

# Analysis of Two Genomes from the Mitochondrion-Like Organelle of the Intestinal Parasite *Blastocystis*: Complete Sequences, Gene Content, and Genome Organization

Vicente Pérez-Brocal and C. Graham Clark

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Acquisition of mitochondria by the ancestor of all living eukaryotes represented a crucial milestone in the evolution of the eukaryotic cell. Nevertheless, a number of anaerobic unicellular eukaryotes have secondarily discarded certain mitochondrial features, leading to modified organelles such as hydrogenosomes and mitosomes via degenerative evolution. These mitochondrion-derived organelles have lost many of the typical characteristics of aerobic mitochondria, including certain metabolic pathways, morphological traits, and, in most cases, the organellar genome. So far, the evolutionary pathway leading from aerobic mitochondria to anaerobic degenerate organelles has remained unclear due to the lack of examples representing intermediate stages. The human parasitic stramenopile *Blastocystis* is a rare example of an anaerobic eukaryote with organelles that have retained some mitochondrial characteristics, including a genome, whereas they lack others, such as cytochromes. Here we report the sequence and comparative analysis of the organellar genome from two different *Blastocystis* isolates as well as a comparison to other genomes from stramenopile mitochondria. Analysis of the characteristics displayed by the unique *Blastocystis* organelle genome gives us an insight into the initial evolutionary steps that may have led from mitochondria to hydrogenosomes and mitosomes.

## Introduction

The acquisition of the mitochondrion was a crucial step, perhaps even the decisive event, in the evolution of the eukaryotic cell from an archaeal ancestor (Martin and Müller 1998; Embley and Martin 2006). It is generally accepted that this organelle is derived from a symbiotic alpha-proteobacterium that over time became so integrated into the host cell that almost all its several hundred constituent proteins are now encoded in the nucleus. Only a small and variable subset of genes has been retained on a separate organellar genome that resides in the mitochondrial matrix (Gray et al. 2004). The genes present on this genome are almost exclusively involved in either electron transport or translation.

A significant number of eukaryotic metabolic pathways include steps that occur within the mitochondrion, including the tricarboxylic acid cycle, fatty acid and amino acid metabolism, and iron–sulfur center assembly. Despite their important role in metabolism, a number of anaerobic eukaryotes appear to lack mitochondria. Although previously thought to represent basal lineages that had diverged prior to the integration of the alpha-proteobacterial symbiont (Cavalier-Smith 1987), it is now known that all these putatively primitive eukaryotes contain organelles of mitochondrial ancestry, albeit in dramatically modified form in most cases (van der Giezen et al. 2005).

This secondary modification of the organelle has occurred independently many times and has followed at least two different paths. One path has resulted in conversion of the mitochondrion into a hydrogenosome. These organelles occur in a number of unrelated eukaryotic lineages. Hydrogenosomes were first described and are best known from trichomonad flagellates, but they have also been found in certain anaerobic fungi and ciliates, indicating convergent evolution (van der Giezen et al. 2005). In the latter two groups, the organelle conversion in the anaerobes must

be relatively recent as closely related aerobic species with classical mitochondria can be identified. The terminal electron acceptor in hydrogenosomes is a proton, the end product is hydrogen, and the enzyme responsible is hydrogenase, with limited energy being produced by substrate-level phosphorylation (van der Giezen et al. 2005). Because the only specific defining characteristic of hydrogenosomes is the production of hydrogen, it remains to be seen how similar the organelles in these three groups are in other respects.

The second path of mitochondrial modification in anaerobic eukaryotes results in conversion into mitosomes. Like hydrogenosomes, mitosomes are found in unrelated lineages (including diplomonads, amoebae, and microsporidia), indicating multiple independent origins (Mai et al. 1999; Tovar et al. 1999, 2003; Williams et al. 2002). These organelles are defined largely by exclusion as they have no unique shared characteristic like hydrogen production. Mitosomes are therefore defined as organelles of mitochondrial ancestry that are not known to produce energy or hydrogen, and in some cases (e.g., in *Entamoeba*), their function remains completely unknown (van der Giezen et al. 2005).

Eukaryotes with true mitochondria that can function under anaerobic conditions also exist and include some fungi, some worms, and *Euglena*. These organisms are only facultatively anaerobic as they use oxygen when it is present and, therefore, presumably possess a normal complement of mitochondrial genes. In such organisms, nitrate or fumarate, for example, may be used as the terminal acceptor under anaerobic conditions (Tielens et al. 2002).

