Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi

Guifré Torruella\(^1\), Xavier Grau-Bové\(^2\), David Moreira\(^1\), Sergey A. Karpov\(^1,3\), John A. Burns\(^4\), Arnau Sebé-Pedrós\(^5\), Eckhard Völcker\(^6\) & Purificación López-García\(^1\)

Aphelids are little-known phagotrophic parasites of algae whose life cycle and morphology resemble those of the parasitic rozellids (Cryptomycota, Rozellomycota). In previous phylogenetic analyses of RNA polymerase and rRNA genes, aphelids, rozellids and Microsporidia (parasites of animals) formed a clade, named Opisthosporidia, which appeared as the sister group to Fungi. However, the statistical support for the Opisthosporidia was always moderate. Here, we generated full life-cycle transcriptome data for the aphelid species *Paraphelidium tribonemae*. In-depth multi-gene phylogenomic analyses using several protein datasets place this aphelid as the closest relative of fungi to the exclusion of rozellids and Microsporidia. In contrast with the comparatively reduced *Rozella allomycis* genome, we infer a rich, free-living-like aphelid proteome, with a metabolism similar to fungi, including cellulases likely involved in algal cell-wall penetration and enzymes involved in chitin biosynthesis. Our results suggest that fungi evolved from complex aphelid-like ancestors that lost phagotrophy and became osmotrophic.
fungi are an extremely ecologically successful group of organisms involved in the saprophytic degradation of organics, with some parasitic species. They feed by osmotrophy or direct uptake of dissolved organics through their chitin cell-wall. Traditionally, fungi encompassed several lineages of filamentous species, most of which belong to the Dikarya (Ascomycota and Basidiomycota), and a group of flagellated species known as chytrids (Chytridiomycota). The first molecular phylogenetic analyses based on one or a few genes confirmed the monophyly of Dikarya but also revealed that some deep-branching fungal lineages were not monophyletic, notably the chytrids and the zygomycetes. In parallel, environmental studies have uncovered a wide diversity of sequences belonging to new lineages of unicellular eukaryotes branching at the base of the classical fungi. One of those lineages (known by several names: Rozellida, Cryptomycota, Rozellomycoa) includes the flagellated phagotrophic unicellular parasite Rozella allomyces, which branches as a sister lineage to classical fungi in phylogenetic trees. Another group corresponds to the Microsporidia, well-known fast-evolving parasites, whose deep phylogenetic branches are progressively being populated by new, slowly-evolving species. The third clade of organisms that appear to relate to these lineages is the aphelids (Aphelida). While some authors include all of these protist lineages as fungi (fungi sensu lato), others do not. Delimiting what defines a fungus is not easy, especially when the phylogenetic relationships among fungi-related protists remain undetermined. Here, we refer to fungi as encompassing the classical core fungi (filamentous and chytrid fungi), which feed by osmotrophy through chitinous cell walls.

Aphelids constitute a group of diverse, yet poorly known, parasites of algae, their life cycle and morphology resemble those of zoospores (chytrids) and rozellids (Cryptomycota/Rozellosporea), another species-rich group of parasites of fungi and oomycetes. Unlike fungi, which are osmotrophs, aphelids and rozellids are phagotrophs, feeding on the host’s cytoplasm. Combined RNA polymerase and rRNA gene trees suggested that aphelids and rozellids relate to Microsporidia, extremely reduced parasites with remnant mitochondria unable to perform phagotrophy. Accordingly, aphelids, rozellids, and Microsporidia were proposed to be monophyletic and form a sister group to fungi called Opisthosporidia. However, the limited phylogenetic signal of those genes combined with fast-evolving sequences from microsporidians have resulted in incongruent tree topologies, showing either rozellids or aphelids as the earliest-branching lineages of Opisthosporidia. Furthermore, the support for the monophyly of Opisthosporidia was always moderate, leaving the relative order of emergence of the highly diverse fungal relatives unresolved. To improve the phylogenetic signal for an accurate placement of aphelids in the opisthokont branch containing fungi, nuclearids, rozellids, and Microsporidia (usually referred to as Holomycota), we have generated the first full life-cycle transcriptome data for one aphelid species, Paraphelidium tribonemae, to our knowledge. Phylogenomic analyses of this almost-complete transcriptome including, in particular, data from the Rozella allomyces genome, show that Opisthosporidia are paraphyletic and that P. tribonemae branches as the sister lineage to Fungi. Comparison of gene sets involved in metabolism and major cellular functions strongly suggests that fungi evolved from complex aphelid-like ancestors that were phagotrophic.

Results
Aphelids occupy a deep pivotal position and are the sister group to fungi. To generate the aphelid transcriptome and because P. tribonemae has a complex life cycle (Fig. 1), we maximized transcript recovery by constructing two cDNA libraries corresponding to young and old enrichment cultures of its host, the yellow-green alga Tribonema guayanum, infected with the aphelid. Accordingly, transcripts corresponding to zoospores, infective cysts, trophohyphs, plasmidia, sporangia and sporocysts were represented. After paired-end Illumina HiSeq2500 sequencing, we assembled a metatranscriptome of 68,130 contigs corresponding to the aphelid, its host and bacterial contaminants. After a sequential process of supervised cleaning, including comparison with a newly generated transcriptome of the algal host, we obtained a final dataset of 10,439 protein sequences that were considered free of contaminants. The final predicted proteome (Paraphelidium tribonemae version 1.5; Supplementary Data 1) was 91.4% complete according to BUSCO. We found no stop codons interrupting coding sequences. Therefore, in contrast to Ameobophidium protococcarum, for which the TAG and TAA stop codons appear to encode glutamine, P. tribonemae does not possess these modifications from the canonical genetic code.

