





# A Genome Sequence Assembly of the Phototactic and Optogenetic Model Fungus *Blastocladiella emersonii* Reveals a Diversified Nucleotide-Cyclase Repertoire

Guy Leonard <sup>1</sup>, Luis Javier Galindo <sup>1</sup>, David S. Milner <sup>1</sup>, Gabriela Mol Avelar<sup>2</sup>, André L. Gomes-Vieira<sup>3</sup>, Suely L. Gomes<sup>4</sup>, and Thomas A. Richards <sup>1,\*</sup>

<sup>1</sup>Department of Biology, University of Oxford, United Kingdom

<sup>2</sup>Division of Molecular Microbiology, School of Life Sciences, University of Dundee, United Kingdom

<sup>3</sup>Departamento de Bioquímica, Instituto de Química, Universidade Federal Rural do Rio de Janeiro, Seropédica, Brazil

<sup>4</sup>Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Brazil

\*Corresponding author: E-mail: thomas.richards@biology.ox.ac.uk.

Accepted: 12 October 2022

## Abstract

The chytrid fungus *Blastocladiella emersonii* produces spores with swimming tails (zoospores); these cells can sense and swim toward light. Interest in this species stems from ongoing efforts to develop *B. emersonii* as a model for understanding the evolution of phototaxis and the molecular cell biology of the associated optogenetic circuits. Here, we report a highly contiguous genome assembly and gene annotation of the *B. emersonii* American Type Culture Collection 22665 strain. We integrate a PacBio long-read library with an Illumina paired-end genomic sequence survey leading to an assembly of 21 contigs totaling 34.27 Mb. Using these data, we assess the diversity of sensory system encoding genes. These analyses identify a rich complement of G-protein-coupled receptors, ion transporters, and nucleotide cyclases, all of which have been diversified by domain recombination and tandem duplication. In many cases, these domain combinations have led to the fusion of a protein domain to a transmembrane domain, tying a putative signaling function to the cell membrane. This pattern is consistent with the diversification of the *B. emersonii* sensory-signaling systems, which likely plays a varied role in the complex life cycle of this fungus.

**Key words:** chytrid, fungi, sensory system, light perception.

## Significance

A high-quality genome sequence is important for the development of *Blastocladiella emersonii* as a model system for understanding phototaxis in eukaryotic microbes. Such data will underpin: 1) the identification of the components of the type I rhodopsin optogenetic circuit, 2) the identification of the wider diversity of sensory systems encoded by this fungus, 3) the exploration of how the optogenetic system is transcriptionally and developmentally regulated, and 4) the understanding of the diversity of the pathways potentially targeted by pharmaco-screens of phenotypic function (such as phototaxis). Here, we use a new, highly contiguous genome assembly to assess the diversity of genes that putatively encode sensation-associated signaling systems, identifying a diversified repertoire of nucleotide cyclase genes.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Introduction

*Blastocladiella emersonii* is a constituent species of the fungal phylum Blastocladiomycota (James et al. 2006; Powell 2017). This “chytrid” fungus has a complex life cycle (fig. 1A–G) that results in the production of swimming spores (zoospores; fig. 1D–G) with a single posterior flagellum (James et al. 2006). Zoospores seek out environments suitable for colonization, triggering a developmental cycle leading to networked growth (through septate hyphae and rhizoids) and the development of reproductive structures called zoosporangia where the flagellated zoospores develop (fig. 1G). The earliest branches of the fungi include two major zoosporic radiations, the Blastocladiomycota and Sanchytriomycota forming one clade, and the Chytridiomycota forming the other (Galindo et al. 2021). Genomic studies of Blastocladiomycota and Chytridiomycota species are limited, with a few key genome data sets across diverse taxa (Ruiz-Trillo et al. 2007; Mondo et al. 2017; Amses et al. 2022). *Blastocladiella emersonii* has been found in soil and freshwater, with the American Type Culture Collection (ATCC) 22665 strain isolated from pond water in Pennsylvania (USA) over 70 years ago (Cantino 1951). We have completed genome sequencing of this strain to high contiguity using PacBio HiFi long-read sequencing. We include a description of *B. emersonii* genome characteristics in comparison with published Blastocladiomycota genomes and an analysis of genes that putatively function in sensory perception.

## Results and Discussion

### Genome Assembly and Assessment

We calculated a SSU-LSU rDNA gene phylogeny of the Blastocladiomycota and confirmed the placement of our sequenced strain, and the paraphyly of the genus *Blastocladiella* (fig. 1H). A *B. emersonii* genome assembly was constructed from PacBio HiFi long-read sequencing (no. of reads: 605,547, mean read length: 10,511.8 bp, and no. of bases: 6,365,413,747), with assembly polishing using Illumina GAllx short-read sequences (no. of paired reads: 17,249,925, mean read length: 65 bp, and total no. of bases: 2,196,251,798). After some manual refinement to remove duplicated contigs, this process resulted in an assembly of 34.27 Mb, smaller in size than published Blastocladiomycota species (see [supplementary table S1, Supplementary Material](#) online) for example; *Allomyces macrogynus* (NCBI: PRJNA20563; 57.06 Mb) and *Catenaria anguillulae* (NCBI: PRJNA330705; 41.34 Mb) (Ruiz-Trillo et al. 2007; Mondo et al. 2017; Amses et al. 2022). The *B. emersonii* genome assembly has an N50 of 2.02 Mb (see fig. 1K). It also showed a high level of contiguity with an L50 of 6 and a total of 21 contigs. Blob-plot analyses (Challis et al. 2020) demonstrate no clear pattern of contamination (fig. 1).