One characteristic shared by most hydrogenosomes and all mitosomes is the absence of the cristae that give classical mitochondria their instantly recognizable morphology under the electron microscope. Additionally, with one published exception, an organelle genome is absent from all these secondarily modified mitochondria. Although it is still incompletely sequenced, the cristate hydrogenosome of the ciliate *Nyctotherus* contains a genome that is known to encode several components of complex I of the mitochondrial electron transport chain as well as ribosomal proteins (Boxma et al. 2005). No trace of an organellar genome has been detected in any mitosome (León-Avila and Tovar 2004).

**Key words:** *Blastocystis*, hydrogenosome, mitochondrion, mitosome, reductive evolution, stramenopile.

E-mail: graham.clark@lshtm.ac.uk.

*Mol. Biol. Evol.* 25(11):2475–2482. 2008

doi:10.1093/molbev/msn193

Advance Access publication September 2, 2008

© 2008 The Authors

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eukaryotes with morphologically classical mitochondria that are obligate anaerobes are rare. *Blastocystis* is a genetically diverse anaerobic parasite and a member of the diverse eukaryotic group known as the stramenopiles or heterokonts, which includes aerobic organisms like diatoms, oomycetes, and brown algae (Silberman et al. 1996). *Blastocystis* is a common member of the mammalian gut flora although whether it causes symptoms in humans is still a point of active debate (Tan et al. 2002). The mitochondrion of *Blastocystis* resembles what might be postulated as an intermediate stage during the conversion of classical mitochondria into hydrogenosomes or mitosomes. Electron microscopy showed that cristae are present (Stenzel and Boreham 1996), as are a transmembrane potential (Zierdt et al. 1988; Nasirudeen and Tan 2004) and DNA (Matsumoto et al. 1987; Nasirudeen and Tan 2004), but cytochrome-mediated electron transport and certain typical metabolic pathways (Zierdt 1986) appear to be absent.

In a recent publication, we reported a preliminary characterization of the *Blastocystis* mitochondrion-like organelle (MLO) through the analysis of expressed sequence tags (ESTs) sequences, confocal microscopy, and the partial sequence of the organellar DNA (Stechmann et al. 2008). Although hydrogenase protein was detected, hydrogen production has not been demonstrated, and so the definition of the organelle as a mitochondrion, mitosome, or hydrogenosome remains uncertain. As expected from the early work of Zierdt (1986), no evidence for complexes III and IV of the electron transport chain was found, in either the EST sequences or the MLO genome data. However, only a small segment of the MLO genome was sequenced (less than 6 kb), which encoded two subunits of reduced nicotinamide adenine dinucleotide (NADH) dehydrogenase, seven transfer RNAs (tRNAs), and the ribosomal RNAs. As a result, our understanding of the genome was very incomplete. For example, the presence of genes encoding cytochrome or adenosine triphosphatase (AT-Pase) subunits on the MLO genome could not be ruled out even though they were not detected among the ESTs, as they may not have been expressed in the cells used for library construction or could have been present as pseudogenes, indicating a recent loss of function.

To complete our picture of the *Blastocystis* MLO and its genome, we here report the entire annotated sequence of the organelle genome from isolates representing two subtypes of *Blastocystis* and compare their structure and content with the mitochondrial genomes from aerobic stramenopiles sequenced to date. These are the first complete organellar genomes from obligately anaerobic protists, and the results illustrate one possible intermediate stage along the evolutionary pathway from mitochondrion to hydrogenosome or mitosome.

## Materials and Methods

### *Blastocystis* Origin and Culture

DNA from *Blastocystis* sp. NandII (ATCC50177) was a gift from Alexandra Stechmann and Tetsuo Hashimoto (Dalhousie University, Canada). *Blastocystis* sp. DMP/02-328 was grown at 36 °C with a mixed bacterial flora

in LYSGM with 5% adult bovine serum. Subtyping of *Blastocystis* sp. DMP/02-328 indicated that this isolate belongs to subtype 4, whereas *Blastocystis* sp. NandII is subtype 1 (Stechmann et al. 2008). Both are of human origin.

### DNA Extraction

Harvested *Blastocystis* sp. DMP/02-328 cells were concentrated and partially separated from bacteria using Histopaque-1077 (Sigma-Aldrich Ltd, Gillingham, Dorset, UK). The *Blastocystis*-enriched band at the medium–Histopaque interphase was recovered and washed with phosphate-buffered saline. Total DNA extraction was carried out on the resulting pellet using the Gentra Puregene Cell Kit (QIAGEN Ltd, Crawley, UK).

