



HAL
open science

Ancient Adaptive Lateral Gene Transfers in the Symbiotic Opalina - Blastocystis Stramenopile Lineage

Naoji Yubuki, Luis Xavier Galindo, Guillaume Reboul, Purificacion Lopez-Garcia, Matthew W Brown, Nicolas Pollet, David Moreira

► **To cite this version:**

Naoji Yubuki, Luis Xavier Galindo, Guillaume Reboul, Purificacion Lopez-Garcia, Matthew W Brown, et al.. Ancient Adaptive Lateral Gene Transfers in the Symbiotic Opalina - Blastocystis Stramenopile Lineage. *Molecular Biology and Evolution*, 2019, 37, pp.651-659. 10.1093/molbev/msz250 . hal-02367918

HAL Id: hal-02367918

<https://hal.science/hal-02367918v1>

Submitted on 18 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ancient Adaptive Lateral Gene Transfers in the Symbiotic *Opalina* - *Blastocystis* Stramenopile Lineage

Naoji Yubuki¹, Luis Javier Galindo¹, Guillaume Reboul¹, Purificación López-García¹, Matthew W. Brown^{2,3}, Nicolas Pollet⁴, and David Moreira^{1*}

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

²Department of Biological Sciences, Mississippi State University, MS Mississippi State, MS, USA

³Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS, USA

⁴Laboratoire Evolution Génomes Comportement et Ecologie, CNRS, IRD, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

***Corresponding author:** E-mail: david.moreira@u-psud.fr.

Abstract

Lateral gene transfer (LGT) is a very common process in bacterial and archaeal evolution, playing an important role in the adaptation to new environments. In eukaryotes, its role and frequency remain highly debated, although recent research supports that gene transfer from bacteria to diverse eukaryotes may be much more common than previously appreciated. However, most of this research focused on animals and the true phylogenetic and functional impact of bacterial genes in less-studied microbial eukaryotic groups remains largely unknown. Here, we have analyzed transcriptome data from the deep-branching stramenopile Opalinidae, common members of frog gut microbiomes and distantly related to the well-known genus *Blastocystis*. Phylogenetic analyses suggest the early acquisition of several bacterial genes in a common ancestor of both lineages. Those LGTs most likely facilitated the adaptation of the free-living ancestor of the Opalinidae-*Blastocystis* symbiotic group to new niches in the oxygen-depleted animal gut environment.

Key words: Opalinids, *Blastocystis*, lateral gene transfer, gut microbiome.

Lateral gene transfer (LGT) plays an important role in prokaryotic evolution. LGT provides bacteria and archaea with the possibility to adapt, sometimes very rapidly, to new environments by obtaining genes from organisms already living in those environments. Although the significance of this phenomenon is widely recognized in prokaryotes, LGT-mediated gene acquisition from distant donors remains a contentious issue in eukaryotes (Martin 2017; Leger et al. 2018). Nevertheless, there is increasing evidence for LGT in eukaryotes from prokaryotes as well as from other eukaryotes (e.g., Keeling and Palmer 2008; Karnkowska et al. 2016; Eme et al. 2017; Husnik and McCutcheon 2017). A recent example concerns the stramenopile *Blastocystis*, which experienced LGTs from both eukaryotic and prokaryotic donors (Denoëud et al. 2011; Eme et al. 2017).

Blastocystis is recognized as the most widespread human gut eukaryotic parasite (Clark et al. 2013). This strict anaerobic and single-celled protist displays some unique and interesting biological features, such as the presence of unusual mitochondrion-related organelles (MRO) that display functions of mitochondria, hydrogenosomes and mitosomes (Stechmann et al. 2008). Some *Blastocystis* enzymes crucial for life in oxygen-depleted conditions were acquired by LGT from prokaryotes. For instance, the sulfur-mobilization (SUF) machinery involved in Fe-S protein maturation in the cytoplasm appears to have been acquired from archaeal Methanomicrobiales (Tsaousis et al. 2012). Furthermore, Eme et al. (2017) reported 74 purported cases of LGT mostly from prokaryotes to various subtypes of *Blastocystis* and suggested that several of the new LGT-acquired functions facilitated the metabolic adaptation of *Blastocystis* to the human gut in terms of metabolism but also to escape the immune defense mechanisms. The origins of those 74 gene families were very diverse. Although many of them were already present in the common ancestor of several *Blastocystis* subtypes, the time of their acquisition remained unclear due to the poor taxon sampling available for closely related stramenopile lineages.

Together with Alveolata and Rhizaria, Stramenopiles (or Heterokonta) constitute one of the main clades of the eukaryotic super-group SAR (Burki et al. 2007; Adl et al. 2019). Stramenopiles mostly encompass free-living phagotrophs or photosynthetic algae, but some are well-known parasites, such as the oomycetes and *Blastocystis*, or commensals, such as the Opalinidae (Patterson 1989; Andersen 2004). Ribosomal RNA phylogenetic analyses suggested a close relationship

between *Blastocystis* and Opalinidae, supporting the existence of a deep-branching symbiotic (parasitic/commensal) clade adapted to live in the gut of very diverse vertebrates (Silberman et al. 1996; Kostka et al. 2004; Li et al. 2018). However, despite the phylogenetic affinity of *Opalina* and *Blastocystis*, their morphological characteristics and lifestyles are very different. *Blastocystis* is characterized by a round unflagellated cell largely filled by a large vacuole. The cytoplasm and organelles are concentrated in the thin peripheral area between the vacuole and the cell membrane. Members of the genus *Blastocystis* live in the intestines of humans, birds, cows and pigs, most likely as parasites (Tan 2004). By contrast, members of the genus *Opalina* have a leaf-like cell shape with numerous nuclei and hundreds of short flagella on the cell surface, which is reminiscent of the cellular organization of ciliates. They live mainly in the intestine of anurans (frogs and toads) but seem to be innocuous to their hosts being therefore most often reported as commensal symbionts (Kostka 2016). Using the numerous flagella, *Opalina* members actively move in the intestine. All other known Opalinidae species are also commensal symbionts (Kostka 2016). Phylogenetic analyses have supported the monophyly of the Opalinidae-*Blastocystis* clade with the Placidida, a lineage of small free-living marine flagellates such as *Wobblia* and *Placidia* (Li et al. 2018; Shiratori et al. 2015, 2017; Derelle et al. 2016). Another free-living marine flagellate, *Cantina marsupialis*, is an anaerobic deep-branching relative that also possesses MROs (Yubuki et al. 2015). Since the closest relatives of Opalinidae and *Blastocystis* are all free-living, their ancestor was most likely free-living as well.