Because resolving the phylogenetic relationships in this part of the eukaryotic tree is challenging due to the occurrence of lineages with very different evolutionary rates, notably the fast-evolving Microsporidia, we used three different datasets and carried out a battery of phylogenomic analyses including and excluding the fastest-evolving species and amino acid positions from our datasets. The three datasets have been previously used to study the phylogeny of animals and related protists (Holozoa, the sister group to Holomycota) within Opisthokonts (SCPD), Microsporidia (including highly conserved genes for this fast-evolving group, BMC), and Microsporidia and related deeper-branching lineages (maximizing the putative orthologs in the whole clade, GBE). First, we incorporated P. tribonemae orthologs to a dataset of 93 single-copy protein domains (SCPD dataset) previously used for phylogenomic analyses of basal-branching opisthokonts, updating it with several microsporidian species including the early-branching Mitosporidium daphniae, the metchnikovellid Amphibiomyces sp., and the recently released genome of Paramicrosporidium saccamoebae. We then generated Bayesian inference trees reconstructed under the CAT-Poisson mixture model and maximum likelihood (ML) trees constructed under C60, the closest mixture model to CAT, for a taxon sampling including 49 species (22,976 positions). The two phylogenomic analyses yielded the same topology. We recovered the monophyly of all opisthokont lineages previously reported. In contrast with previous analyses based on 18S rRNA and RNA polymerase genes, Opisthosporidia appeared as paraphyletic and the aphelid was placed as the sister lineage to fungi. However, the ML bootstrap support was moderate (ultrafast bootstrapping; 96% ubs) and the Bayesian probability was low (0.66 pp; even if the two Bayesian chains converged in terms of likelihood, the consensus topology was not distinctively supported by the other chain). To limit potential long-branch attraction artefacts derived from the inclusion of fast-evolving Microsporidia, we built a reduced dataset without the thirteen fastest-evolving microsporidian species (SCPD36, 36 species; 24,413 positions). Bayesian and ML trees again reproduced the same topology (Supplementary Fig. 1) with higher pp (0.88) and ubs (95%) values, but still not fully supporting the consensus topology (Fig. 1; Table 1).

To resolve the phylogenetic position of the aphelid with higher support, we analyzed two additional datasets previously used for phylogenomic analyses of Microsporidia with the same two sets of 49 and 36 opisthokont species. The resulting datasets were called BMC49 and BMC36 for a set of 53 protein markers, which, after concatenation, yielded 15,744 and 16,840 amino acid positions, respectively; and GBE49 and GBE36 for datasets with 259 protein markers, which resulted in 84,810 and 96,640
amino acids, respectively. For all datasets, Bayesian and ML trees recovered the monophyly of Paraphelidium and fungi (Supplementary Fig. 1) and, for the two datasets without fast-evolving Microsporidia, posterior probabilities and ML bootstrap support were maximal (Fig. 1, Table 1, Supplementary Data 2). In addition, we carried out alternative topology tests for the three datasets containing 49 species. All of the alternative topology tests supported P. tribonemae as the sister group to fungi (AU test $p$-value significantly excluded if $<0.1$), and both BMC49 and GBE49 significantly excluded the Opisthosporidia monophyly (AU $p$-value 0.0039 and 0.0000, respectively; Supplementary Data 2).

![Diagram](https://example.com/diagram.png)

**Fig. 1** Phylogenomic analyses and cell cycle of the aphelid Paraphelidium tribonemae. a Bayesian phylogenetic tree based on single-copy protein domains for 49 species (SCPD49) inferred using a CAT-Poisson model. Statistical supports indicated at some crucial nodes correspond, from left to right, to PhyloBayes Bayesian posterior probabilities and IQ-TREE ML ultrafast-bootstrap support using the C60 model. Branches with maximum support values (pp = 100%) are indicated by black circles. The support for the monophyly of Paraphelidium and fungi (blue star) using the three datasets (SCPD, BMC and GBE) with the 49 species and after removing the fastest-evolving microsporidian sequences (36 species) is shown in Table 1. b Schematic cell cycle of P. tribonemae. Briefly, infecting cysts (red wall), deliver an ameboid trophont to an algal filament cell via an infection tube; the trophont engulfs the algal cytoplasm by phagocytosis, leaving a growing residual body (dark red particle); after nuclear and cell division, a mature sporangium releases amoeboid or flagellated zoospores (occasionally amoeboid only) that get out the algal cell wall and close the life cycle11. c Evolution of IQ-TREE ML ubs support for the monophyly of aphelids and fungi ($A+F$) and the monophyly of Opisthosporidia ($O$) as a function of the proportion of fast-evolving sites removed from the dataset. All the phylogenomic trees can be seen in Supplementary Fig. 1.

**Table 1** Statistical support for the monophyly of Paraphelidium and fungi (blue star in Fig. 1) in multi-gene phylogenetic trees using three datasets (SCPD, BMC, and GBE) and two sets of species including (49 species) or excluding (36 species) the fastest-evolving microsporidian sequences.

| Dataset   | 49 Species | 36 Species |
|-----------|------------|------------|
| SCPD      | 0.66/96    | 0.88/95    |
| BMC       | 1/100      | 1/100      |
| GBE       | 0.5/100    | 1/100      |
Enzymes involved in cell-wall synthesis and degradation.