### DNA Amplification by Polymerase Chain Reaction

The small segments of MLO DNA previously sequenced (Stechmann et al. 2008) were used as the starting points to design primers for inverse polymerase chain reaction (PCR) at both ends of the segment for both isolates.

Based on the restriction map of the available sequence, a set of restriction enzymes generating overhanging ends (*Bst*BI, *Cla*I, *Eco*RI, *Hind*III, *Mfe*I, *Nde*I, *Xba*I) or blunt ends (*Rsa*I) was used for Southern blot hybridization (data not shown). Those enzymes generating short- to medium-sized fragments were used to digest the *Blastocystis* DNA. The resulting restriction fragments were self-ligated with T4 DNA ligase (Promega UK Ltd, Southampton, UK) and used as templates for the subsequent inverse PCRs.

In addition, degenerate primers for direct PCR were designed for other genes likely to be present in the genome (e.g., those encoding NADH dehydrogenase [*nad*] subunits), based on an alignment of gene sequences from other stramenopiles (see below).

Two types of PCR amplification were performed: a standard reaction using BioTaq (Bioline Ltd, London, UK) or one using TaKaRa Premix *Taq* (Lonza Biologics PLC, Wokingham, UK) for those reactions giving a negative result in the standard reaction as well as to obtain long PCR products (>2 kb). Because different combinations of primers were employed, some of which were degenerate, a touchdown amplification strategy was used. Normal touchdown PCR conditions consisted of 94 °C for 2 min, followed by 10 cycles of 94 °C for 15 s, 60 °C (decreasing by 0.5 °C per cycle) for 30 s, and 72 °C for 2 min, and followed by 20 cycles of 94 °C for 15 s, 55 °C for 30 s, and 72 °C for 2 min. Long PCR conditions consisted of 94 °C for 2 min, followed by 10 cycles of 94 °C for 15 s, 60 °C (decreasing by 0.5 °C per cycle) for 30 s, and 68 °C for 4–5 min, and followed by 20 cycles of 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 4–5 min (increasing by 5 s per cycle).

After PCR product sequencing, the process was repeated with new specific primers to “walk” along the MLO DNA in two directions: outward through inverse PCR and inward through direct PCR to cover the gaps between separate regions. To obtain the latter, pairs of specific primers were combined in all the possible ways and direct long PCR carried out in order to generate a PCR product linking two contigs.

## Sequencing, Assembly, and Genome Analysis

PCR products were gel purified (Qiaquick Gel Extraction kit; QIAGEN Ltd), and direct sequencing of PCR products was performed on an ABI3730 with ABI Prism BigDye Terminator v3.1 reagents (Applied Biosystems, Warrington, UK). Sequences were assembled and analyzed with the Staden Package software programs (Staden et al. 2000). Blast searches were conducted to identify genes, pseudogenes, and structural RNAs. tRNA gene prediction was performed under the different search modes and models implemented in tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>; Lowe and Eddy 1997). Genes were annotated manually, based on Blast and alignment similarity (Altschul et al. 1997) to genes in other sequenced stramenopile mitochondrial genomes, and G + C content was calculated using GeneRunner v3.05 (<http://www.generunner.com>) (supplementary table S1, Supplementary Material online). The presence of group I and II introns was tested using the intron prediction tool RNAweasel (<http://megasun.bch.umontreal.ca/RNAweasel>; Lang et al. 2007). The GenBank accession numbers of the *Blastocystis* sp. DMP/02-328 and NandII MLO genomes analyzed in the present work are EF494739 and EF494740, respectively.

## Sequence Alignments

For mitochondrial gene sequence alignments, all the stramenopile mitochondrial genomes sequenced to date were used: *Thalassiosira pseudonana* (DQ186202), *Cafeteria roenbergensis* (AF193903), *Chrysodidymus synuroideus* (AF222718), *Ochromonas danica* (AF287134), *Phytophthora infestans* (U17009), *Phytophthora sojae* (DQ832717), *Phytophthora ramorum* (DQ832718), *Saprolegnia ferax* (AY534144), *Desmarestia viridis* (AY500367), *Dictyota dichotoma* (AY500368), *Fucus vesiculosus* (AY494079), *Laminaria digitata* (AJ344328), and *Pylaiella littoralis* (AJ277126).

Nucleotide sequences were either directly aligned, in the case of structural RNAs, or first translated into amino acid sequences, in the case of protein-encoding genes, using the ClustalW tool implemented in the MEGA4 package (Kumar et al. 2004). The latter alignment was in turn used as a template to align the corresponding nucleotides to reduce ambiguities.