Here, we report the first transcriptome sequences from two Opalinidae strains, *Opalina* sp. OP10 and Opalinidae sp. Opal32, from two different continents (Europe and North America). OP10 and Opal32 cells were collected manually from the intestine of a *Xenopus tropicalis* frog and a *Lithobates sphenoccephalus* tadpole, respectively. After transcriptome sequencing and assembly, we decontaminated the translated protein sequences inferred from the two transcriptomes to remove host and bacterial sequences (see Materials and Methods) and kept 7,232 and 18,765 proteins for OP10 and Opal32, respectively. Using BUSCO (Simão et al. 2015), we determined 33.3% transcriptome completeness for OP10 and 57.4% for Opal32. For comparison, we also applied BUSCO on the near-complete genome of *Blastocystis hominis* and determined a completeness of 75.2%, indicating a reduced genome as expected for a derived symbiont. We found in our datasets 44.3% (OP10) and 76.3%

(Opal32) of the *Blastocystis* proteome (supplementary table S1, Supplementary Material online), suggesting a rather good coverage especially for Opal32.

After transcriptome decontamination we searched with BLAST (Camacho et al. 2009) Opalinidae homologues of the 74 gene families likely acquired by LGT in *Blastocystis* (Eme et al. 2017). We recovered 37 and 38 of those LGT candidates in OP10 and Opal32, respectively. Thirty genes were common in both OP10 and Opal32, and seven and eight genes were unique in OP10 and Opal32, respectively. In total, 45 different candidate LGT genes were found in the two Opalinidae species. To verify that these sequences did not derive from the genomes of other gut microbes, we carried out two types of analyses. First, we investigated the codon usage of the coding sequences of both decontaminated transcriptomes and those of the LGT candidates and measured the frequency of optimal (F_{OP}) codons, which indicates the ratio of optimal (most frequent) codons to synonymous codons. The proportion of synonymous codons is unique to each genome and often results in a unimodal distribution of the F_{OP} score (Ikemura 1985), whereas the presence two F_{OP} peaks has been linked to contamination with bacterial sequences (Heinz et al. 2012). We obtained single-peak F_{OP} plots for our transcriptomes, indicating homogeneous codon usage and absence of contamination. All our LGT candidates fitted into these unimodal distributions supporting that they represent bona fide opalinid genes (supplementary figure S1, Supplementary Material online). Furthermore, their fit into the unimodal distribution supports an ancient integration of these LGT genes since they have adapted to the codon usage of the recipient genome. Second, we conducted phylogenetic analyses for all the LGT protein sequences. Phylogenetic trees showed that 29 of these proteins clustered robustly with their respective *Blastocystis* homologues (supplementary table S2 and supplementary figures S2-S30, Supplementary Material online). Those 29 proteins belonged to different functional families including carbohydrate metabolism, lipid metabolism, amino acid metabolism, and transporters. The phylogenetic analyses also allowed the identification of the donors of these sequences. Most of them had prokaryotic donors belonging to the Archaea, Proteobacteria and Actinobacteria, which are major components of frog gut microbiomes (Colombo et al. 2015). In some cases, the two Opalinidae species grouped with other eukaryotes belonging to the Amoebozoa, Excavata and Metazoa, suggesting eukaryote-to-eukaryote LGT, although it was impossible to infer from these trees whether the Opalinidae species were donors or recipients. Several of the LGT proteins most likely play important functions in the adaptation of Opalinidae to the anaerobic gut environment. One example is the mitochondrial iron-sulfur cluster (ISC) biogenesis system, essential for the assembly of iron-sulfur-containing proteins. These proteins are involved in a variety of metabolisms, including electron transport, nitrogen

fixation, and photosynthesis. In some protists living in low-oxygen environments, the canonical eukaryotic ISC machinery has been replaced by alternative bacterial machineries acquired via LGT, such as the nitrogen fixation (NIF) system and the bacterial sulfur mobilization (Suf) machinery. For instance, *Entamoeba histolytica* has a bacterial NIF system (van der Giezen et al. 2004), whereas *Monocercomonoides exilis*, which has completely lost mitochondria and the mitochondrial ISC pathway, contains a bacterial Suf system (Karnkowska et al. 2015). By contrast, *Blastocystis* has an archaeal-like SufC+SufB fused protein (Tsaousis et al. 2012). Similar fused *sufCB* genes related to Methanomicrobiales homologues were also identified in anaerobic flagellates such as the jakobid *Stygiella incarcerate* and the breviate *Pygsuia biforma* (Leger et al. 2016; Stairs et al. 2014). In prokaryotes, the *suf* operon is upregulated under oxidative stress (Outten et al. 2004), suggesting that the Suf machinery can be important for living in oxygen-depleted environments. We only identified an incomplete *sufB* gene in *Opalina*, which lacked a mitochondrial target signal. Similarly, SufCB is inferred to function in the cytosol in *Blastocystis*, *Pygsuia* and *Stygiella* (Tsaousis et al. 2012; Stairs et al. 2014; Leger et al. 2016). Our phylogenetic analysis showed that *Opalina* was closely related to these other anaerobic protists within a clade of Methanomicrobiales with robust support (fig. 1). These eukaryotes belong to three unrelated supergroups (*Opalina* and *Blastocystis* to SAR, *Pygsuia* to Breviatea, and *Stygiella* to Excavata). Therefore, one parsimonious explanation for this uneven distribution of SufCB is that one of these eukaryotic lineages first obtained the *sufC* and *sufB* genes from Methanomicrobiales, then both genes fused and, finally, the fused gene was transferred by eukaryote-to-eukaryote LGT to the other eukaryotic lineages. Since we only identified the *sufB* part in *Opalina*, it seems that it secondarily lost *sufC* after branching off from the lineages with fused *sufCB*. In fact, the well-supported separation of *Opalina* and *Blastocystis* in our tree (fig. 1) suggests that they have followed different evolutionary histories for the *sufCB* gene. Interestingly, the SufB and SufC proteins of *M. exilis* and *Paratrimastix pyriformis* are not related with the clade of *Opalina*, *Blastocystis*, *Pygsuia*, and *Stygiella*, indicating that they acquired these genes by independent LGT events from other prokaryotic donors. These genes were not identified in *C. marsupialis*.

