The Mitochondrial Genome of the Entomoparasitic Green Alga *Helicosporidium*

Jean-François Pombert*, Patrick J. Keeling

Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

**Abstract**

**Background:** Helicosporidia are a chlrophyllous, non-photosynthetic protists that are obligate parasites of invertebrates. Highly specialized, these pathogens feature an unusual cyst stage that dehisces inside the infected organism and releases a filamentous cell displaying surface projections, which will penetrate the host gut wall and eventually reproduce in the hemolymph. Long classified as incertae sedis or as relatives of other parasites such as Apicomplexa or Microsporidia, the Helicosporidia were surprisingly identified through molecular phylogeny as belonging to the Chlorophyta, a phylum of green algae. Most phylogenetic analyses involving Helicosporidia have placed them within the subgroup Trebouxiophyceae and further suggested a close affiliation between the Helicosporidia and the genus *Prototheca*. *Prototheca* species are also achlorophyllous and pathogenic, but they infect vertebrate hosts, inducing protothecosis in humans. The complete plastid genome of an *Helicosporidium* species was recently described and is a model of compaction and reduction. Here we describe the complete mitochondrial genome sequence of the same strain, *Helicosporidium* sp. ATCC 50920 isolated from the black fly *Simulium jonesi*.

**Methodology/Principal Findings:** The circular mapping 49343 bp mitochondrial genome of *Helicosporidium* closely resembles that of the vertebrate parasite *Prototheca wickerhamii*. The two genomes share an almost identical gene complement and display a level of syntenesis that is higher than any other sequenced chlorophyte mitochondrial DNAs. Interestingly, the *Helicosporidium* mtDNA feature a trans-spliced group I intron, and a second group I intron that contains two open reading frames that appear to be degenerate maturase/endonuclease genes, both rare characteristics for this type of intron.

**Conclusions/Significance:** The architecture, genome content, and phylogeny of the *Helicosporidium* mitochondrial genome are all congruent with its close relationship to *Prototheca* within the Trebouxiophyceae. The *Helicosporidium* mitochondrial genome does, however, contain a number of novel features, particularly relating to its introns.

**Citation:** Pombert J-F, Keeling PJ (2010) The Mitochondrial Genome of the Entomoparasitic Green Alga *Helicosporidium*. PLoS ONE 5(1): e8954. doi:10.1371/journal.pone.0008954

**Editor:** Jason E. Stajich, University of California, Riverside, United States of America

**Received** November 19, 2009; **Accepted** January 12, 2010; **Published** January 29, 2010

**Copyright:** © 2010 Pombert, Keeling. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by a Natural Sciences and Engineering Research Council (NSERC) operating grant to PJK. JFP is the recipient of the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT)/Genome Québec Louis-Berlinguet Postdoctoral Fellowship. PJK is a Fellow of the Canadian Institute for Advanced Research (CIFAR), and a Senior Scholar of the Michael Smith Foundation for Health Research (MSFHR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: jpombert@interchange.ubc.ca

**Introduction**

Helicosporidia are single cell parasitic eukaryotes infecting a wide range of insects ([1] and references therein). These entomopathogens feature three different life stages: cysts, filamentous cells and vegetative cells. When the infectious cysts burst open within the gut of their host, they release a filamentous cell with surface barbs along with three egg-shaped accessory cells [2]. The barbed filaments proceed to invade the gut cells, passing through them and emerging into the hemolymph [3]. In their vegetative state within the hemolymph, Helicosporidia reproduce by several rounds of autosporegenic division within the pellicle of the mother cell, with each autosporation producing up to eight daughter cells [4]. Generally, the infection leads to the death of the host, but the exact mode of transmission remains poorly known.

First described in 1921 by Keilin [5], the Helicosporidia were long ignored in classification systems due to their mysterious origins ([3,6] and references therein). Initially, they were ascribed to the Protozoa, then transferred to the Fungi, before being reclassified as Protozoa, specifically within the Cnidosporidia. Cnidosporidia are a long-standing group consisting of Helicosporidium, Microsporidia, and Myxosporidia. The latter two groups are now known to be fungi and animals, respectively, so it is fitting that the Helicosporidia should eventually be determined to be closely related to plants, or more specifically to green algae. This was first suggested based on the astute observation that the morphology and *in vitro* development of Helicosporidia resemble that of the achlorophyllous non-photosynthetic green algae of the genus *Prototheca* [3]. This taxonomic affiliation was quickly supported by molecular phylogeny of several nucleus-encoded genes [6,7,8,9], and further supported by the subsequent finding that Helicosporidia harbour a functional yet heavily reduced chloroplast genome [10,11].