Amino acid sequence alignments were used to design degenerate primers based on conserved regions within the proteins. This allowed us to obtain additional sequences outside the existing contigs and, thus, additional sites for inverse PCR “walking.”

## Phylogenetic Analyses

From the orthologous open-reading frames (ORFs) shared by *Blastocystis* MLO DNA and the other stramenopile genomes (see above), those producing unambiguous alignments (i.e., *nad* genes) were selected. For phylogenetic analyses, a range of other mitochondrial genomes were selected, based on the similarity of their *nad* gene content, as well as five alpha-proteobacterial genomes. Nucleotide se-

quences were concatenated, translated into amino acids, and aligned with ClustalW and MUSCLE v3.6 (Edgar 2004). The resulting alignments were edited manually to remove regions of ambiguous homology. Maximum likelihood (ML) and Bayesian inference analyses were carried out using Phylml\_v2.4.5 (Guindon and Gascuel 2003) and MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001), respectively, using the best of the amino acid substitution models derived from ProtTest 1.4 (Abascal et al. 2005): a CpREV model using a gamma distribution, with the proportion of invariable sites estimated from the data and eight categories of substitution. In all, 400 bootstrap resamplings were performed to estimate statistical support for branches in the optimal ML tree. Bayesian analysis used four Markov chain Monte Carlo (MCMC) strands, 1,000,000 generations, with trees sampled every 100 generations. A consensus tree was produced after excluding an initial burn-in of 25% of the samples, as recommended.

## Selection Analyses

To evaluate the type of selection operating on each gene (i.e., positive, purifying, or absence of either), we carried out an analysis using a Z-test of selection, implemented in MEGA3. The model used was modified Nei–Gojobori, one-tail with  $\alpha = 0.025$ , and Jukes–Cantor as a nucleotide substitution model.

## Results

### Physical and Gene Map, Gene Content, and Structure of *Blastocystis* sp. DMP/02-328 and NandII MLO DNA

The MLO genome consists of a circular mapping molecule of 27,719 and 28,382 bp in *Blastocystis* sp. DMP/02-328 and NandII, respectively. The synteny between the two MLO genomes is complete, with no evidence of rearrangements having occurred since the divergence of subtypes 1 and 4, represented by *Blastocystis* sp. NandII and DMP/02-328, respectively. The genome consists of 45 genes, of which 27 are ORFs (23 of them with identifiable orthologs) and the rest structural RNA genes (the small and large subunit ribosomal RNAs, plus a set of 16 tRNAs, including two distinct tRNA-Met<sub>c</sub> genes). The structure of the *Blastocystis* sp. DMP/02-328 MLO genome is displayed in figure 1. It is a highly compact genome, with coding regions representing 96% of its total length and overlapping genes being a common feature.

The extraordinary genetic diversity in *Blastocystis* previously reported (Stensvold et al. 2007) has also been detected in the MLO genome. The sequence divergence, which for some genes exceeds 50% between the two subtypes, is higher than that found between the mitochondrial genomes of many congeneric species or even distinct genera. The overall nucleotide identity between the *Blastocystis* sp. DMP/02-328 and NandII MLO genomes is 75.4%, or 71.6% and 64.3% for nucleotides and amino acids, respectively, when only protein-coding genes are considered. The gene-by-gene percent identity is shown in supplementary table S1 (Supplementary Material online).



