Phylogenomics Supports the Monophyly of Aphelids and Fungi and Identifies New Molecular Synapomorphies

Luis Javier Galindo, Guifré Torruella, Purificación López-García, Maria Ciobanu, Ana Gutiérrez-Preciado, Sergey A Karpov, David Moreira

Title

Phylogenomics Supports the Monophyly of Aphelids and Fungi and Identifies New Molecular Synapomorphies

Authors:

LUIS JAVIER GALINDO[1¶], GUIFRÉ TORRUELLA[1¶], PURIFICACIÓN LÓPEZ-GARCÍA[1], MARIA CIOBANU[1], ANA GUTIÉRREZ-PRECIADO[1], SERGEY A. KARPOV[2], AND DAVID MOREIRA[1*]

[1]Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Orsay, France

[2]Zoological Institute RAS, Universitetskaya emb. 1, and St Petersburg State University, Universitetskaya emb. 7/9, St Petersburg 199034, Russia

*Corresponding author: E-mail: david.moreira@universite-paris-saclay.fr

[¶]Both authors contributed equally to this work.

Running title: Phylogenomics and Megasystematics of Holomycota

Abstract.- The supergroup Holomycota, composed of Fungi and several related lineages of unicellular organisms (Nucleariida, Rozellida, Microsporidia, and Aphelida), represents one of the major branches in the phylogeny of eukaryotes. Nevertheless, except for the well-established position of Nucleariida as the first holomycotan branch to diverge, the relationships among the other lineages have so far remained unresolved largely owing to the lack of molecular data for some groups. This was notably the case aphelids, a poorly known group of endobiotic phagotrophic protists that feed on algae with cellulose walls. The first molecular phylogenies including aphelids supported their sister relationship with Rozellida and Microsporidia which, collectively, formed a new group called Opisthosporidia (the 'Opisthosporidia hypothesis'). However, recent phylogenomic analyses including massive sequence data from two aphelid genera, *Paraphelidium* and *Amoeboaphelidium*, suggested that the aphelids are sister to fungi (the 'Aphelida+Fungi hypothesis'). Should this position be confirmed, aphelids would be key to understanding the early evolution of Holomycota and the origin of Fungi. Here, we carry out phylogenomic analyses with an expanded taxonomic sampling for aphelids after sequencing the transcriptomes of two species of the genus *Aphelidium* (*A. insulamus* and *A. tribonematis*) in order to test these competing hypotheses. Our new phylogenomic analyses including species from the three known aphelid genera strongly rejected the Opisthosporidia hypothesis. Furthermore, comparative genomic analyses further supported the Aphelida+Fungi hypothesis via the identification of 19 orthologous genes exclusively shared by these two lineages. Seven of them originated from ancient horizontal gene transfer events predating the aphelid-fungal split and the remaining 12 likely evolved *de novo*, constituting additional molecular synapomorphies for this clade. Ancestral trait reconstruction based on our well-resolved phylogeny of Holomycota suggests that the progenitor of both fungi and rozellids, was aphelid-like, having an amoeboflagellate state and likely preying endobiotically on cellulose-containing, cell-walled organisms. Two lineages, which we propose to call Phytophagea and Opisthophagea, evolved from this ancestor. Phytophagea, grouping aphelids and classical fungi, mainly specialized in endobiotic predation of algal cells. Fungi emerged from this lineage after losing phagotrophy in favour of osmotrophy. Opisthophagea, grouping rozellids and Microsporidia, became parasites, mostly of chitin-containing hosts. This lineage entered a progressive reductive process that resulted in a unique lifestyle, especially in the highly derived Microsporidia.

Keywords: Fungi, Aphelida, Holomycota, phylogenomics, horizontal gene transfer, synapomorphy

The Aphelida is a diverse group of endobiotic phagotrophic unicellular protists that feed on algae that have cellulose walls (green algae and stramenopiles) (Gromov 2000; Karpov et al. 2014, 2017). Despite being known since the end of the 19th century (Zopf 1885), aphelids were neglected until recent years when, thanks to classical culturing efforts, they gained attention due to their idiosyncratic cell biology and phylogenetic position within the Holomycota (one of the two main branches of the Opisthokonta) (Karpov et al. 2014; Adl et al. 2019), resulting in a significant increase in their known biodiversity (Karpov et al. 2020).

The morphology of the aphelid zoospores, which can be flagellated and/or ameboid, has been widely used for species classification (Karpov et al. 2019; Letcher and Powell 2019). From the ~20 described species, only 11 freshwater species belonging to the genera *Aphelidium*, *Paraphelidium* and *Amoeboaphelidium* have been molecularly characterized by SSU rRNA gene sequencing (Letcher and Powell 2019; Karpov et al. 2020). The fourth described genus, represented by the marine species *Pseudaphelidium drebesii*, has not been reported since its original description (Schweikert and Schnepf 1996, 1997) and sequence data is missing. SSU rRNA gene phylogenies showed that aphelids belong to the Holomycota, together with nucleariids, rozellids, microsporidia, and fungi (fungi defined here as the monophyletic group of holomycotan species that excludes all the previous taxa) (Karpov et al. 2013, 2014, 2017; Letcher et al. 2015). These phylogenies, also including environmental sequences from diverse habitats, determined that the aphelids are monophyletic with moderate to high bootstrap support (72-100%) (Simon et al. 2015; Karpov et al. 2020; Seto et al. 2020). Within the Holomycota, some of these phylogenies grouped, albeit with weak support, aphelids with *Rozella* (Letcher and Powell 2018) and Microsporidia (Vávra and Lukeš 2013) in a clade sister to fungi that has been named Opisthosporidia (Karpov et al. 2014).

The phylogenetic signal of SSU rDNA is too limited to resolve ancient speciation events in eukaryote evolution, not only at supergroup level (Adl et al. 2019) but also within clades, including opisthokont phyla (e.g. Galindo et al. 2019). Transcriptome and genome sequencing has facilitated multi-marker phylogenetic analysis that is critical for resolving the phylogeny of all eukaryotes (Sibbald and Archibald 2017; Lax et al. 2018). Recently, phylogenomic analyses using three datasets derived from transcriptome data of the aphelid *Paraphelidium tribonematis* supported the monophyly of this aphelid and fungi, rendering the Opisthosporidia clade paraphyletic (Torruella et al. 2018). Later, an independent phylogenomic study including markers from two *Amoeboaphelidium* genomes again supported the monophyly of aphelids and fungi (Tikhonenkov et al. 2020). However, these studies included limited taxon sampling,

so the monophyly of aphelids and the sister relationship between the aphelids and the fungi requires further validation.

To test the two competing phylogenetic hypotheses, "Opisthosporidia" versus "Aphelida+Fungi", we sequenced transcriptomes of two species from a third aphelid genus, *Aphelidium* (*A. insulamus* and *A. tribonematis*) (Karpov et al. 2016, 2020) and we used phylogenomic methods to analyze all of the available aphelid genomes and transcriptomes, together with representatives of all of the known Holomycota lineages. We resolved the monophyly of these three aphelid genera and confirmed the position of aphelids as the sister lineage to fungi, thereby rejecting the Opisthosporidia hypothesis. We also found several molecular synapomorphies that further support the Aphelida+Fungi hypothesis, including shared genes ancestrally obtained by horizontal gene transfer (HGT) and genes likely evolved *de novo*. Based on these results, we propose a reorganization of the internal systematics of the Holomycota with the creation of two new major taxa, the Phytophagea, containing Aphelida and Fungi, and the Opisthophagea, containing Microsporidia and Rozellida.

MATERIALS AND METHODS

Culturing, RNA Extraction, Transcriptome Sequencing, Assembly and Decontamination

Aphelidium insulamus O-14 (X-134) (Karpov et al. 2020) and *Aphelidium tribonematis* P-2 (X-102) (Karpov et al. 2016) cultures were maintained with their prey, the xantophyte alga *Tribonema gayanum* (Strain 20 CALU), in mineral medium BG11 at 15ºC with a 12h-12h night-day cycle. The cultures were initially cleaned from other contaminant eukaryotes by picking zoospores of both aphelid species by micromanipulation and transferring them into pure *T. gayanum* cultures.

