**Understanding the evolution of parasitic oomycete and fungal-like traits through genome analysis of the free-living pseudofungus *Hyphochytrium catenoides***

Guy Leonard1, Aurélie Labarre1, David Milner1, Adam Monier1, Darren Soanes1, Jeremy G. Wideman1, Finlay Maguire1, Sam Stevens1, Divya Sain2, Xavier Grau-Bove3, Arnau Sebe-Pedros4, Jason E. Stajich2, Konrad Paszkiewicz1, Matthew W. Brown5, Neil Hall6, Bill Wickstead7 & Thomas A. Richards1

1 Living Systems Institute, Department of Biosciences, University of Exeter, Exeter, EX4 4QD, UK.

2 UC Riverside

3 Institute of Evolutionary Biology. CSIC-UPF. Barcelona, Catalonia. Spain.

4 Weizman Institute of Science. Rehovot, Israel.

5 Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762, USA.

6 Earlham Institute, Norwich. Aha!

7 School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH, UK

Corresponding Author: Thomas Richards

Living Systems Institute,

Biosciences,

University of Exeter,

Geoffrey Pope Building,

Exeter, EX4 4QD, UK.

Running Title: **genome sequence of the ‘Pseudofungus’ *Hyphochytrium catenoides***

Keywords: Pseudofungi, polarised filamentous growth, large DNA virus, oomycete pathogenic traits, plastid.

**Abstract**

Eukaryotic microbes have three primary mechanisms for obtaining nutrients and energy: phagotrophy, photosynthesis and osmotrophy. Traits associated with the latter two functions arose independently multiple times in the eukaryotes. The Fungi successfully coupled osmotrophic functions with filamentous growth, similar traits are also manifest in the so-called Pseudofungi (oomycetes and hyphochytriomycetes). Both the Fungi and the Pseudofungi encompass a wide diversity of parasites, including pathogens of plants and animals. Genome sequencing efforts in both groups have largely focused on host-associated microbes (mutualistic symbionts or parasites), providing limited comparisons with free-living relatives. Here we report the draft genome sequence of *Hyphochytrium catenoides*;the first genome sequence for a free-living hyphochytriomycete ‘pseudofungus’. This protist has a complex lifecycle involving diverse filamentous structures and a flagellated zoospore with a single atypical anterior tinselated flagellum. Phylogenomic analysis confirms that *Hyphochytrium* branches sister to the oomycetes. We identify a ‘footprint’ of genes that plausibly have a phylogenetic ancestry arising from a plastid endosymbiotic event, leaving the case for, or against, plastid ancestry in the Pseudofungi controversial. The *H. catenoides* genome assembly also includes a set of genes of recent ancestry from a virus, with examples of viral derived genes confirmed as expressed. Related viral derived genes are also present on the genomes of a number of oomycetes, suggesting a complex history of viral co-evolution and integration across the Pseudofungi.  We use the genomic data, drug sensitivity studies and high-throughput culture arrays to investigate the ancestry of oomycete/pseudo-fungal characteristics and metabolic traits, demonstrating that these features are, given current genome sampling, specific to the oomycete clade. Comparative genomics also reveals clear differences in the repertoire of gene families associated with filamentous growth such as; vesicle trafficking systems, cell wall synthesis pathways and motor proteins between the Fungi and the Pseudofungi, providing a picture of the gene repertoire evolution that underpinned the convergent evolution of filamentous osmotrophic growth within the eukaryotes.

**Introduction**

Stramenopiles (1) (also known as heterokonts (2)) are a highly diverse branch of protists that encompass a multitude of biological forms including; huge multicellular kelps (‘seaweeds’), abundant marine micro-algae through to a variety of microbial pathogens, some of which --oomycetes for example-- feed and grow like fungi and cause important diseases of aquatic animals and algae, and plants (3, 4). The stramenopiles are a phylogenetically robust group (e.g. (5)) defined by the presence of two motile flagella, a ‘standard’ smooth posterior flagellum and a ‘tinselated’ anterior flagellum with a tripartite rigid tubular mastigoneme (“hairs”) (2). However, examples of secondary flagellum loss have occurred during the radiation of this group, for example, in the hyphochytrids (6), which have lost a smooth posterior flagellum but retained a tinsel anterior flagellum. As a consequence, comparisons of the genomes of hyphochytrids to other stramenopiles provides a tool for understanding the diversification of the protein repertoire associated with eukaryotic flagellum characteristics.

Many stramenopiles are photosynthetic and retain a plastid of secondary endosymbiotic ancestry (7, 8). Some commentators have hypothesised that this plastid is derived from an ancient event and as such the plastid has the same derivation as the plastid organelles present in dinoflagellates and Apicomplexa (see (9)), which would mean that the Pseudofungi were ancestrally photosynthetic but have lost both the plastid genome and organelle. This hypothesis has been challenged on a number of levels; either by phylogenomic data which supports a pattern of multiple (or serial) eukaryote–eukaryote endosymbioses (10-13), or alternatively evidence suggesting the oomycetes lack a statistically significant footprint of genes of plastid endosymbiotic ancestry (14). As such, the ancestry of secondary ‘plastid’ endosymbiosis/es relative to the radiation of the stramenopiles remains controversial.

Environmental sequencing, specifically of marine environments (e.g. (15)), has increased the known phylogenetic diversity of the stramenopiles, suggesting that this group is one of the most diverse higher-level groups within the eukaryotes (16). Representatives of these groups remain uncultured with little gene/genome sampling; therefore, the biology of these groups is under-explored. Furthermore, genome-sequencing efforts in the stramenopiles have largely focused on photosynthetic algae (e.g. (17, 18)) or oomycete pathogens (e.g. (19, 20)), leaving the large diversity of heterotrophic free-living stramenopiles under sampled.

Here, we describe the sequencing and comparative genomics analysis of *Hyphochytrium catenoides* (ATCC 18719) originally isolated by D. J. Barr from pine tree pollen in Arizona USA (there is no direct reference in ATCC that accompanies this culture, see (21)). We propose this organism and associated data as a tool to investigate the evolution of stramenopile characteristics (i.e. flagellum remodelling and plastid presence/absence) and for the purpose of comparing and contrasting the evolution of traits between free-living and parasitic pseudofungi. *H. catenoides* is a free-living hyphochytrid protist that forms hyphal-like networks and spores with only a single anterior tinsel flagellum (6, 22). The hyphochytrids are thought to branch sister to the oomycetes (4, 23) and both these groups grow as filamentous/polarized cells feeding osmotrophically by extracellular secretion of digestive enzymes coupled to nutrient uptake (4, 6, 22). These characteristics mean that they resemble fungi (4). We used the genome sequence data to confirm the phylogenetic position of the hyphochytrids, investigate characters shared with oomycete parasites and identify the genes involved in cellular characteristics absent or shared with fungi that typify filamentous/osmotrophic growth. We also use the genome data to investigate protein repertoire putatively associated with loss of the posterior flagellum in the hyphochytrids including the diversity of putatively encoded photoreceptors. Similar to oomycetes, hyphochytrids seem to lack a plastid organelle and photosynthetic function so we also conducted a phylogenomic survey to search for a remnant footprint of plastid endosymbiosis in the *Hyphochytrium* genome. These data provide unique genomic sampling of a free-living stramenopile in order to facilitate further evolutionary and cellular research.

**Results and discussion**

**Genome assembly and gene model prediction**

Using a range of methods, we assembled and tested the completeness of the *H. catenoides* genome assembly (see methods). Comparisons using BLAST of the transcriptome data, CEGMA and BUSCO demonstrated that the genome assembly was predicted to be respectively; 97.8%, 91.5%, and 52% complete in terms of gene sampling (Fig. S1). We suspect that both CEGMA and BUSCO (v1.2) underestimate the completeness of genomes as the core gene list is derived from a subset of eukaryotic genomes that does not fully sample a diverse collection of eukaryotic genomes, (e.g. BUSCO version 1.2 only samples fungi and metazoan genomes) which inevitably gives a much lower estimation of completion. Specifically, Fig. 1a/b shows the CEGMA/BUSCO scores across the stramenopiles, demonstrating a 27/16 complete/partial gene family absence in the *H. catenoides* assembly according to CEGMA. This is a moderate level of absence compared with other stramenopile genomes (e.g. *Blastocystis hominis* and *Ectocarpus siliculosus* with 76/53 and 71/31 complete/partial gene family absences respectively). Using the CEGMA and transcriptome comparison approaches described, we then investigated the completeness of the assembly when scaffolds below 1 kbp, 5 kbp, and 10 kbp were removed (Fig. S1). This analysis demonstrates that the removal of scaffolds less than 1 kbp has a negligible effect on the predicted completeness of the genome. As such we have submitted the >= 1 kbp scaffold assembly along with the predicted proteome as a draft genome to the EMBL EBI database for community access. See Fig. 1b for details of assembly in-comparison to other eukaryotic genome sequences. Details of genome contamination assessment and genome assembly are given in the Materials and Methods.

Analysis using RepeatMasker (24) demonstrated that the > 1 kbp genome assembly contained 9.53% repeat regions where 1.79% of the repeat sequence was assigned to transposable elements. The *Hyphochytrium* genome contained a full set of tRNAs including an additional tRNA for selenocysteine. Putative genes and their predicted proteins were identified using a combined pathway as described in the methods. This approach identified 18,481 putative gene models (406 of which represented different splice forms), a total gene number similar to the mean (15,946) for other sequenced stramenopiles (Fig. 1b). The number of introns and exons reported by the program Genome Annotation Generator (GAG) was 67,332 and 85,813 respectively, with an average exon length of 228 and intron length of 208.

**Genome size, ploidy and evidence of sexual reproduction**

K-mer counting was used to predict a haploid genome size (25) of between 54.1 Mbp and 68.6 Mbp with follow up analysis focusing specifically on the >= 1 kbp assembly suggesting a genome size of 65.7 Mbp across 4,758 scaffolds and a scaffold N50 size of 35.57 kbp (L50 of 399). The average sequencing coverage of the total assembly was estimated to be 312x, and the average coverage over the >= 1 kbp scaffolds is 610x. Extraction and purification of long strands of DNA was not achieved using multiple DNA extraction protocols preventing sequencing using a long-read technology and/or pulsed-field gel electrophoresis. As such, we used a RT PCR method for estimation of genome size (26) suggesting a haploid genome size of 46.9 Mb (S.E.M. = 1.5). Consequently, these data suggest we recovered a tractable draft genome sequence of *H. catenoides*.

As mentioned previously, the N50 of the genome assembly was much improved by the use of Platanus - an assembly algorithm optimised for multi-ploidy genomes. To further investigate evidence of ploidy in our *H. catenoides* culture we mapped ~101 million reads to the 65.7 Mbp assembly identifying 1,393,505 single nucleotide polymorphisms (SNPs) with 1,332,610 (96%) of the SNPs identified consisting of a two-way nucleotide polymorphism (i.e. 58.8/41.2% mean character split). We also took all scaffolds and plotted SNP frequency against scaffold size demonstrating the majority of the scaffolds are clustered around a SNP frequency of ~0.0275 (Fig. S2) suggesting that this variation is consistent and not specific to a subset of chromosomes for example in the case of an aneuploidy. Interestingly this analysis showed two scaffolds with very low SNP frequency compared to the rest of the assembly. These scaffolds contain a number of genes recently derived from a large DNA virus suggesting the presence of a viral genome or evidence of a recent introgression, discussed below. K-mer mapping (25) showed two peaks in the coverage frequency, which is also consistent with the reads mapping to a diploid genome (Fig. S3). Taken together these data are consistent with the *H. catenoides* sampled for sequencing containing a diploid variant during its multifaceted life-cycle (6, 22).

We also checked the transcriptome to determine if the seven meiosis-specific gene families previously shown to be present across a wide diversity of eukaryotes (27) were expressed in *H. catenoides*. Using reciprocal BLAST searches, we found evidence that genes putatively encoding all seven meiosis-specific genes were present on the assembled genome and that these genes were also present in the transcriptome gene set (see Table S1) and therefore likely expressed in the culture conditions used to grow *H. catenoides* prior to sequencing. Sexual recombination has only been observed once in Hyphochytriomycota cultures, with Johnson (1957) identifying cellular forms suggestive of zygote production as a result of fusion in the resting spore development of *Anisolpidium ectocarpii* (28). However, a range of different sexual reproductions systems have been identified in the oomycetes (29), collectively these data suggest meiosis is present in representative taxa across the Pseudofungi.

**Phylogenetic position of *Hyphochytrium* and traits shared across the Pseudofungi.**

*Hyphochytrium* has been shown to branch as a sister-group to the oomycetes in rRNA gene phylogenies (e.g. (3, 23)). To test this phylogenetic relationship, we generated a concatenated amino acid alignment of 245 putative orthologues (85 taxa and 71,862 amino acid sites) and including a comprehensive sampling of eukaryotic taxa based on previously published papers (30, 31).  The resulting tree topology (Fig. 1a) demonstrated that *H. catenoides* forms a sister-branch to the oomycete radiation with 100% bootstrap support consistent with the Pseudofungi hypothesis, i.e. the hyphochytriomycetes are a sister-group to the oomycetes and therefore share a common evolutionary trend towards fungal-like osmotrophic feeding and polarized cell growth (3, 4). Our tree places the Pseudofungi as a sister group to the photosynthetic stramenopiles (i.e. the Ochrophyta) consistent with previously published phylogenetic analysis based on three nuclear encoded genes (32) and wider phylogenomic analysis of 339 protein sequences (8) and in contradiction to analyses of mitochondrial gene phylogenies (concatenation of ten genes, 7,479 positions) which have demonstrated that a separate stramenopile group the Labyrinthulida (i.e. Bigyra) form a sister-group to the oomycetes (33). The tree recovered here is broadly similar to that reported by Derelle *et al.* which uses a large phylogenomic dataset from different taxa. This work convincingly argued for monophyly of Bigyra (e.g. *Blastocystis + Aplanochytrium* and *Schizochytrium*) although our tree shows that this group is paraphyletic. Derelle *et al.* also recovered paraphyly of this group in a subset of their Bayesian analysis and in their ML analysis but then went on to demonstrate that this relationship is likely due to a long branch attraction artefact (e.g. (34)) associated with the *Blastocystis* branch leading to the misplacement of Opalozoa (e.g. *Blastocystis*). Interestingly, sisterhood of the Pseudofungi and Ochrophyta implies a minimum of two losses of photosynthesis (8) and independent specialization of ‘osmotrophic lifestyles’ in the Bigyra (e.g. *Aplanochytrium* and *Schizochytrium*) and the Pseudofungi (e.g. *Hyphochytrium* and *Phytophthora*)within the stramenopiles. However, this scenario implies that the stramenopile lineage was ancestrally photosynthetic (9) a subject of debate (13, 14), which is discussed below.