Despite secondary losses in some fungi, the presence of chitin in cell walls was long considered a typical fungal trait. However, chitin is also present in the cell wall of many other protists across the eukaryotic tree, implying that the machinery for chitin synthesis/remodeling originated prior to the radiation of fungi and other eukaryotic lineages. Microsporidia and Rozella are also able to synthesize chitin, but, unlike fungi, which possess chitin cell walls during the vegetative stage, they produce chitin only in cysts or resting spores. Staining with fluorescently-labeled Wheat Germ Agglutinin (WGA) showed that Paraphelidium also possesses chitin in the wall of infecting cysts, but not in zoospores or, unlike fungi, in the vegetative stage (i.e., the trophonts). In agreement with this observation, we identified homologs of chitin synthases, chitin deacetylases, chitinases and 1,3-beta-glucan synthases in the transcriptome of Paraphelidium (Supplementary Fig. 2a–d). Specifically, we detected seven homologous sequences (including all alternative transcripts such as alleles or splice variants) of division II chitin synthases in Paraphelidium corresponding to at least six distinct peptides (Rozella contains only four). Three of them clustered with class IV chitin synthases, including Microsporidia and Rozella homologs (Supplementary Fig. 2a). The remaining four sequences branched within class V/VII enzymes, two of them (probably corresponding to a single polypeptide) forming a deep-branching group with fungal, mostly chytrid, sequences (Supplementary Fig. 2a). Class V enzymes include a myosin motor thought to intervene in polarized fungal hyphal growth that has been hypothesized to take part in the formation of the germ tube in aphelids and rozellids. Class V chitin synthases were lost in Microsporidia (with the exception of Mitosporidium), still retaining, like Rozella, one homolog, endowed instead with highly specialized polar tube extrusion mechanisms. Neither spore wall nor polar tube proteins specific to Microsporidia occurred in the Paraphelidium transcriptome. Therefore, our data lend credit to the hypothesis that class V chitin synthases are involved in germ tube polar growth.

Among the rest of chitin-related enzymes, we identified twelve sequences (at least five different homologs) of chitin deacetylas. We detected at least three class II chitinase homologs (eight total sequences), which are ancestral in opisthokonts, containing the Glyco_hydro_18 (PF00704) domain, and a class I chitinase (CTS1) with a Glyco_hydro_19 (PF00182) domain. The latter included the catalytic site, an N-terminal predicted signal peptide and a transmembrane region, suggesting an extracellular chitinase activity. CTS1 has a peculiar phylogenetic distribution in eukaryotes, occurring only in Viridiplantae, Fungi, Microsporidia and Ecdysozoa (Supplementary Fig. 2c). Rozella contains two homologs and Microsporidia at least one; they have N-terminal signal peptides and are predicted to localize extracellularly but lack transmembrane domains. In our phylogenetic tree,
opisthosporidian sequences appeared scattered within metazoan and fungal sequences. This might be the result of hidden paralogy and/or horizontal gene transfer (HGT) (Supplementary Fig. 2c). Regardless of its evolutionary origin, aphelid CTSI might be involved in the self-degradation of resting spore and cyst wall chitin. This might happen both, during their release from chitin-containing resting spores or at the tip of the germ tube during infection, as previously suggested for *Rozella* (Fig. 1).

Although not found in chytrids, 1,3-beta-glucan is also considered an idiosyncratic fungal cell-wall component. Surprisingly, we identified two 1,3-beta-glucan synthase (FKS1) homologs (four sequences), with split Glucan synthase (PF02364) and FKS1_dom1 (PF14288) domains (fused for the phylogenetic analysis) (Supplementary Fig. 2d). The presence of FKS1, absent in *Rozella* and Microsporida, in aphelids traces its origin back to the ancestor of fungi and aphelids.

To feed on the algal cytoplasm, aphelids need to traverse the algal cell wall but, so far, the specific penetration mechanism, whether mechanical (germ-tube penetration through the gap between the two algal cell-wall halves) or enzymatic (digestion of algal cell-wall components) was uncertain. Scanning electron microscopy (SEM) observations showed both, clear cases of mechanical penetration between the two algal cell-wall halves, and also infecting cysts scattered on the algal cell-wall surface (Fig. 2c). SEM images additionally confirmed WGA-epifluorescence observations of multiple parasitoid cysts co-infecting the same host cell (Fig. 2a–c). Multiple infections, implying the coexistence of distinct trophonts feeding within the same algal cell-wall delimited compartment, open the intriguing possibility for aphelids to engage in sexual reproduction within the algal host. Our *Paraphelidium* transcriptome has some genes involved in meiosis (e.g., MutS protein homolog 4), but not all (Supplementary Data 1). However, the transcriptome is not complete and, at the same time, many of the detected genes are also involved in more general recombination (e.g., meiotic recombination protein DMCI) and DNA repair (e.g., multiple helicases homologous to HFMI) processes. Therefore, whether the aphelid has the potential to carry out meiosis and whether it might indeed occur at the trophont stage will require future study. *Trichonema* cell walls contain cellulose II based on 1,6-linked glucan (alkali soluble cellulose), 1,3 and 1,4-linked xylose, 1,3-linked rhamnose and, mostly, 1,3; 1,4; and 1,6-linked glucose. We performed sequence similarity searches of known fungal cellulases using the database mycoCLAP, which contains functionally characterized proteins, to search for these enzymes in aphelids, followed by phylogenetic analyses. In support of an enzymatic algal cell-wall penetration, we identified various cellulases in the *Paraphelidium* transcriptome belonging to glycoside-hydrolase families GH3 and GH5. We detected three homologs of the GH3 cellulase beta-glucosidase/xylanase and, which is noted in other opisthosporidian but is present in fungi, amoebozoans, several opisthokonts and other protists, as well as bacteria. Our phylogenetic analysis shows that the three aphelid sequences are most closely related to deep-branched opisthokont protists (respectively, *Capsaspora*, *choanoflagellates*, *nucleariids*) (Supplementary Fig. 2e). Additionally, we identified at least three GH5 cellulase homologs (seven sequences in total) in *P. trichonema*, which were related to GH5 widespread in fungi (GH5_11, GH5_12 and GH5_24) (Supplementary Fig. 2f). Collectively, these observations strongly suggest that these cellulases are involved in the alga cell-wall penetration, but direct proof will only be obtained by purifying or heterologously expressing those cellulases and testing their activity in vitro.