In anoxic conditions, some eukaryotes use rhodoquinone instead of ubiquinone to receive electrons from NADH in the mitochondrial complex I of the electron

transport chain (ETC) and generate rhodoquinol (Castro-Guerrero et al. 2005; Sakai et al. 2012; Takamiya et al. 1999). Rhodoquinol is then reoxidized by the mitochondrial complex II catalyzing the reverse reaction as a fumarate reductase (van Hellemond and Tielens 1994; Tielens et al. 2002). This pathway helps to produce ATP and to reduce the respiratory chain without using the mitochondrial complexes III to V. The putative methyltransferase RquA is required for rhodoquinone biosynthesis (Lonjers et al. 2012) and its distribution among eukaryotes suggests that it is important for the adaptation of the mitochondrial metabolism to low-oxygen environments. In *Blastocystis*, RquA was suggested to be targeted to the MRO (Eme et al. 2017). We identified RquA homologues in both OP10 and Opal32 that also contained predicted mitochondrial-targeting sequences. By contrast, this protein seemed to be absent in *C. marsupialis*. RquA is not very common in eukaryotes and previous phylogenetic analyses demonstrated that RquA-containing eukaryotes are scattered among prokaryotic lineages, mostly Proteobacteria. Stairs et al. (2018) proposed that LGT of *rquA* genes from bacteria to eukaryotes occurred at least twice before subsequent multiple independent LGTs among eukaryotes. Our updated RquA phylogeny (fig. 2) is consistent with this proposal. We retrieved two major clades, A and B: *Opalina* spp. branched together with *Proteromonas* and *Blastocystis* in clade A, composed mostly of alpha- and beta-proteobacteria, and several other eukaryotes (Breviata, Amoebozoa and Euglenida). Group B also contained some eukaryotes (choanoflagellates, diatoms, and ciliates) embedded among bacteria, again mostly alpha- and beta-proteobacteria. The presence of alphaproteobacteria close to the eukaryotic sequences opens the possibility of a mitochondrial origin by endosymbiotic gene transfer (EGT). Nevertheless, several observations argue against this hypothesis: (i) the eukaryotic sequences are not monophyletic, (ii) several eukaryotic sequences appear to be closer to betaproteobacteria than to alphaproteobacteria, and (iii) if *rquA* was present in the last eukaryotic common ancestor (which already had mitochondria), it must have been lost independently many times to result in its current patchy distribution. Thus, the available data so far rather support the origin of eukaryotic *rquA* by LGT from bacteria followed by subsequent LGTs among eukaryotes.

In most mitochondria, coenzyme A is transferred from acetyl-CoA to succinate by two types of acetate:succinate CoA-transferases (ASCT1B and ASCT1C). The resulting succinyl-CoA is used for ATP production by succinyl-CoA synthetase

(SCS). This ASCT/SCS system plays a crucial role in MROs of protists living in anoxic environments, such the human urogenital parasite *Trichomonas vaginalis*, for the production of ATP by substrate-level phosphorylation independent of the mitochondrial Krebs cycle (van Grinsven et al. 2008). In the case of the free-living amoeboflagellate *Naegleria gruberi*, which contains classical mitochondria and transiently experiences low-oxygen conditions, ASCT was predicted to function in mitochondria (Fritz-Laylin et al. 2010). We identified an ASCT/SCS system in our *Opalina* transcriptomes. In contrast with the *Blastocystis* ASCT, which has an MRO-targeting sequence, the *Opalina asct1C* and *asct1B* were incomplete ORFs and did not contain any recognizable mitochondrial targeting signal. The ASCT1C phylogenetic tree (fig. 3) recovered *Opalina* and *Blastocystis* grouped within a large clade also containing trichomonads, *Naegleria*, fungi, and dictyostelid cellular slime molds (Amoebozoa). This eukaryotic clade was closely related to Deltaproteobacteria and Firmicutes. As in the previous cases described above, this tree suggests a bacterial origin of the gene followed by eukaryote-to-eukaryote LGT.

To carry out a more comprehensive comparison of the mitochondrial metabolism of *Opalina* with that of other MRO-containing anaerobic stramenopiles (the parasitic *Blastocystis* and the free-living *C. marsupialis* (Stechmann et al. 2008; Noguchi et al. 2015)), we used BLAST to search for homologues of MRO proteins of these organisms in *Opalina*. We also manually annotated the *Opalina* mitochondrial proteins involved in major energy metabolism pathways. As shown above, *Opalina* obtained many genes for typical MRO anaerobic metabolism by LGT from either prokaryotes or other eukaryotes, but it also contains typical mitochondrial genes vertically inherited (supplementary tables S2 and S3, Supplementary Material online). *Blastocystis* spp. and *C. marsupialis* completely lack complexes III and IV, and F1Fo ATPase (complex V) (Gentekaki et al. 2017; Noguchi et al. 2015). *Opalina* possesses some genes of the tricarboxylic acid (TCA) cycle, complex I (NADH:ubiquinone oxidoreductase), and complex II (succinate dehydrogenase) of the ETC, but does not seem to encode any other recognizable canonical components such as complexes III and IV or the F1Fo ATPase (supplementary table S4, Supplementary Material online). This suggests that *Opalina* has a partial ETC that does not appear to function in energy generation. Data from *Blastocystis* and *Pygsuma* suggest that complex II functions in reverse as a fumarate reductase to regenerate the quinone pool under anaerobic conditions without using complex III, IV