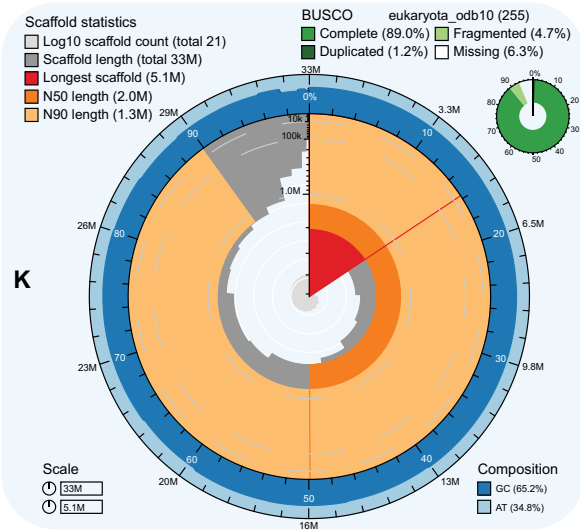
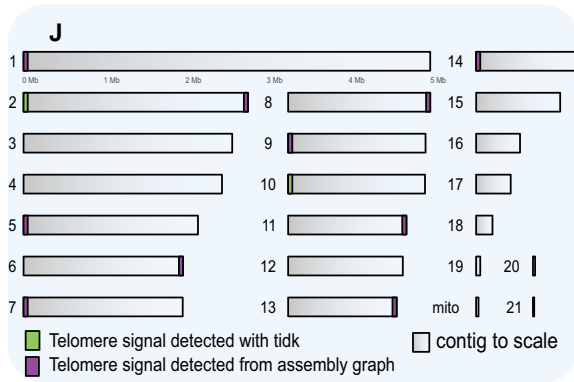
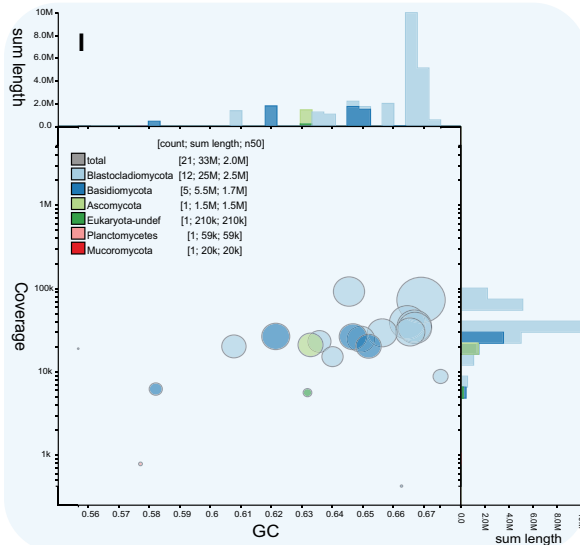
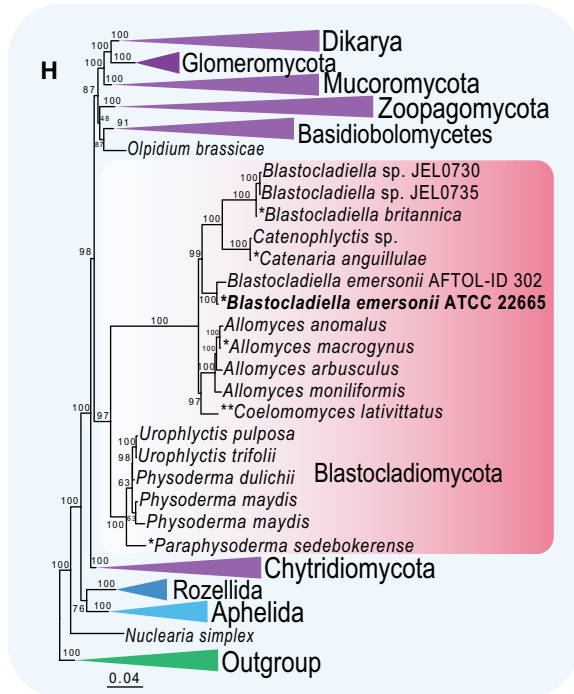
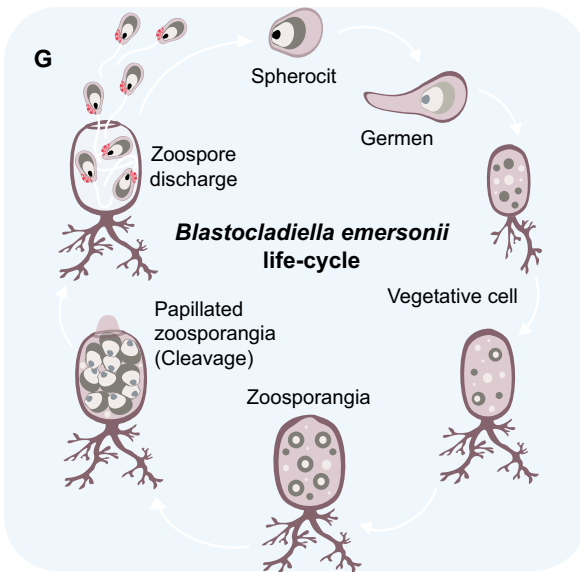
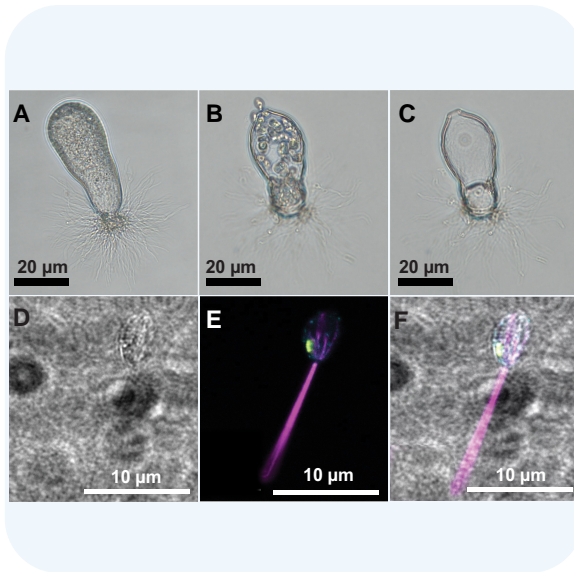
Next, we sought to investigate genome completeness. BUSCO v5.3.2 analyses (Manni et al. 2021) indicated