Phylogenetic analyses have, where the sampling diversity was sufficient, consistently suggested an affiliation to *Prototheca* [6,7,8,9,12], in agreement with their morphological characters.
Prototheca is a member of the green algal class Trebouxiophyceae, and is also achlorophyllous and pathogenic. This association is intriguing since protothecans infect only vertebrates inducing protothecosis in humans [13], whereas the Helicosporidia are known so far to invade only invertebrates.

To learn more about these intriguing but poorly studied parasites, and further compare them with their likely closest relatives in the genus Prototheca, here we report the complete sequence of the Helicosporidium sp. ATCC 50920 mitochondrial genome. The mitochondrial genome is a useful tool for such comparisons because complete mitochondrial genomes are available from representatives of most major groups of green algae, including Prototheca. The architecture of the 49343 pb-long Helicosporidium mitochondrial genome and the 60 genes it encodes are highly similar to those of Prototheca wickerhamii and display a level of synteny that have not been previously observed between any two chlorophyte mitochondrial DNAs (mtDNAs). The Helicosporidium mtDNA also has several interesting characteristics that are not only absent from Prototheca, but are rare in mitochondria as a whole, including a rare case of group I intron spliced in trans and introns that encode multiple ORFs.

Results

Main Features of the Mitochondrial Genome

The Helicosporidium mitochondrial genome (GenBank: GQ339576) was sequenced as part of an ongoing genome project on Helicosporidium sp. strain ATCC 50920 in which 402658 reads totaling 146.7 Mbp were generated by 454 Titanium pyrosequencing. Over 87.5% of these reads (364 bp average) were assembled into 4360 contigs representing about 10.6 Mbp. The mitochondrial genome was represented by a single contig comprising 53785 reads, amounting to 1.96 Mbp, or 396-fold coverage of the genome.

The mitochondrial genome maps as a circular molecule of 49343 bp (Figure 1) featuring an overall A+T content of 74.4% (Table 1). The 60 genes it encodes are distributed with a marked coverage of the genome comprising 53785 reads, amounting to 1.96 Mbp, or 396-fold coverage of the genome.

The mitochondrial genome contains a total of four introns, all of which split the rnl and car1 genes in three exons each. The Helicosporidium mtDNA also features three intron 1 open reading frames (ORFs) and two freestanding ORFs that are longer than 150 codons. Intergenic regions in the Helicosporidium mitochondrial genome range from 0 to 2355 bp, with an average of 183 bp, and no overlapping genes. The Helicosporidium mtDNA is more densely packed than that of Prototheca and is leaner by about 6 kbp despite maintaining a near-identical gene complement, differing only by a single tRNA, trnGleuc (Tables S1 and S2). Both genomes feature trnT(ugu), a tRNA-encoding gene also found within the mtDNA of the ulvophytes (Table S2). Like Prototheca and Pseudendoclonium mtDNAs, the Helicosporidium mitochondrial genome harbors a self-sufficient tRNA gene complement able to encode all codons assuming super Wobble codon/anticodon interactions. Codon usage in Helicosporidium mtDNA (Table S3) is also similar to that of Prototheca mtDNA, which parallels their very similar A+T content.

The two freestanding ORFs in Helicosporidium mtDNA (orf160b and orf185) showed no significant homology in BLAST searches (E-values ≤1E-05). Although orf160b shares no identifiable similarity to any known ORF, it is located at the same genetic locus as ynf45 (orf174) in Prototheca mtDNA, i.e. between the nad2-rps10 and rps3 genes. Given their small size, the two Helicosporidium mtDNA freestanding ORFs might not encode any relevant biological product and rather represent random open reading frames, but the conserved position of Helicosporidium orf160b and Prototheca orf174 does suggest they are rapidly diverging homologues.

Synteny

The Helicosporidium mitochondrial genome features a high level of synteny with that of Prototheca. The two genomes share 12 gene clusters encompassing a total of 45 genes (Figure 1). This level of synteny has not been previously observed between mitochondrial genomes of any other chlorophytes. The mitochondrial genomes of the prasinophytes Nephroselmis and Ostreococcus share 10 clusters comprising a total of 36 genes, those of the ulvophytes Pseudendoclonium and Oltmannsiellopsis share only two gene pairs (4 genes) despite displaying a similar gene complement [14], whereas in the Chlorophyceae the two more similar mtDNAs (Chlorogonium and G. eugametos) share 3 clusters (8 genes). The Helicosporidium mtDNA shares six clusters (13 genes) and five clusters (11 genes) with the prasinophycean mtDNAs of Ostreococcus and Nephroselmis, respectively, and none with the ulvophycean or chlorophycean mitochondrial genomes.