**Primary metabolism reminiscent of free-living lifestyles.** Analysis of the *Rozella allomycis* genome showed that, like microsporidian parasites, it has considerably reduced metabolic capabilities. To comparatively assess the metabolic potential of aphelids, we investigated the presence of orthologous groups related to eight primary metabolism categories (Gene Ontology) in the transcriptome of *Paraphelidium*, using eggNOG annotation. We thus identified 1172 orthologous groups in *Paraphelidium* and a set of 41 eukaryotic species including representatives of major fungal lineages, opisthokont protists and other eukaryotic parasites (Supplementary Data 3). Based on their orthologous group distribution, we built a dissimilarity matrix that was analyzed by Principal Coordinate Analysis. The first axis clearly separated *Paraphelidium* from Microsporida, *Mitosphidium*, *Paramicrosporidium* and *Rozella*, the latter two positioned near one another and having an intermediate position similar to other protist parasites (e.g., *Trypanosoma*, *Leishmania*, *Toxoplasma*) (Fig. 3a and Supplementary Fig. 3a). *Paraphelidium* was positioned at the same level as fungi, *Capsaspora*, *Corallochytrium* and *Pavularia*, along axis 1. However, axis 2 separated *Paraphelidium* and fungi from the rest of eukaryotes. These relationships were further visualized in a cluster analysis of otherwise species comparisons (Supplementary Fig. 3b). The PCoA suggested that *Paraphelidium* has a rich metabolic gene complement, which was made evident by the orthologous group presence/absence heatmap showing that aphelids have a metabolic potential fully comparable to that of (especially chytrid) fungi (Fig. 3b).

The most distinctive metabolic categories when comparing *Paraphelidium* and other Opisthosporida were energy production and conversion, followed by amino acid, nucleotide and lipid transport and metabolism. In all metabolic categories, the aphelid clustered with fungi, and more specifically with chytrids, and sometimes with other free-living opisthokonts (e.g., *nucleariids*, *Capsaspora*). By contrast, *Rozella* always clustered with *Mitosphidium* and *Paramicrosporidium* either together with other Microsporida or with other parasitic protists (Supplementary Fig. 3c–f). The only exception corresponded to *Paramicrosporidium*, which, for energy production and conversion, clustered with *nucleariids* (Supplementary Fig. 3c), in agreement with their rich energy-related gene set.

To check whether these differences between *Paraphelidium* and *Rozella* affected particular metabolic pathways, we compared the annotated proteins in the two organisms based on KEGG annotation. A comparison of the two species in the KEGG general metabolic pathway map showed that, even accounting for the possibility that we missed some genes in the *Paraphelidium*’s transcriptome (e.g., mitochondrial-encoded proteins), the aphelid map contained 200 orthologous groups more than *Rozella* (548 vs. 348 orthologous groups) (Supplementary Fig. 3g). In agreement with previous observations, major differences were observed in energy production and conversion, and amino acid, nucleotide and lipid transport and metabolism. In particular, contrary to *Rozella*, which lacks most subunits of the mitochondrial electron transport chain complex I (NADH dehydrogenase; ETC-I) (Fig. 3a), *Paraphelidium* possesses a practically complete ETC-I as inferred from the nuclear encoded transcripts (the *P. trichonema* transcriptome is biased against mitochondrial transcripts, which lack polyA) (Supplementary Fig. 3b). *Paraphelidium* also possesses wide capabilities related to nucleotide (e.g., purine, uridine, and inosine biosynthesis) and amino acid (serine, threonine, methionine, lysine, ornithine, histidine, shikimate pathway) metabolism, which *Rozella* has lost (Supplementary Fig. 3i). Likewise, the aphelid has also retained pathways for phosphatidylcholine, cholesterol and fatty acid biosynthesis that were subsequently lost in *Rozella*. Most carbohydrate metabolic pathways were conserved in the two species, except for the galactose synthesis and degradation that was lost in *Rozella* (Supplementary Data 3).
By contrast, compared to Rozella, and under the assumption that our transcriptome is rather complete, the aphelid seems to lack several enzymes involved in catecholamine biosynthesis (Supplementary Data 3). However, some of these are also absent from Capsaspora, Monosiga, Salpingoeca, or Spizellomyces. These compounds are likely involved in cell-cell communication in microbes\(^{40}\), e.g., parasite-host signaling, suggesting that they might have a role in rozellid parasitism. The aphelid seems to lack other parasite-specific proteins, such as crinkler, nucleoside H\(^+\)-symporters or ATP/ADP-antiporters, which occur in Rozella and/or Microsporidia\(^{16}\).

These observations suggest that Paraphelidium has a complex metabolic profile, being functionally closer to free-living protists than to parasites and having affinities with fungi and, to a lesser extent, nucleariaids and holozoan protists.

