Protist, Vol. 156, 181—190, August 2005
http://www.elsevier.de/protis
Published online date 11 July 2005

Protist

ORIGINAL PAPER

Expressed Sequence Tag (EST) Survey of the Highly Adapted Green Algal Parasite, *Helicosporidium*

Audrey P. de Koning[a], Aurélien Tartar[b,2], Drion G. Boucias[b], and Patrick J. Keeling[a,1]

[a]Botany Department, Canadian Institute for Advanced Research, University of British Columbia, 3529-6270 University Boulevard, Vancouver, Canada BC, V6T 1Z4
[b]Entomology and Nematology Department, University of Florida, PO Box 110620, Gainesville, FL 32611-0620, USA

Submitted November 12, 2004; Accepted February 6, 2005
Monitoring Editor: Michael Melkonian

Helicosporidia are obligate invertebrate pathogens with a unique and highly adapted mode of infection. The evolutionary history of Helicosporidia has been uncertain, but several recent molecular phylogenetic studies have shown an unexpectedly close relationship to green algae, and specifically to the opportunistic pathogen *Prototheca*. To date, molecular sequences from Helicosporidia are restricted to those genes used for phylogenetic reconstruction and genes related to the existence and function of its cryptic plastid. We have therefore conducted a small expressed sequence tag (EST) project on *Helicosporidium* sp., yielding about 700 unique sequences. We have examined the functional distribution of known genes, the distribution of EST abundance, and the prevalence of previously unknown gene sequences. To demonstrate the potential utility of large amounts of data, we have used ribosomal proteins to test whether the phylogenetic position of *Helicosporidium* inferred from a small number of genes is broadly supported by a large number of genes. We conducted phylogenetic analyses on 69 ribosomal proteins and found that 98% supported the green algal origin of Helicosporidia and 80% support a specific relationship with *Prototheca*. Overall, these data multiply the available molecular information from *Helicosporidium* 100-fold, which should provide the basis for new insights into these unusual but interesting parasites.

Key words: green algae; expressed sequence tag; helicosporidia; parasite; phylogeny; ribosomal protein; entomopathogen.

Introduction

The Helicosporidia are an enigmatic group of obligate pathogens that are found in a variety of insect hosts. The defining feature of the group is the infective cyst stage. Cysts consist of a pellicle, surrounding three internal and relatively undifferentiated cells, around which is wrapped a long

[1]Corresponding author;
fax 1 604 822 6089
e-mail pkeeling@interchange.ubc.ca (P.J. Keeling)
[2]Present address: Biology Department, Clark University, 950 Main Street, Worcester, MA 01610, USA

doi:10.1016/j.protis.2005.02.005

and highly differentiated helical cell with tapered and barbed ends. When the cysts burst in the gut lumen of insects, the helical cell is expelled. It pierces the gut epithelial cells and sticks, apparently aided by barbed ends. The helical cell then migrates through the epithelium wall and into the host hemolymph where it differentiates into the vegetative stage (Boucias et al. 2001). The vegetative stage undergoes a 2—4 cell asporogenic division with both cell division and daughter cell wall formation occurring within the mother cell. By 7—12 days post-infection vegetative cells completely fill the hemocoel and will begin to differentiate into the cyst form. It is likely that the cysts are released into the environment upon the death of the host or possibly transovum transmitted by infected females (Bläske and Boucias 2004).

The evolutionary origin of Helicosporidia has been unclear since their first discovery. The parasites are highly adapted and do not closely resemble any other group of eukaryotes, and they have therefore been excluded from most large taxonomic schemes of eukaryotes. When they are considered, they have been placed in various phylogenetic positions at different times, including protozoans (Keilin 1921; Lindegren and Hoffmann 1976) and fungi (Kellen and Lindegren 1973; Weiser 1970). Recently, the study of Helicosporidia has been greatly advanced by the ability to grow *Helicosporidium* sp. axenically in vitro. This has allowed the parasite to be purified in substantial quantities, and led to the first molecular data from Helicosporidia. These data provided a very different and unexpected picture of the evolution of this group: analysis of nuclear small subunit ribosomal RNA (SSU rRNA), actin and beta-tubulin all showed *Helicosporidium* to be a member of the green algae, and specifically related to the opportunistic parasite *Prototheca* (Tartar et al. 2002), which is consistent with some features of the vegetative stage of *Helicosporidium* (Boucias et al. 2001). The green algal origin of *Helicosporidium* has subsequently been supported by phylogenies based on plastid SSU rRNA (Tartar et al. 2003), elongation factor Tu (Tartar and Boucias 2004), the EFL protein (Keeling and Inagaki 2004), and numerous plastid-targeted proteins (de Koning and Keeling 2004).