and F1Fo ATPase to conduct oxidative phosphorylation. RquA, acquired by LGT in *Opalina* (see above), is the crucial enzyme for this alternative electron transport machinery. *Opalina* also possesses genes involved in classical mitochondrial activities, including transporters, fatty acid metabolism, amino acid metabolism, pyruvate metabolism, and [2Fe-2S] ferredoxin for FeS cluster assembly, some of which are lost in *Blastocystis*. (supplementary table S2, Supplementary Material online). By contrast, we did not identify in *Opalina* some essential mitochondrial proteins, such as those involved in the eukaryotic iron-sulfur cluster (ISC) synthesis system and several enzymes (pyruvate:ferredoxin oxidoreductase (PFO), [FeFe] hydrogenase (HydA), the HydA hydrogenase maturases HydE, HydF and HydG, and two subunits of the NADH:ubiquinone oxidoreductase (NuoE and NuoF)) that are hallmarks of the MROs found in many anaerobic protists, including *Blastocystis* and *Cantina*. In those organisms, PFO oxidizes pyruvate to acetyl-CoA and CO₂. The reduced ferredoxin is reoxidized by HydA that reduces protons to H₂ gas. In *Opalina*, which lacks HydA, the pyruvate:NADP⁺ oxidoreductase (PNO), instead of PFO, presumably oxidizes pyruvate to acetyl-CoA and, then, acetyl-CoA can be utilized by the ASCT/SCS system to generate ATP by substrate-level phosphorylation. Since PFO and HydA are present in *Blastocystis*, we can propose two evolutionary scenarios: First, these two enzymes were present in the common ancestor of *Opalina* and *Blastocystis* and secondarily lost in the *Opalina* lineage or, second, they were obtained in *Blastocystis* independently after it diverged from the *Blastocystis-Opalina* common ancestor. As in the case of *Blastocystis* and *Cantina*, we did not identify a pyruvate carrier in *Opalina*. Glycolysis is described as a cytosolic process in eukaryotes and its product, pyruvate, is imported into the mitochondrion by the pyruvate carrier. However, the second half of glycolysis in some stramenopiles has been predicted to occur in both the cytosol and mitochondria/MRO (Abrahamian et al. 2017). Moreover, in *Blastocystis* this second half of the glycolysis is solely localized in the MRO (Rártulos et al. 2018). Similarly, we identified in *Opalina* several enzymes of the second half of the glycolysis (glyceraldehyde phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK), and enolase (ENO)) with mitochondria-targeting signals (supplementary table S3, Supplementary Material online). Despite these similarities and other shared key adaptations to the oxygen-depleted gut environment, *Opalina* appears to have kept a less derived version of

the mitochondrial metabolism than its sister lineage *Blastocystis* and the stramenopile relative *Cantina*.

Conclusion

Our examination of two *Opalina* transcriptomes based on sequence similarity searches and phylogenetic analyses identified 29 genes likely acquired by LGT by a common ancestor of both *Blastocystis* and Opalinidae (supplementary table S1, Supplementary Material online). Among these genes, those coding for the Suf, RquA and ASCT proteins play important roles in anaerobic metabolism in MROs. It is unclear when a common ancestor of these organisms entered the animal gut but some of the LGTs most likely facilitated the adaptation to this new oxygen-deprived environment before the divergence of these two lineages. *Blastocystis* MROs combine metabolic properties of both mitochondria and hydrogenosomes and contain PFO and [FeFe] hydrogenase as well as incomplete TCA cycle and the complexes I and II (Gentekaki et al. 2017; Stechmann et al. 2008). Although *Opalina* shares with *Blastocystis* many enzymes involved in anaerobic metabolisms acquired via LGT and both lineages have several metabolic modifications in common (incomplete TCA cycles and absence of complexes III and IV and F1Fo ATPase), our data suggest the absence of the typical hydrogenosomal enzymes PFO and [FeFe] hydrogenase. This important difference indicates that *Blastocystis* has achieved a more derived adaptation to hypoxic condition than Opalinidae. *Opalina* represents therefore an excellent model of intermediate adaptation between conventional aerobic mitochondria and derived anaerobic MROs and can help to understand the initial steps in the evolutionary path between both types of organelles.

Materials and Methods

Isolation of *Opalina* sp. Cells

For OP10 strains, the gut content of a *Xenopus tropicalis* frog was collected and resuspended in sterile PBS buffer. Eight *Opalina* cells were manually isolated under an inverted Leica DMI3000 microscope equipped with an Eppendorf TransferMan 4r micromanipulator. The cells were rinsed twice in sterile PBS and finally resuspended in 1.5 μ l of sterile water. For Opal32 strain, a smear of ca. 100 μ l of *Lithobates sphenoccephalus* tadpole gut contents was placed onto a sterile Petri dish and 500 μ l of sterile amphibian Ringer's solution (ARS: in 1 L distilled water, 6.6 g NaCl, 0.15 g KCl, 0.15 g CaCl₂, and 0.2 g NaHCO₃) was added to the drop of gut content. Roughly 10 μ l of this solution was examined under a Zeiss AxioSkop Plus upright microscope, and cells were imaged. A single cell was manually isolated using a micropipetter and washed six times in 100 μ l of fresh and sterile ARS. The cell was then transferred to a 0.5 μ l to nuclease-free PCR tube and processed as below.

***Opalina* sp. Transcriptome Sequencing and Assembly**

For *Opalina* sp. OP10, RNA extraction, cDNA synthesis and amplification were done using the REPLI-g WTA Single Cell kit following the manufacturer's protocol (Qiagen). The resulting cDNA was sequenced using Illumina HiSeq 2500 paired-end sequencing (2x125 bp). For Opal32, the cell was subjected to a modified version of SmartSeq-2 (Picelli et al. 2014, Kang et al. 2017) and full-length cDNA was constructed. This cDNA was then sheared using a Covaris focused-ultrasonicator (Duty% 10, Intensity 5, Burst Cycle 200, Time 30s, Frequency Sweeping Mode). This sheared cDNA was prepped using NEBnext Ultra DNA library kit for Illumina (New England Biolabs) and sequenced on an Illumina MiSeq paired-end (2x300 bp) sequencing run. For both datasets, Illumina adapters were removed using Trimmomatic v. 0.36 (Bolger et al. 2014) and paired-end sequences were assembled using Trinity v.2.2.0 (Haas et al. 2013) with default parameters. A total of 24,170 assembled transcripts were obtained from OP10 and 16,943 from Opal32.