**Shared derived traits across the Pseudofungi**

Given the placement of *H. catenoides* as a sister branch to the oomycetes, we were interested to investigate the conservation of cellular, biochemical and genetic traits shared across these Pseudofungal taxa*.* Oomycete plant pathogens, for example *Phytophthora* spp., are sterol auxotrophs and appear to have lost many of the enzymes involved in sterol biosynthesis (35). Sterol biosynthesis pathway has been predicted to function in *Saprolegnia* and a putative CYP51 sterol-demethylase encoding gene was identified from the *Saprolegnia parasitica* genome and transcriptome data (20, 36). The protein encoded by this gene is a target of antimicrobial drugs such as clotrimazole and therefore has been suggested as therapeutic target for treatment of *Saprolegnia* infections of fish (37). Reciprocal BLASTp searches and phylogenetic analyses demonstrated that *H. catenoides* also possesses a putative orthologue (Hypho2016\_00003038 - Fig. S4a) of the *S. parasitica* CYP51 sterol-demethylase which appears to be lost in plant pathogenic oomycetes. To confirm that this is a viable drug target we grew *H. catenoides* in the presences of two azole ‘antifungals’ - clotrimazole and fluconazole – to assess effectiveness of these compounds in inhibiting *H. catenoides* growth. Both ‘antifungal’ agents were able to inhibit growth of *H. catenoides* (MIC100: clotrimazole 0.25 µg ml-1; fluconazole 4 µg ml-1  Fig. S4b concentrations were as used in (37), while the MIC showed considerably increased potency as compared to *S. parasitica*), indicating that the *H. catenoides* is susceptible to azole compounds, consistent with *H. catenoides* having a functioning CYP51 protein.

There has been considerable effort to sequence a number of oomycete genomes, which has largely focused on parasitic taxa (e.g. (19, 20, 38-41)). This work has also in-part focused on identifying candidate effector proteins (secreted proteins that perturb host function for the benefit to the invading parasite (42) and which often contain N-terminal RxLR amino acid motifs (43-45)) or lectin proteins that bind host molecules. Searches of the *H. catenoides* genome demonstrate there is only one putative protein of unknown function with a candidate RxLR motif (Table 1). In addition, *H. catenoides* lacked several gene families linked with the evolution of plant pathogenic traits in the oomycetes, i.e.: NPP1 or NEP-like (necrosis inducing *Phytophthora* protein (46, 47)), Elicitin proteins (48), Cutinase (49), pectin esterase and pectin lyase (50, 51). The animal parasite *S. parasitica* was noted to show enrichment of Notch proteins and Ricin lectins, as well as presence of other galactose-binding lectins and the bacterial toxin-like gene family (haemolysin E) (20). While the Notch protein and Ricin lectin gene families are present in *H. catenoides*,they show no evidence of enrichment comparable to *S. parasitica* and the galactose-binding lectin and haemolysin E gene families are absent. Protease gene families show no general enrichment in comparison to other stramenopiles (Table 1).

Comparative analysis of candidate secreted proteins defined by in silico identification of putative N-terminal secretion sequences demonstrated that*H. catenoides* contains a lower proportion of secreted proteins compared with many other stramenopiles, comparable with the paraphyletic obligate biotrophs *Albugo laibachii* and *Hyaloperonospora arabidopsidis* (Fig. 2)*.* The *H. catenoides* predicted proteome contains a moderate-to-low proportion of carbohydrate active enzymes relative to other stramenopiles (52). Interestingly, *H. catenoides* has very few secreted carbohydrate active enzymes in comparison to other stramenopiles, suggesting that *H. catenoides* has a low diversity of extra-cellular carbohydrate processing functions and is therefore dependent on a limited subset of extracellular fixed carbon sources (Fig. 2). To test this observation, we grew *H. catenoides* cultures in 190 different carbon sources using OmniLog PM1 and PM2 plates which allows investigation of growth and respiration rate across a diversity of different carbon sources (53). These data demonstrated (Fig S5a/b) a significant increase in growth rate upon the addition of only α- or β- cyclodextrin (*p* value = 0.01 & 0.01), dextrin (*p* value = 0.02), Tween 40 or 80 (*p* value = 0.03 & 0.03), or melibionic acid (*p* value = 0.03). Of note, dextrin/cyclodextrins are products of enzymatic activity upon starch, a typical component of *H. catenoides* growth medium (YpSs), and may be indicative of the environment in which this organism is typically found. The addition of Tween 40 or Tween 80 has been shown to improve yield in other organisms (54) and may result from direct accumulation of fatty acids, or altered membrane permeability affecting nutrient uptake. In contrast to many oomycetes (e.g. (55)), *H. catenoides* demonstrates a limited utilisation of diverse carbon sources. These data are consistent with the hypothesis that the evolution of a wide diversity of secreted carbohydrate active enzymes is associated with evolution of parasitic lifestyle within the oomycete lineages (e.g. (20, 56-58)), although this pattern could also be the product of secondary loss in the *H. catenoides* lineage.

Seidl et al. have demonstrated a significantly higher number of genes encoding unique protein domain combinations in the plant pathogenic oomycetes compared to fungi with similar lifestyles (59). This analysis identified 53 domain architectures that were unique to and conserved across *P. infestans, P. ramorum, P. sojae* and *H. arabidopsidis.* Domains are often recombined by a process of gene fusion and/or domain ‘shuffling’ (60), as such gene fusion characters, although subject to sources of homoplasy such as gene fission (61), can represent synapomorphic traits, i.e. characters that define evolutionary groupings. We searched the *H. catenoides* genome for evidence of the 53 multi-domain gene fusions previously identified as conserved in the oomycetes *P. infestans, P. ramorum, P. sojae* and *H. arabidopsidis* (59). We found that 12 of these 53 domain architectures were also present in *H. catenoides* (Table S2). Of note, we found a fusion of a putative β-glucan synthase enzyme domain and a putative membrane transporter gene unique to the Pseudofungi given current genome sampling (Table S3 & GenBak ‘nr’ protein database), suggesting that domain fusion has led to a unique coupling of substrate transportation and enzymatic processing prior to the radiation of the Pseudofungi.

Using OrthoMCL (62) combined with a custom pipeline we identified candidate orthologue gene sets specific to the currently sampled Pseudofungi (Table S4). Specificity was confirmed using BLAST searches of the GenBank nr and EST databases. This analysis identified six gene families that encode Pseudofungi-specific domain combinations, five of which were newly identified and one that overlapped with the data reported above (see (59) and Table S2). The analysis of Pseudofungi-specific orthologues and comparison of conserved novel domain protein domain architectures reported in Seidl et al. (2011) identifies four group-specific protein domain architectures encoding an AGC family Serine/Threonine Kinase domain. These results demonstrate a novel diversification in this domain family and functional Serine/Threonine Kinase repertoire in the Pseudofungi. Indeed, previous work has demonstrated large expansions of kinase families present in oomycete genomes (20) (Table 1).

**Evidence of viral integration across the Pseudofungi**

The comparative genomic analysis of Pseudofungi described above identified that *H. catenoides, Phytophthora cinnamomi, Phytophthora parasitica* and *Pythium ultimum* harbor genes putatively encoding viral major capsid proteins (MCP) bearing phylogenetic affinities with nucleocytoplasmic large DNA viruses (NCLDV). These proteins have high sequence identity to each other and phylogenetically branch with MCP proteins from African swine fever virus (Asfarviridae lineage of NCLDVs), but which are divergent when compared with other NCLDV MCP proteins (Fig. 3a+b). Exploring the *H. catenoides* genome assembly to determine the presence of viral-like genes, we identified 45 putative genes, 38 of which are present on two scaffolds which were shown to have very low SNP frequency in the assembly (Table S5). All of these 38 genes showed highest similarity to NCLDV families such as Mimiviridae, Marseilleviridae, Phycodnaviridae, Asfarviridae and Poxviridae (Table S5). The genome assembly in these regions was confirmed by nested PCR and sequencing from both the 5’ and 3’ end of the *polB*, *mcp*, *mg96* viral genes (Table S6). The genes of putative viral ancestry were found in linkage with genes of *H. catenoides* provenance in our assembly. For example, the genome assembly demonstrated that the viral *mcp* gene was on the same DNA contig as a putatively native *H. catenoides* histone-encoding gene Fig S6. To confirm this assembly and linkage between ‘host’ and viral gene we conducted a bridging PCR resulting in an amplicon of 2,837 bp and sequenced this amplicon confirming that the *mcp* and histone genes are linked and on the same stretch of DNA (Table S6).

Both viral gene containing contigs consist of 145 predicted genes. BLASTx analyses suggests that the contigs contained 37 (26%) and 18 (12 %) genes of highest similarity to genes of known viral genomes (Table S7). The BLASTx results for the remaining 235 putative genes showed a wide variation of top BLASTx hits including both prokaryotic- and eukaryotic-like genes. The frequency of putative exons for the two contigs was 1.62 and 1.49 respectively, a lower intron frequency observed for the wider genome (4.22), suggesting the genes encoded on both contigs have introns. Indeed, multiple viral-like genes show evidence of intron/exons suggesting either these genes have been: incorrectly modelled, subject to intronization, exon shuffling during integration, or these genes are undergoing pseudogene-isation and are broken ORFs which are being reported as intron/exon structures, however, we note that gene of viral provenance Hypho2016\_00000945-RA (scaffold 5419) is expressed and contain multiple putative coding regions. The low SNP frequency of these contigs suggest they represent a unique haploid portion of the genome, a viral genome captured in our assembly or alternatively a site of viral introgression in the *H. catenoides* genome. We currently favour the hypothesis that this is a site of viral introgression due to the presence of introns in the contig and the low relative proportion of genes of clear viral provenance.

Products from *polB, mg96* and *rps3* were detected by RT-PCR in our culture conditions, suggesting that viral-like genes are transcriptionally active (Fig. 3b)*.* In contrast, a lack of transcript from the *mcp* gene suggests that a complete virus or a viral factory is not being manufactured in the culture conditions tested (Fig. 3b). Electron microscopy also failed to observe icosahedral structures typical of viral particles or an intracellular viral factory (69) (see Fig. S7).

These data combined with evidence of viral genes present in oomycetes genome assemblies (Fig. 3a) (63) suggests a hitherto un-sampled diversity of large DNA viruses found infecting, or integrated within the genome of Pseudofungi. This is consistent with other data suggesting the Pseudofungi have been subject to viral transduction (70). It has also been shown that many different lineages of the stramenopiles have similarly retained fragments of viral genomes (64) suggesting a wider and under sampled diversity of stramenopile infecting large DNA viruses. It is tempting to speculate that this may be a mechanism driving HGT seen in the oomycetes (see (65)), given that NCLDVs have been shown to harbor host-derived and foreign genes (66, 67) and that fragments of large DNA viruses have now been shown to be present in fungi (63), a group shown to be a donor of HGT-genes to the oomycetes (58, 65) and we note that the two contigs containing the viral derived genes also contain two genes with top BLASTx hits to fungi (Table S7). The Pseudofungi lack the capabilities to perform phagotrophy (4), a mechanism thought to be important for HGT in eukaryotes (68). However, There is evidence of gene transfer into the oomycetes from both fungi and prokaryotes (49, 58, 69-73) although, we note the extent of ancient HGTs in eukaryotes has recently been questioned (74), these transfers are likely of relatively recent provenance. Indeed, Ku et al. also identified genes uniquely present in oomycetes and bacteria which are described as ‘recent lineage specific acquisitions’ (see Fig. 1 in (74), marked as *b*). Evidence of viral introgression within the Pseudofungi therefore identifies a possible mechanism driving HGT and recent lineage specific acquisition in the Pseudofungi, which cannot perform phagotrophy. It is important to note that viral transduction as a vector for HGT would likely produce a very different profile of gene transfer compared to mechanisms such as phagocytosis (in eukaryotes), transformation (prokaryotes), or conjugation (prokaryotes). This is because gene transfer via a virus would likely transfer a lower number of gene families for two reasons; 1) genes carried by the virus would have been passaged by selection within the viral lineage, and, 2) the limited DNA carrying capacity of the viroid. Such a mechanism of HGT is therefore consistent with the results of Ku et al. (74), which suggests HGT is much less frequent in eukaryotes compared to prokaryotes. However, this does not exclude the possibility that infrequent HGTs can lead to the acquisition of novel and positively selected traits.