We have conducted an expressed sequence tag (EST) project on the blackfly (*Simulium jonesii*) isolate of *Helicosporidium* sp. to accelerate gene discovery in this enigmatic group of pathogens. Prior to the initiation of this survey, the only molecular data from the group were nuclear and plastid SSU rRNA, actin, beta-tubulin, and a fragment of the plastid genome. We sequenced approximately 1,400 cDNA clones resulting in approximately 700 unique sequences, increasing the available molecular data from the group approximately 100-fold. The survey has already led to descriptions of the *Helicosporidium* EFL homologue (Keeling and Inagaki 2004) and 20 cDNAs for plastid-targeted proteins (de Koning and Keeling 2004). To further demonstrate the potential utility of these data, we have reanalyzed the phylogenetic position of *Helicosporidium* using the ribosomal proteins. We identified 69 ribosomal proteins and conducted phylogenetic analyses on each protein individually. Phylogenetic evidence for the green algal origin of *Helicosporidium* and its close relationship to *Prototheca* is currently based on a small number of genes, each of which provides strong support. The analyses of ribosomal proteins and the ESTs in general complement these data by providing broad and consistent support for this conclusion based on results from a large number of genes.

Results and Discussion

EST Sequencing

An axenic culture of vegetative *Helicosporidium* cells was harvested, and a directionally cloned cDNA library was constructed from poly A purified mRNA. From a mass excision, 1,536 clones were isolated and sequenced from the 5' end, resulting in 1,432 readable sequences and 1,188 sequences passing quality checks and vector-trimming. These 1,188 sequences were assembled into clusters of homologous sequences resulting in 700 clusters. Six clusters were found to correspond to *Escherichia. coli* contaminant (all six were represented by a single EST), which we interpret to have arisen during the library construction since this *Helicosporidium* sp. was cultured in the presence of gentamycin and no contaminant was observed by microscopy prior to RNA isolation. These clusters were removed and are not considered further, leaving 694 unique clusters.

The method used to generate clones for EST sequencing is not strictly quantitative, but the representation of ESTs in samples such as this bears some relationship to the expression levels of genes in a general sense. Certain genes can be over- or underrepresented, but the overall trends do convey some sense of expression patterns.

This may be interpreted with some caution, but the information is also important to gauge the effectiveness of the sampling for gene discovery. From the distribution of the number of ESTs in each cluster (not shown), it is apparent that the majority of ESTs are present in low copy number, while a relatively small number of sequences are over-represented. Almost three-quarters of the clusters were represented by a single EST, while the most highly represented cluster consisted of 49 ESTs, 2.5 times that of the next largest cluster of 20. Overall, the distribution of EST number per cluster shows that the approach was very favorable for gene discovery, since the ultimate rate of new gene discovery was one unique sequence per two ESTs. Although the abundance of ESTs does not correspond exactly to expression levels, the large variation in EST number likely does correspond to high-expression levels of the most extreme outliers.

Distribution of Functional Classes of Expressed Genes

The consensus sequences of the 694 unique clusters were compared with public sequence and motif databases using PEPdb (http://megasun.bch.umontreal.ca/pepdb/pepdb.html) to identify potentially homologous sequences (Fig. 1). A large proportion of clusters (299 or 43%) were not detectably similar to any known protein, and a further 5 clusters were only found to contain similarity to known domains. An additional 57 clusters were found to be similar to hypothetical proteins in other organisms: 8 and 16 being hypothetical proteins with and without known domains respectively, and another 33 being similar only to other ESTs from other organisms. Six clusters corresponded to fragments of the ribosomal RNA operon (individual inspection showed these to be non-overlapping fragments