between 71.3% and 81.8% genome completion using the “Fungi\_ODB10” data set and the metaeuk or augustus options, respectively. This is comparable to the *A. macrogynus* and *C. anguillulae* genome assemblies (72.6–84.5% and 65.4–75.2%). Secondly, we used BLAT to align the 23,370 *B. emersonii* expressed sequence tag (EST) sequences (Ribichich et al. 2006) to the genome assembly. This process identified that 98% of these ESTs mapped to a candidate position in the assembly. Thirdly, we searched the genome assembly for repeats that could potentially represent telomeric regions, using tidk—the Telomere Identification Toolkit (<https://github.com/tolkkit/telomeric-identifier>). This analysis identified a “AAACCT” repeat region on both ends of the c\_be\_22\_2 contig. Further, identification of this motif directly from the assembly graph (<https://github.com/asl/BandageNG> and Davey et al. 2020) allowed recovery of telomere-like motifs on one end of ten additional contigs (fig. 1J). Overall, this suggests a minimum of 6 candidate *B. emersonii* chromosomes in contrast to the 21 contigs identified, suggesting that the genome assembly is at a high level of completion.

Gene annotation, using funannotate v1.8.9 (<https://zenodo.org/record/2604804>), identified 10,031 candidate genes with a gene density of 0.57 per kb; including a total of 29,406 exons, 19,375 introns and 2,438 single exon genes. The 10,031 candidate genes represent a smaller number of predicted genes compared with both the *A. macrogynus* (18,773 genes) and *C. anguillulae* (14,188 genes) genome assemblies. We note that cursory analysis of select gene families has shown that *Allomyces* species have a high number of gene duplications per family (e.g., Swafford and Oakley 2018). This seems to be a wider pattern; notably the BUSCO analysis demonstrates that of the 255 genes surveyed, 152 (60%) and 71 (28%) genes show duplication in *A. macrogynus* and *C. anguillulae* compared with 3 (1.2%) in *B. emersonii*. To investigate this further, we searched all putative genes from *B. emersonii*, *A. macrogynus* and *C. anguillulae* against the Pfam-A v35 database (19,632 total Pfams) using hmmscan (Finn et al. 2011) comparing the average number of hits per Pfam domain. *Blastocladiella emersonii* had a diversity of 7,341 Pfam domains with an average of 4.36 genes per Pfam. In comparison, *Allomyces* possessed a diversity of 8,791 Pfam domains with an average of 7.55 hits per Pfam, while *Catenaria* possessed 7,779 with an average of 4.40 genes per Pfam. This suggests a higher level of gene duplication present in *Allomyces*, indicative of concerted patterns of genome expansion (possibly due to genome duplication/ploidy events [Amses et al. 2022]) that are absent or lost in *B. emersonii*.

### Survey of Sensory System Genes

*Allomyces reticulatus* was the first Blastocladiomycota fungus shown to manifest phototaxis associated with rhodopsin



Downloaded from https://academic.oup.com/gbe/article/14/1/2/evac157/6770634 by Universidad de Granada - Biblioteca user on 24 January 2025

functions (Saranak and Foster 1997). A unique organellar structure composed primarily of lipid droplets called the side-body complex was suggested to function as a light-perception organelle system in Blastocladiomycota zoospores (Chambers et al. 1967; Kazama 1972; Powell 1978; Saranak and Foster 1997). In *B. emersonii*, we identified an optogenetic circuit composed of a type I (microbial) rhodopsin domain fused to a guanylyl cyclase (GC) catalytic domain (NCBI: AIC07007.1) with this protein localized to the external surface of the lipid-filled side-body complex (Avelar et al. 2014). This gene fusion, named BeGC1 (Avelar et al. 2014) (or CyclOp [Gao et al. 2015]), is particularly interesting because it couples the light-sensing rhodopsin domain to an enzyme domain which generates a chemical message, cyclic guanosine monophosphate (cGMP), to activate flagellum beating (swimming). As such, this single gene/protein contains multiple key constituents of an optogenetic circuit (Gao et al. 2015; Scheib et al. 2015; Kumar et al. 2017; Trieu et al. 2017). We have recently shown that this optogenetic circuit and the subcellular lipid organellar structures are widespread in chytrid fungi, implying the ancestral fungus possessed this light-sensing system (Galindo et al. 2022). However, we note that within the Blastocladiomycota this gene has been subject to considerable duplications; for example, *A. macrogynus* possesses four paralogues (fig. 2A and B), *A. reticulatus* two, and *Allomyces arbusculus* a single paralogue with a curtailed GC catalytic domain (Swafford and Oakley 2018).