**Organelle genomes in the *H. catenoides* assembly**

Using common mitochondrial marker genes as query sequences for tBLASTx searches we searched the *H. catenoides* for fragments of the mitochondrial genome. Theses analyses showed that the organelle genome assembly was highly fragmented. By searching both the Platanus and an alternative Ray assembly (75) we identified and re-assembled, two mitochondrial genome contigs using Sequencher (76). We then used multiple primers for targeted PCR amplification and joined the two fragments of the mitochondrial genome, resulting in a linear fragment of 48,124 bp in size (Fig. S8 inner red semi-circle). The *H. catenoides* mitochondrial genome is reported to be circular based on mapping analyses (77). We used multiple combinations of PCR primers to try and join the other sides into a circular genome, but could not amplify a bridging region suggesting that our assembly is missing a large section of sequencing, likely an inverted repeat. A mitochondrial genome size of 54 kbp with a ~14 kbp inverted repeat (a feature also present in some oomycetes) had previously been suggested using physical genome mapping (77, 78). Somewhat similarly, based on the sequence similarity at the ends of our mitochondrial assembly to internal sequences (see methods), we inferred a 19 kb inverted repeat which completes the putatively circular genome (Fig. S8 inner green semi-circle).

Annotation showed that the *H. catenoides* mitochondrial genome encodes 33 unique mitochondrial protein-coding genes (plus six putative orfs with no similarity to other mitochondrial genes), 21 unique tRNAs for 18 amino acids, and two ribosomal RNA genes. Except for the lack of rps7 and TatC genes, this gene set is very similar to reported oomycete mitochondrial genomes (78). No introns were predicted.

The assembly demonstrated the presence of genes similar to mitochondrial *atp1* in both the mitochondrion and also the putative nuclear genome assembly (Hypho2016\_00005578). The *atp1* gene has been endosymbiotically transferred from the mitochondrial genome to the nuclear genome independently in several lineages but is retained in the mitochondrial genomes of a range of stramenopiles including the eustigmatophytes (e.g. *Nannochloropsis*), oomycetes and *Cafeteria* (78, 79). We reconstructed the phylogeny of the *atp1* genes from all domains of life to test the hypothesis that the nuclear *atp1* gene in *H. catenoides* may be the result of an endosymbiont gene transfer. We show that while the mitochondria-encoded *atp1* branches within a clade including all other bona-fide mitochondrial *atp1* genes (regardless of which genome- nuclear or mitochondrial -they are encoded), surprisingly the *H. catenoides* nuclear-encoded *atp1*-like gene branched with oomycete and Nannochloropsis sequences as a separate paralogue of uncertain ancestry, sister to all bacterial-derived (i.e. bacterial, mitochondrial, and chloroplast) *atp1* genes (Fig. S9). With the exception of *Nannochloropsis*, these divergent *atp1*-like proteins have putative N-terminal mitochondrial targeting peptides, indicating that these proteins may represent a stramenopile-specific nuclear encoded mitochondrial *atp1* genes.

Many stramenopile lineages harbour plastid organelles with an associated genome while others, including the oomycetes, appear to have lost (9) or, alternatively, never possessed (14) the organelle based on genome analyses. To search for possible plastid genomes, we took the proteome encoded by the plastid of *Chlamydomonas reinhardtii* (69 proteins) and *Thalassiosira pseudonana* (141 proteins) as query sequences for a tBLASTn search of the entire genome assembly (including contigs and scaffolds below 1 kbp in size). Only a minority of searches recovered any hits in *H. catenoides* and these were shown by reciprocal BLAST searches to be putative mitochondrial-located ribosomal genes, suggesting that there is no plastid genome present in ourgenome assembly. However, this analysis does not exclude the possibility that the organelle and plastid genome has been lost leaving a trace of plastid endosymbiotically derived genes present on the nuclear genome.

**Footprint of a plastid gene ancestry in the *H. catenoides* genome assembly**

The established approach for identifying endosymbiotic ancestry of protist lineages that appear to have no endosymbiotic organelles is to identify genes of endosymbiotic ancestry present on the nuclear genome. This approach has worked well for identifying cryptic endosymbiotic ancestry for both the plastid (80, 81) and mitochondrial (82-84) organelles. Genes derived from the endosymbiont that gave rise to the primary plastid organelle should have eukaryotic orthologues branching with or within a cluster of cyanobacterial genes on a phylogenetic tree. Using a phylogenomic pipeline described in the methods, we constructed a phylogeny for all clusters which included both a *H. catenoides* gene and a gene of a eukaryote with a plastid, and included a wide sampling of putative homologues from both eukaryotic and prokaryotic taxa. This process resulted in 8970 preliminary phylogenetic trees, which we searched for a phylogenetic relationship that demonstrated *H. catenoides* genes branching with photosynthetic eukaryotes or eukaryotes with photosynthetic ancestry, which in turn branched within a cluster of bacterial sequences. Following the example of others (85, 86) we reasoned that cyanobacterial signature was likely mixed up by horizontal gene transfer within the prokaryotes and so any *H. catenoides* gene cluster that showed the relationship described above (i.e. [[*H. catenoides* + Plastid-bearing-Eukaryotes]bacteria]) potentially represented a relict gene ancestrally derived from the endosymbiosis that gave rise to the plastid. For 101 preliminary phylogenies that showed this relationship, the amino acid sequence alignment was edited and masked manually and the taxon sampling checked using additional BLAST searches of NCBI nr protein database. This process identified four gene families where *H. catenoides* protein sequences (in three cases branching with other Pseudofungi/Stramenopiles taxa) branched with photosynthetic eukaryotic taxa sister to bacterial taxa. In all four cases (Fig. S10a-d) we did not detect a putative orthologue of these gene families in eukaryotic taxa thought not to have had a plastid endosymbiotic ancestry (i.e. Fungi, Metazoa and Amoebozoa). In only one case did the eukaryotic sequences branch with the cyanobacteria, although the bootstrap support for this relationship was weak (Fig. S10d). We then used TargetP (87) searches to investigate if the putative orthologues of the *H. catenoides* protein sequence encoded by Archaeplastida taxa had putative plastid targeting N-terminal peptide, demonstrating that none of these protein sequences contained such evidence. TargetP identified both the *Arabidopsis thaliana* and *Oryza sativa* putative Heme oxygenase domain containing protein (Fig. S10d) had evidence of an N-terminal mitochondrial targetingpeptide.

Taken together these data suggest that the *H. catenoides* protein sequences contain a subset of genes that could be putatively of plastid endosymbiotic ancestry. It is possible that our survey protocol missed a number of genes of endosymbiotic ancestry. This is because pipeline generated trees often have inappropriate taxon sampling (i.e. too large), with the pipeline-generated alignments too noisy to allow effective phylogenetic analysis. As such, many of the trees generated rarely show meaningful branching relationships or significant resolution. To partially control for this problem, we reran many of our candidate genes, where the resulting alignment was very large, using a subset of genome sampling. It is difficult to conclude if four gene families identified here, which in all cases do not show a strongly supported cyanobacterial/plastid ancestry for the eukaryotic genes, represents a significant result which confirms plastid ancestry for the Pseudofungi. Indeed, others have argued that similar numbers of gene phylogenies indicate a plastid endosymbiotic ancestry for different protist groups that have no evidence of a plastid organelle (80, 81). Conversely, it is difficult to prove a negative, because it is certainly possible that the majority of plastid/cyanobacterial derived lineages have been lost while any genes retained may be so small in number and with hugely complicated phylogenetic histories that it would be near impossible to recover using pipeline tree building methods. As such these results are inconclusive with regards to photosynthetic/plastid ancestry of *H. catenoides*.

**Evolution of protein repertorie that corresponds to loss of the posterior flagellum in *Hyphochytrium*.**

The stramenopiles (also known as Heterokonta, meaning possessing of two unequal flagella) were formally described as a phylum based on the presence of two motile flagella: a ‘standard’ smooth posterior flagellum and an anterior flagellum with tripartite rigid tubular mastigonemes ( ‘tinselate’) (2). *Hyphochytrium* builds only a single, anterior tinselate flagellum (6) while the oomycetes build the standard stramenopile flagella pair. The phylogenetic placement of the *Hyphochytrium* (Fig. 1b) therefore pinpoints a loss of the posterior smooth flagellum in the ancestor of the hyphochytrids. To explore the consequence of the loss of this organelle in *H. catenoides*, in terms of gene/protein repertoire, we used a comprehensive list of proteins putatively associated with flagellar function (88) to survey the *Hyphochytrium* genome. This list comprises 592 amino acid sequences, 355 are found in both the major eukaryotic phylogenetic groupings of Opimoda and Diphoda (89) suggesting they are universal flagellar proteins (Table S8 & Fig. 4A). 330 of the 355 universal flagellar proteins (UFPs) are also present in the predicted proteome of *H. catenoides*, suggesting that the majority (93%) of the UFPs have been retained and likely encode a function associated with the anterior tinselate flagellum.

Flagellum-specific proteomic analysis of the stramenopile brown alga *Colpomenia bullosa* identified fourteen proteins specific to the posterior flagellum and three specific to the anterior flagellum (88). BLAST similarity searches suggest that the three anterior flagellum proteins are also present in *H. catenoides,* as are twelve of the fourteen posterior flagellum proteins from *C. bullosa*. Conservation of these “posterior-specific” proteins suggests that they have functions associated with the anterior tinselate flagellum in *H. catenoides* (Fig. 4a)*.* One of the *C. bullosa* posterior specific flagellum proteins absent in *H. catenoides* and the oomycetes is the PAS/PAC sensor hybrid histidine kinase (also known as a Helmchrome, CBJ26132.1), a putative photo-sensor associated with a swelling in posterior flagellum of brown algae (88), discussed further below.

Twenty-nine of the UFPs (8%) were present in oomycetes and other eukaryotic groups but absent in *H. catenoides*. These may represent genuine gene losses, although absences in our draft genome may also be due to incomplete genome sequencing and assembly. If these genes are genuine losses from the *H. catenoides* genome these results suggest a recent loss, putatively consistent with loss of the posterior flagellum without the function of these proteins being integrated into the anterior-tinsel flagellum (Fig. 4a)*.* Interestingly, these losses include a putative homologue of the dynein regulatory complex 1 (DRC1) protein, which regulates inner dynein motor activity in *Homo sapiens* and *Chlamydomonas reinhardtii* (90), and Radial Spoke protein 7 (RSP7), a protein that functions in flagellum structure and beating in *Chlamydomonas reinhardtii* (91).Further, analysis of the radial spoke protein repertoire encoded by *H. catenoides* identified a number of other components of the radial spoke complex which are putatively absent in *H. catenoides*. However, RSP7 was the only radial spoke proteome loss specific to the loss of the posterior flagellum in the *Hyphochytrium* lineage (Fig. 4a/b); this protein is putatively encoded in the oomycetes but has been separately lost within the Opisthokonta (e.g. *Homo sapiens*)*.* In *Chlamydomonas* (91), RSP11 and RSP7 have been shown to contain a RIIa domain (92). Association between RIIa and AKAP domains and RSP3 at the spoke stalk are suggested to be important for flagellar function (91). Interestingly, comparative analysis suggests that neither RSP7 or RSP11 are conserved across flagellum-bearing eukaryotes with only *Chlamydomonas*, *Batrachochytrium* and *H. catenoides* retaining RSP11 in our comparative dataset (Fig. 4a/b)*.* Domain analysis (93) of the putative *H. catenoides* RSP3 and RSP11 confirmed these proteins contain an AKAP and a RIIa domain respectively suggesting that *H. catenoides* has retained only RSP3-RSP11 protein-protein interaction at the base of radial spoke proximate to the outer doublet.

Phylogenomic analysis of motor protein repertoire, specifically Kinesins and Dyneins (Fig. 4c/d), confirmed that the *H. catenoides* genome assembly has retained many of the motor proteins associated with flagellum function, these include --like the oomycetes-- the full set of axonemal dyneins such as the intracellular transport Dyneins (94) (Fig. 4c) while the presence of Kinesin-2, -9 and -16 also present in the *H. catenoides* genome assembly and the oomycetes, is also consistent with flagellum function (95) (Fig. 4d), suggesting that the modified tinselated *H. catenoides* anterior flagellum has retained these motor protein functions. Wickstead & Gull have also proposed that kinesin-17 has a flagellum function based on phylogenetic distribution of this paralogue family (95), while the reported genome assembly and subsequent phylogenomic analysis suggests that, unlike the oomycetes, *H. catenoides* has lost kinesin-17, possibly associated with the loss of the posterior-smooth flagellum.

**Photoreceptors**

Stramenopile species have been shown to encode a range of photoreceptor proteins and to initiate a series of responses to light including phototaxis (96). Specifically, the zoospores of some stramenopile algae can show positive and negative phototaxis (97) associated with a flavoprotein photoreceptor (98) putatively the ‘helmchrome’ located in the posterior flagellum (88) and associated with ‘flagellar swelling’ and a stigma (97). Consistent with the loss of the anterior flagellum *H. catenoides* (Fig. 4a) also lacks a gene putatively encoding a helmchrome protein*.*

A number of additional putative photo-responsive proteins have also been reported from *Ectocarpus* (18). Using these data and other seed sequences (e.g. (88, 99)) we searched the *H. catenoides* genome for putative homologues of photo-responsive protein. Reciprocal BLAST searches demonstrated that the *H. catenoides* genome contained putative homologues of the flavoproteins; Cryptochrome (Hypho2016\_00016188), Cryptochrome DASH (Hypho2016\_00004514) and Photolyase (Hypho2016\_00002462) gene families (Fig. S11a)and transcriptome data demonstrate that these genes are transcribed. This analysis also identified three putative type I (microbial) rhodopsins (Hypho2016\_00006030, Hypho2016\_00006031 & Hypho2016\_00010050), the first putative representative of this gene family from a stramenopile (Fig. S11a/b). The three rhodopsins all contain a conserved 11-*cis*- retinal binding pocket, specifically the Schiff base (Fig. S11a/b). Furthermore, reciprocal BLAST searches of both the genome and the transcriptome sequence datasets confirmed the presence of genes putatively encoding the latter two steps of the retinal biosynthesis pathway (e.g. a putative β-carotene-15, 15’-dioxygenase (Hypho2016\_00004122) and a putative retinol dehydrogenase (Hypho2016\_00000702) the steps in the pathway that converts the vitamin β-carotene into 11-*cis*-retinal, the critical cofactor for rhodopsin function as a light responsive protein. These data suggest *H. catenoides* is capable of light-triggered development, physiological responses and/or phototaxis.