Transcriptome Decontamination and Completeness

The decontamination of the two transcriptomes was carried out by a three-step process. First, the transcriptome sequences were subjected to two rounds of assembly, before and after bacterial sequence removal by BlobTools v0.9.19 (Laetsch et al. 2017). Second, open-reading frames were predicted and translated from the assembled transcripts using Transdecoder v2 (<http://transdecoder.github.io>) to produce protein sequences for OP10 and Opal32. Finally, to remove possible host sequences, the predicted protein sequences were searched by BLASTp (Camacho et al. 2009) against two predicted anuran proteomes. We used *Xenopus tropicalis* v9.1 for OP10 and, because of the lack of a proteome from the host species of Opal32 (*Lithobates sphenoccephalus*) we used *Rana catesbeiana* RCv2.1, which is the closest member of the same Ranidae family with available sequence data. At the end, we obtained 8,432 and 11,480 protein sequences from OP10 and Opal32, respectively.

To assess transcriptome completeness, we used BUSCO v2.0.1 (Simão et al. 2015) on the decontaminated predicted proteins with the eukaryote_odb9 dataset of 303 near-universal single-copy orthologs. As an additional step of quality completeness comparison, we calculated the completeness value of the near-complete genome of *Blastocystis hominis* (ASM15166v1) and compared it with the opalinid data.

Codon usage for the coding sequences of both transcriptomes and their LGT candidates were measured using the index of frequency of optimal (F_{OP}) codons (Ikemura 1985). We calculated F_{OP} values using CodonW (Peden 2005) with default settings and generated F_{OP} plots using R (<http://www.r-project.org>).

Identification of LGT Candidates and Phylogenetic Analysis

We used the 74 LGT proteins of *Blastocystis* sp. ST1 Nand II (Eme et al. 2017) as queries to identify Opalinidae homologs using BLASTp searches (Camacho et al. 2009) with an e-value cutoff of $1e-05$. 37 and 38 proteins yielded hits in the OP10 and Opal32 protein databases, respectively. Of these, 30 were found in both transcriptomes and 7 and 8 were unique to OP10 and Opal32, respectively. In total, 45 proteins were recovered from the two strains as LGT candidates. To reconstruct their phylogenies, we searched these proteins by BLASTp against the non-redundant GenBank database with an e-value cutoff of $1e-05$ and maximum of 2,000

hits. To reduce the dataset size for subsequent phylogenetic analysis, hit sequences were clustered by CD-HIT (Limin et al. 2012) at 95% similarity. The resulting 45 protein sequence datasets were aligned using MAFFT v7.388 with default settings (Kato and Stanley 2013). Ambiguously aligned sites were removed using trimAl v1.4.rev15 (Capella-Gutierrez et al. 2009) with -automated1 setting prior to phylogenetic analyses. Preliminary phylogenies were reconstructed using FastTree 2.1.7 (Price et al. 2010) and inspected manually to reduce the size of the data set by keeping only a few representatives for the prokaryotic clades distantly related to the eukaryotic sequences. We thus identified 29 proteins from the two *Opalinidae* strains as LGT candidates. The final datasets were aligned and trimmed as described above. Maximum likelihood phylogenetic trees for each dataset were constructed using IQ-TREE (Nguyen et al. 2015) with the best fitting model determined by applying the Bayesian Information Criterion (BIC) with the -m MFP (model selection) with default settings for each dataset. Branch supports were calculated with 1,000 ultrafast bootstrap replicates.

Protein cellular localization was predicted using TargetP 1.1 (Emanuelsson et al. 2000), MitoFates (Fukasawa et al. 2015) and TPpred 2.0 (Savojardo et al. 2014) with default settings. Homologs of mitochondrial proteins in *Opalina* sp. OP10 were searched with BLASTp using MRO sequences from two close relatives: *Blastocystis* (Stechmann et al. 2008) and *Cantina marsupialis* (Noguchi et al. 2015) (supplementary table S3, Supplementary Material online).

Data Availability

Protein sequence data sets used in this work, including complete and trimmed alignments and phylogenetic trees, are available for download at figshare (10.6084/m9.figshare.9746360). *Opalina* sequences have been submitted to GenBank (for accession numbers, see supplementary tables S2 and S3, Supplementary Material online).

Supplementary Material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments

This study was supported by European Research Council grants ProtistWorld (P.L.-G., agreement no. 322669) and Plast-Evol (D.M. agreement no. 787904), the Agence Nationale de la Recherche (D.M., project ANR-15-CE32-0003 “ANCESSTRAM”) and the Institut Diversité Ecologie et Evolution du Vivant (D.M. and N.P.). This project was supported in part by the United States National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 1456054 (<http://www.nsf.gov>), awarded to M.W.B. We thank Prof. Ronald Altig (Mississippi State University) for collection of tadpoles and Andrew J. Roger, Courtney Stairs, and Marlena Dlutek (Dalhousie University, Canada) for *Opalina* sp. Opal32 transcriptome sequencing, which was supported by grant MOP-142349 from the Canadian Institutes of Health Research awarded to A.J. Roger. We thank two anonymous reviewers for constructive comments.

References

- Abrahamian M, Kagda M, Ah-Fong, AMV, Judelson HS 2017. Rethinking the evolution of eukaryotic metabolism: novel cellular partitioning of enzymes in stramenopiles links serine biosynthesis to glycolysis in mitochondria. *BMC Evol Biol.* 17:241.
- Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, et al. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol.* 66(1):4-119.
- Andersen RA. 2004. Biology and systematics of heterokont and haptophyte algae. *Am J Bot.* 91:1508–1522.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114-2120.
- Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2:e790.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLASTp: architecture and applications. *BMC Bioinformatics* 10:421.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973
- Castro-Guerrero NA, Jasso-Chávez R, Moreno-Sánchez R. 2005. Physiological role of rhodoquinone in *Euglena gracilis* mitochondria. *Biochim Biophys Acta.* 1710:113-121.
- Clark CG, van der Giezen M, Alfellani MA, Stensvold CR. 2013. Recent developments in *Blastocystis* research. *Adv Parasitol.* 82:1-32
- Colombo BM, Scalvenzi T, Benlamara S, Pollet N. 2015. Microbiota and mucosal immunity in amphibians. *Front Immunol.* 6:111.
- Denoëud F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, Diogon M, Viscogliosi E, Brochier-Armanet C, Couloux A, Poulain J, et al. 2011. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biol.* 12:R29.
- Derelle R, López-García P, Timpano H, Moreira D. 2016 A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol Biol Evol.* 33:2890-2898.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 300:1005-1016.
- Eme L, Gentekaki E, Curtis B, Archibald J, Roger A. 2017. Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol.* 27:807–820.
- Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140:631–642.