Given the level of synteny between the Helicosporidium and Prototheca mitochondrial genomes, a minimum of 24 permutations by inversion between the 60 genes they share would be sufficient to convert the structure of one genome into that of the other. In contrast, at least 30 permutations (63 genes shared) and 46 permutations (50 genes shared) would be required to interconvert the structure of the Nephroselmis and Ostreococcus mtDNAs and of the Pseudendoclonium and Oltmannsiellopsis mtDNAs, respectively. However this number does not account for the creation of the inverted repeats in Ostreococcus due to the limitations of the GRIMM algorithm and is therefore an underestimate. In the Chlorophyceae, the gene-poor mtDNAs of Chlorogonium and G. eugametos are more closely related to each other (12 genes shared, 7 permutations) than to that of C. reinhardtii (12 genes shared, 19 and 18 permutations, respectively). The 42-gene mtDNA of Scenedesmus was not compared to the other gene-poor mtDNAs from the Chlorophyceae.

Introns

The Helicosporidium mitochondrial genome contains a total of four group I introns inserted into the carx and rnl genes (Figures 2 and S1). Although Prototheca mtDNA also has five introns within these two genes, none are located at cognate sites within Helicosporidium mtDNA. Interestingly, the Hsp. car1.1 intron (Figure 2A) contains two distinct open reading frames, orf166 and orf239, located in different variable loops (L4 and L5, respectively). Both ORFs display a single dodecapeptide LAGLIDADG motif, indicative of a putative endonuclease function for these proteins. However, functional LAGLIDADG endonucleases contain two dodecapeptide motifs [15], raising the interesting possibility that the two ORFs generate an heterodimer constituted of one product in the N-terminal domain and of the other product in the C-terminal portion of the endonuclease. Alternatively, the two ORFs may also code for two independent homodimeric endonucleases.

Homodimeric LAGLIDADG endonucleases are commonly found in group I introns, although the presence of two endonucleases in a single intron is extremely rare.

Perhaps the most surprising finding is that the Helicosporidium mtDNA contains a trans-spliced group I intron. The Hsp.carx.1 intron (Figure 2B) is fragmented into two pieces that are located on different strands and separated by three genes (trnKucu), nad5, rps19) spanning over 3 kb. This fragmentation is a genuine feature of the genome, supported by an assembly in which the mean coverage is about 400X. Because the carx gene is conserved among...
all chlorophyte mtDNAs and because its product, the first subunit of the mitochondrial cytochrome oxidase, is likely essential to the mitochondrion, we expected the trans-splicing of this intron to occur at the mRNA level so that a functional protein could be produced. To test this, we performed RT-PCRs on Helicosporidium RNA treated with DNase using primers specific for the second and third cox1 exons. As expected, we obtained a 343 bp fragment consisting of the spliced exons (Figure 3). Sequencing of the 343 bp amplicon confirmed the 5' and 3' exon/intron splice junctions of the Hsp.cox1.2 intron illustrated in Figure 2B. When DNA was used as template for PCRs with the same primers, no amplicon was observed.

To our knowledge, trans-spliced group I introns were never reported until this year, and presently only three other instances are known: all in the mitochondrial cox1 gene. The first of these is in the mtDNA of the lycophyte Isoetes engelmannii [16] while the other two were identified by reinvestigation of the Trichoplax adhaerens mitochondrial genome [17]. However, the Hsp.cox1.2 intron is not inserted at a cognate site with any of these trans-spliced introns (Figure S2). Also, aside from its canonical group I intron structure, the Hsp.cox1.2 intron does not display strong identity with the Isoetes and Trichoplax introns even though it does share the same L8 trans-splicing location as the two Trichoplax introns.

Repeated Elements

The Helicosporidium mtDNA is not rich in repeated elements. Most of the repeated nucleotide strings detectable in the Helicosporidium mitochondrial genome are confined to AT-rich intergenic regions, are short, and are composed of adenosine and/
or thymine residues arranged either in stretches or as alternating bases. The distribution of repeated elements observed in Helicosporidium mtDNA (Figure S3) parallels that of Prototheca mtDNA in which the presence of A+T-rich repeats arrayed in tandem has been previously reported [18]. As in Prototheca mtDNA, the repeated elements are dispersed throughout the whole genome sequence in intergenic regions and in introns.