**Distinct and ancestral-like phagotrophy-related machinery.**

Like rozellids, aphelids are phagotrophs, but their global metabolism resembles that of osmotrophic fungi. What does their phagocytosis-related proteome look like? The core phagocytic molecular machinery, already present in the last eukaryotic common ancestor\(^{41}\), is involved in multiple dynamic cell processes (endocytosis, exocytosis, autophagy, protein recycling, etc.). Structurally, the phagocytic machinery encompasses various endomembrane organelles (phagosomes, lysosomes, peroxisomes, endoplasmic reticulum, Golgi apparatus), and multiple membrane-trafficking components (signaling pathways, transporters, specific enzymes, cytoskeleton elements, motors, etc.)\(^{42}\). To examine phagotrophy-related genes in Paraphelidium and the 41 additional eukaryotes used for comparison, we built a presence/absence matrix of 695 KEGG orthologs from 11 functional categories (5 KEGG maps and 6 BRITE categories) that aimed at including all necessary proteins to perform phagotrophy; i.e., phagolysosome biogenesis, membrane trafficking and the actin cytoskeleton\(^{42}\) (Supplementary Data 4). A PCoA showed that Paraphelidium and Rozella are much closer to one another in terms of phagotrophy-than metabolism-related genes, grouping with fungi and far from Microsporidia (Fig. 3c and Supplementary Fig. 4a). This pattern was also evident in the presence/absence matrix (Fig. 3d and Supplementary Fig. 4b). In both, the two species clustered with early-
branching fungi: chytridiomycetes (Spizellomyces, Batrachochytrium, Gonapodya) and blastocladiomycetes (Alomyces, Catenaria, Blastocladiella). The clustering of Rozella and Paraphelidium with fungi based on 11 phagotrophy-related KOs (695 orthologs) might seem puzzling, since fungi are osmotrophs. However, this clustering is not unexpected, since their phagotrophic function likely involves many genes also involved in other cellular processes (e.g., cytoskeletal movement, endocytosis) that they share with fungi. Only a subset of genes in Rozella and Paraphelidium (less than 100 genes), seem to be specific and slightly different between Rozella and the aphelid. These genes are those likely responsible, together with the genes that they share with fungi, for the phagotrophic function. Furthermore, since Paraphelidium and Rozella have similar, but not identical, protein sets involved in phagosome, lysosome and endocytic processes, different protein subsets specialized in each lineage for phagocytosis function from their ancestral phagocytic ancestor (Supplementary Fig. 4c–e).

In order to gain more insights into the diversification of the actin cytoskeleton toolkit of fungi and opisthosporidians, we analyzed the evolution of myosin motor proteins. The myosin toolkit of Paraphelidium contains a mixture of classical fungal families and others previously identified in holozoans (animals and their unicellular relatives; Supplementary Data 5). We recovered diversified class I myosins in Paraphelidium, Spizellomyces and nucleariids (Supplementary Fig. 4g), with paralogs of Ic/h and Ik families, previously described only in holozoans.

Paraphelidium, nucleariids and Gonapodya also possess homologs of class XV myosins, formerly thought holozoan-specific. In addition, the aphelid not only possesses homologs of the V/VII myosin-motor family associated to chitin-synthase (see above; Supplementary Fig. 2a), but also myosins of If (pan-eukaryotic), II (amorphean), and XXII (opisthokont) classes, which clustered with previously described fungal homologs (Supplementary Data 5; Supplementary Fig. 4f). Thus, compared with the ancestral opisthokont myosin complement, Paraphelidium retains all but one myosin class (VI), with homologs of all myosin families present in fungi (If, II, V, XVII - chitin synthase) plus four additional families (Ic/h, Ik, XV and XXII) that were lost in eumycotan fungi (i.e., fungi to the exclusion of chytrids). This suggests an independent step-wise simplification of the myosin complement in fungi, Rozella and Microsporidia, with Paraphelidium, nucleariids and chytrids retaining ancestral classes.

Recently, it has been proposed that WASP and SCAR/WAVE genes, encoding activators of branched actin assembly proteins (Supplementary video 1), resembling that of chytrid fungi. Likewise, its features support a free-living opisthosporidian ancestor that had a complex life cycle including chitin-containing resting cysts, amoeboflagellate zoospores and a phagotrophic amoeba stage possibly specialized in endobiotic predation (Fig. 4). By contrast, the fungal ancestor was a free-living osmotrophic amoeba stage possibly specialized in endobiotic predation. From their aphelid-like opisthosporidian stage and are endobiotic with a unique mode of penetration into their host.

Methods
Cultures. Paraphelidium tribonemae was maintained in an enriched culture with its host Tribonema gayanum in mineral medium or in Volvic™ water.
temperature in the presence of white light. The culture was progressively cleaned from other heterotrophic eukaryotes by micromanipulation and transfer of infected host filaments to uninfected T. suecicus cultures.

WGA staining, epifluorescence, and scanning electron microscopy. To detect chitin in different cell cycle stages of Paraphelidium tribonemae, we incubated actively growing cultures with 5 μg/ml WGA conjugated to Texas Red (Life Technologies) for 10 min at room temperature. After rinsing with Volvic® water, cells were then observed under a LEICA DM2000 LED fluorescence microscope with an HCX PL FLUOTAR 100×/1.30 oil PH3 objective. Pictures were taken with a LEICA DFC3000G camera using the LEICA Application Suite v4.5 and edited with ImageJ (http://imagej.nih.gov/ij/). For SEM observations, we transferred 1-week-old cultures to a Petri dish containing mineral medium and two cover slips. Cells were let to settle overnight before fixation (1% OsO4 and 1% HgCl2 for 45 min). After washing (3 × 10 min in distilled water), samples were dehydrated in ethanol series (30, 50, 70, 90, 96, and 100%) for 10 min each. After critical-point drying and sputter-coating with platinum, cells were visualized with a Zeiss Sigma FE-SEM at 1 kV acceleration voltage.

Transcriptome sequencing. To obtain a complete representation of the aphelid transcriptome across different cell-cycle stages, we extracted RNA from a young culture (3 days after inoculation) rich in aphelid zoospores and cysts, and eight old cultures (5–7 days after inoculation) with few remaining T. suecicus living cells and rich in aphelid plasmodia, resting spores and zoospores. To minimize algal overrepresentation, we maintained the cultures in the dark. Total RNA was extracted with the RNeasy mini Kit (Qiagen), quantified with a NanoDrop 2000 (ThermoFisher Scientific) and sent to Eurofins Genomics (Germany) for de novo transcriptome sequencing. Two cDNA Illumina libraries for, respectively, the young and old enrichment cultures were constructed after polyA mRNA selection and paired-end (2 × 125 bp) sequenced with Illumina HiSeq 2500 Chemistry v4.