**Gene families encoding hallmarks of fungal characteristics in the Pseudofungi**

One purpose for the sequencing the *H. catenoides* genome was to investigate the genes that underpin the fungal/pseudofungal lifestyle. Most fungi grow as filamentous cells, reinforced by robust cell walls composed of sugar polysaccharides such as chitin. These characters are not unique to fungi but are typical in many fungal lineages (100). A suite of cellular systems allowing fungi to grow as polarised cells, laying down cell wall and feeding on extracellular substrates by a combination of exocytosis of enzymes and cell wall material and endocytosis and transporter mediated uptake of target nutrients. Fungal filamentous structures such as hyphae grow almost exclusively from the tip of the hyphal structure (101) allowing fungi to ‘grow as they feed.’ This feature combined with a robust cell wall means they can generate high turgor pressures, ramify into recalcitrant material, feed osmotrophically and maximize metabolic rates (100, 102, 103). Homologous cellular systems also drive bud growth in *S. cerevisiae* allowing researchers to use *S. cerevisiae* to study proteome function involved in polarized growth in fungi (for review see (101, 104)). The proteins that are known to control this system are illustrated on Fig. 5a and involve key complexes, the Exocyst and the Polarisome. These systems are therefore very important for establishing the temporal and spatial control of polarized cell growth in (101, 104). Comparative analyses show the Exocyst system and Sec4 orthologues are conserved across a diversity of eukaryotes including *H. catenoides*, while the Polarisome and associated proteins seem to be specific to fungi, given current taxon sampling, and demonstrate that specific elements of polarized cell growth control are not present in Pseudofungi, suggesting these filamentous microbes accomplish polarized growth using a combination of different proteome functions.

Motor protein evolution has been suggested to be an important factor in the acquisition of filamentous growth phenotypes in the fungi, with a specific focus on the roles of type Myosins and Kinesins and in polarised cell growth, vesicle-transit and chitin synthesis (105-107). Phylogenomic analysis of the motor head domain of all three motor types (Fig. 4 c, d, e) demonstrates no expansion in motor paralogues uniquely shared by the filamentous fungi and the filamentous Pseudofungi. In addition Pseudofungi lack the Myosin V and XVII shown to be important in fungal growth and chitin synthesis (106) (Fig. 4e). The lack of shared/unique motor repertoire between fungi and Pseudofungi, is again consistent with the idea that these groups evolved filamentous polarized growth characteristic separately and based on different cellular systems. It has been noted that oomycetes contain a diverse complement of myosin paralogues, some of which show additional duplication and paralogue family expansion (108, 109). The analyses reported here demonstrates that elements of these oomycete motor protein gene family expansion are also present in the genome of *H. catenoides,* specifically; Myosin XXX and XXI and Kinesin 14 and 20 show high degrees of expansion by duplication specific to the Pseudofungi (Fig. 4 c, e, f), suggesting that these motor proteins may be linked to unique filamentous polarized growth characteristic present in the Pseudofungi, although further work is needed to test the function of these motor protein paralogues.

Like fungi (110) and many other eukaryotes (111-117), *H. catenoides* also produce chitin as cell wall material (118). Oomycetes have also been shown to produce chitin in their cell walls (119). This is consistent with previous data that suggest that chitin synthesis and deposition as a cell wall material predates the diversification of many major lineages of the eukaryotes (100, 118). *H. catenoides* has a similar repertoire of chitin synthesis and digestion as found in the oomycetes (i.e. chitin synthase division I gene family), while another group of stramenopiles, the diatoms, which also produce chitin (120), have a variant chitin gene repertoire, namely chitin synthase division II and a chitinase (GH19) not present in Pseudofungi (Fig. 6). This suggests that chitin production as a cell wall component is universal and anciently acquired in the eukaryotes, but the genes that control the synthesis and remodelling of this structural sugar have been reconfigured numerous times. Specifically, Pseudofungi seem to lack all chitin synthase division II which are numerous and diversified in fungi, suggesting another key difference, between the Fungi and Pseudofungi.

**Conclusion**

The draft genome of the free-living stramenopile pseudofungus *Hyphochytrium catenoides* provides an important reference for comparative biology specifically with a view to understanding the evolution of filamentous growth and osmotrophic feeding. *H. catenoides* branches sister to the oomycetes that contains many important pathogenic groups. These data demonstrate that *H. catenoides* does not encode many of the gene families found in oomycetes that have been associated with pathogenic function, suggesting that these characteristics are more recent adaptation/acquisitions within the oomycetes (Table 1). Our data also demonstrates that *H. catenoides* and the Pseudofungi more widely possess the genes which encode a range of features associated with filamentous growth and osmotrophic feeding in fungi. These include the Exocyst vesicle trafficking system, sterol biosynthesis pathway, and a repertoire of chitin cell wall synthesis systems common to fungi. In contrast, Pseudofungi do not possess the genes encoding a Polarisome complex or chitinase I, chitin synthase II / Myosin V, Myosin XVII, identifying clear differences between these two filamentous osmotrophic groups. Figure 7 summarises how various features associated with filamentous growth and osmotrophic feeding arose relative to the branching position of the fungi and the Pseudofungi. We hope the *H. catenoides* draftgenome will provide a useful dataset for comparative biology within the Pseudofungi and across the eukaryotes especially with regards to understanding the evolution of filamentous osmotrophic characteristics.

**Materials and Methods**

**Cell Culture in preparation for sequencing.**

*Hyphocytrium catenoides* (ATCC 18719) was inoculated onto Emerson YpSs agar and cultured and cell mass prepared for DNA and RNA extraction as described previously (58). DNA samples were checked for contamination using an environmental DNA SSU PCR approach (58) using both eukaryotic 18S PCR primers 1F (5’CTGGTTGATCCTGCCAG-3’) and 1520R (5’-CTGCAGGTTCACCTA-3’) (e.g. (121) and prokaryotic 16S PCR primers PA (5’-AGAGTTTGATCCTGGCTCAG-3’) and PH (5’-AAGGAGGTCATCCAGCCGCA-3’) (e.g. (122).

**Genome and Transcriptome Sequencing, Assembly and Validation and ORF Calling**

One lane of paired-end (100 bp) Illumina HiSeq data was generated along with two lanes of paired-end (76bp) Illumina GAiix at the Exeter Sequencing Service generating 2x 212,760,559 HiSeq reads along with 2x 15,266,599 and 2x 16,274,715 GAiix reads. After trimming and cleaning (using TagCleaner (123) and PRINSEQ (124)) of the data, we subsequently digitally normalised it with KHMER (30) in order to discard redundant data, sampling variation and remove errors. This reduced the number of reads to 415,241,668 HiSeq along with 28,964,302 and 30,961,514 GAiix; a reduction of 13,436,262 reads in total. The raw reads are deposited in NCBI and EBI with accessions as: Illumina GAiix = SRX033129 and Illumina HiSeq = ERS1151585 respectively. An initial assembly, using the program Ray v2.2.1 (125) was generated (see tps://github.com/guyleonard/hyphochytrium/tree/master/manuscript/data for details of commands used), and generated 29,448 scaffolds, with a total of 107,387,882bp, and an N50 of 8,746bp. Next, we investigated the possibility that *H. catenoides* was a diploid using an assembly program that allows for multiple ploidy. The program Platanus v1.2.1 was used to produce an assembly with 53,358 scaffolds incorporating 68,330,525bp, and an N50 of 29,450bp (see tps://github.com/guyleonard/hyphochytrium/tree/master/manuscript/data for details of commands used).

The Platanus assembly was subsequently filtered into 4 datasets; all scaffolds, scaffolds >= 10 kbp, scaffolds >=5 kbp and scaffolds >=1 kbpin order to test the effects or the N50 statistic and gene recovery rate by removing short and erroneous scaffolds/contigs (Fig. S1). We determined that the set of scaffolds >= 1kbp did not affect our predicted proteome complement and increased the N50. The filtered >=1 kbp Platanus assembly, along with the mitochondrial assembly, are deposited in EBI with the accessions; Study ID: PRJEB13950, Scaffolds: FLMG01000001-FLMG01004758 and Mitochondria: LT578416. The full assembly and other filtered datasets, can be accessed here: <https://github.com/guyleonard/hyphochytrium>.

K-mer counting analysis was conducted using Jellyfish along with two publically available scripts (estimate\_genome\_size.pl and the website GenoScope [see https://github.com/josephryan/estimate\_genome\_size.pl and (126). The average sequencing coverage of this assembly was estimated using the ‘estimate\_genome\_size.pl’ tool for the total assembly and using the ‘genomeCoverageBed’ from BEDTOOLS (127) for the >=1kbp subset of scaffolds.

Gene prediction was conducted by using CEGMA to predict which of the 246 core genes are present in our *Hyphochytrium* >=1kbp scaffolds, these predicted CEGs are then used in the training step of the program SNAP (see http://korflab.ucdavis.edu/software.html) to generate a set of *ab initio* gene models. The program GeneMark-ES (128) is also run independently on the >=1kbp scaffold data which produces another set of gene models. Both these sets of gene models are in the form of a hidden Markov model (HMM). A first pass of the pipeline MAKER is then run with the default settings, incorporating the gene models from SNAP & GeneMark-ES whilst also deriving alignment statistics from the 454-transcriptome assembly with tBLASTn, repeatmasker and exonerate. The output is a set of gene models in GFF3 format. A second round of SNAP is then performed with the new predictions (after the GFF3 has been converted to a HMM) and the program AUGUSTUS (129) is run in *ab initio* mode using the MAKER first pass predictions (i.e. AUGUSTUS is not run against any 'similar/related' taxa). Both outputs of SNAP (run 2) and AUGUSTUS are then fed back into MAKER for a second run with stricter settings (gene predictions are available here: <https://github.com/guyleonard/hyphochytrium/tree/master/gene_predictions>). The final output is a GFF3 file, transcripts and protein FASTA files. The resulting gene predictions were then BLAST searched against the SwissProt database along with InterproScan to assign functions. The results were then used with the program ANNIE (130) to provide the correct format of annotation information to the program GAG (131) which is used to convert genome information (scaffold nucleotides, gene prediction, annotation, etc.) for database deposition. The resulting genome data is therefore submitted as an update of a prior BioProject sequence submission (58), to do this we used the ‘gff3toembl’ program from PROKKA (132)

Previously we had sequenced a transcriptome from the same culture strain of *Hyphochytrium* (58) using 454 FLX sequencing of cDNA reads and assembled it with Newbler 2.5 (133) using the default cDNA settings. We cleaned 70 sequences from this assembly by removing contigs of less than 100 bp in length (excluding the polyA regions) and/or contigs that consisted of predominately repeat motifs. This resulted in 6,202 transcript sequences with an N50 of 1,044 bp. The transcriptome assembly is based on 454 FLX+ cDNA reads. Assembled in Newbler 2.5 using the standard settings for cDNA. The reads were also assembled in Trinity but resulted in significantly [nearly double] more contigs and a lower N50 score.

**Assessment of contamination of the genome sequence**

To identify any prokaryotic contamination in the >=1 kbp scaffold assembly we first conducted BLASTn searches of the assembly using prokaryotic SSU and LSU rDNA sequences as search seeds (*Escherichia coli* taken from [CP012802] and *Sulfolobus acidocaldarius* [NR\_043400 & NR\_076363]). This analysis only returned sequences of similarity to the *H. catenoides* mitochondria genome assembly (discussed below), suggesting that no, or very limited, prokaryotic sequences contamination was present. To support this, we subjected all 4,758 genome scaffolds to a BLASTx analysis against a database of 65 eukaryotic and 164 representative prokaryotic complete predicted proteomes (Table S9) with a gathering threshold of 1e-10. This approach did not identify any scaffolds that did not have at least one top hit to a eukaryotic genome for a subsection of the scaffold. Indeed, only 87 of the scaffolds had > 50% of the subsections with a top BLAST hit to a prokaryotic genome and only 20 of the scaffolds had > 70% of their top BLAST hits to a prokaryotic genome. These 20 scaffolds were inspected manually, 11 of these showed the presence of putative spliceosomal introns and/or other genes more similar to other eukaryotic genes. For the remaining nine scaffolds (totalling 31.8 kbp) we could not exclude them as possible prokaryotic contamination (listed in Table S10).

Comparisons of GC content versus read coverage coupled with BLASTn analysis to identify likely aberrant genomic affiliation of assembly scaffolds (e.g. “blobology” (134)) has emerged as useful tool for identifying contamination of genome sequencing projects (135). We undertook this approach on both the > 1 kbp scaffold assembly and the total assembly, and the graphs did not identify any suspect traces of contamination, however, they do show the presence of the mitochondrial genome as an aberrant cluster of ‘blobs’, i.e. with lower than average GC content (Fig. S12a-d).

A fourth round of checks for contamination were conducted by using tetramer counting of the >=1 kbp scaffold dataset for the building of Emergent Self Organising Maps (ESOMs) (136). These use similarities in the 4-mer frequencies to build, by way of an artificial neural network, an emergent ‘map’ of the input space properties of the data. Two runs of the software developed in (136) were completed (see Fig. S13a+b): a) the *Hyphochytrium* scaffolds only, and b) the *Hyphochytrium* scaffolds along with the scaffolds from eight other ‘small’ genomes which were added to the tetramer frequency dataset, (Bacteria (blue): *E. coli, Mycobacterium tuberculosis;* Archaea (grey): *Methanococcus vanniellii, Sulfolobus solfataricus;* Fungi (purple): *Encephalitozoon intestinalis, Saccharomyces cerevisiae;* Archaeplastida *Ostreococcus tauri;* Protist (red): *Cryptosporidium hominis*). The maps produced in Fig. S13 show no indication of overlap or features indicative of contamination.