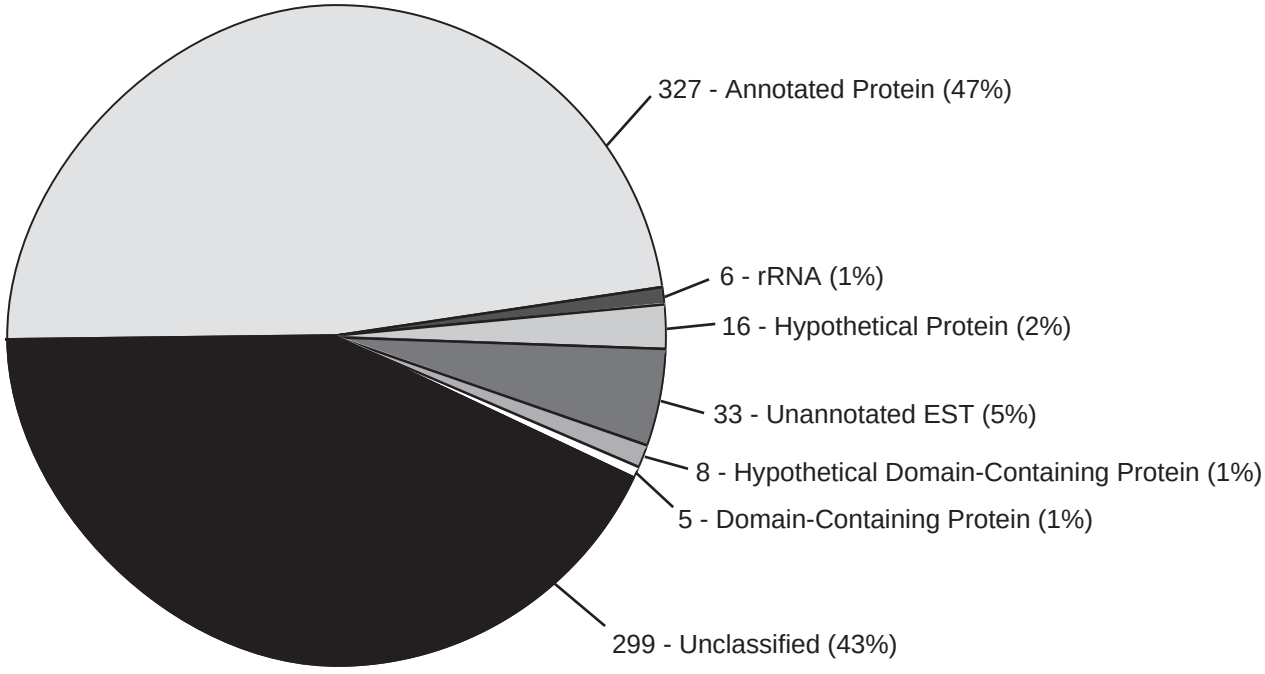

Figure 1. Distribution of clusters by similarity to known genes according to the annotation protocol. For each fraction, the label includes the number of clusters (left), the designation of the cluster (centre) and the percent of clusters in this fraction (left). Designations are, clockwise from the top: "Annotated Protein" clusters are similar to sequences of annotated genes in public databases. "rRNA" clusters are homologous to ribosomal RNA. "Hypothetical Protein" clusters are similar to proteins of unknown function in other organisms. "Unannotated EST" clusters are similar to EST sequences of unknown function from other organisms in public databases. "Hypothetical Domain-Containing" clusters are similar to hypothetical proteins, but also contain a known functional domain. "Domain-Containing" clusters contain known functional domains but are otherwise not similar to any known protein. Lastly, "Unclassified" clusters are those with no significant similarity to any sequence is extant databases.