- Fukasawa Y, Tsuji J, Fu S, Tomii K, Horton P, Imai K. 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics*.14:1113-1126.
- Gentekaki E, Curtis BA, Stairs CW, Klimeš V, Elias M, Salas-Leiva DE, Herman EK, Eme L, Arias MC, Henrissat B, et al. 2017. Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. *PLoS Biol*. 15:e2003769–42.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8:1494-1512.
- Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg AV, Harris SR. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog*. 8:e1002979.
- Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Micro*. 16:67–79.
- Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 2:13–34.
- Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Čepička I, Kostka M, Kosakyan A, Alcântara DM, Roger AJ, Shadwick LL, Smirnov A, Kudryavstev A, Lahr DJG, Brown MW. 2017. Between a pod and a hard test: the deep evolution of amoebae. *Mol Biol Evol*. 34:2258-2270.
- Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, Soukal P, Hroudová M, Dolezal P, Stairs CW, Roger AJ, Elias M, Dacks JB, Vlček C, and Hampl V. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol* 26:1274–1284.
- Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780
- Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 9:605–618.
- Kostka M. 2016. Opalinidae. In: Archibald JM, Simpson AGB, Slamovits CH, Margulis L, editors. Handbook of the Protists. Springer, Charm, Switzerland.
- Kostka M, Hampl V, Cepicka I, Flegr J. 2004. Phylogenetic position of *Protoopalina intestinalis* based on SSU rRNA gene sequence. *Mol Phylogenet Evol*. 33:220–224.
- Laetsch DR, Blaxter ML, Leggett RM. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*. 6:1287.

- Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. *Mol Biol Evol.* 33:2318–2336.
- Leger MM, Eme L, Stairs CW, Roger AJ. 2018. Demystifying eukaryote lateral gene transfer. *Bioessays.* 40:e1700242.
- Li M, Ponce-Gordo F, Grim JN, Li C, Zou H, Li W, Wu S, Wang G. 2018. Morphological redescription of *Opalina undulata* Nie 1932 from *Fejervarya limnocharis* with molecular phylogenetic study of Opalinids (Heterokonta, Opalineae). *J Eukaryot Microbiol.* 65:783-791
- Limin F, Beifang N, Zhengwei Z, Sitao W, Weizhong L, 2012. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* 28:3150-3152.
- Lonjers ZT, Dickson EL, Chu TPT, Kreutz JE, Neacsu FA, Anders KR, Shepherd JN. 2012. Identification of a new gene required for the biosynthesis of rodoquinone in *Rhodospirillum rubrum*. *J Bacteriol.* 194:965–971.
- Martin WF. 2017. Too much eukaryote LGT. *Bioessays.* 39:1700115.
- Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura K, Takishita K. 2015. Metabolic capacity of mitochondrion-related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*. *Protist* 166:534–550.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268-274.
- Outten FW, Djaman O, Storz G. 2004. A suf operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol Microbiol.* 52:861–872.
- Peden J 2005. CodonW version 1.4.2. <http://codonw.sourceforge.net/>
- Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-Seq from single cells using Smart-seq2. *Nat. Protocols* 9:171-181.
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
- Patterson DJ. 1989. Stramenopiles: Chromophytes from a protistan perspective. In: Green JC, Leadbeater BSC, Diver WL, editors. The chromophyte algae: problems and perspectives. Clarendon Press, Oxford. p357–379.
- Río Bártulos C, Rogers MB, Williams TA, Gentekaki E, Brinkmann H, Cerff R, Liaud M-F, Hehl AB, Yarlett NR, Gruber A, Kroth PG, van der Giezen M. 2018. Mitochondrial glycolysis in a major lineage of eukaryotes. *Genome Biol Evol.* 10:2310-2325.
- Sakai C, Tomitsuka E, Esumi H, Harada S, Kita K. 2012. Mitochondrial fumarate reductase as a target of chemotherapy: from parasites to cancer cells. *Biochim Biophys Acta.* 1820:643-651.

- Savojardo C, Martelli PL, Fariselli P, Casadio R. 2014. TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. *Bioinformatics* 30:2973-2974.
- Shiratori T, Nakayama T, Ishida K-I. 2015. A new deep-branching stramenopile, *Platysulcus tardus* gen. nov., sp. nov. *Protist* 166:337–348.
- Shiratori T, Thakur R, Ishida K-I. 2017. *Pseudophyllomitus vesiculosus* (Larsen and Patterson 1990) Lee, 2002, a poorly studied phagotrophic biflagellate is the first characterized member of stramenopile environmental clade MAST-6. *Protist* 168:439–451.
- Silberman JD, Sogin ML, Leipe DD, Clark CG. 1996. Human parasite finds taxonomic home. *Nature* 380:398–398.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212.
- Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygusua*. *Curr Biol*. 24:1176-1186.
- Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, Fawcett JP, Roger AJ. 2018. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. *eLife* 7:e34292
- Stechmann A, Hamblin K, Perez-Brocal V, Gaston D, Richmond GS, van der Giezen M, Clark CG, Roger AJ. 2008. Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. *Curr Biol*. 18:580–585.
- Takamiya S, Matsui T, Taka H, Murayama K, Matsuda M, Aoki T. 1999. Free-living nematodes *Caenorhabditis elegans* possess in their mitochondria an additional rhodoquinone, an essential component of the eukaryotic fumarate reductase system. *Arch Biochem Biophys*. 371:284-289.
- Tan KSW. 2004. *Blastocystis* in humans and animals: new insights using modern methodologies. *Vet Parasitol*. 126:121–144.
- Tielens AGM, Rotte C, van Hellemond JJ, Martin W. 2002. Mitochondria as we don't know them. *Trends Biochem Sci*. 27:564–572.
- Tsaousis AD, de Choudens SO, Gentekaki E, Long S, Gaston D, Stechmann A, Vinella D, Py B, Fontecave M, Barras F. 2012. Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. *Proc Natl Acad Sci USA* 109:10426–10431.
- van der Giezen M, Cox S, Tovar J. 2004. The iron-sulfur cluster assembly genes *iscS* and *iscU* of *Entamoeba histolytica* were acquired by horizontal gene transfer. *BMC Evol Biol*. 4:7.