For each species, total RNA was extracted from three cultures at different growing stages, 8, 16 and 30 days after being transferred to healthy *T. gayanum* cultures. These times corresponded to peaks in the abundance of different aphelid cell cycle stages (cysts, trophonts, and zoospores) and were used to maximize the gene coverage of the transcriptomes (see Torruella et al. 2018). The grown cultures were filtered together through 0.22 µm pore-diameter filters (Millipore) using a vacuum filtration pump. Filters were incubated in 2 ml Eppendorf tubes filled with lysis buffer from the RNeasy mini Kit (Qiagen), following the manufacturer protocol. Extracted RNA was quantified with a Qubit fluorometer (ThermoFisher Scientific). One cDNA Illumina library was constructed for each species after polyA mRNA selection and paired-end (2 × 125 bp) sequenced with Illumina HiSeq 2500 chemistry v4 (Eurofins Genomics, Germany). We obtained 9,261,168 reads for *A. tribonematis* P-2 and

7,271,713 reads for *A. insulamus* O-14. Raw read sequences have been submitted to GenBank under BioProject accession numbers PRJNA719988 (*A. insulamus* O-14) and PRJNA720686 (*A. tribonematis* P-2).

Paired-end read maximum quality scores were checked with FastQC v0.11.8 (Andrews 2010), and each transcriptome was *de novo* assembled using rnaSpades v3.15.0 (Bankevich et al. 2012) with default parameters. Predicted proteins were obtained using Transdecoder v2 (http:transdecoder.github.io) with default parameters and Cd-hit v4.6 (Li and Godzik 2006) with 100% identity. Decontamination of host transcripts was done using the *T. gayanum* transcriptome as local Blastp (Altschul et al. 1990) query to exclude identical sequences, as previously done by Torruella et al. (2018). To assess genome completeness, we used BUSCO v2.0.1 on the decontaminated predicted proteomes with the eukaryota_odb9 dataset of 303 near-universal single-copy orthologs (Simão et al. 2015).

Phylogenomic Analyses

The dataset of 351 conserved proteins from Lax et al. (2018) was updated by Blastp search of each protein against the inferred proteomes of representatives of all known major eukaryotic lineages, enriched in opisthokonts but also in stramenopiles to ensure the elimination of *T. gayanum* contaminants. Each protein marker was aligned with MAFFT v7 (Katoh and Standley 2013) and trimmed with TrimAl with the automated1 option (Capella-Gutiérrez et al. 2009). Alignments were manually inspected and edited with AliView (Larsson 2014) and Geneious v6.06 (Kearse et al. 2012). Single-protein trees were reconstructed with IQ-TREE v1.6.11 (Nguyen et al. 2015) with the corresponding best-fitting model. Each single-protein tree was manually inspected to discard contaminants and possible cases of horizontal gene transfer or hidden paralogy. At the end of this curation process, we kept a final taxon sampling of 25 species, including Nucleariida, Aphelida, Rozellida, Microsporidia, and Fungi (Table S1, Supplementary Material available on Dryad), and 175 protein markers which were present in all aphelid species (including the species with the least amount of available data, *Amoeboaphelidium protococcarum*). All proteins were realigned, trimmed and concatenated using Alvert.py from the package Barrel-o-Monkeys (http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm), creating a final supermatrix with 59,889 amino acids. For details on the origin of the sequence data see Table S4 (Supplementary Material available on Dryad).

Bayesian inference (BI) analyses were done with PhyloBayes-MPI v1.5a (Lartillot et al. 2009) on CIPRES Science Gateway (Miller et al. 2010) with both CAT-Poisson and CAT-GTR

models (Lartillot and Philippe 2004), with two MCMC chains and run for 10,000 generations, saving one every 10 trees. Analyses were stopped once convergence thresholds calculated using bpcomp were reached (i.e. maximum discrepancy <0.1 and minimum effective size >100) and consensus trees were constructed after a burn-in of 25%. Maximum Likelihood (ML) analyses were done with IQ-TREE v1.6.11 (Nguyen et al. 2015) with the best-fitting model (LG+F+R6) identified by ModelFinder (Kalyaanamoorthy et al. 2017). Additional ML trees were done using the PMSF model (Wang et al. 2018) to account for among-site rate heterogeneity and with the C60 profile mixture model. ML statistical support was calculated with 1000 ultrafast bootstraps (Minh et al. 2013). All trees were visualised and edited with FigTree v1.4.3 (Rambaut 2016).

In addition to the phylogenetic analysis of concatenated protein markers, we examined the 175 ML phylogenetic trees of individual proteins (inferred with RAxML v.8.1.6 with the PROT+CAT+LG+F model and 100 bootstrap replicates (Stamatakis 2014)) using a coalescent approach with ASTRAL-III under the multi-species coalescent model with default parameters (Zhang et al. 2018).To minimize possible systematic bias in our concatenated dataset due to the presence of fast-evolving sites, known to be more prone to homoplasy and compositional bias (Rodríguez-Ezpeleta et al. 2007), we reconstructed a series of trees by progressively removing the fastest-evolving sites, 5% of sites at a time. For that, among-site substitution rates were inferred using IQ-TREE under the -wsr option and the best-fitting model for a total of 19 new data subsets (Table S2, Supplementary Material available on Dryad). We then reconstructed phylogenetic trees for all these subsets using IQ-TREE with the same best-fitting model as for the whole concatenated dataset.

Detection of Molecular Synapomorphies

We inspected 23 opisthokont proteomes (including the three aphelids *P. tribonematis*, *A. tribonematis* and *A. insulamus*) to look for proteins that were unique to aphelids and fungi. We first carried out clustering of orthologous genes with OrthoFinder v1.1.20 (Emms and Kelly 2015) with default parameters. We then identified the orthologs shared by aphelids, chytrids, blastocladiomycetes and the rest of fungal representatives. Representative protein sequences of each candidate ortholog were blasted against the non-redundant GenBank database using Blastp (Altschul et al. 1990) with and without Fungi (taxid:4751). The first 100 hits of both searches were downloaded and merged into single-protein datasets together with positive hits from our 23 opisthokonts. The putative identity and function of each ortholog was assigned using domain searches both in Blastp and HMMER using hmmscan (http://hmmer.org/). Cd-

hit v4.6 with 100% identity was used to eliminate redundant sequences from the datasets. To verify that the Blast searches did not miss distant homologs of the orthologs found only in aphelids and fungi, we reconstructed HMM profiles for each of those orthologs using the hmmbuild program from the HMMer package v3.2.1 (Mistry et al. 2013). These profiles were used to scan the GenBank RefSeq database with the hmmsearch program from the same package. The taxonomy (NCBI taxID identifiers) of the retrieved hits was obtained using the protein accession number as reference to query the NCBI taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy). The hits for each ortholog and their taxonomy were sorted by e-value and printed in tabular output files.

Each protein dataset was aligned using MAFFT v7 (Katoh et al. 2002) and trimmed with TrimAl in automated1 mode. Single-protein ML trees were inferred using IQ-TREE (Nguyen et al. 2015) with the best-fitting model selected with the IQ-TREE TESTNEW algorithm as per BIC. All datasets and trees were checked manually for possible contaminating sequences.

Assembled transcriptomes, predicted proteins, HMM profiles and hmmsearch results, phylogenomic datasets, and molecular synapomorphies have been deposited in figshare at https://figshare.com/projects/Aphelida_Extended_Data/111539 and in the Dryad Digital Repository: https://doi.org/10.5061/dryad.j3tx95xdn.

Ancestral State Reconstruction

We inferred ancestral states for the following five traits in various clades within the Holomycota: flagellum (presence / absence), pseudopodia (presence / absence), feeding mode (osmotrophy / phagotrophy), preferred food/host (decaying organic matter / bacteria / chitinous organisms / plants or algae), and habitat (freshwater or soil / marine). We used the Bayesian inference phylogenetic tree of 25 species as reference phylogeny (see above) and two approaches for ancestral state reconstruction implemented in Mesquite 3.70 (Maddison and Maddison 2021): a parsimony approach with the unordered states model of evolution for categorical characters, and a likelihood approach with the Markov k-state 1 parameter (Mk1) model.