**Table 1**  
**Comparison of Some Genomic Features among 13 Mitochondrial Genomes from Stramenopiles**

	<i>Thalassiosira pseudonana</i>	<i>Cafeteria roenbergensis</i>	<i>Chrysochromidymus synuroideus</i>	<i>Ochromonas danica</i>	<i>Phytophthora infestans</i>	<i>Saprolegnia ferax</i>	<i>Desmaretia viridis</i>	<i>Dictyota dichotoma</i>	<i>Fucus vesiculosus</i>	<i>Laminaria digitata</i>	<i>Pylaiella littoralis</i>	<i>Blastocystis</i> sp. DMP/02-328	<i>Blastocystis</i> sp. NandII
Total length (bp)	43,827	43,159	34,119	41,035	37,957	46,930	39,049	31,617	36,392	38,007	58,507	27,719	28,382
G + C content (%)	30.1	27.3	24.1	26.2	22.3	23.1	36.6	36.5	34.5	35.1	38.0	21.9	19.9
Total Shared genes <sup>a</sup>	30.0	27.7	25.6	27.1	24.4	24.7	38.7	37.4	36.7	37.5	38.0	22.9	21.2
Intergenic regions	30.1	18.0	12.5	21.9	10.7	6.8	26.5	24.4	23.4	23.7	37.4	10.8	10.9
Protein-coding regions	29.0	25.9	22.9	24.7	21.0	20.9	35.6	35.4	33.3	34.0	37.1	20.7	17.9
Structural RNA genes	35.3	37.9	33.9	36.8	35.6	35.2	45.1	42.9	43.5	45.2	42.9	28.7	30.0
Gene number	61	58	62	75	67	77	68	66	67	67	79	45	45
Protein coding	35	34	37	44	40	43	39	38	38	39	52	27	27
Structural RNAs	27	24	25	31	27	24	28	27	28	27	26	18	18
Average length	804	1049	718	692	706	743	784	645	743	756	855	793	813
ORFs (nt)	160	29	38	62	58	57	46	20	39	50	127	28	33
Average length													
Intergenic region (nt)													

<sup>a</sup> Consists of all *Blastocystis* genes except *orf143*, *orf160*, *orf175*, *orf192*, which have no clear homologs, and one of the two tRNA-Met<sub>c</sub> genes. Additionally, *rps11* and *rps10* are excluded due to their absence in certain other stramenopiles.

Because the mitochondrial gene order is extremely varied among stramenopile mtDNAs, it is not surprising that their organization in *Blastocystis* does not resemble any of the others, despite the synteny within *Blastocystis*. However, a few traces of conserved blocks of ribosomal protein genes, previously reported in distantly related bacteria and mtDNA (Hauth et al. 2005), still remain in *Blastocystis*. These include the cluster *rpl2-rps19-rps3-rpl16*, as well as the block formed by *rps14-rps8-rpl6*. The latter is preceded in *Blastocystis* by *orf175*, which has weak similarity to *rpl5*, the gene normally (although not universally) found upstream of *rps14-rps8-rpl6* in stramenopiles.

### Phylogenetic Position of *Blastocystis*

The ML tree based on the amino acid sequences of nine *nad* genes is shown in figure 2 (*nad11* was excluded as it is not present in all organelle genomes analyzed). The topology of the Bayesian tree was identical. Trees using either method strongly support the clustering of *Blastocystis* sp. DMP/02-328 and NandII as a monophyletic clade, with weaker support for the Peronosporomycetes, represented by *S. ferax* and *Phytophthora* spp., as a sister group. This result suggests that oomycete mitochondrial genomes are the closest relatives of the *Blastocystis* MLO, in agreement with the analysis of Stechmann et al. (2008), which was based on a smaller number of MLO genes. Our trees support the monophyly of the Phaeophyceae, the Peronosporomycetes, and the clade consisting of the Chrysophyceae and Synurales. The branch lengths also agree with the relative rate test of substitution (Moran 1996; data not shown) in reflecting intermediate values of acceleration in substitution rates in *Blastocystis* sp., compared with the more accelerated *C. roenbergensis* and the slower remaining stramenopiles.

### Discussion

The completion of the MLO genome from two different *Blastocystis* subtypes has revealed some distinctive features that combine to make this organism's organelle unique. The MLO genome is missing several protein-coding genes that are present in the mtDNAs of all other stramenopiles studied to date. However, these are not a random selection but primarily consist of all the genes encoding cytochrome and ATPase subunits. The absence of MLO-encoded subunits of the ubiquinol-cytochrome *c* reductase complex (complex III) and cytochrome *c* oxidase (complex IV) confirms the metabolism inferences made by Stechmann et al. (2008), which were based on the absence of any evidence for these complexes among their EST sequences. The absence of complexes III and IV in *Blastocystis*, together with the absence of F<sub>0</sub>F<sub>1</sub>-ATPase subunits, also agrees with previous biochemical studies (Zierdt 1986; Zierdt et al. 1988; Nasirudeen and Tan 2004) and represents a unique case to date among completely sequenced mitochondrial genomes. Our motivation for sequencing two MLO genomes from distinct subtypes was to see whether evidence for ongoing