### Table 1. Main features of *Helicosporidium* and other chlorophyte mtDNAs.

| Chlorophyte mtDNA | Size (bp) | A+T content (%) | Gene content | Gene density | Coding seq. (%) | Introns (I/II) | Intron ORFs (I/II) |
|-------------------|-----------|-----------------|--------------|--------------|----------------|---------------|-----------------|
| **Prasinophyceae** |           |                 |              |              |                |               |                 |
| Nephroselmis      | 45223     | 67.2            | 65           | 1/696        | 80.6           | 4/0           | 4/0             |
| Ostreococcus      | 44237     | 61.8            | 65           | 1/590        | 92.1           | 0/0           | 0/0             |
| **Trebuolixyphaceae** |         |                 |              |              |                |               |                 |
| Helicosporidium   | 49343     | 74.4            | 60           | 1/822        | 75.9           | 3/0           |                 |
| Prototheca        | 55328     | 74.2            | 61           | 1/907        | 70.6           | 5/0           | 2/0             |
| **Ulvophyceae**   |           |                 |              |              |                |               |                 |
| Oltmannsiellopsis | 56761     | 66.6            | 54           | 1/1051       | 68.7           | 2/1           | 2/1             |
| Pseudendoclonium  | 95880     | 60.7            | 57           | 1/1682       | 58.7           | 7/0           | 6/0             |
| **Chlorophyceae** |           |                 |              |              |                |               |                 |
| C. eugametos      | 22897     | 65.4            | 12           | 1/1908       | 84.6           | 9/0           | 7/0             |
| C. reinhardtii    | 15758     | 54.8            | 12           | 1/1313       | 83.1           | 0/0           | 0/0             |
| Chlorogonium      | 22704     | 62.2            | 12           | 1/1892       | 89.1           | 6/0           | 6/0             |
| Scenedesmus       | 42919     | 63.7            | 42           | 1/1022       | 60.6           | 2/2           | 1/0             |
| **Uncertain affiliation** | | | | | | | |
| Pedinomonas       | 25137     | 77.8            | 22           | 1/1143       | 60.5           | 0/1           | 0/0             |

*Duplicated genes, unique ORFs and intron ORFs were not taken into account.*

*Duplicated genes were taken into account (size/number of genes).*

*Conserved genes (unique and duplicated), ORFs, introns and introns ORFs were considered as coding sequences.*

doi:10.1371/journal.pone.0008954.t001

Phylogeny

The availability of the *Helicosporidium* mitochondrial genome provides us with another opportunity to probe the phylogenetic position of the Helicosporidia, and a useful opportunity because all major subgroups of green algae are available for analysis (this is not true for many nuclear genes, and the plastid genome does not contribute substantially to the question since the Prototheca plastid...
The mitochondrial genome of the obligately parasitic green alga *Helicosporidium* stands out in two different ways. First, it strongly supports the relationship between *Helicosporidium* and *Prototheca*, but not just because it provides a large molecular data set from which phylogenies can be inferred, but also because of their shared genomic structure. Based on the low levels of gene order conservation in other green algal mitochondrial genomes, we might expect to see few blocks of conservation between *Helicosporidium* and *Prototheca*. Clearly this was not the case for *Helicosporidium* and *Prototheca*, because their mtDNAs display a surprisingly high level of similarity in form. This close resemblance is probably best explained by a recent split between these two species. The only chlorophyte mtDNAs that display a comparable level of similarity are those of the prasinophytes *Nephroselmis* and *Ostreococcus*, but even here the level of conservation is much lower, with one featuring an inverted repeat that is missing from the other. In the Chlorophyceae, the gene-poor mtDNAs of *Chlorononium* and *C. eugametos* also display an appreciable level of synteny, with 8 of their 12 genes (66%) being located in shared clusters, although this percentage is still lower than that observed between *Helicosporidium* and *Prototheca* mtDNAs (75%), and there are far fewer combinations of 12 genes than of 60.

A second standout feature of the *Helicosporidium* mtDNA is in its introns, and in particular the presence of a group I intron that splices in trans. Although trans-splicing in various group II introns has been known to occur for some time (reviewed in [19,20]), the first examples of trans-spliced group I introns have only been described recently [16,17]. Like other known trans-spliced group I introns, the predicted secondary structure of the *Hsp.cox1.2* intron (Figure 2) closely conforms to a canonical group I intron, suggesting it most likely arose from a cis-spliced group I intron that was broken into two pieces, but that could still fold at the mRNA level to produce a functional ribozyme. Because this fragmentation occurred within a variable loop, its effect on the intron self-splicing capability may have been minimal (although presumably if such an event had little impact it would occur more frequently than it does). It is unclear what effect such a fragmentation might have on the viability of the intron if it

---

**Figure 3.** Electrophoretic analysis of RT-PCRs performed on *Helicosporidium* total RNA. The amplicon in lane A corresponds to the expected size (343 bp) between the two internal *cox1* primers in exons 2 and 3 (*Hecox1F & Hecox1R*) after trans-splicing of the *Hsp.cox1.2* group I intron. The amplicon in lane B corresponds to the in cis positive control (166 bp) performed with internal *atp1* primers (07487R & 00085R).