RESULTS AND DISCUSSION

The Aphelid Clade is the Sister Lineage to Fungi

After sequencing, assembly and decontamination of the two aphelid transcriptomes, we obtained 45,543 transcripts for *A. insulamus* O-14 and 49,586 transcripts for *A. tribonematis* P-2. Comparison with the BUSCO eukaryotic dataset of 303 near-universal single-copy

orthologs (Simão et al. 2015) indicated very high transcriptome completeness for the two species (97,7% for *A. insulamus* O-14; 96,7% for *A. tribonematis* P-2). We used 175 conserved protein markers after updating a previous dataset (Lax et al. 2018) to include the aphelid sequences and representatives of major holomycotan groups for a final taxon sampling of 25 species (Table S1, Supplementary Material available on Dryad) and 59,889 amino acid positions. Both maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses of this data set supported the overall same tree topology (Fig. 1A). Nucleariida emerged as the first branch among Holomycota, followed by two sister clades, one containing *Rozella* and Microsporidia, in agreement with previous analyses (Brown et al. 2009; Liu et al. 2009; Galindo et al. 2019), and another containing Aphelida and Fungi. Within fungi, despite their reduced taxon sampling included in this tree, we found support for the "chytrid first" hypothesis already recovered in recent phylogenomic studies (Spatafora et al. 2016; Mikhailov et al. 2017; Tedersoo et al. 2018; Galindo et al. 2021). Both ML and BI methods yielded maximum support (bootstrap (BS) 100% and posterior probability (PP) 1, respectively) for the monophyly of aphelids and their placement as sister lineage to fungi (Fig. 1A and Supplementary Fig. S1a-d, Supplementary Material available on Dryad). These analyses confirmed with a richer taxon sampling and stronger support the Aphelida+Fungi hypothesis previously proposed by Torruella et al. (2018). We tested the possible influence of fast-evolving sites on this result by applying a slow-fast approach (Brinkmann and Philippe 1999) to progressively remove the fastest-evolving sites (in 5% steps, see Fig. 1B and Table S2, Supplementary Material available on Dryad). The monophyly of aphelids+fungi obtained maximum support (BS >99%) in all steps until only 35% of the sites remained and the phylogenetic signal was too scarce to resolve any deep-level relationship (Fig. 1B). In fact, the monophyly of aphelids+fungi received higher statistical support than the monophyly of fungi. On the contrary, the monophyly of Opisthosporidia (aphelids+rozellids+microsporidia) was never well supported. These results were confirmed by the combined analysis of the 175 ML trees based on individual protein markers with a multi-species coalescent model (Zhang et al. 2018). The resulting species tree had a similar topology as the BI and ML trees based on the concatenation of these markers and showed full support for the Aphelida+Fungi hypothesis (Fig. 1C and Supplementary Fig. S1e, Supplementary Material available on Dryad).

Despite their overall congruence, the ML and BI trees based on the concatenated dataset showed conflicting topologies for the internal relationship of aphelids. The two *Amoeboaphelidium* species were paraphyletic in all phylogenies, as observed in Tikhonenkov et al. (2020), with *A. occidentale* sister to *P. tribonematis* (e.g., Fig. 1A), whereas *A.*

protococcarum was the first aphelid branch to diverge in ML trees (with BS between 85 and 94%, Supplementary Fig. S1c-d, Supplementary Material available on Dryad) as well as in the coalescent-based species tree (Supplementary Fig. S1e, Supplementary Material available on Dryad). In the BI trees, *Aphelidium* spp. were the earliest diverging aphelids with maximum support under both CAT-Poisson and CAT-GTR models (Supplementary Fig. S1a-b, Supplementary Material available on Dryad).

Molecular Synapomorphies of Aphelids and Fungi

Given the phylogenetic evidence for the monophyly of aphelids and fungi, we looked for proteins unique to these two groups but not present in other eukaryotes. These unique proteins could be considered molecular synapomorphies and would therefore provide further support for their monophyly. We identified a total of 80,441 orthogroups in our dataset of opisthokont proteomes. 93 of these orthologs were shared by at least two species of aphelids, chytrids, blastoclads and the rest of fungal representatives. We then verified if these 93 orthologs were absent in other eukaryotic groups by Blast against the non-redundant GenBank database. 74 orthologs were present in other eukaryotes and 19 were found only in aphelids and fungi. We further confirmed the absence of these 19 orthologs in other eukaryotes using a more sensitive HMM-based search against the NCBI Reference Sequence (RefSeq) database. The 19 orthologs found only in aphelids and fungi were likely already present in the last common ancestor of both groups. To test this idea, the sequences of these orthologs were aligned, trimmed and used for ML tree reconstruction (Table S3 and Supplementary Figs. S2-S20, Supplementary Material available on Dryad).

After manual checking of these trees, we identified 7 genes that appeared to have been acquired by HGT from bacteria and 12 that likely evolved *de novo* in a common ancestor of aphelids and fungi. Among the 7 putative HGT cases, all but one ortholog (OG0000718) had functional domains assigned (Table S3, Supplementary Material available on Dryad). The remaining 12 proteins did not have homologs in other organisms and, therefore, we considered them as *de novo* evolved proteins, although the possibility that they were also acquired by HGT from not yet sequenced prokaryotic or eukaryotic donors cannot be discarded. Two of these proteins (OG0007603 and OG0011744) did not contain any known domain (Supplementary Figs. S17 and S20, Supplementary Material available on Dryad), and another two (OG0007040 and OG0009707) carried the conserved domains of unknown function (DUF) DUF2418 and DUF2423, respectively (Supplementary Figs. S14 and S19, Supplementary Material available

on Dryad). For the other nine proteins, we could infer a putative function based on their domain composition (Table S3, Supplementary Material available on Dryad).

The seven proteins derived from ancestral HGT events in the aphelids+fungi clade had various putative functions and cellular localizations, from intramembrane mechanosensitive channels (OG0001250) to cytoskeletal components (OG0002101) and extracellular enzymes (OG0002372) (Fig. 2 and Supplementary Figs. S3, S4, and S6, Supplementary Material available on Dryad). One interesting case was that of OG0003785, a GTP cyclohydrolase II enzyme (GTPCH2) involved in riboflavin synthesis. GTPCH2 has been described only in fungi (mainly yeasts), bacteria and plants (Oltmanns et al. 1969; Bereswill et al. 1998; Herz et al. 2000; Spoonamore et al. 2006; Yadav and Karthikeyan 2015). However, our analysis of the OG0003785 ortholog (Supplementary Fig. S7, Supplementary Material available on Dryad) showed that the plant sequences are extremely divergent (Supplementary Fig. S7B, Supplementary Material available on Dryad) and only their C-terminal half is homologous to the fungal and bacterial sequences. By contrast, the GTPCH2 proteins from aphelids and fungi are highly similar to the bacterial ones along the whole protein length. Thus, the gene encoding GTPCH2 (*RIB1* or *RibA*) was most likely acquired from bacteria by an ancestor of aphelids and fungi, which gained this function independently from plants. GTPCH2 catalyzes the first step of riboflavin biosynthesis by converting GTP into 2,5-diamino-6-(5-phospo-D-ribosylamino)-pyrimidin-4(3H)-one (DARP), then DARP is processed by several enzymes to generate the final product (Anam et al. 2020). It has been shown that GTPCH2 activity is required for *RIB1*-dependent protection of yeast cells from nitrosative stress (Anam et al. 2020). This stress is caused by excessive levels of nitric oxide (NO) leading to cellular damage or death, as it has been demonstrated in yeast (Almeida et al. 2007; Yoshikawa et al. 2016). Defence against NO is essential also for pathogenicity, as shown in *Candida albicans*, which uses flavohemoglobin to weaken the toxicity of NO released by the host (Hromatka et al. 2005; Zakikhany et al. 2007; Martin et al. 2011; Liao et al. 2016). Riboflavin biosynthesis is essential in microbes but absent in humans and therefore this pathway is an ideal antimicrobial drug target with minimum host toxicity (Yadav and Karthikeyan 2015). The horizontal transfer of this gene from bacteria to an ancestor of aphelids and fungi could have played an important role in the evolution of pathogenicity in this eukaryotic clade.

In addition to these shared ancestral HGTs in aphelids and fungi, we found 12 genes that appear to have evolved *de novo* in their ancestor. They encoded proteins with various putative cellular localizations, including the nucleus (e.g., OG0007045) and mitochondrion (e.g.,

OG0008189) (Fig. 2, and Supplementary Figs. S15 and S18, Supplementary Material available on Dryad). Among them, it is worth highlighting OG0006536 (Supplementary Fig. S11, Supplementary Material available on Dryad), which corresponds to a S25-like mitochondrial ribosomal protein from the 37S small ribosomal subunit (RSM25-like). Mitoribosomes are attached to the mitochondrial inner membrane and, in yeast, consist of a small (37S) and a large (54S) subunit. The 37S subunit contains at least 33 different proteins and one rRNA (15S) (Saveanu et al. 2001; Pfeffer et al. 2015; Desai et al. 2017). Similarly to some other yeast mitoribosomal proteins (Saveanu et al. 2001), the RSM25-like protein had no known homolog in any other eukaryotic or prokaryotic lineage, making it a conspicuous synapomorphy for the aphelids+fungi group.