**Table 2**  
**Gene Content in Stramenopile mtDNAs<sup>a</sup>**

Genes <sup>b</sup>	<i>Thalassiosira pseudomana</i>	<i>Cafeteria roenbergensis</i>	<i>Chrysochromidymus synuroideus</i>	<i>Ochromonas danica</i>	<i>Phytophthora infestans</i>	<i>Saprolegnia ferax</i>	<i>Desmareestia viridis</i>	<i>Dicryota dichotoma</i>	<i>Fucus vesiculosus</i>	<i>Laminaria digitata</i>	<i>Pyraliella littoralis</i>	<i>Blastocystis</i> sp.
<i>rns, rnl</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>trnA-W</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>atp6, atp8</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>cob</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>cox1-cox3</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>nad1-nad7, nad9</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>nad11</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>atp1</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>atp9</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps2</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps3</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps4</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps7</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps8</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps10</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps11</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rps12, rps14, rps19</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rpl2</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rpl5</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>rpl6, rpl14, rpl16</i>	■	■	■	■	■	■	■	■	■	■	■	■
<i>ymf16</i>	■	■	■	■	■	■	■	■	■	■	■	■
ORFs	1	4	5	10 (12)	4	4	4	3	3	4	16	4

<sup>a</sup> The table is based on the gene content reported by Chesnick et al. (2000), with the addition of more recently sequenced stramenopile mtDNAs and the MLO.

<sup>b</sup> Filled boxes—genes present; open boxes—genes absent.

degeneration existed, but, at least in the MLO genome, no evidence for this was found. The absence of pseudogenes in the MLO genomes suggests that the gene loss is not recent.

Stechmann et al. (2008) proposed a metabolic map of the *Blastocystis* MLO based primarily on a library of EST sequences. The six nuclear encoded subunits of complex I they identified among the ESTs and the total of 10 now identified in the completed MLO genome reinforce the apparent importance of maintaining this complex in *Blastocystis*. It is likely to be responsible for generating the proton gradient in these organelles detected with MitoLight and Rhodamine 123 staining (Zierdt et al. 1988; Nasirudeen and Tan 2004) and MitoTracker (Hamblin et al. 2008; Stechmann et al. 2008). In *Blastocystis*, it is likely that complex I pumps protons derived from NADH as it passes electrons via an unidentified quinone to alternative oxidase (AOX).

Despite being incompletely sequenced, some traits of the *Nyctotherus ovalis* hydrogenosome genome (Boxma et al. 2005) are known to be shared with the MLO of *Blastocystis*. In both cases, a set of subunits from complexes I and II of the electron transport chain are present (Boxma et al. 2005; Hamblin et al. 2008; Stechmann et al. 2008), whereas no components of complexes III and IV or the F<sub>0</sub>F<sub>1</sub>-ATPase have been found. The fact that the latter are typically present in classical mitochondria but are absent from the MLO, hydrogenosomes, and mitosomes suggests that the loss of these three complexes could be linked to the first steps in degenerative evolution of mitochondria in anaerobic organisms, whatever the ultimate outcome.

We propose that a likely scenario would be one in which the organism passed through a facultatively anaerobic stage. In contrast to known examples where nitrate or fumarate may be used as an alternative, the absence of the normal terminal electron acceptor (O<sub>2</sub>) would result in all electrons being passed through existing pathways, either to AOX or hydrogenase via a quinone or ferredoxin, respectively, or potentially both (Stechmann et al. 2008). Subsequent mutations affecting any component of complexes III and IV would no longer be lethal, resulting in a removal of selection for maintaining those genes and producing a now obligately anaerobic organelle. The reduced transmembrane proton gradient generated by complex I alone may no longer be sufficient to support both solute transport and adenosine triphosphate synthesis via the F<sub>0</sub>F<sub>1</sub>-ATPase, leading to loss of the latter. All trace of the now unnecessary genes would disappear from the MLO genome (and presumably the nucleus also).

This appears to be the stage of degeneration at which *Blastocystis* is presently found, and in the absence of evidence for hydrogen production, the MLOs are probably most accurately described as obligately anaerobic mitochondria. Whether all other mitochondrion-derived organelle-bearing organisms went through an identical stage is impossible to know, as is whether the *Blastocystis* organelle will eventually degenerate further. Compared with the trichomonad hydrogenosome (Carlton et al. 2007) and known mitosomes, the *Blastocystis* MLO has a much more complete mitochondrion-like metabolism (Stechmann et al. 2008). This suggests that electron transport is one