***H. catenoides* genome qPCR size estimation**

50 ml of a *H. catenoides* culture, grown in YpSs for 7 days at 25°C, was centrifuged for 3 mins at 3200 x *g*. The supernatant was removed and genomic DNA was extracted from the remaining cells using a PowerSoil DNA isolation kit (MO BIO Laboratories). The genome size *H. catenoides* was estimated using a qPCR based method to quantify the haploid genome size (26). An *rps3* PCR standard was amplified using primers Hcat\_rps3\_F (5’- CGAGGGCTACATGGTCAAGA-3’) and Hcat\_rps3\_R (5’-CCTTTGGCTCGATGATGGTG-3’). Each 25 μl reaction consisted of 0.5 U Phusion polymerase (New England Biolabs), 1x HF buffer, 400 μM dNTPs, 2 μM each primer and 1 μl *H. catenoides* genomic DNA (11.6 ng μl-1). Cycling conditions consisted of an initial denaturation of 5 mins at 98°C, followed by 30 cycles of 10 s at 98°C, 30 s at 61.0°C and 30 s at 72°C, then a final extension of 5 mins at 72°C. The 185 bp PCR product was purified by gel extraction (Thermo Scientific GeneJET Gel Extraction kit) and eluted using elution buffer. Concentration of the purified product was determined using a Qubit dsDNA HS assay kit (Thermo Fisher Scientific). Real-time PCR was used to quantify the number of copies of *rps3* present in each genomic DNA sample. Quantitative PCR was performed in a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific). Reaction conditions were optimised using a gradient PCR and a standard curve was determined using dilutions of *H. catenoides* genomic DNA and analysed using StepOne Software v2.3 (slope: -3.367; y-intercept: 33.841; efficiency: 98.15%). Each 20 μl PCR contained 10 μl PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), 500 nM each primer (Hcat\_rps3\_F and Hcat\_rps3\_R, sequences as above) and 1 μl template DNA. Template was either *H. catenoides* genomic DNA or the PCR standard. Standards were diluted (10-1 to 10-7) from an initial concentration of 24.7 ng μl-1 and performed in triplicate, while three independent genomic DNA samples were run in quintuplicate. Cycling conditions were as follows: UDG activation for 2 mins at 50°C and DNA polymerase activation for 2 mins at 95°C, followed by 40 cycles of 15 secs at 95°C and 1 min at 60°C. ROX was used as a reference dye for analysis of CT values. Each reaction was followed by melt-curve analysis, with a temperature gradient of 60°C to 95°C at 0.3°C s-1, to ensure presence of only a single amplicon. The PCR standards were used to create a calibration curve (*y = 8x1010 x e-0.67x;* R2 = 0.99992); CT values from amplifications of genomic DNA templates were then applied to this curve and the ‘mass’ of a haploid genome was calculated (26). This value was then used to calculate the haploid genome size, using 660 g mol-1 as the mean molar mass of a base pair.

**Mitochondrial genome assembly**

As indicated above the genome of *Hyphochytrium catenoides* was sequenced using Illumina HiSeq with Illumina GAiix and assembled using both Ray and Platanus. Contigs of putative mitochondrial origin, from both assemblies, were identified by BLAST searches against the mitochondrial genome of *Phytophthora infestans* (NC\_002387.1). The contigs from the genome assemblies were visualized, linked and edited using the program SEQUENCHER (<https://www.genecodes.com>) resulting in two contigs. However, we were unable to circularise the genome using these two fragments. Therefore, regions spanning the gaps in the mtDNA super-contigs were amplified by polymerase chain reaction (PCR) with primers specific to the flanking sequences. Purified PCR products were sequenced using Sanger chemistry (externally at Eurofins Genomics, Ebersberg). This allowed the two contigs to be joined, resulting in a linear genome flanked on one end with rpl16 and atp8 on the other. These genes were identical to the other rpl16 and atp8 genes found in the assembled mitochondrial genome; we therefore inferred that these represented the beginning and end of a 19 kb inverted repeat (Fig S8). Mitochondrial genes were identified and annotated using mfannot (<http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl> last accessed 20 June, 2017) followed by manual inspection. The putatively circular genome was visualized using CGView (137) (Figure S8).

**Search for *H. catenoides* representatives of key oomycete gene families**

Using Pfam searches (Pfam release 29.0) with default defined e-value cut-offs we searched the *H. catenoides* predicted proteome for: NPP1-like proteins (Pfam domain: PF05630), elicitin (PF00964), cutinase (PF01083), pectin methyl esterases (PF01095), pectate lyase (PF03211), polygalacturonase (PF00295) PAN lectin (PF00024), Ricin lectin (PF00652), Jacalin lectin (PF01419) Galactose binding lectin (PF00337), Legume lectin (PF00139), Legume-like lectin (PF03388), ABC transporters (PF00005), Protein kinase (PF00069 & PF07714), Notch protein (PF00066) and Haemolysin E (PF06109). In addition, the *H. catenoides* predicted proteome was searched against the MEROPS database <https://merops.sanger.ac.uk/> to identify putative protease inhibitors and proteases and the CAZymes analysis (138) toolkit (using Pfam) at <http://mothra.ornl.gov/cgi-bin/cat/cat.cgi?tab=PFAM1> to identify putative carbohydrate interacting proteins**.** Predicted proteins containing putative RxLR motifs and Crinkler domains were identified using the pipelines described in the following publications (139, 140).

**Secretome analysis**

Putatively secreted proteins were predicted using a custom pipeline (<https://github.com/fmaguire/predict_secretome/tree/refactor>) which predicts sequences as secreted by taking the intersection of those sequences predicted to have a signal peptide (via SignalP 4.1 (141)), no TM domains in their mature peptide (via TMHMM 2.0c (142, 143)), a signal peptide that targets for secretion (via TargetP (87)) and belonging to the extracellular ‘compartment’ (as predicted by WoLFPSort 0.2 (144)).  The CAZY database (145) was downloaded, converted into a BLAST-DB and searched using the predicted proteome and secretomes using BLASTp with an expectation of 1e-5. Hit tallies were then summed, proportions calculated and data plotted in Python via the Pandas and Seaborn packages (Fig. 2).

**Phylogenetic analysis of individual gene families**

Unless otherwise stated in the figure legends all phylogenetic analyses were conducted using the following protocols. Using BLASTp we used the seed sequence to identify putative homologues across a locally maintained database of eukaryotic and prokaryotic genome-derived protein datasets (Table. S3) with a gather threshold of 1e-10. The Multiple Sequence Comparison by Log-Expectation (MUSCLE) program (v3.8.31) (146) was used to produce a multiple sequence alignment for each set of proteins. Alignments were then manually corrected and masked in SeaView (version 4.2.4) (147). Sequences that required a high level of site exclusion (due to the sequence not aligning or not masking well) or where they formed long-branches in preliminary analysis were removed. The phylogenies were calculated using parallelised-PTHREADS RAxML (version 7.7 -Pfeiffer and Stamatakis 2010) with 1,000 (non-rapid) bootstrap replicates and using the substitution matrix and gamma distribution identified using ProtTest3 (version 3.2.1). In some cases, the invariant sites parameter was also included in the model (if indicated in the ProtTest3 analysis).

To identify putative orthologues that arose at the base of the Pseudofungi, gene clusters identified from 74 genomes (Table S11) were mapped onto the species phylogeny (e.g. Fig. 1a) using a pipeline described here (https://github.com/guyleonard/orthomcl\_tools & <http://dx.doi.org/10.5281/zenodo.51349)>.

**Multi-gene concatenated phylogenetic analysis to identify the branching position of *H. catenoides***

Using the methods established in (148) we built a concatenated amino acid alignment of 245 putative orthologues resulting in a masked data matrix of 85 taxa and 71,862 amino acid sites. This dataset encompassed a wide sampling of eukaryotes based on previous analysis (e.g. (31, 148)).  Using this alignment, we calculated a maximum likelihood (ML) tree using  RAxML (149) with 60 ML searches under PROTGAMMALGF model of sequence substitution. 1000 boostrap replicates were then calculated using the IQ-TREE software using the LG\_G4+C20+F substitution model.

**Identification of genes of plastid ancestry**

We used a sequence similarity search combined with phylogenetic reconstruction pipeline (e.g. (150)) to identify gene families which include *H. catenoides* genes which have a phylogenetic ancestry demonstrating the gene was derived from the cyanobacterial ancestor that gave rise to the plastid organelle. Initially we constructed a database of (Table S5) genomes and clustered these genes into putative orthologues using OrthoMCL (62) finally retaining only the groups containing *H. catenoides* genes. Next we resampled sequences from a wider database of 1205 taxa (Table S3) using BLASTp searches (151) to recover up to three sequences from each genome using a gathering threshold of 1e-10. We then filtered these clusters identifying only those only containing both a *H. catenoides* gene and genes from photosynthetic or ancestrally-photosynthetic eukaryotic taxa. These sequences were then aligned using MAFFT (152), masked using TRIMAL (153) and a phylogeny was calculated from the data matrix using FASTTREE2 (154). The resulting phylogenies were manually inspected for a phylogeny that showed *H. catenoides*/Pseudofungal/Stramenopile genes which: a) branched within the Archaeplastida radiation, b) branched with genes of photosynthetic eukaryotes and within a bacterial radiation, or c) branched with cyanobacterial genes. This process required re-running of the phylogenetic pipeline for many gene clusters either reducing gene sampling or removing long-branch sequences. A subset of 101 gene cluster phylogenies putatively showed a phylogenetic relationship consistent with criteria *a-c* described above. The alignments from these clusters were then manually refined, the taxon sampling checked using manual BLAST searches of the NCBI nr database and phylogenies recalculated using the RAxML approach described above.

**Testing for CYP51 sterol-demethylase drug sensitivity**

Azole susceptibility was assessed using a modified protocol of (37). Briefly, fluconazole and clotrimazole were dissolved in dimethyl sulfoxide (DMSO) to a stock concentration of 25.6 mg ml-1. Dilutions were then made with DMSO to prepare 100x stock solutions. These stocks were diluted in PYG medium to a final volume of 5 ml, each containing 100 µl of *H. catenoides* liquid culture (grown in YpSs at 25°C shaking for 7 days) to achieve final azole concentrations of 256, 128, 64, 32, 16, 8, 4, 2, 1, 0.5, and 0.25 μg ml-1 and with control samples containing 1% (v/v) DMSO. Cultures were incubated, in triplicate, for 7 days at 25°C with 200 rpm shaking and MIC100 was scored manually by assessing for presence/absence of hyphal growth.

**OmniLog ‘Phenotype Microarrays’**

100 ml *H. catenoides* cultures were grown in PYG (1.25 g L-1 peptone, 1.25 g L-1 yeast extract, 3 g L-1 glucose) in baffled flasks, 25°C 170 rpm shaking, to minimise aggregation. Cells were recovered by centrifugation at 3200 x *g*, washed twice with water and re-suspended in PY-G (as above, no carbon-source) to a final concentration of approximately 1.5 x 103 cells ml-1. Cells were allowed to recover at 25°C with shaking for 30 minutes before Dye mix D (Biolog) was added to a 1x final concentration. 100 µl of cells was inoculated into each well of PM1 and PM2 carbon-source plates and incubated for 7 days at 25°C. Each growth assay was performed in triplicate from independent cultures.

OmniLog Phenotype Microarray outputs were analysed using OPM (155). Data were aggregated using the ‘opm-fast’ method, analysed using the A parameter (maximum value of OmniLog units reached) and tested by t-test. Significant p-values were extracted if they resulted in increased growth in comparison to the negative control well A01.

**Confirmation of viral genes in the *H. catenoides* assembly and reverse transcriptase PCR of viral genes**

To confirm that the viral genes were assembled correctly and were resident in the *H. catenoides* genome fragments, PCRs across the 3’ and 5’ junctions of the putative viral open reading frame for three of the viral genes *polB*, *MCP*, and *mg96* were performed. 25 µl PCR reactions (1x Phusion HF buffer, 400 µM dNTP mix, 200 nM each primer, 0.5 U Phusion polymerase) were performed with the following cycling conditions: initial denaturation of 5 mins at 98°C, followed by 30 cycles of 10 s at 98°C, 30 s at 56-64°C and 1 min at 72°C, then a final extension of 5 mins at 72°C. These were purified using a GeneJET PCR Purification Kit or GeneJET Gel Extraction kit (Thermo Scientific) and sequenced to confirm that each product matched the expected amplicon. To confirm that the *mcp* gene was on the same contig as the histone H3 gene, we performed a PCR across these two genes (expected amplicon of 2837 bp) using the same conditions as above, except with an annealing temperature of 64 °C and with a 3-min extension. The PCR product was purified and A-tailed using *Taq* polymerase; then cloned using the StrataClone PCR Cloning Kit (Agilent Technologies). The resulting vector was sequenced using T3/T7 primers, with primer-walking to confirm the entire 2. 8 kb sequence.

To investigate if the viral derived genes are actively transcribed in our culture conditions, we conducted RT-PCRof the *polB, mcp*, *mg96* and *rps3* virus confirming *polB, mg96 and rps3* are expressed in our culture conditions suggesting that the virus is transcriptionally active. RNA was extracted from *H. catenoides* using RNA PowerSoil Total RNA Isolation (MoBio). Residual genomic DNA was removed using RQ1 RNase-Free DNase (Promega) and Taq PCR was performed to confirm absence of DNA. Reverse-transcriptase PCR was then performed using a Qiagen OneStep kit according to the manufacturer’s instructions, alongside genomic DNA positive and no-template controls. The following cycling conditions were used: reverse-transcriptase of 30 mins at 50°C and initial denaturation 15 minutes at 94°C, followed by 32 cycles of 1 min at 94°C, 1 min at 50°C and 1 min at 72°C, then a final extension of 10 mins at 72°C. Samples were then analysed on a 2% (w/v) agarose gel.