In the absence of reliable forward or reverse-genetic methods for *B. emersonii*, Avelar et al. (2014) explored the role of BeGC1 function using a range of drug screens to inhibit the type I rhodopsin and the GC catalytic domain functions. These experiments provided evidence that the BeGC1 gene was functioning in the light-sensing cascade. Using the genome assembly reported here we used BLAST and Pfam sequence similarity searches to identify the repertoire of variant type I rhodopsin, type II rhodopsin and nucleotide cyclase (i.e., GC and adenylyl cyclase [AC]) domain-encoding genes. This process identified a complex diversity of these gene families. We used OrthoFinder (Emms and Kelly 2019) to identify how these genes putatively cluster into distinct paralogue families. This process identified one type I rhodopsin gene: the fusion gene that

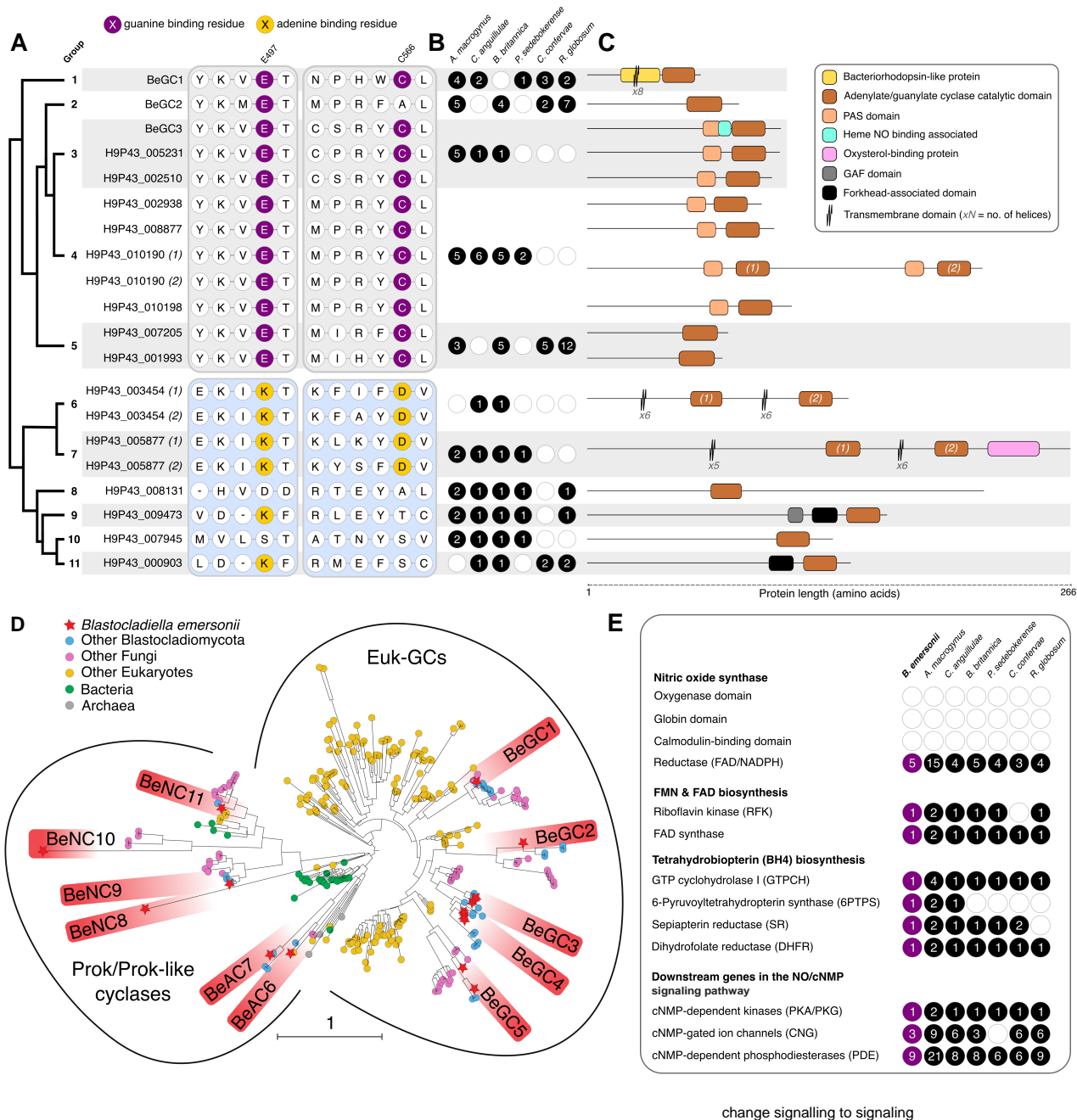
encodes the BeGC1 (H9P43\_007630) protein, and eight putative orthogroups (encapsulating 54 genes) predicted to encode proteins of the seven transmembrane (TM) G-protein-coupled receptor (GPCR) domain protein family. This gene family includes type II rhodopsin, along with a range of GPCRs responsible for a diversity of functions (Pfam: PF00001, InterPro: IPR000276—the eight OrthoFinder-clusters are detailed in [supplementary table S2, Supplementary Material](#) online) and, in *B. emersonii*, was found to possess a range of gene architecture variations and tandem duplicated forms ([supplementary table S2, Supplementary Material](#) online).

