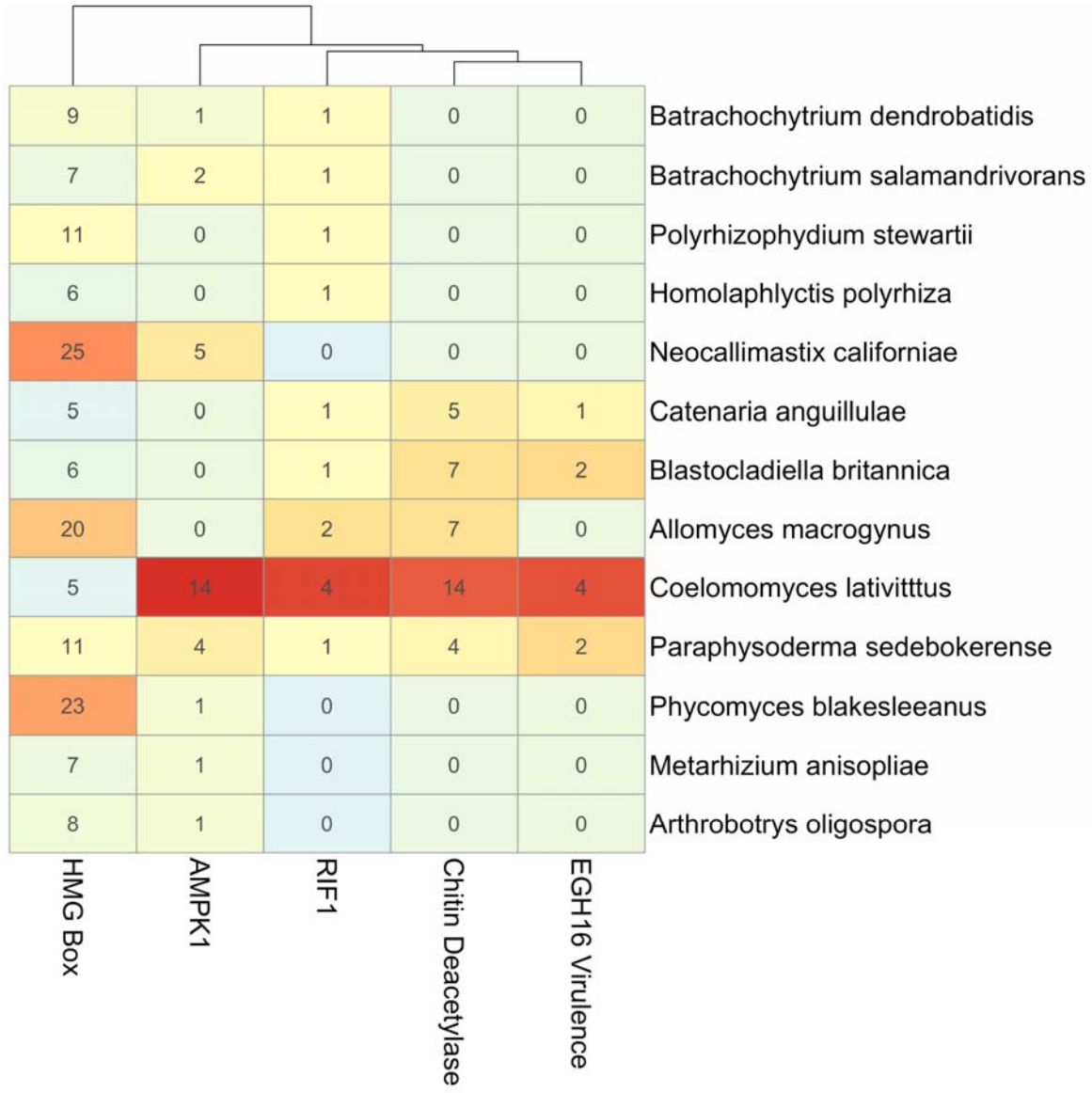



Figure 5. *C. lativittatus* copy number for HMG box, *RIF1*, chitin deacetylase, *Egh16-like*, and *AMPK1* gene orthogroups compared to other fungi. Here we depict a heatmap, organized by fungal phylogenetic relationships, depicting the copy number of orthologous genes, representing differentially expressed transcripts, with expanded gene counts in *C. lativittatus* relative to other fungi. The colors of the gene counts are normalized per gene family with red indicating a high normalized count and blue indicating a low count. Validated HMG boxes were found in two orthogroups which are shown combined here for simplicity.