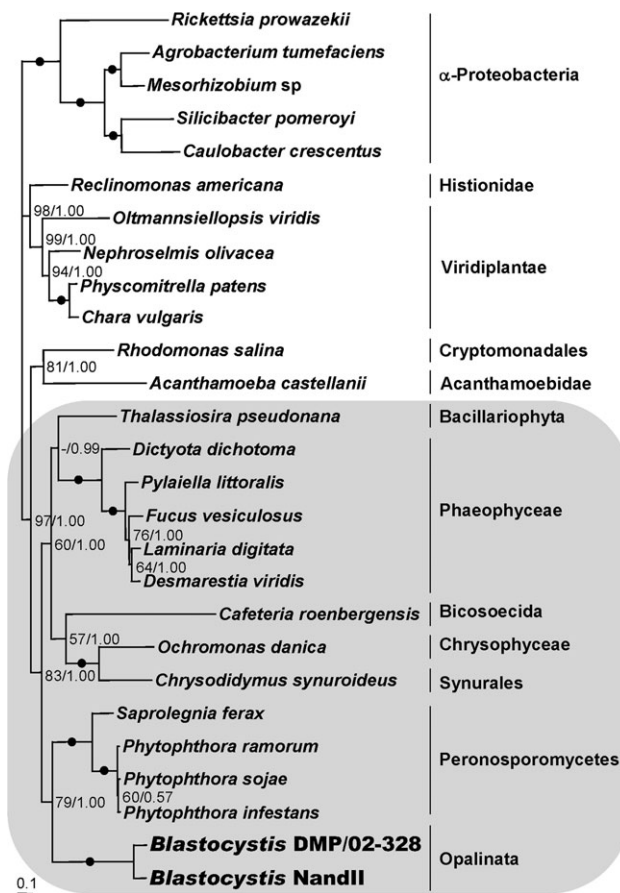


FIG. 2.—ML tree of *nad* proteins. This tree was obtained using the edited alignment produced by MUSCLE v. 3.6, the CpREV+I+G+F amino acid substitution model, and is based on the concatenated amino acid sequences of the nine NADH dehydrogenase genes present in all of the stramenopiles included (i.e., *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*; see Materials and Methods) as well as in the 12 taxa used as outgroups. Numbers beside the internal nodes are the ML bootstrap values from 400 resamplings obtained with Phym1 and the Bayesian MCMC posterior probability values. Black circles indicate 100% bootstrap support and 1.00 posterior probability values. Support values over 50% are shown adjacent to the corresponding nodes. Values below 50% are represented by a dash (-). The clade containing all the stramenopiles is shaded. Taxonomic classification follows the nomenclature of Adl et al. (2005).

of the first, if not actually the first, pathways to be affected during the degenerative evolution of the mitochondrion into the derived organelles.

## Note

While this article was in revision, a paper appeared describing the MLO genome of another subtype of *Blastocystis* sp. (Wawrzyniak et al. 2008).

## Supplementary Material

Supplementary tables S1, S2, and S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

This work was supported by grant 078566 from the Wellcome Trust awarded to CGC. We would like to thank Drs Andrew Roger and Alexandra Stechmann for providing us with the DNA from *Blastocystis* sp. NandII and also Dr Mark van der Giezen for his advice and for commenting on the manuscript.

## Literature Cited

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21: 2104–2105.
- Adl SM, Simpson AGB, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol*. 52:399–451.
- Altschul SE, Madden TL, Schäffer AA, Zhang J, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.
- Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci USA*. 101:17747–17752.
- Boxma B, de Graaf RM, van der Staay GW, et al. (15 co-authors). 2005. An anaerobic mitochondrion that produces hydrogen. *Nature*. 434:74–79.
- Carlton JM, Hirt RP, Silva JC, et al. (65 co-authors). 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. 315:207–212.
- Cavalier-Smith T. 1987. Eukaryotes with no mitochondria. *Nature*. 326:332–333.
- Chesnick JM, Goff M, Graham J, Ocampo C, Lang BF, Seif E, Burger G. 2000. The mitochondrial genome of the stramenopile alga *Chrysodidymus synuroideus*. Complete sequence, gene content and genome organization. *Nucleic Acids Res*. 28:2512–2518.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature*. 440:623–630.
- Feagin JE. 2000. Mitochondrial genome diversity in parasites. *Int J Parasitol*. 30:371–390.
- Fontaine JM, Rousvoal S, Leblanc C, Kloareg B, Loiseaux-de Goër S. 1995. The mitochondrial LSU rDNA of the brown alga *Pylaiella littoralis* reveals alpha-proteobacterial features and is split by four group IIB introns with an atypical phylogeny. *J Mol Biol*. 251:378–389.
- Ghai R, Hain T, Chakraborty T. 2004. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics*. 5:198.
- Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu Rev Genet*. 38:477–524.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.
- Hamblin K, Standley DM, Rogers MB, Stechmann A, Roger AJ, Maytum R, van der Giezen M. 2008. Localization and nucleotide specificity of *Blastocystis* succinyl-CoA synthetase. *Mol Microbiol*. 68:1395–1405.
- Hancock K, Hajduk SL. 1990. The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. *J Biol Chem*. 265:19208–19215.