We also identified 17 genes which encode nucleotide cyclase domains (Pfam: PF00211, InterPro: IPR001054) across 11 different “deeply divergent” cluster groups (fig. 2A–D and [supplementary table S8, Supplementary Material](#) online for OrthoFinder-clusters). Of these 17 genes, 11 are predicted to encode GCs based on conserved amino acid motifs; these GCs catalyze the conversion of guanosine triphosphate to cGMP, and include the previously identified BeGC1, BeGC2 (H9P43\_007348) and BeGC3 (H9P43\_009044) genes (Vieira et al. 2009). The remainder are predicted to encode adenylyl cyclases, which catalyze the conversion of adenosine triphosphate to cyclic adenosine monophosphate, or have no recognizable motifs, so no specific function has been assigned. Interestingly, the ACs all belong to a group which consists of bacterial-like sequences (groups 6–11; fig. 2D) and include multiple forms that consist of a tandemly duplicated AC domain and multiple TM domain regions. Intriguingly, levels of membrane-associated AC proteins, such as those in groups 6 and 7 (fig. 2C), are known to increase during the formation and release of *B. emersonii* zoospores (Gomes et al. 1978), suggesting that these ACs play roles in specific *B. emersonii* life cycle stages.

Nitric oxide (NO) signaling tied to GC function has been demonstrated to control specific movement behaviors in some choanoflagellate protists (Reyes-Rivera et al. 2022). We note that although the *B. emersonii* genome does not contain a predicted NO synthase (NOS) gene (fig. 2E), GC activity tied to NO signaling has been linked to zoospore biogenesis as part of the *B. emersonii* life cycle (Vieira et al. 2009). The nucleotide cyclases identified here are diverse across the Blastocladiomycota and Chytridiomycota sampled

**Fig. 1.**—*Blastocladiella emersonii* life cycle and genome assembly information. (A) Bright field micrograph depicts early zoosporangium showing rhizoid filamentous feeding structures arising from the base. (B) Late zoosporangium showing the release of zoospores. (C) Finished zoosporangium with all zoospores released. (D) Brightfield phase-contrast micrograph of a zoospore. (E) Fluorescent microscopy of the zoospore in D with tubulin structures stained with magenta, actin stained with cyan, and the lipid body proposed to function as a light spot stained with yellow. (F) Merged image of E and D. Microscopy images were taken and stained using the methods described in Galindo et al. (2022). (G) Cartoon of the asexual life cycle of the *B. emersonii* adapted from Bongiorno et al. (2012). (H) SSU-LSU rDNA phylogeny of the Blastocladiomycota and other fungi showing the relative position of *B. emersonii*. One asterisk (\*) next to the species names indicates an available genome, and two asterisks (\*\*) indicate a genome with uncertain availability. (I) Blob-plot analysis of the *B. emersonii* genome assembly showing no evident patterns of contamination. (J) Contig map showing a distribution of putative telomere sequence motifs. (K) Snail plot report summarizing assembly data.





**FIG. 2.**—Nucleotide cyclase and NO signaling pathway component analysis. (A) Amino acid sequence conservation of YKVE and MPRYCL motifs (described by Vieira et al. 2009) for each *B. emersonii* GC. Eleven cluster groups (i.e., orthogroups identified using OrthoFinder) were identified, with groups 1–5 predicted to encode GCs, and groups 6 and 7 predicted to encode adenylyl cyclases. Key guanine (residues E497 and C566 in BeGC1; accession AIC07007.1) and adenine binding residues are indicated. Protein sequences with two predicted cyclase domains have been split into (1) and (2). (B) Presence (black fill) or absence (white fill) of each nucleotide cyclase cluster-group in four Blastocladiomycota and two Chytridiomycota species, demonstrating mosaic phylogenetic distribution across these clades. Numbers denote the number of sequences found within each orthogroup. (C) PFAM domain architecture of each *B. emersonii* nucleotide cyclase, showing extensive variation in domain arrangement. Transmembrane domains were predicted using TOPCONS. (D) Phylogeny of all *B. emersonii* nucleotide cyclases, clustered into the eukaryotic GCs “Euk-GCs” and the cluster containing the prokaryote cyclases “Prok/Prok-like cyclases,” which also contains the predicted *B. emersonii* adenylyl cyclases (BeAC6–7). (E) Presence (solid fill) or absence (no fill) of NOS domains, NOS cofactors (FMN/FAD/BH4), and downstream signaling pathway components (Reyes-Rivera et al. 2022), in *B. emersonii*, four Blastocladiomycota species, and two Chytridiomycota. Numbers denote the number of sequences identified for each orthogroup.