doi:10.1371/journal.pone.0008954.g003

---

**Figure 4.** Phylogenetic position of *Helicosporidium* sp. ATCC 50920 as inferred from amino acid sequences derived from the seven protein-encoding genes that are shared between all sequenced chlorophyte mtDNAs. The best ML-tree inferred with PHYML 3.0 under the LG+4+F+I model of amino acid substitutions is shown here (13 taxa, 2362 positions, 1373 phylogenetically informative). The charophyte green algae *Mesostigma viride* and *Chaetosphaeridium globosum* were used as outgroups. Bootstrap values over 80% are shown above the corresponding nodes. Branch lengths are drawn to scale. *Pedinomonas minor* has not yet been assigned unambiguously to one of the four classes of chlorophyte green algae.

doi:10.1371/journal.pone.0008954.g004
occurred in core regions like the P7 pairings. However, as ribozymes derived from group I introns can catalyse trans-excision-splicing reactions in other RNA molecules [21,22,23], their functional core may be somewhat malleable.

The four trans-spliced group I introns known so far most likely arose independently. Not only do they appear dissimilar at the nucleotide level outside of their canonical group I intron structure, but despite their shared location within the mitochondrial coxl gene, none are inserted at cognate sites. Also to be transferred horizontally from one organism to another, at least two recombinational events would be required, one for each trans-spliced segment. This appears unlikely, especially considering that such recombinational events would likely involve the adjacent exons. As Helicosporidium, Trichoplax and Isoetes are evolutionary distant and belong to very different lineages, recombination between their genes, even as conserved as coxl, is not a very compelling hypothesis. It is perhaps not surprising that these rare introns were first discovered within the coxl gene, given its conservation and its importance for the mitochondrion. Other existing instances of trans-splicing group I introns in less conserved genes may have been overlooked, and reinvestigation of sequenced mitochondrial genomes, as performed by Burger and coauthors on the Trichoplax mtDNA [17], may reveal more of these segmented yet functional ribozymes.

The homing endonucleases encoded within intronic ORFs confer mobility to the intron host by permitting double strand breaks of a target DNA. Very often, these endonucleases are lost yet functional ribozymes.

Conclusions

The structure and content of the Helicosporidium mitochondrial genome, as well as the phylogenetic inferences derived from the sequences it encodes, support the specific relationship to the genus Prototheca. The introns of this genome also have a number of interesting characteristics rarely seen in other organelle genomes. Our results, combined with the previously published plastid genome sequence of Helicosporidium sp. ATCC 50920 [21], complete the deciphering of this peculiar species’ organellar genetic imprint. The sequencing of the Helicosporidium sp. ATCC 50920 nuclear genome would provide us with a global picture that, hopefully, would yield clues into the adaption of this alga from a free-living entity to that of an entomoparasite. Also, a comparative approach with its protocellian relatives would give us interesting insights into the nature of their selective parasitism.

Materials and Methods

PCR amplifications and RT-PCRs

PCR amplifications were performed using 22- to 24-mers and the EconoTaq PLUS GREEN kit from Lucigen (Middleton, WI, USA) with 35 cycles of denaturation (1 min at 94°C), annealing (1 min at 55°C) and elongation (3 min at 72°C). RT-PCRs were performed using the SuperScript One-Step RT-PCR with Platinum Tag kit from Invitrogen (Carlsbad, CA, USA) with an initial cDNA synthesis cycle (30 min at 50°C) followed by a 2 min denaturation cycle at 94°C. A total of 35 amplification cycles (15 sec at 94°C), annealing (30 sec at 53°C), and elongation (1 min at 72°C) were then performed. The Hecox1F (5’-CTGTTCCTGTAT-TAGCTGGTTG-3’) and Hecox1R (5’-GAATATCGATTG-TAGCTGGCAG-3’) and the 00085R (5’-GTTTGCAATAGTGTGG-CTTACAG-3’) primers were used in RT-PCRs.

Genome Sequencing

The Helicosporidium mtDNA was sequenced using the massively parallel GS-FLX DNA pyrosequencing platform from Roche 454 Life Sciences ( Branford, CT, USA).

The creation of the Helicosporidium mtDNA GS-FLX shotgun library and the GS-FLX 434 pyrosequencing (using the GS-FLX Titanium reagents) were carried out by the McGill University and Génome Québec Innovation Centre. The Newbler assemblies obtained from Génome Québec were converted to, edited, and assembled with CONSED 19 [24]. Ambiguous regions in the assemblies were either (1) edited according to their conceptual translations or (2) amplified by PCR with 22-mers primers flanking the ambiguous regions, sequenced using traditional Sanger chemistry by Macrogen (Seoul, Korea) and then edited according to Sanger base calling.