**WGA staining**

*H. catenoides* was grown for 7 days at 25°C and 100 µl of mycelial growth was removed and suspended in 1 ml PBS. 5 µg ml-1 calcofluor white (Fluka) and 10 µg ml-1 WGA, Alexa Fluor 488 conjugate (Invitrogen) were added and cells were incubated for 30 minutes in the dark. Cells were washed twice in PBS and imaged using an Olympus IX73 microscope on a 40x objective. Unstained cells were also checked to confirm the absence of autofluorescence.

**Figure Legends**

**Fig. 1. Concatenated multi gene phylogeny and genome statistics of representative stramenopiles. (A)** Subsection of a 245 orthologous multi-gene concatenated phylogeny (85 taxa and 71,862 amino acid sites) demonstrating the branching position of *Hyphochytrium* within the stramenopiles calculated using Maximum Likelihood methods*.* See <https://github.com/guyleonard/hyphochytrium/blob/master/manuscript/data/data_s1_fig_1a_hyphochytrium_phylogeny.tree> for the full tree file. Dots on the branches of the phylogenetic tree indicate 100% bootstrap support, values below 50% have been removed. Cartoons of cells indicate change in only stramenopile flagellum morphology over the tree. Key indicates generalized lifestyle of the taxa sampled. **(B)** Table of genome statistics for a range of different stramenopiles.

**Fig. 2 Comparison of secreted proteome and putative carbohydrate active proteins across the Pseudofungi including an outgroup of photosynthetic stramenopile taxa.** The schematic phylogeny at the top indicates the relationship between different oomycete species with ‘lifestyle’ of each species indicated by text colour; green (*Phytophthora species*) indicates hemibiotrophy, blue (*Hyaloperonospora* and *Albugo*) obligate biotrophy, teal (*Pythium*) necrotrophy, orange (*Saprolegnia*) saprotrophy/necrotrophy, and black indicates free living (e.g. *Hyphochytrium*, *Ectocarpus* and *Thalassiosira*). The first heat map in white/purple indicates the proportion of proteome of each organism which was identified as belonging to a particular CAZY ([www.cazy.org](http://www.cazy.org/)) category using BLASTP with an expectation of 1e-5. The number listed is the proportion and the colour relates to magnitude of the listed number (as shown by scale bar). The second heatmap, in blue/yellow, indicates the proportion of the secretome (predicted via a custom pipeline <https://github.com/fmaguire/predict_secretome/tree/refactor>) that is identified as belonging to each of these CAZY categories. The bar chart shows the proportion of the proteome for each organism which is predicted to be secreted by the above-mentioned pipeline.

**Fig. 3. Phylogeny of viral MCP proteins indicating the branching position of the Pseudofungal genes and evidence of transcription of viral derived genes in *H. catenoides*. (A)** Homologous sequences were identified using 3 psi-BLAST iterations with *H. catenoides* putative MCP as query; to remove sequence redundancies, retrieved sequences were clustered at 90% amino-acid identity with cd-hit v4.6. Sequences were then aligned using MAFFT v7 iterative, global homology mode (G-INS-i); alignment sites retained for subsequent phylogenetic analysis were selected using trimAL (153) gap distribution mode. Final MCP multiple sequence alignment was composed of 386 sites. ML tree was inferred using iq-tree v1.3 and LG+I+G4+F model (determined as the best-fitting model by Bayesian information criterion). Node supports were evaluated with 100 non-parametric bootstrap replicates. The Mimiviridae clade was used to root the ML tree (unrooted version displayed on the lower left part). B. Reverse transcriptase (RT) PCR showing expression of polB and mg96 viral genes alongside an rps3 positive control. No expression of the *mcp* gene was detected. RT-PCR was performed on *H. catenoides* RNA alongside genomic DNA (+) and no template (-) controls, with PCR products run on an agarose gel alongside a 1kb ladder (Promega; 250 bp shown).

**Fig. 4. Comparative genomic analysis of *H. catenoides* flagellum proteome and motor protein repertoire. (A)** Heat map showing sequence identity profiles for flagella proteins with putative homologues present across the eukaryotes (see Table S8 for full dataset). The proteins show proteome conservation patterns consistent with retention or loss of proteins that function in the anterior-tinselated flagellum. The heat map identifies 29 proteins present in the oomycetes but absent in *H. catenoides,* suggesting that this gene had been lost at the same proximate point to the loss of the posterior flagellum. The analysis also shows 12 proteins (marked\*) identified as posterior flagellum specific in *C. bullosa* that are retained in *H. catenoides***.** Three *C. bullosa* anterior flagellum specific proteins are also retained in *H. catenoides*. The putative radial spoke proteome also shows numerous losses similar to *Homo sapiens* (\*\*), this includes the loss of RSP7 (\*\*\*). Only changes in flagella cytology relevant to the evolution of the stramenopiles are sketched on the top tree. **(B)** Shows a cartoon of the radial spoke protein complex identified in *Chlamydomonas* with each shape number referring to the RPS number (91)*.* Black shapes illustrate conserved proteins of the spoke complex, grey are non-conserved proteins, while the white complex refers to RPS7 which although absent in *Homo sapiens* has been lost separately when the posterior flagellum was also lost in *H. catenoides*. **(C)** Distribution of major kinesin paralogue families. Kinesin-2, -9, -16 & -17 have been suggested to have function associated with the flagellum (95). **(D)** Distribution of major Dynein paralogue families. Paralogues are grouped by functional annotation category: Dynein Heavy Chain (DHC), Intermediate Chain (IC), Light Intermediate Chain (LIC) and Intra-flagellar Transport (IFT). **(E)** Distribution of major Myosin paralogue families focusing on variation between fungi and Pseudofungi. **(F)** Different myosin domain architectures identified in the Stramenopiles including *H. catenoides*.

**Fig. 5. Comparative genomic analysis of gene families that function in polarised filamentous growth in the Fungi. (A)** Cartoon outlining proteins and complexes involved in polarised growth in *Saccharomyces cerevisiae* (this is a variation of a figure shown in (156)).Vesicles are delivered from the Golgi (Fig. 5A-i) along cytoskeleton tracks to predetermined sites on the plasma membrane. Cdc42p is activated by Cdc24p (Fig. 5A-ii) promoting (104) assembly of the Polarisome complex (Fig. 5A-iii) resulting in the formin Bni1p radiating actin cables (157, 158). Msb3p and Msb4p interact with Spa2 in the Polarisome (Fig. 5A-iv) which is thought to recruit Cdc42 from the cytosol at the site of tip growth (159). Post-Golgi secretory vesicles are transported along actin cables using a type V myosin motor protein (160, 161) (Fig. 5A-v), to dock with the Exocyst complex in a process dependent on Sec4 and its GEF Sec2 (162, 163)(Fig. 5A-vi) and so the vesicle is guided to its target site on the plasma membrane (164). Cdc42p and Rho1 are required for localisation of Sec3p, which together form a spatial marker for the Exocyst (Fig. 5A-vii) and Rho3p and Cdc42p mediate vesicle docking (Fig. 5A-viii). Cdc42p plays a key role in regulating these processes in *S. cerevisiae* but in Pezizomycotina and basidiomycete fungi equivalent functions are performed by Rac1p (165, 166). **(B)** The domain architecture of the 17 proteins associated with polarised growth in fungi. **(C)** The taxon distribution of putative homologues of polarised growth proteins across a representative set of taxa including the Pseudofungi. ‘P’ indicates a putative paralogue relationship as identified using phylogenetic analysis.

**Fig. 6. Comparative genomic analysis of gene families that function in cell wall synthesis. (A)** Micrographs showing the Wheat Germ Agglutinin fluorescentstaining of a chitin cell wall on *Hyphochytrium* developmental structures. **(B)** The domain architecture of eight proteins that function in cell wall synthesis. **(C)** The taxon distribution of putative gene families associated with cell wall synthesis across a representative set of taxa including the Pseudofungi.

**Fig. 7. Phylogenetic tree summarising the evolution of cell and genomic characters relevant to the evolution of the Pseudofungi*.***Only changes in flagella complement relevant to the evolution of the stramenopiles are sketched.

**Fig. S1. Comparison of transcriptome to genome assembly.** Scaffolds were sorted in to four groups; all, >1 kbp, >5 kbp and >10 kbp in length. For each dataset, the N50 statistic was computed, along with gene predictions, CEGMA results and ‘BLAT’ hits against the transcriptome. The data suggests that removal of scaffolds <1 kbp does not have an appreciable effect on the “completeness” of the *Hyphochytrium* gene set.

**Fig. S2. Identification of SNP frequency mapped to individual scaffolds.** A plot of SNP frequency for each scaffold of the draft genome arranged by length. SNPs are more common in shorter scaffolds. Blue line indicates trend-line with standard deviation. Two aberrant scaffolds with low SNP frequency are identified.

**Fig. S3. K-mer coverage of genome assembly showing two peaks indicative of a diploid genome.** A GenomeScope graph of the k-mer profile generated by using Jellyfish displaying k-mer coverage vs frequency. The two k-mer peaks are indicative of a diploid genome.

**Fig. S4. Phylogeny indicating the branching position of the *Hyphochytrium* putative CYP51 sterol-demethylase protein and results of drug sensitivity assay.**

**Fig. S5. Omnilog.**

**Fig. S6. Histone H3 phylogeny.**

ML reconstruction based on a gene alignment (too conserved at the aa level). Purple for alveolates and blue for stramenopiles, H3 from Hyphochytrium scaffolds displayed in red with scaffold id and H3 coordinates on scaffold. A red diamond indicates the Hyphochytrium scaffold harboring Asfarviridae-like MCP gene is marked with Other accession numbers are RefSeq. Lots of homologs, reduced redundancy by 95% global identity clustering. Alignment with mafft (ginsi mode) curated with trimal gap distribution algo (411 nt sites, 164 sequences). Reconstruction with IQTREE v1.5 with 100 non-parametric bootstrap replicates. Best fit model was GTR+I+G4.

**Fig. S7. Representative *Hyphochytrium* TEM images showing absence of evidence of viral particles or factory.** This figure shows six imaged cells (A, C, G, J, K & L, with micrographs: B, D, E, F, H, I) after twenty cells on one slide were subject to cursory inspection. The cultures were grown in PYG medium at 25°C, 170 rpm and samples then underwent high pressure freezing using a Leica EM RTS and were left to undergo free substitution using an infiltration mixture of 0.1% uranyl acetate and 1% osmium tetroxide as the fixative overnight. The fixative was removed and the concentration of EPON resin was increased in over 6 increments until embedding occurred at 100% EPON. The resultant palate was cut using a diamond knife and exposed to lead citrate before being mounted on a standard copper grid. All images were obtained using a Jeol JEM 1400 transmission electron microscope.

**Fig. S8. The mitochondrial genome of *H. catenoides*.** Ribosomal genes (pink), tRNA genes (red), and protein coding genes (blue) are represented around the mitochondrial genome (black). Genes represented on the outside are encoded clockwise whereas genes represented on the inside are encoded counterclockwise. Small and large ribosomal protein genes have been abbreviated ‘s’ and ‘L’, respectively. The inferred inverted repeat regions are overlayed with green arrows. The inner circle depicts the original 48,124 bp assembly (red), the inferred ~19 kb inverted repeat (green), and the overlap (grey).

**Fig. S9. ATP1 gene phylogeny showing the presence of a divergent nuclear and mitochondrially encoded proteins in the stramenopiles.** The nuclear encoded *H. catenoides* Atp1-like protein sequence was used as a query in BLAST and pHMMer searches. Sequences retrieved were aligned with sequences from (79) using MUSCLE. Sequences were subjected to phylogenetic analysis using RaxML with 500 bootstraps using the LG substitution matrix. Atp1-like sequences predicted to be targeted to mitochondria by TargetPv1.1 are indicated by an encircled "M". Support represented as inset. Bar represents substitutions per site. *H. catenoides* sequences are in red.

**Fig. S10. Four phylogenies showing pseudofungal protein sequences branching with plastid carrying eukaryotes nested within a cluster of bacterial sequences potentially representing a gene derived from the plastid endosymbiosis.** All phylogenies were calculated using Maximum Likelihood methods with 1000 bootstrap replicates as described in the methods.Sequences from all taxa possessing plastid of primary endosymbiotic ancestry (i.e. Archaeplastida) were searched using TargetP for evidence of an N-terminal organelle targeting peptide. Alignments are provided at [https://github.com/guyleonard/hyphochytrium/blob/master/manuscript/data/](https://github.com/guyleonard/hyphochytrium/blob/master/manuscript/data/data_s1_fig_1a_hyphochytrium_phylogeny.tree). **(A)** Pas -Histidine Kinase domain containing protein phylogeny calculated from a data matrix of 46 sequences and 281 alignment positions. **(B)** Cytidylytransferase family domain containing protein phylogeny calculated from a matrix of 38 sequences and 178 alignment positions. **(C)**  Phylogeny of a conserved hypothetical protein calculated from a data matrix of 41 sequences and 143 alignment positions. **(D)** Phylogeny of a putative Heme oxygenase domain calculated from a data matrix of 49 sequences and 186 alignment positions.

**Fig. S11 Photoreceptors putatively encoded on the *Hyphochytrium* genome. (A)** Distribution of known photoreceptor families across a subset of stramenopile genomes. **(B)** Subsection of the Rhodopsin alignment showing the conserved Schiff base. **(C)** Phylogeny of the rhodopsin gene family showing phylogenetic position of the *Hyphochytrium* putative rhodopsins.

**Fig. S12. ‘Blobology’ analysis of genome assembly of composition and likely taxonomic association. ‘**Blobplots’ visualize genomes by displaying putative taxon-annotated GC-coverage plots. **(A/B)** These plots show all scaffolds from the draft assembly with BLAST annotation from the NCBI ‘nt’ database, in **(B)** the transcriptome of *Hyphochytrium* and its mitochondria have been included. Notice the small circle of *Hyphochytrium* annotated blobs in the upper-left of the image, these aberrant GC contigs are the mitochondria. **(C/D)** Shows the same images, however, only scaffolds greater than 1kbp have been included.