Transcriptome assembly, decontamination and annotation. A total of 71,015,565 and 58,826,379 reads, respectively, were obtained for the libraries of young and old cultures. From these, we obtained 34,844,871 (young) and 25,569,845 (old) paired-end reads. After quality/Illumina Hiseq adapter trimming and RNeasy mini Kit (Qiagen), quantified with a NanoDrop 2000 (ThermoFisher Scientific) and sent to Eurofins Genomics (Germany) for de novo transcriptome sequencing. Two cDNA Illumina libraries for, respectively, the young and old enrichment cultures were constructed after polyA mRNA selection and paired-end (2 × 125 bp) sequenced with Illumina HiSeq 2500 Chemistry v4.

Phylogenetic analyses. We used three different protein datasets previously used for phylogenomic analyses of Opisthokonta and, more specifically, Microsporida that included, respectively, 93 single-copy protein domains (SCPD)10, 53 proteins (BMC)11 and 259 proteins (GBE)12. We updated these datasets with sequences from the Paraphelidium tribonemae transcriptome, thirteen derived Microsporida as well as Mitosporidium daptomatis13, Paramecrinus purpurascens14, and Amphianthus sp15. Some stramenopile taxa such as Eucaryotic supergroup, Thaissiassa pseudonana16 or Phytophthora infestans (GenBank NW_003303758.1) were also included to further prevent any previously undetected Paraphelidium tribonemae contamination. Markers were aligned with MAFFT17, using the method L-INS-i with 1000 iterations. Unambiguously aligned regions were trimmed with TrimA18 and the automated algorithm, then approximate ML trees were inferred using FastTree19, all visually inspected in Geneious 20 to check for possible contamination. Contamination-free alignments were retrieved from distant outgroup taxa. For in-depth analyses, we used opisthokont, mainly holomycotan, sequences, including apicomplexa and breviate sequences as outgroup. We then made phylogenetic analyses including or excluding thirteen fast-evolving microsporian parasites (49 or 36 species). Sequences of the selected taxa were concatenated using AlverTPy from the package Barrel-o-Monkeys (http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm#Monkeybarrel). This resulted in alignments containing the following number of amino acids as a function of dataset and taxon sampling: 22,976 (SCPD49); 15,744 (BMC49); 84,810 (GGE49); 24,413 (SCPD36); 16,840 (BMC36); 96,640 (GBE36). Bayesian phylogenetic trees were inferred using PhyloBayes-MPI v1.2.69 under the CAT-Poisson evolutionary model, applying the Dirichlet process and removing constant sites. Two independent MCCM chains for each dataset were run for >15,000 generations, saving one every 10 trees. Phylogenetic analyses were stopped once convergence thresholds were reached after a burn-in of 25% (i.e., maximum discrepancy <0.1 and minimum effective size >100 using bcpomp). We also applied ML phylogenetic reconstruction using IQ-TREE20 with the profile mixture model C60. Statistical support was obtained with 1000 ultrafast bootstraps90 and 1000 replicates of the SH-like approximate likelihood ratio test71. To alleviate the local computational burden, many of these analyses were carried out using CIPRES Science Gateway21. Trees were visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree). To test if the best topology obtained was significantly better than other alternatives (Supplementary Data 2), we tested whether constrained alternative topologies could be rejected for different datasets. We used Mesquite23 to constrain topologies representing the opisthospidian monophyly (aphelids +...
rozelids + microsporidians) and the monophyly of aphelids with either blastocladiomycetes, chytridiomycetes or chytridiomycetes + blastocladiomycetes. The correct aphelid placement. From these datasets we identified the IQ-TREE option of IQ-TREE and the best-fitting model for each dataset. The resulting trees were then concatenated and AU tests were performed for each dataset with the -z and -au options as described in the advanced documentation of IQ-TREE. To minimize systematic error due to the inclusion of fast-evolving sites in our protein dataset, we progressively removed the fastest evolving sites at steps of 5% sites removed at a time. Among-site evolutionary rates were inferred using IQ-TREE -w option and its best-fitting model for each dataset (Supplemental Data 2). A total of 19 subsets were created for each dataset. We then reconstructed phylogenetic trees using the same fungal phylogenetic outgroup model dataset. To know how supported were alternative topologies in bootstrapped trees, we used CONSENSE from the PHYLIP package74 and interrogated the UFBOOT dataset. For each dataset, the number of species, the number of orthologs and the number of sequences retrieved from previous datasets and sequences identified by BLAST in GenBank or in the mycoCLAP database were calculated. We performed comprehensive alignments including sequences retrieved from previous studies and sequences identified by BLAST in the mycoCLAP database as initial queries for cellulose and related cell-wall proteins (chitin synthases (CHS), chitin deacetylases (CDA), and chitinases). We used sequences obtained from the NCBI protein database; from NCBI in January 2018. [These sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community]. Query sequences to retrieve chitin and related cell-wall proteins (chitin synthases (CHS), chitin deacetylases (CDA), and chitinases) were obtained from the NCBI database75. We used sequences obtained from Saccharomyces cerevisiae S288C. We used sequences from Aspergillus fumigatus as initial queries for cellulose and related cell-wall proteins, using the HMM profile of A. fumigatus. We performed comprehensive alignments including sequences retrieved from previous studies and sequences identified by BLAST in the mycoCLAP database applying the progressive L-INS-i algorithm optimized for global sequence homology. The alignment was run with the highest fraction of explained variance, which were in all cases higher than the fractions expected under the broken stick model. In addition, we plotted binary OG profiles of each species in a presence/absence heatmap, produced using the heatmap2 function in the R gplots library76. Species’ order was defined by a Ward hierarchical clustering of the aforementioned interspecific Pearson’s correlation coefficients. Ortholog order was defined by Ward hierarchical clustering on euclidean distances. All clustering and distance analyses were performed using R libraries77. We represented the raw species clustering (Pearson correlation + Varclus for the presence of 6469 unique orthologous groups as inferred by BLAST in GenBank or in the mycoCLAP database) as a heatmap function in the R gplots library77. Species’ order was defined by a Ward hierarchical clustering as above to cluster species according to their metabolism gene content (based on Pearson + Ward). We carried out similar comparisons and statistical analyses for proteins involved in phagotrophy, membrane trafficking and cytoskeleton. More specifically, we looked for statistically significant overlaps, focusing especially on the fractions of explained variance and the importance of individual proteins. The KEGGgos were thus used to get the KEGG ortholog profile for the 41 taxa. All proteins in 35 eukaryote genomes80 were clustered to a non-redundant set using cd-hit at a 90% identity threshold. Those 35 nr90 genomes were compared in an all-vs-all BLAST analysis with an e-value cutoff of 1 x 10^-3. The all-vs-all BLAST output was clustered using the MCL algorithm with an inflation parameter of 2.0. To avoid lineage specific proteins, the resultant clusters were retained if the proteins in the cluster were derived from 3 or more organisms. Proteins from each cluster were aligned using MAFFT, the alignments trimmed with trimAl, and the best hit bit score among those proteins to the HMM was calculated. All proteins from the UniProt-SwissProt database were searched against the entire set of 14,095 HMMs. Each HMM was assigned a best-hit protein from the UniProt-SwissProt database. KEGG ortholog identifiers (IDs) associated with each HMM were obtained from UniProtKB IDs and KEGG ortholog IDs using the mappings available from UniProtKB. Additional KEGG ortholog IDs were mapped to the HMMs by searching all protein IDs against the KEGG ortholog ID database. The best hit score among those proteins to the HMM was compared to the best-hit bit score of the UniProtKB annotation for that HMM.
If the best hit KEGG ortholog ID bit score was at least 80% of the best hit UniProtKB protein bit score, and there was no KEGG ortholog ID already associated with the UniProtKB annotation, the KEGG ortholog ID was transferred to that HMM. To build a presence-absence matrix associated with KEGG ortholog IDs, the KEGG orthologs were used to perform the same multivariate statistical analyses as per the metatranscriptomic nucleotide contig assembly and version 1.5 of the predicted proteome are deposited in NCBI under accession number gshare82,83.