Genome Annotation and Analysis

Genes were identified by Blast homology searches [25] against a local copy of the National Center for Biotechnology Information (NCBI) nonredundant database using the NCBI BLASTALL suite (http://www.ncbi.nlm.nih.gov/Ftp/blast). Positions of open reading frames and protein-coding genes were determined using GETORF from EMBOSs 6.0.1 [26] and ORFFINDER at NCBI, whereas positions of tRNA-encoding genes were determined with tRNAscan-SE [27]. Insertions sites of group I introns and their predicted secondary structures were determined manually. Codon usage in protein-coding genes was determined with CUSP from the EMBOSS package. Repeated elements were first visualized with PipMaker [28]. Then, repeated elements aligned in tandem were identified with ETANDEM from the EMBOSS package whereas dispersed repeated elements were located with REPuter 2.74 [29]. Potential hairpin structures were screened for with PALINDROME from the EMBOSS package. Minimal number of permutations by inversions between mitochondrial genomes were inferred with GRIMM [30]. For this analysis, the trans-spliced exons of the coxl gene in Helicosporidium mtDNA and the fragmented rRNA genes in chlorophycean mtDNAs were coded as distinct fragments. Also, as GRIMM cannot handle duplicate genes, one copy of the Ostreococcus inverted repeats was removed.

Phylogenetic Analyses

In addition to the Helicosporidium mitochondrial genome sequenced by the authors [GenBank:GQ339575], the following mtDNAs used in this study were retrieved from GenBank: Chaeotphaeridium globosum [GenBank:NC_001118], Chlamydomonas eugametos [GenBank:NC_001872], Chlamydomonas reinhardtii [GenBank:NC_001638], Chlamydomonales elongatum [GenBank:Y13643, Y13644, Y07814], Mesostigma viride [GenBank:NC_00240], Neurospora crassa [GenBank:NC_002839], Oomycetes sp. [GenBank:NC_00256], Ostreococcus tauri [GenBank:NC_00290], Pedinomonas minor [GenBank:NC_000892], Protothea wickerhamii [GenBank:NC_001613], Pseudendoclonium akinatum [GenBank:NC_005926], Scedesmus obliquus [GenBank:NC_002254]. Mitochondrial protein sequences were inferred from the conceptual translation of the seven protein-encoding genes that are share among all chlorophyte mtDNAs. The amino acid sequences were
aligned using T-COFFEE 7.81 [31], the ambiguous regions within these alignments filtered with GBLOCKS 0.91 b [32], and the filtered individual sequences concatenated. Maximum Likelihood computations were performed using PHYML 3.0 [33] under the LG+Γ+F+I model of amino acid substitution selected with ProtTest 2.0 [34]. Bayesian inferences were performed with PhyloBayes 3.2 [35] under the CAT+ F4 model of amino acid substitution running two concurrent chains terminated using PhyloBayes automatic stopping rule (maxdiff <0.3).

Supporting Information

**Figure S1** Predicted secondary structures of Helicosporidium mt group I introns. The *ml* group I introns displayed according to Burke *et al* were classified according to Michel and Westhof. Splice sites between exon and intron residues are denoted by arrows. Canonical Watson-Crick base pairings are denoted by dashes whereas guanine-uracil pairings are marked by dots. Numbers inside variable loops indicate the sizes of these loops. The putative LAGLIDADG endonuclease encoded within the Hsp.ml.1 intron ORF contains a single copy of this motif. 

*Found at: doi:10.1371/journal.pone.0008954.s001 (0.91 MB EPS)*

**Figure S2** Trans-spliced GI introns insertion sites. The insertion sites of the *Helicosporidium*, *Isoetes* and *Trichoplax* trans-spliced group I introns are indicated by arrows on the Cox1 amino acid alignment (positions 225 to 464 shown). Numbers on the right of the alignment indicate the amino acid positions on the corresponding sequences. The different trans-spliced introns are indicated by roman numerals: I, *Helicosporidium* (Hsp.cox1 2); II, *Trichoplax* (cox1 intron 4); III, *Trichoplax* (cox1 intron 5); IV, *Isoetes* (cox1 305). The *Trichoplax* intron insertion sites were taken from Burger *et al.*

*Found at: doi:10.1371/journal.pone.0008954.s002 (4.97 MB EPS)*

**Figure S3** Densities of repeated elements in *Helicosporidium* and other chlorophyte mitochondrial genomes. Repeated elements identified with REPuter are connected by lines on the corresponding mtDNA circular representations (adapted from Kurtz). Repeats of at least 15, 30 and 45 nt are shown on the left, middle and right panels respectively. For this analysis, one copy of the *Ostreococcus* mtDNA inverted repeat has been removed. Also, the corresponding mtDNA circular representations (adapted from Kurtz).