The ortholog OG0007007 (Supplementary Fig. S12, Supplementary Material available on Dryad) provided another interesting synapomorphy. It corresponds to the AAA-ATPase Vps4-associated protein 1 (Vfa1). Controlling and remodelling membranes (e.g., extracellular vesicle formation, retroviral budding, cell abscission during cytokinesis, multivesicular body formation) is a vital process for cellular homeostasis (Oliveira et al. 2013; Vild and Xu 2014). It is especially important in fungi, since two of their canonical traits, apical growth and osmotrophic feeding (Schultzhaus and Shaw 2015; Richards and Talbot 2018), depend on this process. Cell membrane control and remodelling involve a class of proteins collectively known as the "endosomal sorting complexes required for transport" (ESCRT). The ATPase Vps4 is one of the five distinct multimeric protein complexes of the ESCRT machinery (Williams and Urbé 2007) and has a confirmed role in fungal extracellular vesicle formation (Oliveira et al. 2013). Vps4 is recruited to the site of vesicle formation and, through ATP hydrolysis, provides energy and drives the completion of vesicle formation (Baumgärtel et al. 2011; Elia et al. 2011). Recently, the Vps4-binding protein Vfa1 (Vps4-associated-1) was identified in an overexpression study of all putative genes of *Saccharomyces cerevisiae* aiming at exploring vacuole morphology defects (Arlt et al. 2011). Vfa1 overexpression caused altered vacuole size and, later on, it was proved that it is a regulator of the Vps4 ATPase activity (Vild and Xu 2014). Our results showed that the Vfa1/Vfa1-like protein (OG0007007) is unique to the aphelids+fungi clade. Thus, this protein arguably constitutes an important molecular synapomorphy that might have played a role in the evolution of the unique feeding and growth strategies found in this clade. Elucidating the unknown functions of the other exclusive genes ancestral to the clade, either derived from HGT (OG0000718, Supplementary Fig. S2, Supplementary Material available on Dryad) or evolved *de novo* (OG0007603, OG0007040, OG0009707, OG0011744; Supplementary Figs. S14, S17, S19, and S20, Supplementary

Material available on Dryad), should help unravelling unique common cellular activities in the aphelids+fungi clade.

Aphelids and the Early Evolution of Holomycota

Our results strongly support a new phylogenetic framework of Holomycota composed of three main clades: Nucleariida, Rozellida+Microsporidia, and Aphelida+Fungi (Fig. 1). Because they are the closest relatives of Fungi, characterizing aphelid genomes and phenotypes is important to decipher the early evolution of Fungi and, more globally, Holomycota. Although the parasitic rozellids have a more simplified metabolism and, similarly to their microsporidian relatives, have acquired transporters to import nucleotides directly from their hosts (James et al. 2013; Quandt et al. 2017; Dean et al. 2018), they share with aphelids many important structural and biological similarities. Notably, both have free-living flagellated zoospores and feed as endobiotic phagotrophs (trophonts), i.e. preying on the host cytoplasm within the cell wall-defined space. Both aphelids and rozellids penetrate the cell walls of their diverse eukaryotic hosts starting from an extracellular chitinous cyst stage that adheres to the host cell surface. The cytoplasm of this cyst stage then enters the host cell wall-defined space by locally digesting the cell wall and/or exploiting natural spaces between cell wall halves, like in *Tribonema* and diatoms (Karpov et al. 2014; Torruella et al. 2018). Therefore, these characters were most likely present in the last common ancestor of all known Holomycota excluding nucleariids, i.e. the Rozellida+Microsporidia+Aphelida+Fungi group (Fig. 3 and Supplementary Figs. S21 and S22, Supplementary Material available on Dryad). By contrast, nucleariids, the first branch to diverge within Holomycota, typically feed on bacteria and do not exhibit complex life cycles with infective cysts, zoospores, trophonts, and exo- and endobiotic stages (Galindo et al. 2019). Interestingly, the other two holomycotan clades also show a clear feeding preference trend. On the one hand, even if some species parasitize other protists (e.g. amoebae or gregarines), rozellids and microsporidia mostly feed on other opisthokonts (fungi and animals) (Wadi and Reinke 2020). On the other hand, although many fungi have become parasites of animals and many other eukaryotes, aphelids and fungi were ancestrally more specialized in infecting and/or degrading photosynthetic eukaryotes (Naranjo-Ortiz and Gabaldón 2019). Indeed, a wealth of evidence points towards the co-evolution of fungi with green algae and plants as they transitioned to terrestrial environments (Lutzoni et al. 2018; Naranjo-Ortiz and Gabaldón 2019; Berbee et al. 2020). Accordingly, we propose to name the supergroup composed of Microsporidia and Rozellida ‘Opisthophagea’ and the supergroup composed of Aphelida and Fungi ‘Phytophagea’ (Fig. 3).

These different feeding affinities may have been a major driving force in the evolution and metabolic specialization of these two lineages towards the degradation of, respectively, chitin- or cellulose-based polymers (Fig. 3). Whereas the last common ancestor of Holomycota may have been a bacterivore as most contemporary nucleariids, the common ancestor of Opisthophagea and Phytophagea was most likely already adapted to prey endobiotically on other eukaryotes (Fig. 3). However, whether it fed on opisthokonts, algae or both is difficult to determine (Supplementary Figs. S21 and S22, Supplementary Material available on Dryad). Recently, a study using SSU rRNA gene-based fluorescent *in situ* hybridization on natural marine samples identified an organism that fed inside diatoms and that exhibited a similar life cycle to that of aphelids and rozellids (Chambouvet et al. 2019). In SSU rRNA gene phylogenetic trees, its sequence formed a clade with other marine environmental sequences (chytrid-like-clade-1, NCLC1), which branched as sister to rozellids, although with weak support. Should this relationship be confirmed by future multi-marker phylogenies, the hypothesis that the last common ancestor of Opisthophagea and Phytophagea fed on algae would gain support, such that the dependency of rozellids and microsporidia on opisthokonts would be secondary (Fig. 3).

Although nucleariids have a non-polarized, non-flagellated cell type, there is a clear similarity between their multinucleated cells and the plasmodial intracellular trophic stage of aphelids and rozellids. This, likely ancient, capacity to grow plasmodial cells may have been a key aspect in the evolution of Holomycota. The adaptation of the Opisthophagea and Phytophagea common ancestor to endobiotic phagotrophy was probably also critical for the subsequent evolution of fungi and microsporidia. Inside the confined environment of the prey cell wall, some of the ancestral holomycotan lineages may have easily lost phagotrophy in favour of extracellular degradation of the algal cell contents by secreting digestive enzymes, leading to the evolution of osmotrophy, a quintessential trait of fungi (Richards and Talbot 2018; Torruella et al. 2018). The release of extracellular enzymes within the host cell wall space would have prevented their diffusion to the surrounding environment, disfavouring possible cheaters and competitors. In fungi, the loss of phagotrophy implied the loss of the trophont stage, such that only a cyst-like stage persisted, i.e. a vegetative form endowed with chitin walls and able to emit rhizoids that penetrated host/food cell walls delivering digestive enzymes. Subsequent evolution enriched the enzymatic digestive battery of ancient fungi, making them capable to degrade a variety of complex cell wall polysaccharides (Chang et al. 2015; Lutzoni et al. 2018). Many of these digestive capabilities might have been acquired from

bacteria. In fact, bacterial symbionts, known for example in nucleariids, have been suggested to protect from algal toxins (Dirren and Posch 2016; Dirren et al. 2017; Galindo et al. 2019) and might have been donors of digestive enzymes through HGT (Garcia-Vallvé et al. 2000; Hall et al. 2005; Marcet-Houben and Gabaldón 2010; Fitzpatrick 2012).

Our results suggest that, while fungi entered an evolutionary pathway marked by their feeding specialization and co-evolution with algae and plants, aphelids seem to have largely retained the ancestral phenotype inferred for their common ancestor with fungi.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.j3tx95xdn]

Funding

This work was supported by the European Research Council (ERC) Advanced Grants 'Protistworld' and 'Plast-Evol' (322669 and 787904, respectively) and the Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN project SINGEK (http://www.singek.eu/; grant agreement no. H2020-MSCA-ITN-2015-675752) and RSF grant No 21-74-20089. Aphelid cultures of the CCPP ZIN RAS collection were supported by a grant of the Ministry of Science and Higher Education of the Russian Federation (no. 075-15-2021-1069).

Acknowledgements

We thank Kirill V. Mikhailov and Vladimir V. Aleoshin for permission to use the *Amoeboaphelidium* sequences and the UNICELL platform (http://www.deemteam.fr/en/unicell) for help in transcriptome production. We also thank Christina Cuomo and the Broad Institute for allowing the use of the unpublished genome sequence of *Allomyces macrogynus* ATCC 38327.