- van Grinsven KWA, Rosnowsky S, van Weelden SWH, Pütz S, van der Giezen M, Martin W, van Hellemond JJ, Tielens AGM, Henze K. 2008. Acetate:Succinate CoA-transferase in the hydrogenosomes of *Trichomonas vaginalis*. *J Biol Chem*. 283:1411–1418.
- van Hellemond JJ, Tielens AGM 1994. Expression and functional properties of fumarate reductase. *Biochem J*. 304:321–331.
- Yubuki N, Pánek T, Yabuki A, Cepicka I, Takishita K, Inagaki Y, Leander BS. 2015. Morphological identities of two different marine stramenopile environmental sequence clades: *Bicosoeca kenaiensis* (Hilliard, 1971) and *Cantina marsupialis* (Larsen and Patterson, 1990) gen. nov., comb. nov. *J Eukaryot Microbiol*. 62:532–542.

Figure Legends

FIG. 1. Maximum likelihood phylogenetic tree of SufCB (188 sequences). Bootstrap values <50% are not shown. The long branch of *Paratrimastix* and *Monocercomonoides* has been shortened to 1/4. For the complete tree see supplementary figure S2, Supplementary Material online.

FIG. 2. Maximum likelihood phylogenetic tree of RquA (102 sequences). Bootstrap values <50% are not shown. Groups A and B are defined according to Stairs et al. 2018. For the complete tree see supplementary figure S3, Supplementary Material online.

FIG. 3. Maximum likelihood phylogenetic tree of ASCT1C (96 sequences). Bootstrap values <50% are not shown. The branch of *Schizosaccharomyces cryophilus* has been shortened to 1/2. For the complete tree see supplementary figure S4, Supplementary Material online.