*Found at: doi:10.1371/journal.pone.0008954.s003 (2.22 MB PDF)*

**Table S1** Gene repertoires of *Helicosporidium* and other chlorophyte mtDNAs. *Nol*, *Nephroselmis olivacea*; *Ota*, *Ostreococcus tauri*; *Hsp*, *Helicosporidium sp.*; *Pwi*, *Prototexa wickerhamii*; *Ovi*, *Olmannsielopsis viridis*; *Pak*, *Pseudodinobolium akinetum*; *Sob*, *Sedeneusmus obliquus*; *Pmi*, *Pedinomonas minor*; *Cre*, *Chlamydomonas reinhardtii*; *Ceu*, *Chlamydomonas egametosa*; *Cel*, *Chlorogonium elongatum*. Presence/absence of a gene is denoted by +/−. * Gene fragments in corresponding mtDNA.

*Found at: doi:10.1371/journal.pone.0008954.s004 (0.10 MB DOC)*

**Table S2** tRNA gene repertoires of *Helicosporidium* and other chlorophyte mtDNAs. *Nol*, *Nephroselmis olivacea*; *Ota*, *Ostreococcus tauri*; *Hsp*, *Helicosporidium sp.*; *Pwi*, *Prototexa wickerhamii*; *Ovi*, *Olmannsielopsis viridis*; *Pak*, *Pseudodinobolium akinetum*; *Sob*, *Sedeneusmus obliquus*; *Pmi*, *Pedinomonas minor*; *Cre*, *Chlamydomonas reinhardtii*; *Ceu*, *Chlamydomonas egametosa*; *Cel*, *Chlorogonium elongatum*. Presence/absence of a gene is denoted by +/−. * b Anticodon of *Helicosporidium* mtDNA-encoded tRNA recognizing the corresponding codon. The following tRNAs with an uracil in the first position of their anticodon are assumed to deduce all four members of the four-codon families: alanine, GCN; glycine, GGN; leucine, CUN; proline, CCN; serine, UCN; threonine, ACN; valine, GUN. *c* Amino acids are labelled by their one-letter IUPAC code. Termination codons are indicated by asterisks. *d* In chlorophytes mtDNAs, the gene coding for tRNA^Trn^ (ugu) has been found only within the mitochondrial genomes of *Helicosporidium*, *Prototheca* and *Pseudodinobolium*. *e* The initiator and elongator tRNA^Met^ (cau) are encoded by different genes.

*Found at: doi:10.1371/journal.pone.0008954.s005 (0.09 MB DOC)*

**Table S3** Codon usage in the 32 protein-encoding genes of *Helicosporidium* mtDNA. *a* Percentage of each amino acid specified by the specified codon. *b* Anticodon of *Helicosporidium* mtDNA-encoded tRNA recognizing the corresponding codon. The following tRNAs with an uracil in the first position of their anticodon are assumed to deduce all four members of the four-codon families: alanine, GCN; glycine, GGN; leucine, CUN; proline, CCN; serine, UCN; threonine, ACN; valine, GUN. *c* Amino acids are labelled by their one-letter IUPAC code. Termination codons are indicated by asterisks. *d* In chlorophytes mtDNAs, the gene coding for tRNA^Trn^ (ugu) has been found only within the mitochondrial genomes of *Helicosporidium*, *Prototheca* and *Pseudodinobolium*.

*Found at: doi:10.1371/journal.pone.0008954.s006 (0.07 MB DOC)*

**Acknowledgments**

We thank our collaborator on the *Helicosporidium* genome project, Dr Drion Boucias from the Department of Entomology and Nematology at the University of Florida, for providing *Helicosporidium* sp. ATCC 50920 total DNA and RNA.

**Author Contributions**

Conceived and designed the experiments: PJK. Performed the experiments: JFP. Analyzed the data: JFP. Wrote the paper: JFP. Contributed insight into data interpretation: PJK. Helped draft the manuscript: PJK.

References

1. Cooklin T, Bla¨ske-Lietze VU, Becnel JJ, Boucias DG (2005) Infectivity of two isolates of Helicosporidium spp. (Chlorophyta: Trebouxiophyceae) in heterologous host insects. Fla Entomol 88: 431–439.

2. Bla¨ske-Lietze VU, Shapiro AM, Denton JS, Botts M, Becnel JJ, et al. (2006) Development of the insect pathogenic alga Helicosporidium. J Eukaryot Microbiol 53: 165–176.

3. Boucias DG, Becnel JJ, White SE, Bots M (2001) In vivo and in vitro development of the protist *Helicosporidium* sp. J Eukaryot Microbiol 48: 460–470.

4. Bla¨ske-Lietze VU, Boucias DG (2005) Pathogenesis of *Helicosporidium* sp. (Chlorophyta: Trebouxiophyceae) in susceptible noctuid larvae. J Invertebr Pathol 80: 161–168.