References

Adl S.M., Bass D., Lane C.E., Lukeš J., Schoch C.L., Smirnov A., Agatha S., Berney C., Brown M.W., Burki F., Cárdenas P., Čepička I., Chistyakova L., del Campo J., Dunthorn M., Edvardsen B., Eglit Y., Guillou L., Hampl V., Heiss A.A., Hoppenrath M., James T.Y., Karnkowska A., Karpov S., Kim

464 E., Kolisko M., Kudryavtsev A., Lahr D.J.G., Lara E., Le Gall L., Lynn D.H., Mann D.G., Massana
465 R., Mitchell E.A.D., Morrow C., Park J.S., Pawlowski J.W., Powell M.J., Richter D.J., Rueckert S.,
466 Shadwick L., Shimano S., Spiegel F.W., Torruella G., Youssef N., Zlatogursky V., Zhang Q. 2019.
467 Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol.
468 66:4–119.
469 Almeida B., Buttner S., Ohlmeier S., Silva A., Mesquita A., Sampaio-Marques B., Osório N.S., Kollau
470 A., Mayer B., Leão C., Laranjinha J., Rodrigues F., Madeo F., Ludovico P. 2007. NO-mediated
471 apoptosis in yeast. J. Cell Sci. 120:3279–3288.
472 Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool.
473 J. Mol. Biol. 215:403–410.
474 Anam K., Nasuno R., Takagi H. 2020. A novel mechanism for nitrosative stress tolerance dependent
475 on GTP Cyclohydrolase II activity involved in riboflavin synthesis of yeast. Sci. Rep. 10:6015.
476 Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.
477 Arlt H., Perz A., Ungermann C. 2011. An overexpression screen in *Saccharomyces cerevisiae* identifies
478 novel genes that affect endocytic protein trafficking. Traffic. 12:1592–1603.
479 Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko
480 S.I., Pham S., Prjibelski A.D., Pyshkin A. V., Sirotkin A. V., Vyahhi N., Tesler G., Alekseyev M.A.,
481 Pevzner P.A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell
482 sequencing. J. Comput. Biol. 19:455–477.
483 Baumgärtel V., Ivanchenko S., Dupont A., Sergeev M., Wiseman P.W., Kräusslich H.-G., Bräuchle C.,
484 Müller B., Lamb D.C. 2011. Live-cell visualization of dynamics of HIV budding site interactions
485 with an ESCRT component. Nat. Cell Biol. 13:469–474.
486 Berbee M.L., Strullu-Derrien C., Delaux P.-M., Strother P.K., Kenrick P., Selosse M.-A., Taylor J.W.
487 2020. Genomic and fossil windows into the secret lives of the most ancient fungi. Nat. Rev.
488 Microbiol. 18:717-730.
489 Bereswill S., Fassbinder F., Völzing C., Covacci A., Haas R., Kist M. 1998. Hemolytic properties and
490 riboflavin synthesis of *Helicobacter pylori*: cloning and functional characterization of the *ribA* gene
491 encoding GTP-cyclohydrolase II that confers hemolytic activity to *Escherichia coli*. Med. Microbiol.
492 Immunol. 186:177–187.
493 Brinkmann H., Philippe H. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction
494 artifacts in ancient phylogenies. Mol. Biol. Evol. 16:817–825.
495 Brown M.W., Spiegel F.W., Silberman J.D. 2009. Phylogeny of the "forgotten" cellular slime mold,
496 *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. Mol. Biol. Evol. 26:2699-
497 2709.
498 Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: A tool for automated alignment
499 trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.
500 Chambouvet A., Monier A., Maguire F., Itoïz S., del Campo J., Elies P., Edvardsen B., Eikreim W.,

Richards T.A. 2019. Intracellular infection of diverse diatoms by an evolutionary distinct relative of the Fungi. Curr. Biol. 29:4093-4101.

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

Dean P., Sendra K.M., Williams T.A., Watson A.K., Major P., Nakjang S., Kozhevnikova E., Goldberg A. V., Kunji E.R.S., Hirt R.P., Embley T.M. 2018. Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. Nat. Commun. 9:1709.

Desai N., Brown A., Amunts A., Ramakrishnan V. 2017. The structure of the yeast mitochondrial ribosome. Science. 355:528–531.

Dirren S., Pitsch G., Silva M.O.D., Posch T. 2017. Grazing of *Nuclearia thermophila* and *Nuclearia delicatula* (Nucleariidae, Opisthokonta) on the toxic cyanobacterium *Planktothrix rubescens*. Eur. J. Protistol. 60:87–101.

Dirren S., Posch T. 2016. Promiscuous and specific bacterial symbiont acquisition in the amoeboid genus *Nuclearia* (Opisthokonta). FEMS Microbiol. Ecol. 92:fiw105.

Elia N., Sougrat R., Spurlin T.A., Hurley J.H., Lippincott-Schwartz J. 2011. Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its role in abscission. Proc. Natl. Acad. Sci. U. S. A. 108:4846–4851.

Emms D.M., Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Fitzpatrick D.A. 2012. Horizontal gene transfer in fungi. FEMS Microbiol. Lett. 329:1–8.

Galindo L.J., López-García P., Torruella G., Karpov S., Moreira D. 2021. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. Nat. Commun. 12:4973.

Galindo L.J., Torruella G., Moreira D., Eglit Y., Simpson A.G.B., Völcker E., Clauß S., López-García P. 2019. Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. Philos. Trans. R. Soc. B Biol. Sci. 374:20190094.

Garcia-Vallvé S., Romeu A., Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol. Biol. Evol. 17:352–361.

Gromov B. V. 2000. Algal parasites of the genera *Aphelidium*, *Amoeboaphelidium*, and *Pseudaphelidium* from the Cienkovski's "monadinea" group as representatives of a new class. Zool. Zhurnal. 79:523-525.

Hall C., Brachat S., Dietrich F.S. 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. Eukaryot. Cell. 4:1102–1115.

Herz S., Eberhardt S., Bacher A. 2000. Biosynthesis of riboflavin in plants. The *ribA* gene of *Arabidopsis thaliana* specifies a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-

538 phosphate synthase. Phytochemistry. 53:723–731.
539 Hromatka B.S., Noble S.M., Johnson A.D. 2005. Transcriptional response of *Candida albicans* to nitric
540 oxide and the role of the YHB1 gene in nitrosative stress and virulence. Mol. Biol. Cell. 16:4814–
541 4826.
542 James T.Y., Pelin A., Bonen L., Ahrendt S., Sain D., Corradi N., Stajich J.E. 2013. Shared signatures
543 of parasitism and phylogenomics unite cryptomycota and microsporidia. Curr. Biol. 23:1548–1553.
544 Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Haeseler A. Von, Jermiin L.S. 2017. ModelFinder : fast
545 model selection for accurate phylogenetic estimates. Nat. Methods. 14:587-589.
546 Karpov S., Vishnyakov A., López-García P., Zorina N., Ciobanu M., Tcvetkova V., Moreira D. 2020.
547 Morphology and molecular phylogeny of *Aphelidium insulamus* sp. nov. (Aphelida,
548 Opisthosporidia). Protistol. 14:191–203.
549 Karpov S.A., Cvetkova V.S., Annenkova N. V., Vishnyakov A.E. 2019. Kinetid structure of *Aphelidium*
550 and *Paraphelidium* (Aphelida) suggests the features of the common ancestor of Fungi and
551 Opisthosporidia. J. Eukaryot. Microbiol. 66:911-924.
552 Karpov S.A., Mamkaeva M.A., Aleoshin V. V., Nassonova E., Lilje O., Gleason F.H. 2014.
553 Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for
554 the new superphylum Opisthosporidia. Front. Microbiol. 5:112.
555 Karpov S.A., Mamkaeva M.A., Moreira D., López-García P. 2016. Molecular phylogeny of *Aphelidium*
556 *tribonemae* reveals its sister relationship with *A . aff . melosirae* (Aphelida, Opisthosporidia).
557 Protistology. 10:97–103.
558 Karpov S.A., Mikhailov K. V, Mirzaeva G.S. 2013. Obligately phagotrophic aphelids turned out to
559 branch with the earliest-diverging fungi. Ann. Anat. 164:195–205.
560 Karpov S.A., Tcvetkova V.S., Mamkaeva M.A., Torruella G., Timpano H., Moreira D., Mamanazarova
561 K.S., López-García P. 2017. Morphological and genetic diversity of Opisthosporidia: New aphelid
562 *Paraphelidium tribonemae* gen. et sp. nov. J. Eukaryot. Microbiol. 64:204-212.
563 Katoh K., Misawa K., Kuma K.I., Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence
564 alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
565 Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7:
566 Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
567 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A.,
568 Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic:
569 An integrated and extendable desktop software platform for the organization and analysis of
570 sequence data. Bioinformatics. 28:1647–1649.
571 Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets.
572 Bioinformatics. 30:3276–3278.
573 Lartillot N., Lepage T., Blanquart S. 2009. PhyloBayes 3: A Bayesian software package for
574 phylogenetic reconstruction and molecular dating. Bioinformatics. 25:2286–2288.