**Data availability**

Raw read sequences have been deposited in NCBI under accession number PRJNA402302. *Paraphelidium trioneum* metatranscriptomic nucleotide contig assembly and version 1.5 of the predicted proteome are deposited in figshare82,83 under Creative Commons 4.0 licence. All trees are available as Supplementary file all_trees.txt.

Received: 19 September 2018 Accepted: 22 November 2018

**References**

1. Richards, T. A., Leonard, G. & Wideman, J. G. What defines the "Kingdom" fungus? In Heitman et al. (ed.), *The fungal Kingdom, p. 57–77* (ASM Press, Washington, 2017).
2. James, T. Y. et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443, 818–822 (2006).
3. Jones, M. D. et al. Discovery of novel intermediate forms redeeming the cryptomycota. *J. Eukaryot. Microbiol.* 69, 183 (2015).
4. Lara, E., Moreira, D. & Lopez-Garcia, P. The environmental clade LKM11 and the nematode parasite new to the Cryptomycota isolated from an outdoor algal pond. *Proc. Natl Acad. Sci. USA* 111, 15480–15485 (2014).
5. Mikhailov, K. V., Simdyanov, T. G. & Aleoshin, V. V. Genomic survey of a eukaryote new to the Cryptomycota and proposal for the new superphylum Amphiamblys. *Ann. Rev. Microbiol.* 69, 451–474 (2015).