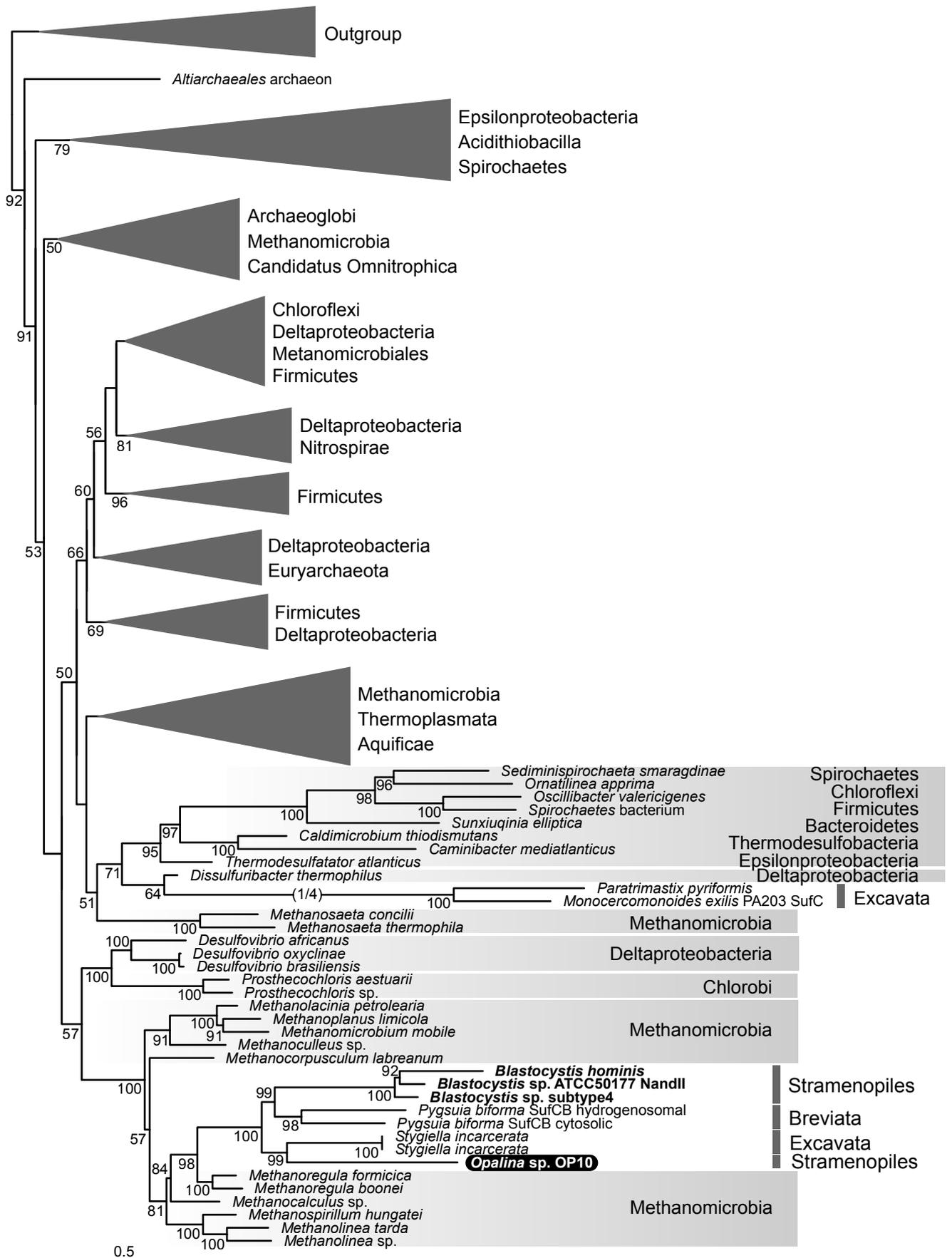


Figure 1

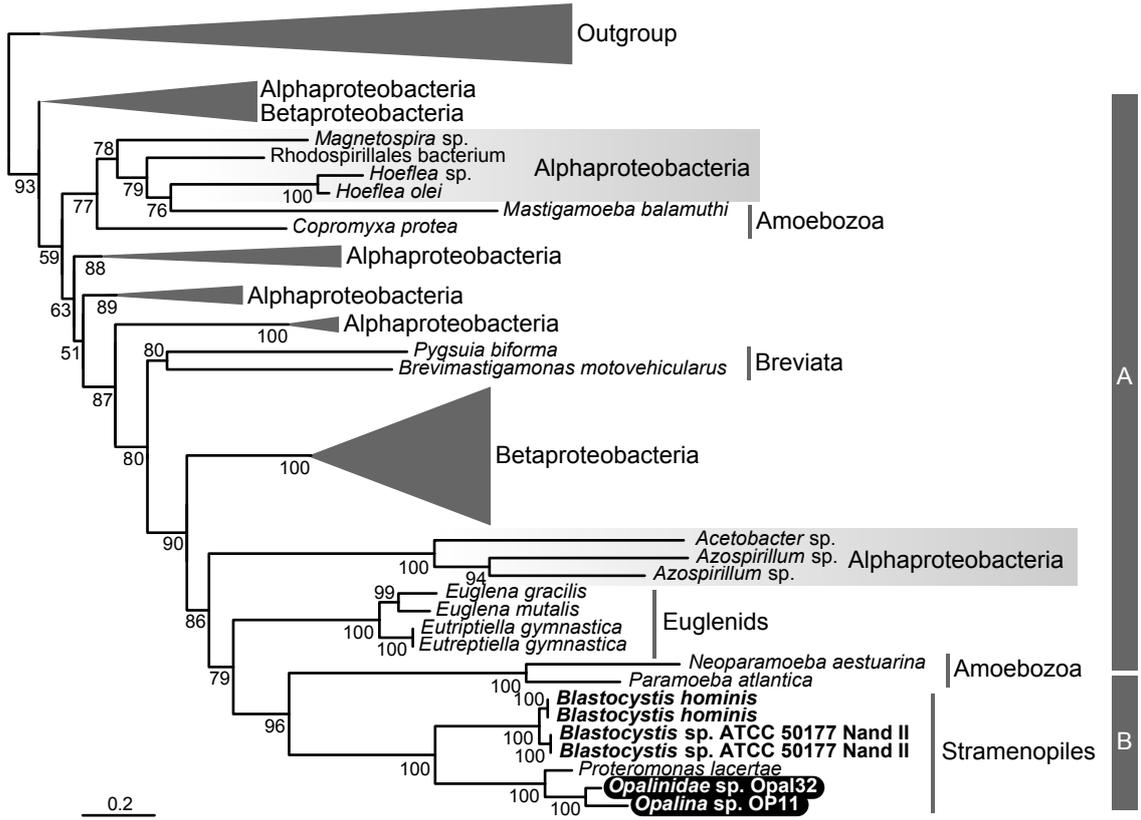


Figure 2

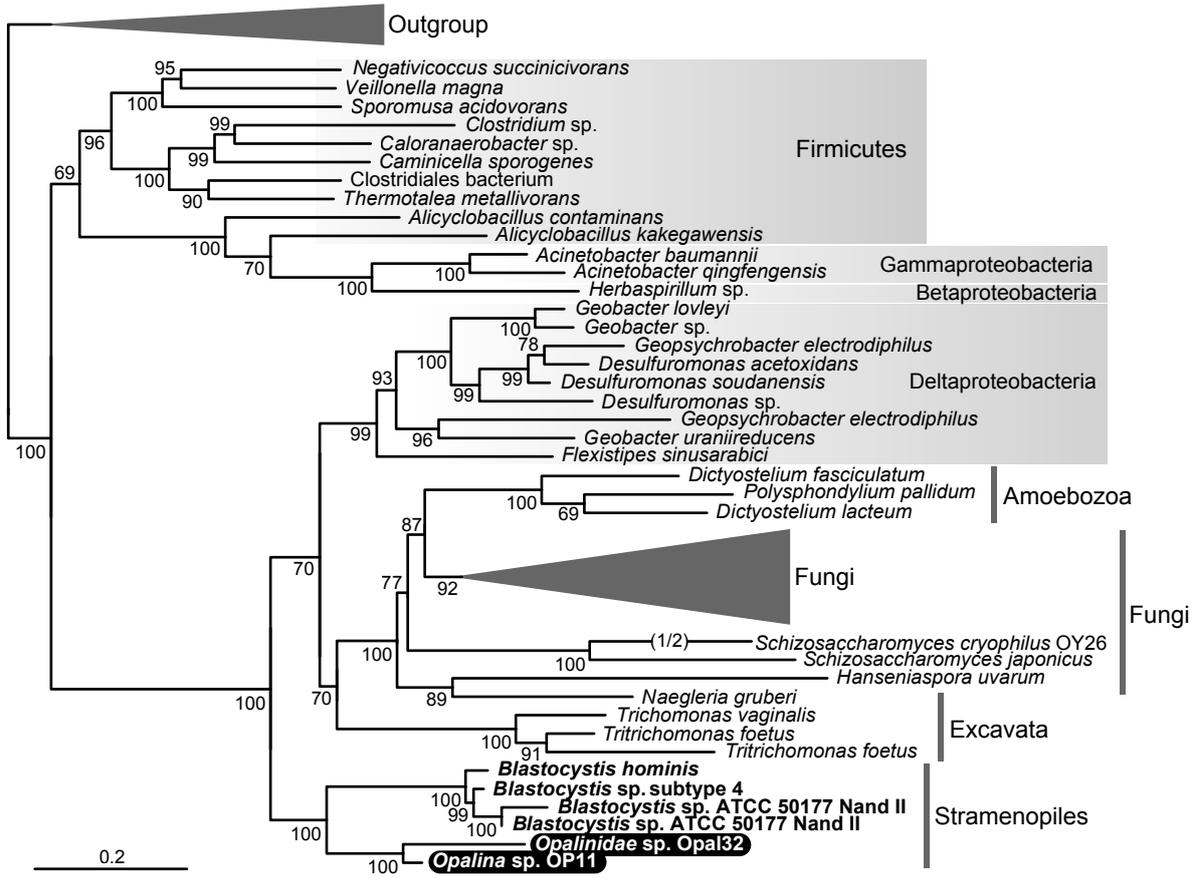


Figure 3