575 Lartillot N., Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-
576 acid replacement process. Mol. Biol. Evol. 21:1095–1109.
577 Lax G., Eglit Y., Eme L., Bertrand E.M., Roger A.J., Simpson A.G.B. 2018. Hemimastigophora is a
578 novel supra-kingdom-level lineage of eukaryotes. Nature. 564:410–414.
579 Letcher P.M., Powell M.J. 2018. A taxonomic summary and revision of *Rozella* (Cryptomycota). IMA
580 Fungus. 9:383–399.
581 Letcher P.M., Powell M.J. 2019. A taxonomic summary of Aphelidiaceae. IMA Fungus. 10:1–11.
582 Letcher P.M., Powell M.J., Lopez S., Lee P.A., McBride R.C. 2015. A new isolate of
583 *Amoeboaphelidium protococcarum*, and *Amoeboaphelidium occidentale*, a new species in phylum
584 Aphelida (Opisthosporidia). Mycologia. 107:522–531.
585 Li W., Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or
586 nucleotide sequences. Bioinformatics. 22:1658–1659.
587 Liao Z., Yan Y., Dong H., Zhu Z., Jiang Y., Cao Y. 2016. Endogenous nitric oxide accumulation is
588 involved in the antifungal activity of Shikonin against *Candida albicans*. Emerg. Microbes Infect.
589 5:e88.
590 Liu Y., Steenkamp E.T., Brinkmann H., Forget L., Philippe H., Lang B.F. 2009. Phylogenomic analyses
591 predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with
592 significant support. BMC Evol. Biol. 9:272.
593 Lutzoni F., Nowak M.D., Alfaro M.E., Reeb V., Miadlikowska J., Krug M., Arnold A.E., Lewis L.A.,
594 Swofford D.L., Hibbett D., Hilu K., James T.Y., Quandt D., Magallón S. 2018. Contemporaneous
595 radiations of fungi and plants linked to symbiosis. Nat. Commun. 9:5451.
596 Maddison W.P., Maddison D.R.. 2021. Mesquite: a modular system for evolutionary analysis. Version
597 3.70 http://www.mesquiteproject.org
598 Marcet-Houben M., Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends
599 Genet. 26:5–8.
600 Martin R., Wächtler B., Schaller M., Wilson D., Hube B. 2011. Host–pathogen interactions and
601 virulence-associated genes during *Candida albicans* oral infections. Int. J. Med. Microbiol.
602 301:417–422.
603 Mikhailov K.V., Simdyanov T.G., Aleoshin V.V. 2017. Genomic survey of a hyperparasitic
604 microsporidian *Amphiamblys* sp. (Metchnikovellidae). Genome Biol. Evol. 9:454–467.
605 Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of
606 large phylogenetic trees. Gatew. Comput. Environ. Work. GCE 2010.
607 Minh B.Q., Nguyen M.A.T., Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap.
608 Mol. Biol. Evol. 30:1188–1195.
609 Mistry J., Finn R.D., Eddy S.R., Bateman A., Punta M. 2013. Challenges in homology search:
610 HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41:e121.
611 Naranjo-Ortiz M.A., Gabaldón T. 2019. Fungal evolution: major ecological adaptations and

612 evolutionary transitions. Biol. Rev. 94:1443–1476.
613 Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective
614 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.
615 Oliveira D.L., Rizzo J., Joffe L.S., Godinho R.M.C., Rodrigues M.L. 2013. Where do they come from
616 and where do they go: candidates for regulating extracellular vesicle formation in fungi. Int. J. Mol.
617 Sci. 14:9581–9603.
618 Oltmanns O., Bacher A., Lingens F., Zimmermann F.K. 1969. Biochemical and genetic classification
619 of riboflavine deficient mutants of *Saccharomyces cerevisiae*. Mol. Gen. Genet. 105:306–313.
620 Pfeffer S., Woellhaf M.W., Herrmann J.M., Förster F. 2015. Organization of the mitochondrial
621 translation machinery studied in situ by cryoelectron tomography. Nat. Commun. 6:6019.
622 Quandt C.A., Beaudet D., Corsaro D., Walochnik J., Michel R., Corradi N., James T.Y. 2017. The
623 genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative
624 adaptations to obligate intracellular parasitism. Elife. 6:e29594.
625 Rambaut A. 2016. FigTree v1.4.3. http://tree.bio.ed.ac.uk/software/figtree/.
626 Richards T.A., Talbot N.J. 2018. Osmotrophy. Curr. Biol. 28:R1179–R1180.
627 Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and
628 overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389-399.
629 Saveanu C., Fromont-Racine M., Harington A., Ricard F., Namane A., Jacquier A. 2001. Identification
630 of 12 new yeast mitochondrial ribosomal proteins including 6 that have no prokaryotic homologues.
631 J. Biol. Chem. 276:15861–15867.
632 Schultzhaus Z.S., Shaw B.D. 2015. Endocytosis and exocytosis in hyphal growth. Fungal Biol. Rev.
633 29:43–53.
634 Schweikert M., Schnepf E. 1996. *Pseudaphelidium drebesii*, gen. et spec. nov. (incerta sedis), a parasite
635 of the marine centric diatom *Thalassiosira punctigera*. Arch. Protistenkd. 147:11–17.
636 Schweikert M., Schnepf E. 1997. Electron microscopical observations on *Pseudaphelidium drebesii*
637 Schweikert and Schnepf, a parasite of the centric diatom *Thalassiosira punctigera*. Protoplasma.
638 199:113–123.
639 Seto K., Matsuzawa T., Kuno H., Kagami M. 2020. Morphology, Ultrastructure, and molecular
640 phylogeny of *Aphelidium collabens* sp. nov. (Aphelida), a parasitoid of a green alga *Coccomyxa* sp.
641 Protist. 171:125728.
642 Sibbald S.J., Archibald J.M. 2017. More protist genomes needed. Nat. Ecol. Evol. 1:145.
643 Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E. V., Zdobnov E.M. 2015. BUSCO:
644 Assessing genome assembly and annotation completeness with single-copy orthologs.
645 Bioinformatics. 31:3210–3212.
646 Simon M., Jardillier L., Deschamps P., Moreira D., Restoux G., Bertolino P., López-García P. 2015.
647 Complex communities of small protists and unexpected occurrence of typical marine lineages in
648 shallow freshwater systems. Environ. Microbiol. 17:3610–3627.

649 Spatafora J.W., Chang Y., Benny G.L., Lazarus K., Smith M.E., Berbee M.L., Bonito G., Corradi N.,
650 Grigoriev I., Gryganskyi A., James T.Y., O'Donnell K., Roberson R.W., Taylor T.N., Uehling J.,
651 Vilgalys R., White M.M., Stajich J.E. 2016. A phylum-level phylogenetic classification of
652 zygomycete fungi based on genome-scale data. Mycologia. 108:1028–1046.
653 Spoonamore J.E., Dahlgran A.L., Jacobsen N.E., Bandarian V. 2006. Evolution of new function in the
654 GTP Cyclohydrolase II proteins of *Streptomyces coelicolor*. Biochemistry. 45:12144–12155.
655 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large
656 phylogenies. Bioinformatics. 30:1312-1313.
657 Tedersoo L., Sánchez-Ramírez S., Kõljalg U., Bahram M., Döring M., Schigel D., May T., Ryberg M.,
658 Abarenkov K. 2018. High-level classification of the Fungi and a tool for evolutionary ecological
659 analyses. Fungal Divers. 90:135–159.
660 Tikhonenkov D. V, Mikhailov K. V, Hehenberger E., Karpov S.A., Prokina K.I., Esaulov A.S.,
661 Belyakova O.I., Mazei Y.A., Mylnikov A.P., Aleoshin V. V, Keeling P.J. 2020. New lineage of
662 microbial predators adds complexity to reconstructing the evolutionary origin of animals. Curr. Biol.
663 30:4500-4509.
664 Torruella G., Grau-Bové X., Moreira D., Karpov S.A., Burns J.A., Sebé-Pedrós A., Völcker E., López-
665 García P. 2018. Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports
666 the phagotrophic origin of fungi. Commun. Biol. 1:231.
667 Vávra J., Lukeš J. 2013. Microsporidia and 'The Art of Living Together.' Adv. Parasitol. 82:253-319.
668 Vild C.J., Xu Z. 2014. Vfa1 binds to the N-terminal microtubule-interacting and trafficking (MIT)
669 domain of Vps4 and stimulates its ATPase activity. J. Biol. Chem. 289:10378–10386.
670 Wadi L., Reinke A.W. 2020. Evolution of microsporidia: An extremely successful group of eukaryotic
671 intracellular parasites. PLoS Pathog. 16:e1008276.
672 Wang H.C., Minh B.Q., Susko E., Roger A.J. 2018. Modeling site heterogeneity with posterior mean
673 site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67:216–235.
674 Williams R.L., Urbé S. 2007. The emerging shape of the ESCRT machinery. Nat. Rev. Mol. Cell Biol.
675 8:355–368.
676 Yadav S., Karthikeyan S. 2015. Structural and biochemical characterization of GTP cyclohydrolase II
677 from *Helicobacter pylori* reveals its redox dependent catalytic activity. J. Struct. Biol. 192:100–115.
678 Yoshikawa Y., Nasuno R., Kawahara N., Nishimura A., Watanabe D., Takagi H. 2016. Regulatory
679 mechanism of the flavoprotein Tah18-dependent nitric oxide synthesis and cell death in yeast. Nitric
680 Oxide. 57:85–91.
681 Zakikhany K., Naglik J.R., Schmidt-Westhausen A., Holland G., Schaller M., Hube B. 2007. In vivo
682 transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination.
683 Cell. Microbiol. 9:2938–2954.
684 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree
685 reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153.