- Hauth AM, Maier UG, Lang BF, Burger G. 2005. The *Rhodomonas salina* mitochondrial genome: bacteria-like operons, compact gene arrangement and complex repeat region. *Nucleic Acids Res.* 33:4433–4442.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5:150.
- Lang BF, Brinkmann H, Koski LB, Fujishima M, Gortz HD, Burger G. 2005. On the origin of mitochondria and Rickettsia-related eukaryotic endosymbionts. *Jpn J Protozool.* 38:171–183.
- Lang BF, Laforest M-J, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23:119–125.
- León-Avila G, Tovar J. 2004. Mitosomes of *Entamoeba histolytica* are abundant mitochondrion-related remnant organelles that lack a detectable organellar genome. *Microbiology.* 150:1245–1250.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J. 1999. Hsp60 is targeted to a cryptic mitochondrion-derived organelle (“crypton”) in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell Biol.* 19:2198–2205.
- Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature.* 392:37–41.
- Matsumoto Y, Yamada M, Yoshida Y. 1987. Light-microscopical appearance and ultrastructure of *Blastocystis hominis*, an intestinal parasite of man. *Zentralbl Bakteriell Mikrobiol Hyg [A].* 264:379–385.
- Moran NA. 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA.* 93:2873–2878.
- Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell.* 108:583–586.
- Nasirudeen AM, Tan KS. 2004. Isolation and characterization of the mitochondrion-like organelle from *Blastocystis hominis*. *J Microbiol Methods.* 58:101–109.
- Schneider A, Maréchal-Drouard L. 2000. Mitochondrial tRNA import: are there distinct mechanisms? *Trends Cell Biol.* 10:509–513.
- Silberman JD, Sogin ML, Leipe DD, Clark CG. 1996. Human parasite finds taxonomic home. *Nature.* 380:398.
- Simpson AM, Suyama Y, Dewes H, Campbell DA, Simpson L. 1989. Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. *Nucleic Acids Res.* 17:5427–5445.
- Staden R, Beal K, Bonfield JK. 2000. The Staden package, 1998. *Methods Mol Biol.* 132:115–130.
- Stechmann A, Hamblin K, Pérez-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.
- Stensvold CR, Suresh GK, Tan KS, Thompson RC, Traub RJ, Viscogliosi E, Yoshikawa H, Clark CG. 2007. Terminology for *Blastocystis* subtypes—a consensus. *Trends Parasitol.* 23:93–96.
- Stenzel DJ, Boreham PF. 1996. *Blastocystis hominis* revisited. *Clin Microbiol Rev.* 9:563–584.
- Tan KS, Singh M, Yap EH. 2002. Recent advances in *Blastocystis hominis* research: hot spots in terra incognita. *Int J Parasitol.* 32:789–804.
- Tielens AGM, Rotte C, van Hellemond JJ, Martin W. 2002. Mitochondria as we don’t know them. *Trends Biochem Sci.* 27:564–572.
- Tovar J, Fischer A, Clark CG. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol Microbiol.* 32:1013–1021.
- Tovar J, León-Avila G, Sánchez L, Sutak R, Tachezy J, van der Giezen M, Hernández M, Müller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature.* 426:172–176.
- van der Giezen M, Tovar J, Clark CG. 2005. Mitochondrion-derived organelles in protists and fungi. *Int Rev Cytol.* 244:175–225.
- Wawrzyniak I, Roussel M, Diogon M, Couloux A, Texier C, Tan KSW, Vivarès CP, Delbac F, Wincker P, El Alaoui H. 2008. Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. *Int J Parasitol.* 38:1377–1382.
- Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature.* 418:865–869.
- Zierdt CH. 1986. Cytochrome-free mitochondria of an anaerobic protozoan-*Blastocystis hominis*. *J Protozool.* 33:67–69.
- Zierdt CH, Donnolley CT, Muller J, Constantopoulos G. 1988. Biochemical and ultrastructural study of *Blastocystis hominis*. *J Clin Microbiol.* 26:965–970.

Prof. Martin Embley, Associate Editor

Accepted August 26, 2008