10. Murphy, C., Powolowski, J., Wu, M., Butler, G. & Tsang, A. Curation of characterized glycose hydrolyases of fungal origin. *Database* 2011, ba020 (2011).
11. McAndrew, R. P. et al. From soil to structure, a novel dimeric beta-glucosidase belonging to glycose hydrolase family 3 isolated from compost metagenomic analysis. *J. Biol. Chem.* 288, 14985–14992 (2013).
12. Aspeborg, H. et al. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12, 186 (2012).
13. Haertter-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293 (2016).
14. Okuda, S. et al. KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.* 36, W423–W426 (2008).
15. Roshchina, V. V. New Trends and perspectives in the evolution of neurotransmitters in microbial, plant, and animal cells. *Adv. Exp. Med. Biol.* 784, 25–77 (2017).
16. Jutin, N., Wolf, M. Y., Wolf, Y. I. & Koonin, E. V. The origins of phagocytosis and eukaryogenesis. *Biol. Direct.* 4, 9 (2009).
17. Boulais, J. et al. Molecular characterization of the evolution of phagosomes. *Mol. Syst. Biol.* 6, 423 (2010).
18. Sebé-Pedrós, A., Grau-Bové, V., Richards, T. A. & Ruiz-Trillo, I. Evolution and classification of myosins, a pan eukaryotic whole-genome approach. *eLife* 6, 290–305 (2017).
19. Sebé-Pedrós, A. et al. Insights into the origin of metazoan filopodia and macrovilli. *Mot. Biol. Evol.* 30, 2013–2023 (2013).
20. Fritz-Laylin, L. K., Lord, S. J. & Mullins, R. D. WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. *J. Cell Biol.* 216, 1673–1688 (2017).
46. Held, A. A. Zoospore of Rozella allomyces—ultrastructure. Can. J. Bot. 53, 2212–2217 (1975).
47. Held, A. A. Development of rozella in allomyces—a single zoospore produces numerous zoosporangia and resistant sporangia. Can. J. Bot. 58, 959–979 (1980).
48. Letcher, P. M. et al. Morphological, molecular, and ultrastructural characterization of Rozella rhizoclosmatii, a new species in Cryptomycota. Fungal genet. Biol. 123, 1–10 (2017).
49. Hehenberger, E. et al. Novel predators reshape Holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of Curr. Biol. 27, 2043–2050.e2046 (2017).
50. Corsaro, D. et al. Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozelomyces. Parasitol. Res. 113, 1909–1916 (2014).
51. Karpov, S. A., Torruella, G., Moreira, D., Mamkaeva, M. A. & López-García, P. Molecular phylogeny of Paraphyllum letcheri sp. nov. (Apheilida, Opisthosporidia). J. Eukaryot. Microbiol. 64, 573–578 (2017).
52. Cavalier-Smith, T. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loutozaecia, Sulcozoa, and Chlamydomonas. Eur. J. Protistol. 49, 115–178 (2013).
53. Karpov, S. A., Mamkaeva, M. A., Benzerara, K., Moreira, D. & Lopez-Garcia, P. Molecular phylogeny and ultrastructure of Aphelidium aff. melosiae (Apheilida, Opisthosporidia). Protist 165, 512–526 (2014).
54. Tedersoo, L. & Lindahl, B. Fungal identification biases in microbiome projects. Environ. Microbiol. Rep. 8, 774–779 (2016).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
56. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652 (2011).
57. Lagesen, K. et al. RNAmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108 (2007).
58. Laetsch, D. R. & Blaxter, M. L. BlobTools: interrogation of genome assemblies using DIAMOND. Nat. Methods 12, 2120 (2014).
59. Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 32, 268–274 (2015).
60. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490 (2010).
61. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649 (2012).
62. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288 (2009).
63. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274 (2015).
64. Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195 (2013).
65. Anisimova, M., Gil, M., Dufayard, J.-F., Descimon, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699 (2011).
66. Miller, M. A., Pfeifer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), p. 1–8 (New Orleans, LA, 2010).
67. Mesquite: a modular system for evolutionary analysis. Version3.40 (2018).
68. PHYLIP - Phylogeny Inference Package v. 3.6 (University of Washington, Seattle, WA, 1999).
69. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37 (2011).
70. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285 (2016).
71. Anisimova, M., Gil, M., Dufayard, J.-F., Desimmon, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699 (2011).
72. Miller, M. A., Pfeifer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), p. 1–8 (New Orleans, LA, 2010).
73. Mesquite: a modular system for evolutionary analysis. Version3.40 (2018).
74. PHYLIP - Phylogeny Inference Package v. 3.6 (University of Washington, Seattle, WA, 1999).
75. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37 (2011).
76. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285 (2016).
77. R: A language and environment for statistical computing. v. http://www.r-project.org (R Foundation for Statistical Computing. Vienna, Austria, 2017).
78. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290 (2004).
79. Warnes, G. R. et al. gplots: various R programming tools for plotting data. R Package Version 2, 1 (2016).
80. Huson, D. H., Pflister, A. A. & Kim, E. Gene-based predictive models of trophic modes suggest Asgard archaea are not phycotrophic. Nature 583, 1–10 (2020).
81. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731 (2016).
82. Farrag, D. et al. https://doi.org/10.6084/m9.ﬁgshare.739469.v1 (2018).
83. Farrag, D. & Ruiz-Trillo, I. Transcriptome of Parvularia atlantis (Nucearia sp. ATCC50694) https://doi.org/10.6084/m9.ﬁgshare.3898485.v4 (2017).

Acknowledgements
We are grateful to Wayne Pfeiffer and Mark Müller from the CIPRES Science Gateway (http://www.phylo.org/) for computational resources and personal assistance. We thank Victoria S. Tvetekova for micromanipulating infected Trichomona sp. filaments in view of establishing clean apheloid cultures, Alex de Mendoza for providing CHS alignments, Miklphile Des Rochamps, Rafael Ponce, Gaaulme Rebolu, Ana Gutierrez-Preciado, Ares Rocatin-Arjo and Ricardo Rodriguez de la Vega for discussion on bioinformatics and statistics aspects, and Peter Kubicke and Bernard Hernissat for helpful discussion about cellulases. Research leading to these results received funding from the European Research Council under the European Union’s Seventh Framework Program (ERC Grant Agreement 322669 ProtistWorld). G.T. was also financed by the European Marie Skłodowska-Curie Action (704566 AlgDates). S.K. was supported by RSF grant No 16–14–10302, ZIN RAS program AAAA–A17–117030130322–3 and Jean d’Alembert’ program (Université Paris-Saclay).

Author contributions
G.T., D.M., and P.I.-G. conceived and coordinated the study. G.T. performed culture cleaning, RNA extraction, de novo transcriptome assembly, phylogenomics and comparative genomic analyses. G.T. and X.G.-B. cleaned the protein set from contamination. X.G.-B. and A.S.-B. performed myosin comparative genomic analyses. G.T. and X.G.-B. carried out multivariate statistical analyses of metabolic genes. G.T. and J.A.B. carried out phagolysosome protein analyses. G.T. and S.K. maintained cultures and performed WGA staining and imaging. S.K. studied morphology and life-cycle aspects, E.V. performed SEM fixation and imaging. P.I.-G. and G.T. wrote the manuscript. All authors commented on the manuscript.

Additional information
Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-018-0235-z.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018