686 Zopf W. 1885. Zur Morphologie und Biologie der niederen Pilzthiere (Monadinen), zugleich ein Beitrag
687 zur Phytopathologie. Leipzig :Veit & Comp.
688
689

FIGURE LEGENDS

FIGURE 1. Phylogenomic analysis of Holomycota. a) Bayesian inference (BI) phylogenomic tree based on 175 conserved proteins from Lax et al. (2018). The tree was reconstructed using 25 species and 59,889 amino acid positions with the CAT-GTR model and the LG+F+R6+C60 model using the PMSF approximation for maximum likelihood (ML). Numbers on branches indicate Bayesian posterior probabilities and ML bootstrap values; bootstrap values <50% are indicated by ---. Branches with support values higher or equal to 0.99 BI posterior probability and 99% ML bootstrap are indicated by black dots; b) ML bootstrap support for the monophyly of Fungi ("Fungi") and two competing hypotheses for the position of Aphelida ("Aphelida+Fungi" and "Opisthosporidia" (Aphelida+Rozellida+Microsporidia)) as a function of the proportion of fast-evolving sites removed from the phylogenomic dataset; c) ASTRAL coalescent-based species tree estimated with the 175 individual protein ML trees. All phylogenomic trees can be seen in Supplementary Fig. 1 a-e, Supplementary Material available on Dryad.

FIGURE 2. Predicted cellular localization of identified synapomorphic proteins shared by aphelids and fungi. Diagram of an aphelid+fungi ancestor in which black lines indicate the localization of each protein orthogroup. Orthogroup numbers labelled with an asterisk (*) indicate proteins derived from horizontal gene transfer events; orthogroup numbers without asterisk correspond to protein shared by aphelids and fungi which likely evolved *de novo* in a common ancestor of these two lineages. For the list of proteins see Table S3, Supplementary Material available on Dryad.

FIGURE 3. Evolution of key phenotypic traits across Holomycota. Boxes show the ancestral traits present in the Phytophagea (fungi+aphelids, green), Opisthophagea (rozellids+microsporidians, orange) and nucleariid (blue) clades and in the last common ancestors (LCA) of Holomycota and of both Phytophagea and Opisthophagea (grey). For details on the ancestral state reconstruction, see Supplementary Figs. S21 and S22, Supplementary Material available on Dryad.

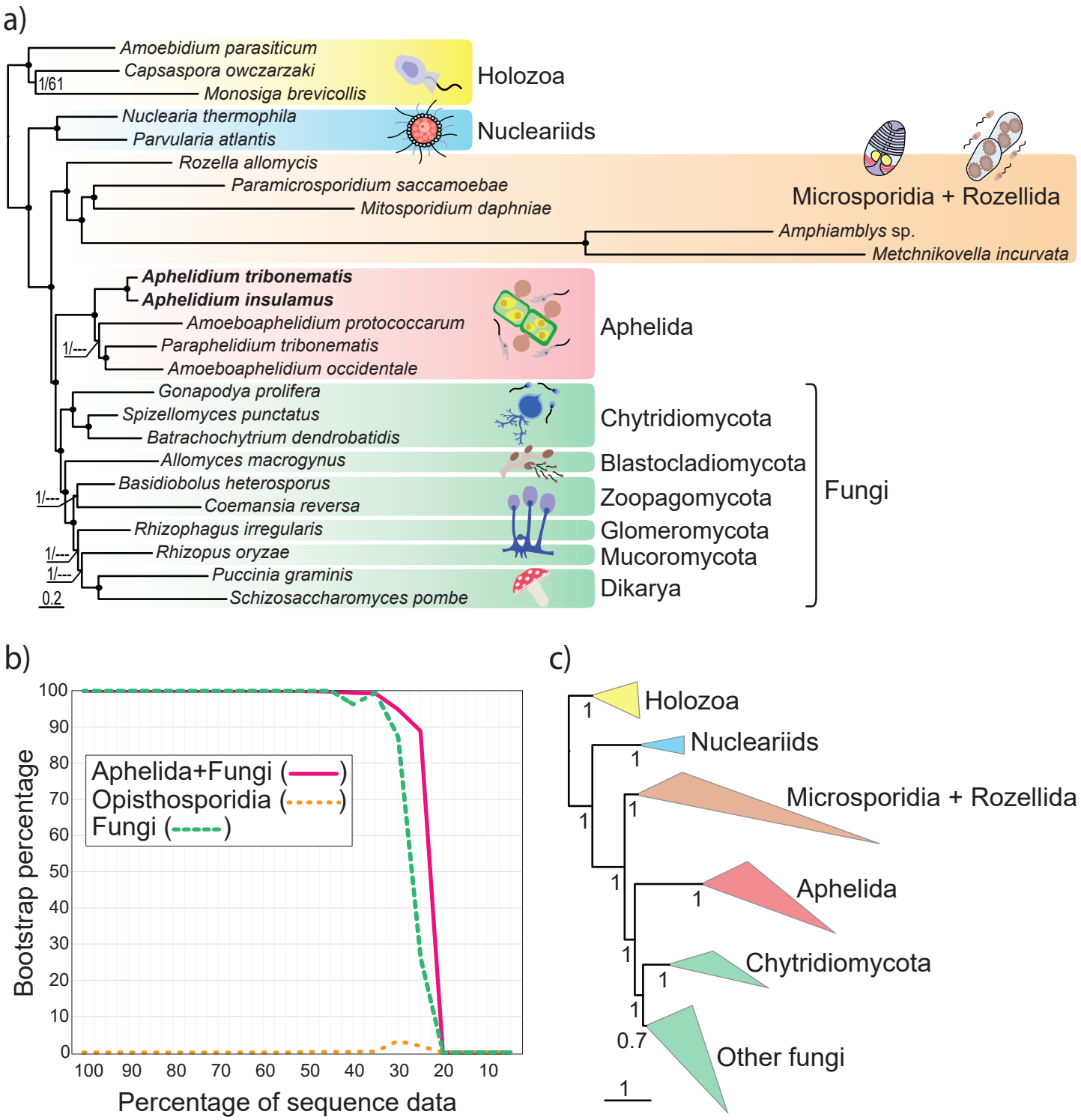

FIGURE 1. Phylogenomic analysis of Holomycota. a) Bayesian inference (BI) phylogenomic tree based on 175 conserved proteins from Lax et al. (2018). The tree was reconstructed using 25 species and 59,889 amino acid positions with the CAT-GTR model and the LG+F+R6+C60 model using the PMSF approximation for maximum likelihood (ML). Numbers on branches indicate Bayesian posterior probabilities and ML bootstrap values; bootstrap values <50% are indicated by ---. Branches with support values higher or equal to 0.99 BI posterior probability and 99% ML bootstrap are indicated by black dots; b) ML bootstrap support for the monophyly of Fungi ("Fungi") and two competing hypotheses for the position of Aphelida ("Aphelida +Fungi" and "Opisthosporidia" (Aphelida+Rozellida+Microsporidia)) as a function of the proportion of fast-evolving sites removed from the phylogenomic dataset; c) ASTRAL coalescent-based species tree estimated with the 175 individual protein ML trees. All phylogenomic trees can be seen in Supplementary Fig. 1 a-e, Supplementary Material available on Dryad.