(fig. 2B) and the group 3 forms include BeGC3, which contains a protein domain architecture with a Heme/NO binding-associated (HNOBA) domain (fig. 2C). This indicates that the GC activity of BeGC3 is likely to be regulated by NO signaling. Consistent with this finding, *B. emersonii* encodes all the genes for NOS cofactors and downstream signaling pathways (fig. 2E and [supplementary table S7, Supplementary Material](#) online for OrthoFinder-clusters), demonstrating the possibility of NO signaling in *B. emersonii* physiological processes. However, an orthologue of BeGC3 with a HNOBA domain is absent in the Chytridiomycota *Chytrium confervae* (van de Vossenberg et al. 2019) and *Rhizoclosmatium globosum* (Mondo et al. 2017) genome assemblies (fig. 2B), suggesting that there are variant responses to NO across the chytrids.

Finally, to explore what other alternative sensing systems may be encoded by the *B. emersonii* genome, we searched the assembly for a curated selection of genes known to encode proteins that function in light response, other sensory cascades, and/or associated signal transduction processing ([supplementary table S2, Supplementary Material](#) online). The *B. emersonii* genome assembly identified two putative photolyases (H9P43\_000523-T1 and H9P43\_001129-T1) ([supplementary tables S2 and S9, Supplementary Material](#) online for OrthoFinder-clusters), a single putative large-conductance mechanosensitive channel gene (H9P43\_007225-T1; [supplementary table S2, Supplementary Material](#) online) and a wide diversity of putative ion transporters ([supplementary table S2, Supplementary Material](#) online) and other GPCRs ([supplementary table S2, Supplementary Material](#) online), again with multiple gene architectural forms.

## Conclusion

Here we report a highly contiguous genome assembly of *B. emersonii*, a Blastocladiomycota fungus with a complex multimodal life cycle. Using the genome data, we identify a diversity of genes encoding cyclic nucleotide-based signal sensory systems. Our results show that these systems have been consistently diversified by domain recombination, tandem duplication and coupling of TM protein domains, effectively tying signaling functions together and, in many cases, to the cell membrane. This pattern is consistent with the diversification of the sensory repertoire of *B. emersonii*. These results are also consistent with the idea that domain architectural recombination, duplication and gene fusion has been a factor, as demonstrated for fungal gene repertoire evolution (Leonard and Richards 2012) and the wider evolution of protein families (Doolittle 1995) including signaling systems. These features also appear to be highly variant across the chytrid fungi sampled suggesting sensory diversification is tied to species diversification and probably speciation (Swafford and Oakley 2018).

## Materials and Methods

### *Blastocladiella emersonii* Culture Preparation and DNA Extraction

Cultures of *B. emersonii* ATCC 22665 were grown vegetatively in Nunc EasYFlask 25 cm<sup>2</sup> flasks (ThermoFisher) filled with 25 ml of peptone yeast glucose liquid media. Zoospore production was initiated as previously described (Galindo et al. 2022). Zoospores were then collected by centrifugation at 1,000×g for 5 mins followed by removal of supernatant. The remaining pellet was used for DNA extraction using the MagAttract HMW DNA Kit (QIAGEN) following the manufacturer's protocol. Extracted DNA (1.7 ng) was quantified by Qubit-fluorometer (ThermoFisher). Quality control on the DNA sample was performed with a 4,200 TapeStation system (Agilent), confirming the presence of 120–250 kb DNA fragments.

### Long-Read Sequencing Using Pacific Biosciences (PacBio™) Methods

The PacBio™ low-input, high-molecular weight DNA sample method was used. Fragmentation was conducted using g-TUBEs (Covaris) with a target size of 15 kb but recovering a mean fragment size of 13.7 kb. Library preparation was completed using the Express TPK 2.0 and SMRTbell Enzyme Clean-up kit v1 (PacBio™). The Sequel II Binding Kit 2.2 (PacBio™) and Sequencing Primer v5 (PacBio™) along with the Sequel II Sequencing Plate v2.0 (PacBio™) was used. Sequencing was performed using the SMRT Link software version 10.2.0.133434 on a Sequel IIe PacBio device. The “Run Design Application” was set to HiFi Reads with default settings. The resulting PacBio HiFi long-read library comprised 605,793 reads; mean read length: 10,511.8; and 6,365,413,747 bases.

### Short-Read DNA Sequencing Using an Illumina™ Method

Genomic DNA was extracted from *B. emersonii* ATCC 22665 using methods previously reported in (Avelar et al. 2014). An Illumina GAlx paired-end 75 bp library was prepared and resulted in: number of paired reads: 17,249,925; mean read length: 65 bp; and total number of bases: 2,196,251,798.

### Long-Read Genome Assembly

The HiFi long reads were filtered using “HiFiAdapterFilter” to remove 246 adapter contaminated reads resulting in 605,547 reads. Using the La Jolla Assembler (Bankevich et al. 2020), we generated an assembly of 247 contigs and an N50 of 2,030,682 bp and then used the workflow “purge\_dups” ([https://github.com/dfguan/purge\\_dups](https://github.com/dfguan/purge_dups)) to remove haplotigs and contig overlaps based on read depth. Further contigs were also removed with minimap2

detecting overlaps based on percent identity (60%) and coverage (10%). On completion this resulted in a “purged” assembly, see [supplementary table S1, Supplementary Material](#) online. One contig represented an exact copy of the 36,503 bp mitochondrion (Tambor et al. 2008) (NCBI: NC\_011360.1) with 8,400x coverage (44x higher than the mean assembly). The genome polishing tool Pilon (Bankevich et al. 2020) was also used, in conjunction with the Illumina short reads, to help “fix” small indels, homopolymers, single nucleotide polymorphism (SNPs), and ambiguous base calls.