**Fig. S13. Emergent properties graphs for assessment of contamination in the *Hyphochytrium* genome assembly.** Two emergent self-organizing maps show the scaffold data plotted as tetra-mer frequencies represented by a neural network mapping approach. **(A)** To determine contamination, the map was run on the *Hyphochytrium* scaffolds and does not show extensive regions of variance. **(B)** To further the contamination check, other eukaryotic genomes and prokaryotes were spiked in to the process, here we see clear separation of all the genomes. *Hyphochytrium scaffolds* (orange), (Bacteria (blue): *E. coli*, *Mycobacterium tuberculosis*; Archaea (grey): *Methanococcus vanniellii, Sulfolobus solfataricus*; Fungi (purple): *Encephalitozoon intestinalis, S. cerevisiae*; Archaeplastida (green): *Ostreococcus tauri*; Protist (red): *Cryptosporidium hominis*)

­

**Table 1.** Comparison of pseudofungal genes with generalised function.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene families** | *Hyphochytrium catenoides* | *Albugo laibachii* | *Hyaloperonospora arabidopsidis* | *Phytophthora infestans* | *Phytophthora ramorum* | *Phytophthora sojae* | *Pythium ultimum* | *Saprolegnia parasitica* | *Ectocarpus siliculosus* | *Thalassiosira pseudonana* |
| RXLR | 1 | 0 | 23 | 317 | 102 | 106 | 0 | 0 | 0 | 0 |
| NPP1-like proteins | 0 | 0 | 21 | 27 | 62 | 74 | 7 | 0 | 0 | 0 |
| Elicitin | 0 | 9 | 14 | 43 | 47 | 53 | 44 | 25 | 0 | 0 |
| **Plant Cell wall**  **degrading:** |  |  |  |  |  |  |  |  |  |  |
| Cutinase | 0 | 3 | 2 | 4 | 4 | 16 | 0 | 0 | 0 | 0 |
| Glycosyl hydrolases | 357 | 384 | 242 | 533 | 838 | 1208 | 436 | 415 | 282 | 264 |
| Pectin methyl esterases | 0 | 0 | 4 | 11 | 13 | 19 | 0 | 0 | 0 | 0 |
| Pectate lyase | 0 | 0 | 8 | 36 | 25 | 24 | 16 | 0 | 0 | 0 |
| Polygalacturonase | 0 | 3 | 3 | 24 | 17 | 25 | 6 | 3 | 0 | 0 |
| **Lectins** |  |  |  |  |  |  |  |  |  |  |
| PAN lectin | 4 | 3 | 2 | 5 | 8 | 5 | 11 | 6 | 1 | 0 |
| Ricin lectin | 1 | 1 | 3 | 5 | 9 | 10 | 5 | 57 | 0 | 1 |
| Jacalin lectin | 2 | 0 | 8 | 15 | 23 | 15 | 3 | 4 | 1 | 0 |
| Galactose lectin | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Leguminous lectin | 2 | 1 | 0 | 2 | 1 | 1 | 2 | 0 | 2 | 1 |
| Legume-like lectin | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 |
| **Protease functions** |  |  |  |  |  |  |  |  |  |  |
| Protease inhibitors | 13 | 11 | 14 | 51 | 35 | 46 | 30 | 28 | 15 | 15 |
| Proteases, all | 428 | 379 | 324 | 450 | 541 | 602 | 482 | 630 | 361 | 367 |
| Serine proteases | 166 | 84 | 106 | 170 | 182 | 189 | 200 | 248 | 112 | 140 |
| Metalloproteases | 92 | 91 | 80 | 98 | 100 | 91 | 107 | 129 | 88 | 101 |
| Cysteine proteases | 115 | 124 | 92 | 140 | 116 | 113 | 121 | 208 | 117 | 85 |
| **Others** |  |  |  |  |  |  |  |  |  |  |
| ABC transporters | 81 | 36 | 49 | 148 | 171 | 175 | 158 | 138 | 70 | 58 |
| Protein kinases | 243 | 305 | 217 | 423 | 398 | 430 | 232 | 690 | 330 | 160 |
| Notch protein | 3 | 0 | 1 | 1 | 1 | 1 | 1 | 18 | 11 | 2 |
| Haemolysin E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |

**Table S1. Meiosis specific gene families and their presence in *H. catenoides***

**Table S2. Pseudofungal domain architectures**

**Table. S3. Genomes used for standard phylogenetic analyses**

**Table S4. Putative orthologue families unique to Pseudofungi**

**Table S5. Putative genes of viral ancestry in the *H. catenoides* assembly**

**Table S6. Results and primers used to confirm viral presence in the *H. catenoides* genome**

**Table S7. Contigs containing viral genes and low number of SNPs.**

**Table S8. Identification of putative homologues of known flagellum associated proteins.** This analyses includes comparison with the *Colpomenia bullosa* flagella proteomes (88)

**Table S9. List of genomes used in contamination comparisons**

**Table S10. Possible contamination scaffolds**

**Table S11. Genomes used for orthologue identification**

**References**

1. Patterson DJ. Stamenopiles: chromophytes from a protistan perspective. In: Green JC, Leadbeater BSC, Diver WL, editors. The chromophyte algae, problems and perspectives. Oxford: Claredon; 1989. p. 357-79.

2. Cavalier-Smith T. The kingdom Chromista: origin and systematics. In: Round FE, Chapman DJ, editors. Progr Phycol Res. 4. Bristol: Biopress Ltd; 1986. p. 309-47.

3. Cavalier-Smith T, Chao EE. Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). J Mol Evol. 2006;62(4):388-420.

4. Cavalier-Smith T. The origin of fungi and pseudofungi. In: Rayer ADM, editor. Evolutionary Biology of the Fungi (British Mycological Society Symposia). Cambridge: Cambridge University Press; 1987. p. 339-53.

5. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc Roy Soc B. 2016;283(1823).

6. Cooney EW, Barr DJS, Barstow WE. The ultrastructure of the zoospore of *Hyphochytrium catenoides*. Can J Bot. 1985;63(3):497-505.

7. Archibald JM. The puzzle of plastid evolution. Curr Biol. 2009;19(2):R81-8.

8. Derelle R, López-García P, Timpano H, Moreira D. A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). Mol Biol Evol. 2016;33(11):2890-8.

9. Cavalier-Smith T. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 2000;5(4):174-82.

10. Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, et al. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol. 2010;27(7):1698-709.

11. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat Commun. 2014;5.

12. Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, et al. Chromera velia, endosymbioses and the Rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH Lineages). Genome Biol Evol. 2014;6(3):666-84.

13. Dorrell RG, Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J., Bowler, C. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. eElife. 2017;6:e23717.

14. Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C. Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? BMC Genomics. 2009;10:484-.

15. Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, et al. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl Environ Microbiol. 2004;70(6):3528-34.

16. Richards TA, Bass D. Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. Curr Opin Microbiol. 2005;8(3):240-52.

17. Armbrust E, Berges J, Bowler C, Green B, Martinez D, Putnam N, et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science. 2004;306:79 - 86.

18. Cock J, Sterck L, Rouze P, Scornet D, Allen A, Amoutzias G, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature. 2010;465:617 - 21.

19. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 2006;313(5791):1261-6.

20. Jiang RHY, de Bruijn I, Haas BJ, Belmonte R, Löbach L, Christie J, et al. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. PLoS Genet. 2013;9.

21. Barr D. *Hyphochytrium catenoides*: a morphological and physiological study of North American isolates. Mycologia. 1970;62:492-503.

22. Karling JS. A new fungus with anteriorly uniciliate zoospores: *Hyphochytrium catenoides*. Am J Bot. 1939;26(7):512-9.

23. Van der Auwera G, De Baere R, Van de Peer Y, De Rijk P, Van den Broeck I, De Wachter R. The phylogeny of the Hyphochytriomycota as deduced from ribosomal RNA sequences of *Hyphochytrium catenoides*. Mol Biol Evol. 1995;12(4):671-8.

24. Jurka J, Kapitonov V, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462 - 7.

25. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764-70.

26. Wilhelm J, Pingoud A, Hahn M. Real-time PCR-based method for the estimation of genome sizes. Nucleic Acids Res. 2003;31(10):e56-e.

27. Ramesh MA, Malik SB, Logsdon JM, Jr. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. Curr Biol. 2005;15:185-91.

28. Johnson TW, Jr. Resting spore development in the marine phycomycete Anisolpidium ectocarpii. Am J Bot. 1957;44(10):875-8.

29. Beakes G, Glockling S, Sekimoto S. The evolutionary phylogeny of the oomycete “fungi”. Protoplasma. 2012;249(1):3-19.

30. khmer-protocols documentation [Internet]. 2013.

31. Noguchi F, Tanifuji G, Brown MW, Fujikura K, Takishita K. Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage stramenopiles. Mol Phylogenet Evol. 2016;101:133-41.

32. Tsui C, Marshall W, Yokoyama R, Honda D, Lippmeier J, Craven K, et al. Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. Mol Phylogenet Evol. 2009;50:129 - 40.

33. Oudot-Le Secq M-P, Loiseaux-de Goër S, Stam WT, Olsen JL. Complete mitochondrial genomes of the three brown algae (Heterokonta: Phaeophyceae) *Dictyota dichotoma*, *Fucus vesiculosus* and *Desmarestia viridis*. Curr Genet. 2005;49(1):47-58.

34. Philippe H. Opinion: long branch attraction and protist phylogeny. Protist. 2000;151(4):307-16.

35. Gaulin E, Bottin A, Dumas B. Sterol biosynthesis in oomycete pathogens. Plant Signal Behav. 2010;5(3):258-60.

36. Gaulin E, Madoui M, Bottin A, Jacquet C, Mathe C, Couloux A, et al. Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways. PLoS ONE. 2008;3:e1723.

37. Warrilow AGS, Hull CM, Rolley NJ, Parker JE, Nes WD, Smith SN, et al. Clotrimazole as a potent agent for treating the oomycete fish pathogen *Saprolegnia parasitica* through inhibition of sterol 14α-demethylase (CYP51). Appl Environ Microbiol. 2014;80(19):6154-66.

38. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, et al. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science. 2010;330(6010):1549-51.

39. Raffaele S, Win J, Cano L, Kamoun S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics. 2010;11(1):637.

40. Levesque CA, Brouwer H, Cano L, Hamilton J, Holt C, Huitema E, et al. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol. 2010;11.

41. Haas B, Kamoun S, Zody M, Jiang R, Handsaker R, Cano L, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. 2009;461:393 - 8.

42. Birch P, Rehmany A, Pritchard L, Kamoun S, Beynon J. Trafficking arms: oomycete effectors enter host plant cells. Trends Microbiol. 2006;14:8 - 11.

43. Birch P, Armstrong M, Bos J, Boevink P, Gilroy E, Taylor R, et al. Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. J Exp Bot. 2009;60:1133 - 40.

44. Dou D, Kale S, Wang X, Jiang R, Bruce N, Arredondo F, et al. RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. Plant Cell. 2008;20:1930 - 47.

45. Dou D, Kale S, Wang X, Chen Y, Wang Q, Wang X, et al. Carboxy-terminal motifs common to many oomycete RXLR effectors are required for avirulence and suppression of BAX-mediated programmed cell death by Phytophthora sojae effector Avr1b. Plant Cell. 2008;20:1118 - 33.

46. Qutob D, Kemmerling B, Brunner F, Kufner I, Engelhardt S, Gust AA, et al. Phytotoxicity and innate immune responses induced by Nep1-like proteins. Plant Cell. 2006;18(12):3721-44.

47. Fellbrich G, Romanski A, Varet A, Blume B, Brunner F, Engelhardt S, et al. NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. Plant J. 2002;32(3):375-90.

48. Yu LM. Elicitins from *Phytophthora* and basic resistance in tobacco. Proc Natl Acad Sci USA. 1995;92(10):4088-94.

49. Belbahri L, Calmin G, Mauch F, Andersson J. Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate Phytophthora virulence factor. Gene. 2008;408:1 - 8.

50. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, et al. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. PLoS Biol. 2011;9.

51. Blackman LM, Cullerne DP, Hardham AR. Bioinformatic characterisation of genes encoding cell wall degrading enzymes in the *Phytophthora parasitica* genome. BMC Genomics. 2014;15(1):1-24.

52. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):D490-D5.

53. Bochner BR, Gadzinski P, Panomitros E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. 2001;11(7):1246-55.

54. Taoka Y, Nagano N, Okita Y, Izumida H, Sugimoto S, Hayashi M. Effect of Tween 80 on the growth, lipid accumulation and fatty acid composition of T*hraustochytrium aureum* ATCC 34304. J Biosci Bioeng. 2011;111(4):420-4.

55. Khalil S, Alsanius BW. Utilisation of carbon sources by *Pythium*, *Phytophthora* and *Fusarium* species as determined by Biolog(®) microplate assay. Open Microbiol J. 2009;3:9-14.

56. Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS One. 2013;8(10):e75072.

57. Zerillo MM, Adhikari BN, Hamilton JP, Buell CR, Lévesque CA, Tisserat N. Carbohydrate-active enzymes in *Pythium* and their role in plant cell wall and storage polysaccharide degradation. PLoS ONE. 2013;8(9):e72572.

58. Richards TA, Soanes DM, Jones MD, Vasieva O, Leonard G, Paszkiewicz K, et al. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc Natl Acad Sci USA. 2011;108(37):15258-63.

59. Seidl MF, Van den Ackerveken G, Govers F, Snel B. A domain-centric analysis of oomycete plant pathogen genomes reveals unique protein organization. Plant Physiology. 2011;155(2):628-44.

60. Doolittle RF. The multiplicity of domains in proteins. Annu Rev Biochem. 1995;64:287-314.

61. Leonard G, Richards TA. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. Proc Natl Acad Sci USA. 2012;109(52):21402-7.

62. Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13(9):2178-89.

63. Gallot-Lavallée L, Blanc G. A Glimpse of Nucleo-Cytoplasmic Large DNA Virus Biodiversity through the Eukaryotic Genomics Window. Viruses. 2017;9(1):17.

64. Delaroque N, Boland W. The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. BMC Evol Biol. 2008;8:110-.

65. Savory F, Leonard G, Richards TA. The role of horizontal gene transfer in the evolution of the oomycetes. PLoS Pathog. 2015;11(5):e1004805.

66. Monier A, Pagarete A, de Vargas C, Allen MJ, Read B, Claverie J-M, et al. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. Genome Res. 2009;19(8):1441-9.

67. Monier A, Welsh RM, Gentemann C, Weinstock G, Sodergren E, Armbrust EV, et al. Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. Environ Microbiol. 2012;14(1):162-76.

68. Doolittle WF. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 1998;14(8):307-11.

69. Misner I, Blouin N, Leonard G, Richards TA, Lane CE. The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. Genome Biol Evol. 2014;7(1):120-35.

70. Luis P, Gauthier A, Trouvelot S, Poinssot B, Frettinger P. Identification of *Plasmopara viticola* genes potentially involved in pathogenesis on grapevine suggests new similarities between oomycetes and true Fungi. Phytopathology. 2013;103(103):10.

71. Torto T, Rauser L, Kamoun S. The pipg1 gene of the oomycete *Phytophthora infestans* encodes a fungal-like endopolygalacturonase. Curr Genet. 2002;40:385 - 90.

72. Morris PF, Schlosser LR, Onasch KD, Wittenschlaeger T, Austin R, Provart N. Multiple horizontal gene transfer events and domain fusions have created novel regulatory and metabolic networks in the oomycete genome. PLoS ONE. 2009;4.

73. Whitaker JW, McConkey GA, Westhead DR. The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biol. 2009;10(4):R36.

74. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, et al. Endosymbiotic origin and differential loss of eukaryotic genes. Nature. 2015;524(7566):427-32.

75. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol. 2012;13(12):R122.

76. GCC. Sequencher. 4.2 ed. Ann Arbor: Gene Codes Corporation; 2003.

77. McNabb SA, Eros RW, Klassen GR. Presence and absence of large inverted repeats in the mitochondrial DNA of Hyphochytriomycetes. Can J Bot. 1988;66(12):2377-9.

78. O'nrien MA, Misner I, Lane CE. Mitochondrial genome sequences and comparative genomics of *Achlya hypogyna* and *Thraustotheca clavata*. J Eukaryot Microbiol. 2014;61(2):146-54.

79. Ševčíková T, Klimeš V, Zbránková V, Strnad H, Hroudová M, Vlček Č, et al. A comparative analysis of mitochondrial genomes in eustigmatophyte algae. Genome Biol Evol. 2016;8(3):705-22.

80. Reyes-Prieto A, Moustafa A, Bhattacharya D. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr Biol. 2008;18(13):956-62.

81. Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. Genome Biol. 2004;5(11):R88.

82. Riordan CE, Ault JG, Langreth SG, Keithly JS. *Cryptosporidium parvum Cpn60* targets a relict organelle. Curr Genet. 2003;44(3):138-47.

83. Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, Gillin FD, et al. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA. 1998;95(1):229-34.

84. Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. A mitochondrial HSP70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Curr Biol. 1997;7(12):995-8.

85. Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol. 1998;118(1):9-17.

86. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA. 2002;99(19):12246-51.

87. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000;300(4):1005-16.

88. Fu G, Nagasato C, Oka S, Cock JM, Motomura T. Proteomics analysis of heterogeneous flagella in brown algae (stramenopiles). Protist. 2014;165(5):662-75.

89. Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, et al. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci USA. 2015;112(7):E693-E9.

90. Wirschell M, Olbrich H, Werner C, Tritschler D, Bower R, Sale W, et al. The nexin-dynein regulatory complex subunit DRC 1 is essential for motile cilia function in algae and humans. Nat Genet. 2013;45(3):10.1038/ng.2533.

91. Yang P, Diener DR, Yang C, Kohno T, Pazour GJ, Dienes JM, et al. Radial spoke proteins of Chlamydomonas flagella. Journal of Cell Science. 2006;119(6):1165-74.

92. Yang C, Compton MM, Yang P. Dimeric novel HSP40 is incorporated into the radial spoke complex during the assembly process in flagella. Mol Biol Cell. 2005;16(2):637-48.

93. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, et al. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. 2005;33(Database issue):D192-6.

94. Wickstead B, Gull K. Dyneins across eukaryotes: a comparative genomic analysis. Traffic. 2007;8(12):1708-21.

95. Wickstead B, Gull K. A "holistic" kinesin phylogeny reveals new kinesin families and predicts protein functions. Mol Biol Cell. 2006;17(4):1734-43.

96. Jékely G. Evolution of phototaxis. Philos Trans R Soc Lond B Biol Sci. 2009;364(1531):2795-808.

97. Kawai H. Green flagellar autofluorescence in brown algal swarmers and their phototactic responses. Bot Mag Tokyo. 1992;105(1):171-84.

98. Kawai H, Nakamura S, Mimuro M, Furuya M, Watanabe M. Microspectrofluorometry of the autofluorescent flagellum in phototactic brown algal zoids. Protoplasma. 1996;191(3-4):172-7.

99. Avelar Gabriela M, Schumacher Robert I, Zaini Paulo A, Leonard G, Richards Thomas A, Gomes Suely L. A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. Curr Biol. 2014;24(11):1234-40.

100. Richards TA, Leonard G, Wideman JG. What defines the “Kingdom” Fungi? Microbiol Spectr. 2017;5(3).

101. Sudbery PE. Regulation of polarised growth in fungi. Fungal Biol Rev. 2008;22(2):44-55.

102. Bartnicki-Garcia S. The cell wall: a crucial structure in fungal evolution. In: Rayner ADM, Brasier CM, Moore D, editors. Evolutionary Biology of the Fungi. Cambridge: Cambridge University Press; 1987. p. 389-403.

103. Richards TA, Talbot NJ. Horizontal gene transfer in osmotrophs: playing with public goods. Nat Rev Microbiol. 2013;11(10):720-7.

104. Park H-O, Bi E. Central roles of small GTPases in the development of cell polarity in yeast and beyond. Microbiol Mol Biol Rev. 2007;71(1):48-96.

105. James TY, Berbee ML. No jacket required – new fungal lineage defies dress code. BioEssays. 2012;34(2):94-102.

106. Schuster M, Treitschke S, Kilaru S, Molloy J, Harmer NJ, Steinberg G. Myosin‐5, kinesin‐1 and myosin‐17 cooperate in secretion of fungal chitin synthase. EMBO J. 2011;31(1):214-27.

107. Steinberg G. Hyphal growth: a tale of motors, lipids, and the Spitzenkörper. Eukaryot Cell. 2007;6(3):351-60.

108. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature. 2005;436:1113-8.

109. Sebé-Pedrós A, Grau-Bové X, Richards TA, Ruiz-Trillo I. Evolution and classification of myosins, a paneukaryotic whole-genome approach. Genome Biol Evol. 2014;6(2):290-305.

110. Bartnicki-Garcia S. Cell wall chemistry, morphogenesis, and taxonomy of fungi. Annu Rev Microbiol. 1968;22:87-108.

111. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, et al. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell. 2010;22(9):2943-55.

112. Fuller MS, Barshad I. Chitin and cellulose in the cell walls of *Rhizidiomyces* sp. Am J Bot. 1960;47(2):105-9.

113. Kneipp LF, Andrade AF, de Souza W, Angluster J, Alviano CS, Travassos LR. *Trichomonas vaginalis* and *Tritrichomonas foetus*: expression of chitin at the cell surface. Exp Parasitol. 1998;89(2):195-204.

114. Lin CC, Aronson JM. Chitin and cellulose in the cell walls of the oomycete, *Apodachlya* sp. Arch Mikrobiol. 1970;72(2):111-4.

115. Arroyo-Begovich A, xe, rabez-Trejo A, Ru, xed, z-Herrera J, et al. Identification of the structural component in the cyst wall of *Entamoeba invadens*. J Parasitol. 1980;66(5):735-41.

116. Das S, Van Dellen K, Bulik D, Magnelli P, Cui J, Head J, et al. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). Mol Biochem Parasit. 2006;148(1):86-92.

117. Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja T, Bonilla-Rosso G, et al. The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. Sci Rep. 2015;5:11153.

118. Clay RP, Benhamou N, Fuller MS. Ultrastructural detection of polysaccharides in the cell walls of two members of the Hyphocytriales. Mycological Research. 1991;95(9):1057-64.

119. Mélida H, Sandoval-Sierra JV, Diéguez-Uribeondo J, Bulone V. Analyses of extracellular carbohydrates in oomycetes unveil the existence of three different cell wall types. Eukaryot Cell. 2013;12(2):194-203.

120. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. Eukaryot Cell. 2009;8(7):1038-50.

121. Lefèvre E, Bardot C, Noël C, Carrias JF, Viscogliosi E, Amblard C, et al. Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. Environ Microbiol. 2007;9(1):61-71.

122. Edwards U, Rogall T, Blocker H, Emde M, Bottger EC. Isolation and direct complete nucleotide determination of entire genes. Characterisation of a gene encoding for 16S ribosomal RNA. . Nucleic Acids Res. 1989;17:7843-53.

123. Schmieder R, Lim YW, Rohwer F, Edwards R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC Bioinformatics. 2010;11:341.

124. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863-4.

125. Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biology. 2012;13(12):R122.

126. Vurture GW, Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J.,, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017:1-3.

127. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

128. 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;18(12):1979-90.

129. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27(6):757-63.

130. Tate R, Hall B, DeRego T, Geib S. Annie: the ANNotation Information Extractor (Version 1.0) <http://genomeannotation.github.io/annie2014>.

131. Hall B, DeRego T, Geib S. GAG: the Genome Annotation Generator (Version 1.0) [Software]. <http://genomeannotation.github.io/GAG2014>.

132. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

133. Margulies M, Egholm M, Altman W, Attiya S, Bader J, Bemben L, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376 - 80.

134. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Frontiers in Genetics. 2013;4:237.

135. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc Natl Acad Sci USA. 2016;X(X):600338113v1-201600338.

136. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, et al. Community-wide analysis of microbial genome sequence signatures. Genome Biology. 2009;10(8):R85.

137. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. Bioinformatics. 2005;21(4):537-9.

138. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 2010;20.

139. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, et al. A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature. 2007;450(7166):115-8.

140. Sharma R, Xia X, Cano LM, Evangelisti E, Kemen E, Judelson H, et al. Genome analyses of the sunflower pathogen *Plasmopara halstedii* provide insights into effector evolution in downy mildews and *Phytophthora*. BMC Genomics. 2015;16(1):741.

141. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785-6.

142. Krogh A, Larrson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.

143. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175-82.

144. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(Web Server issue):W585-7.

145. Cantarel B, Coutinho P, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009;37:D233 - D8.

146. Edgar R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792 - 7.

147. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci. 1996;12(6):543-8.

148. Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, et al. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc Roy Soc B. 2013;280(1769).

149. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014 30(9):1312-3.

150. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell. 2009;21(7):1897-911.

151. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.

152. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005;33(2):511-8.

153. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3.

154. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490.

155. Vaas LAI, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk H-P, et al. opm: an R package for analysing OmniLog® phenotype microarray data. Bioinformatics. 2013;29(14):1823-4.

156. Richards TA, Leonard G, Wiseman JG. What defines the "Kingdom" Fungi? ASMscienceorg/MicrobiolSpectrum. 2017;In Press.

157. Pruyne D, Evangelista M, Yang C, Bi E, Zigmond S, Bretscher A, et al. Role of formins in actin assembly: nucleation and barbed-end association. Science. 2002;297(5581):612-5.

158. Pruyne DW, Schott DH, Bretscher A. Tropomyosin-containing actin cables direct the Myo2p-dependent polarized delivery of secretory vesicles in budding yeast. J Cell Biol. 1998;143(7):1931-45.

159. Tcheperegine SE, Gao X-D, Bi E. Regulation of cell polarity by interactions of Msb3 and Msb4 with Cdc42 and polarisome components. Mol Cell Biol. 2005;25(19):8567-80.

160. Bretscher A. Polarized growth and organelle segregation in yeast: the tracks, motors, and receptors. J Cell Biol. 2003;160(6):811-6.

161. Pruyne D, Legesse-Miller A, Gao L, Dong Y, Bretscher A. Mechanisms of polarized growth and organelle segregation in yeast. Annu Rev Cell Dev Biol. 2004;20(1):559-91.

162. Walch-Solimena C, Collins RN, Novick PJ. Sec2p mediates nucleotide exchange on Sec4p and Is involved in polarized delivery of post-Golgi vesicles. J Cell Biol. 1997;137(7):1495-509.

163. Goud B, Salminen A, Walworth NC, Novick PJ. A GTP-binding protein required for secretion rapidly associates with secretory vesicles and the plasma membrane in yeast. Cell. 1988;53(5):753-68.

164. Walworth NC, Goud B, Kabcenell AK, Novick PJ. Mutational analysis of SEC4 suggests a cyclical mechanism for the regulation of vesicular traffic. EMBO J. 1989;8(6):1685-93.

165. Virag A, Lee MP, Se H, Harris SD. Regulation of hyphal morphogenesis by cdc42 and rac1 homologues in *Aspergillus nidulans*. Mol Micro. 2007;66(6):1579-96.

166. Mahlert M, Leveleki L, Hlubek A, Sandrock B, Bölker M. Rac1 and Cdc42 regulate hyphal growth and cytokinesis in the dimorphic fungus *Ustilago maydis*. Mol Micro. 2006;59(2):567-78.