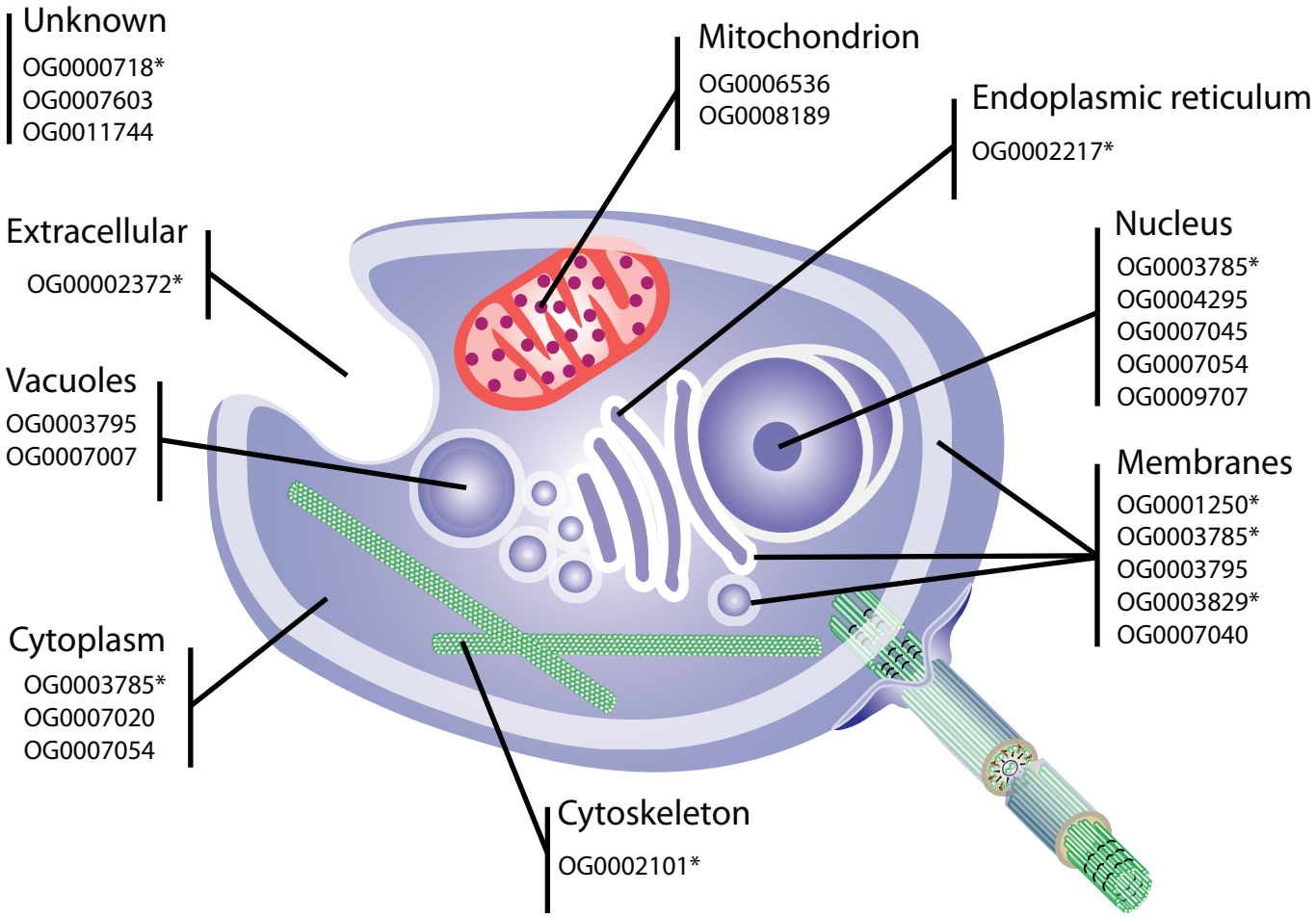

FIGURE 2. Predicted cellular localization of identified synapomorphic proteins shared by aphelids and fungi. Diagram of an aphelid+fungi ancestor in which black lines indicate the localization of each protein orthogroup. Orthogroup numbers labelled with an asterisk (*) indicate proteins derived from horizontal gene transfer events; orthogroup numbers without asterisk correspond to protein shared by aphelids and fungi which likely evolved de novo in a common ancestor of these two lineages. For the list of proteins see Table S3, Supplementary Material available on Dryad.

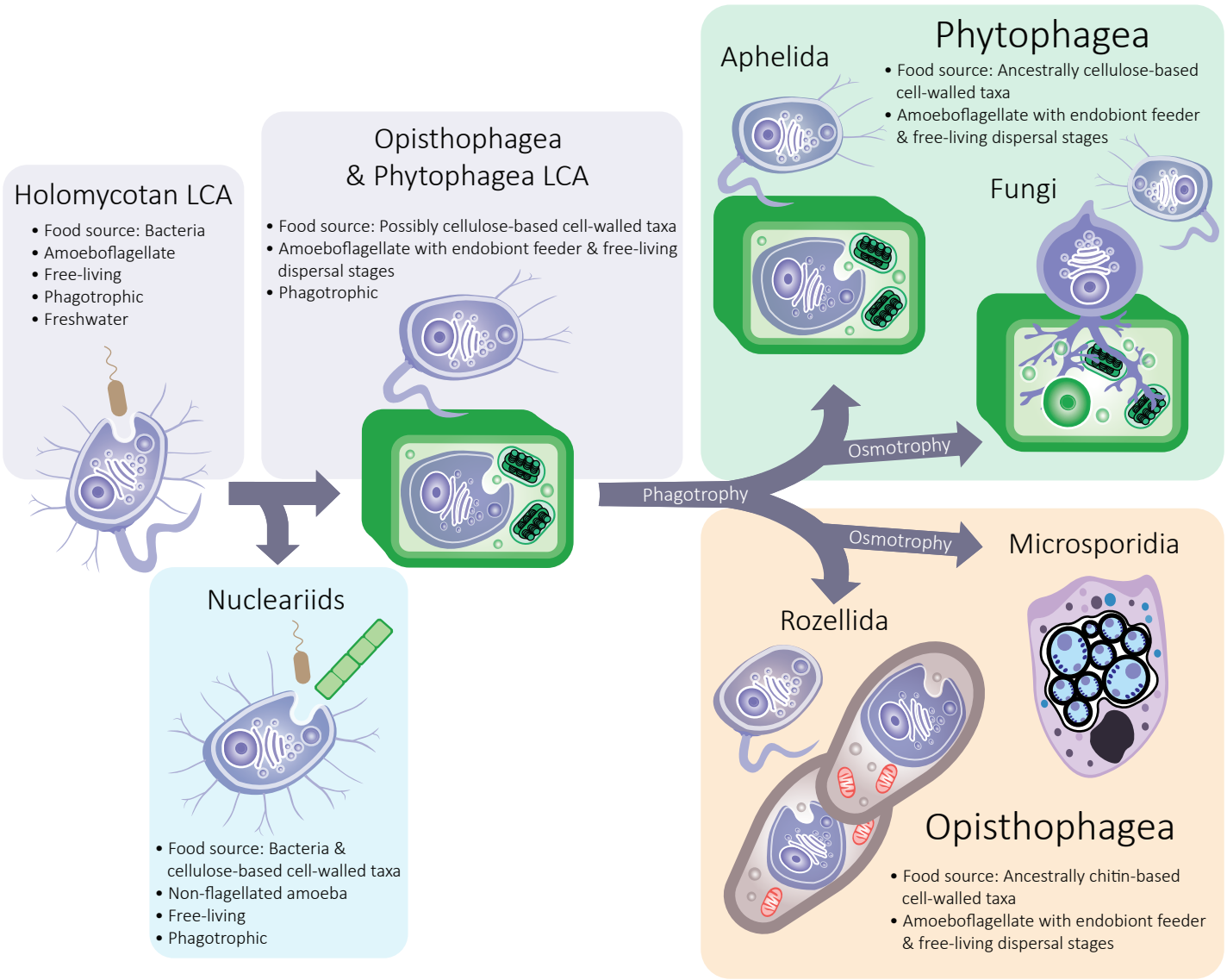

FIGURE 3. Evolution of key phenotypic traits across Holomycota. Boxes show the ancestral traits present in the Phytophagea (fungi+aphelids, green), Opisthophagea (rozellids+microsporidians, orange) and nucleariid (blue) clades and in the last common ancestors (LCA) of Holomycota and of both Phytophagea and Opisthophagea (grey). For details on the ancestral state reconstruction, see Supplementary Figs. S21 and S22, Supplementary Material available on Dryad.