### Genome Analysis and Annotation

Annotation of the PacBio HiFi genome assembly was conducted with v 1.8.9 of the pipeline FUNANNOTATE (<https://github.com/nextgenusfs/funannotate>). Genes were predicted using BUSCO (Manni et al. 2021), AUGUSTUS, SNAP (<https://github.com/KorfLab/SNAP>), GlimmerHMM (Majoros et al. 2004), and GeneMark-ES (Ter-Hovhannisyan et al. 2008) and were annotated using several databases (PFAM [Mistry et al. 2021], EGGNOG [Huerta-Cepas et al. 2019], BUSCO, Phobius [Madeira et al. 2022], antiSMASH [Blin et al. 2021]).

The predicted proteins of *A. macrogynus*, *C. anguillulae*, *Bougainvillia britannica*, *Paraphysoderma sedebokerense*, *B. emersonii*, *C. confervae*, and *R. globosum* were subjected to OrthoFinder analysis (default settings) (Emms and Kelly 2019). Cluster groups (i.e., putative OrthoFinder-orthogroups) were identified and are available at <https://github.com/guyleonard/blastocladiella>.

### Identification of Signaling Genes of Interest

Candidate genes of interest were recovered using BLASTp (Altschul et al. 1990) and/or PFAM searches of the *B. emersonii* predicted proteome. Candidate homologues were investigated using reciprocal BLASTp searches of GenBank “nr” database and/or HMMSCAN (Finn et al. 2011) searches of the PFAM database. HMMSCAN searches were also used to confirm protein domain architectures and assess for the presence of TM domains and signal peptides ([supplementary table S2, Supplementary Material](#) online). TOPCONS was used to predict TM domains for nucleotide cyclases (Tsirigos et al. 2015).

### rRNA Gene Phylogenetic Analysis

We used an updated SSU + LSU rDNA alignment (Karpov et al. 2018) adding sequences from Blastocladiomycota taxa, including sequences from the genera *Paraphysoderma*, *Physoderma*, *Urophlyctis*, *Coelomomyces*, *Allomyces*, *Blastocladiella*, *Catenaria*, and *Catenophlyctis*. The new data set of 215 sequences was aligned with MAFFT (Katoh and Standley 2013) and trimmed with TrimAl (Capella-Gutiérrez et al. 2009) with the “automated1” option. For the maximum

likelihood phylogenetic analyses, we selected the best-fitting model (GTR + F + R7) using IQ-TREE (Nguyen et al. 2015) MFT algorithm with BIC. Topology support was evaluated using 1,000 ultrafast bootstrap replicates and 1,000 replicates of the SH-like approximate likelihood ratio test. Trees were drawn using FigTree (Rambaut 2010).

For the phylogenetic analysis of *B. emersonii*'s nucleotide cyclase gene family, we created a multiple sequence alignment (MSA) of eukaryotic and prokaryotic nucleotide cyclases by merging previously published data sets (Avelar et al. 2014; Reyes-Rivera et al. 2022) together with 17 additional sequences found in *Be* encoding putative homologues of nucleotide cyclase domains. We also included BLASTp searches of these additional 17 sequences against the GenBank non-redundant database and recovered the first 50 sequences. We manually removed redundant sequences and aligned the sequences using MAFFT (Katoh and Standley 2013) followed by trimming using TrimAl (Capella-Gutiérrez et al. 2009) with the “automated1” option producing an alignment of 303 sequences and 120 nucleotide positions. Best-fitting phylogenetic models were selected with the IQ-TREE (Nguyen et al. 2015) MFT algorithm as per BIC, obtaining the LG + R7 model as the best. Statistical support was evaluated using 1,000 ultrafast bootstrap replicates and 1,000 replicates of the SH-like approximate likelihood ratio test, and the resulting trees were visualized with FigTree (Rambaut 2010).

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

We wish to thank the Oxford Genomics Centre for rapid and fulsome support. This project was initially supported by a BBSRC award (BB/G00885X/2). L.J.G. is supported by a Marie Skłodowska-Curie Individual Fellowship H2020-MSCA-IF-2020 (grant agreement no. 101022101—FungEye). T.A.R. is supported by a Royal Society URF (URF/R/191005). S.L.G. was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

### Data Availability

The sequence reads and assembly and annotations have been deposited in NCBI GenBank: BioProject: PRJNA194096, PacBio: SRR19593111, and Illumina: SRR12507012, or alternatively, the genome annotations and other data can be accessed here: <https://github.com/guyleonard/blastocladiella> and/or here: <https://doi.org/10.6084/m9.figshare.c.6221042>. The *B. emersonii* genomic assembly is available at NCBI nuccore JANHCV000000000 and the predicted protein set is available from NCBI protein

KAI9148506 to KAI9188641. All MSAs, trees, and the sequences used in our alignments can be found at Figshare (<https://doi.org/10.6084/m9.figshare.c.6221042>).

## Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Amses KR, et al. 2022. Diploid-dominant life cycles characterize the early evolution of Fungi. *Proc Natl Acad Sci U S A.* 119: e2116841119.
- Avelar GM, et al. 2014. A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. *Curr Biol.* 24:1234–1240.
- Bankevich A, Bzikadze A, Kolmogorov M, Antipov D, Pevzner PA. 2020. Assembling long accurate reads using de Bruijn graphs. *bioRxiv:2020.2012.2010.420448*.
- Blin K, Shaw S, Kautsar S, Medema MH, Weber T. 2021. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* 49:D639–D643.
- Bongiorno VA, Ferreira da Cruz A, Ferreira da Silva A, Corrêa LC. 2012. Phosphate limitation induces sporulation in the chytridiomycete *Blastocladiella emersonii*. *Can J Microbiol.* 58:1104–1111.
- Cantino EC. 1951. Metabolism and morphogenesis in a new *Blastocladiella*. *Antonie van Leeuwenhoek.* 17:325–362.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda)* 10:1361–1374.
- Chambers TC, Markus K, Willoughby LG. 1967. The fine structure of the mature zoosporangium of *Nowakowskiella profusa*. *Microbiology* 46:135–141.
- Davey JW, Davis SJ, Mottram JC, Ashton PD. 2020. Tapestry: validate and edit small eukaryotic genome assemblies with long reads. *bioRxiv:2020.2004.2024.059402*.
- Doolittle RF. 1995. The multiplicity of domains in proteins. *Annu Rev Biochem.* 64:287–314.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39:W29–W37.
- Galindo LJ, López-García P, Torruella G, Karpov S, Moreira D. 2021. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. *Nat Commun.* 12:4973.
- Galindo LJ, Milner DS, Gomes SL, Richards TA. 2022. A light-sensing system in the common ancestor of the fungi. *Curr Biol.* 32(14):3146–3153.e3
- Gao S, et al. 2015. Optogenetic manipulation of cGMP in cells and animals by the tightly light-regulated guanylyl-cyclase opsin CyclOp. *Nat Commun.* 6:8046.
- Gomes SL, Mennucci L, Da Costa Maia J. 1978. Adenylate cyclase activity and cyclic AMP metabolism during cytodifferentiation of *Blastocladiella emersonii*. *Biochim Biophys Acta.* 541:190–198.
- Huerta-Cepas J, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47: D309–D314.
- James TY, et al. 2006. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* 98:860–871.
- Karpov SA, et al. 2018. The chytrid-like parasites of algae *Amoeboradix gromovi* gen. et sp. nov. and *Sanchytrium tribonematis* belong to a new fungal lineage. *Protist* 169:122–140.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kazama FY. 1972. Ultrastructure and phototaxis of the zoospores of *Phlyctochytrium* sp., an estuarine chytrid. *Microbiology* 71: 555–566.
- Kumar RP, et al. 2017. Structure and monomer/dimer equilibrium for the guanylyl cyclase domain of the optogenetics protein RhoGC. *J. Cell Biol.* 292:21578–21589.
- Leonard G, Richards TA. 2012. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc Natl Acad Sci U S A.* 109:21402–21407.
- Madeira F, et al. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50: W276–W279.
- Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878–2879.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 38: 4647–4654.
- Mistry J, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49:D412–D419.
- Mondo SJ, et al. 2017. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet.* 49:964–968.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.
- Powell MJ. 1978. Phylogenetic implications of the microbody-lipid globule complex in zoosporic fungi. *Biosystems* 10:167–180.
- Powell MJ. 2017. Blastocladiomycota. In: Archibald JM, Simpson AGB, Slamovits CH, editors. *Handbook of the protists*. Cham: Springer. p. 1497–1521.
- Rambaut A. 2010. Figtree v1.3.1. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh.
- Reyes-Rivera J, et al. 2022. Nitric oxide signaling controls collective contractions in a colonial choanoflagellate. *Curr Biol.* 32:2539–2547.e5.
- Ribichich KF, Georg RC, Gomes SL. 2006. Comparative EST analysis provides insights into the basal aquatic fungus *Blastocladiella emersonii*. *BMC Genomics* 7:177.
- Ruiz-Trillo I, et al. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet.* 23:113–118.
- Saranak J, Foster KW. 1997. Rhodopsin guides fungal phototaxis. *Nature* 387:465–466.
- Scheib U, et al. 2015. The rhodopsin-guanylyl cyclase of the aquatic fungus *Blastocladiella emersonii* enables fast optical control of cGMP signaling. *Sci Signal.* 8:rs8.
- Swafford AJM, Oakley TH. 2018. Multimodal sensorimotor system in unicellular zoospores of a fungus. *J Exp Biol.* 221:jeb163196.
- Tambor JH, Ribichich KF, Gomes SL. 2008. The mitochondrial view of *Blastocladiella emersonii*. *Gene* 424:33–39.
- Ter-Hovhannisyann V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18: 1979–1990.
- Trieu MM, et al. 2017. Expression, purification, and spectral tuning of RhoGC, a retinylidene/guanylyl cyclase fusion protein and optogenetics tool from the aquatic fungus *Blastocladiella emersonii*. *J Biol Chem.* 292:10379–10389.



Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 43:W401–W407.

van de Vossenberg BTLH, et al. 2019. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. *Sci Rep.* 9:8672.

Vieira ALG, Linares E, Augusto O, Gomes SL. 2009. Evidence of a  $\text{Ca}^{2+}$ -NO-cGMP signaling pathway controlling zoospore biogenesis in the aquatic fungus *Blastocladiella emersonii*. *Fungal Genet Biol.* 46:575–584.

**Associate editor:** Jason Stajich