5. Ke¨lin D (1921) On the life history of Helicosporidium pansuicum, n. g., n. sp., a new type of protist parasitic in the larva of Dasyholus obscura Winn. (Diptera, Ceratopogonidae) and in some other arthropods. Parasitology 13: 97–113.

6. Tartar A, Boucias DG, Adams BJ, Becnel JJ (2002) Phylogenetic analysis identifies the invertebrate pathogen *Helicosporidium* sp. as a green alga (Chlorophyta). Int J Syst Evol Microbiol 52: 273–279.

7. Tartar A, Boucias DG, Becnel JJ, Adams BJ (2003) Comparison of plastid 16S tRNA (rrn16) genes from Helicosporidium spp.: evidence supporting the reclassification of Helicosporidium as green algae (Chlorophyta). Int J Syst Evol Microbiol 53: 1719–1723.

8. de Koning AP, Keeling PJ (2004) Nucleus-encoded genes for plastid-targeted proteins in Helicosporidium: functional diversity of a cryptic plastid in a parasitic alga. Eukaryot Cell 3: 1198–1205.

9. de Koning AP, Tartar A, Boucias DG, Keeling PJ (2005) Expressed sequence tag (EST) survey of the highly adapted green algal parasite, Helicosporidium. Protist 156: 181–190.

10. Tartar A, Boucias DG (2004) The non-photosynthetic, pathogenic green alga *Helicosporidium* sp. has retained a modified, functional plastid genome. FEMS Microbiol Lett 233: 153–157.

11. de Koning AP, Keeling PJ (2006) The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. BMC Biol 4: 12.

12. Keeling PJ, Inagaki Y (2004) A class of eukaryotic GTPase with a punctate factor 1alpha. Proc Natl Acad Sci U S A 101: 15380–15385.
13. Lass-Florl C, Mayr A (2007) Human protothecosis. Clin Microbiol Rev 20: 230–242.
14. Pombert JF, Beauchamp P, Otis C, Lemieux C, Turmel M (2006) The complete mitochondrial DNA sequence of the green alga Oltmannsiellopsis varidis: evolutionary trends of the mitochondrial genome in the Ulvophyceae. Curr Genet 50: 137–147.
15. Silva GH, Belfort M, Wende W, Pingoud A (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. J Mol Biol 361: 744–754.
16. Grewe F, Viehoever P, Weisshaar B, Knoop V (2009) A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. Nucleic Acids Res 37: 5093–5104.
17. Burger G, Yan Y, Javadi P, Lang BF (2009) Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. Trends Genet 25: 381–386.
18. Wolff G, Plante I, Lang BF, Kuck U, Burger G (1994) Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. Gene content and genome organization. J Mol Biol 237: 75–86.
19. Bonen L (2008) Cis- and trans-splicing of group II introns in plant mitochondria. Mitochondrion 8: 26–34.
20. Lambowitz AM, Zimmerly S (2004) Mobile group II introns. Annu Rev Genet 38: 1–35.
21. Dotson PP 2nd, Johnson AK, Testa SM (2008) Tetrahymena thermophila and Candida albicans group I intron-derived ribozymes can catalyze the trans-excision-splicing reaction. Nucleic Acids Res 36: 3281–3289.
22. Dotson PP 2nd, Sinha J, Testa SM (2008) A Pneumocystis carinii group I intron-derived ribozyme utilizes an endogenous guanosine as the first reaction step nucleophile in the trans-excision-splicing reaction. Biochemistry 47: 4780–4787.
23. Einvik C, Fiskaa T, Lundblad EW, Johansen S (2004) Optimization and application of the group I ribozyme trans-splicing reaction. Methods Mol Biol 252: 359–371.
24. Gordon D (2004) Viewing and Editing Assembled Sequences Using Consed. In: Baxevanis AD, Davison DB, eds (2004) Current Protocols in Bioinformatics. New York: John Wiley & Co. pp 11.12.11–11.12.43.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
26. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
27. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 953–964.
28. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, et al. (2000) PipMaker: a web server for aligning two genomic DNA sequences. Genome Res 10: 577–586.
29. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher G, Stoye J, et al. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29: 4635–4642.
30. Teder G (2002) GRIMM: genome rearrangements web server. Bioinformatics 18: 492–493.
31. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205–217.
32. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.
33. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
34. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
35. Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneous in the amino-acid replacement process. Mol Biol Evol 21: 1169–1189.
36. Burke JM, Belfort M, Cech TR, Davies RW, Schoyen RJ, et al. (1997) Structural conventions for group I introns. Nucleic Acids Res 15: 7217–7221.
37. Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216: 585–610.