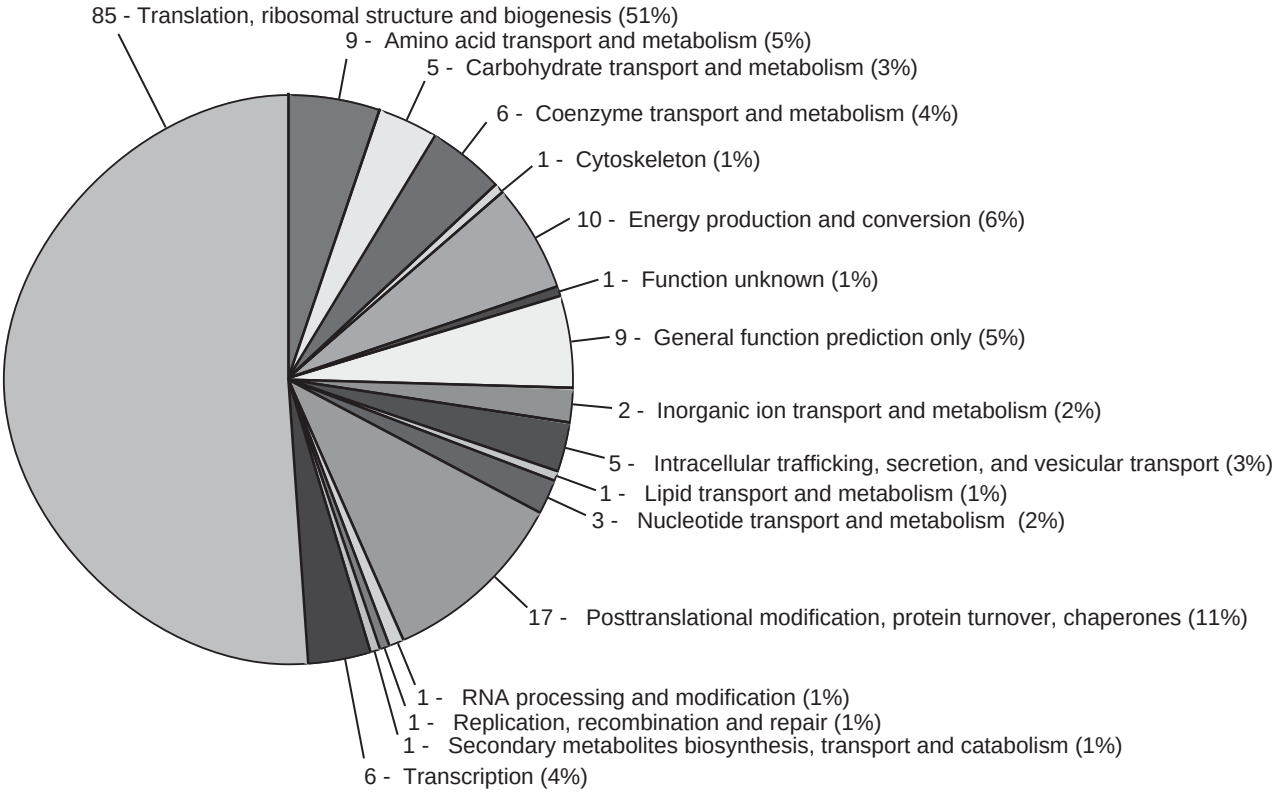

Figure 2. Distribution of 163 annotated protein clusters to functional groups by COG categories. For each fraction, the labels include the number of clusters (left), the COG category name (centre) and the percent of clusters in this fraction (left).

corresponding to the SSU, ITS region, and LSU). The remaining 327 clusters (47%) were predicted to be known genes according to the annotation protocol.

The functional distribution of the 327 clusters matching annotated proteins was examined by classifying clusters by COG (clusters of orthologous genes: (Tatusov et al. 2003)) categories (Fig. 2). Over half of the 163 clusters that may be classified into COG categories are related to translation: many of these encode ribosomal proteins, which represent the single largest class of genes found (see below). The next largest class of proteins (at 11%) is related to protein modification and turnover, followed by energy production (6%), amino acid metabolism (5%) and coenzyme metabolism (4%). Many other categories are represented by a single cluster.

Novel Genes in *Helicosporidium*

At 43%, the proportion of *Helicosporidium* clusters that were not detectably similar to any known gene is not outside the range typically expected when a large sample is acquired from a eukaryotic genome, but there are a few interesting characteristics of this class of genes that deserve note. Helicosporidia are obligate parasites with a highly specialized infection strategy, but they evolved from a free-living, photosynthetic green alga. One aspect of this transition that has been studied to date is the fate of its plastid. There is now evidence for a plastid genome and a variety of proteins targeted to the organelle (although the organelle itself has yet to be visualized), suggesting very strongly that it has been retained in a cryptic form for metabolic pathways other than photosynthesis. Other significant adaptations to parasitism in the Helicosporidia likely involve many of the genes of unknown function, so it is interesting to note that many of the most abundant ESTs are unclassified (Table 1). In fact, 6 out of the top 10 represented ESTs are unclassified proteins, including by far the most highly represented EST at 49 copies. Moreover, of all clusters with more than 5 ESTs, every one but three are either unclassified or related to translation. The other three proteins are: a member of the HSP20 family

Table 1. Identity of clusters containing greater than 5 ESTs, excluding those corresponding to fragments of the ribosomal RNA.

Rank	Identity of cluster	Number of ESTs
1	Unclassified	49
2	Unclassified	20
3	Unclassified	18
4	Unclassified	14
5	HSP20 family	14
6	Unclassified	11
7	Cellular repressor of E1A-stimulated genes CREG	10
8	40S Ribosomal protein S23	9
9	Unclassified	8
10	Putative cell wall protein FLO11p	8
11	40S Ribosomal protein S16	7
12	60S Acidic ribosomal protein P1	7
13	Unclassified	7
14	40S Ribosomal protein S27	6
15	40S Ribosomal protein S19	6
16	40S Ribosomal protein S24	6
17	EFL (EF-like protein)	6

(a ubiquitous family of small proteins with a variety of functions), a putative homologue of a plant cell wall protein, and a putative repressor. Obviously, we cannot speculate on the function of the proteins encoded by these unclassified clusters, but they would be interesting candidates for further investigation as they may represent surface proteins or some other abundant protein of interest.

The overall proportion of ESTs that are not detectably similar to any known gene sequence are normally taken to represent genes that are relatively recent inventions or are evolving sufficiently rapidly to be beyond detection. In either case, if a closely related species is sampled one would expect to find some of these "unknown" genes in that close relative. Indeed, when *Helicosporidium* ESTs were compared with 3,943 EST sequences from its closest known relative, the trebouxiophyte green alga *Prototheca wickerhamii* (Borza et al. 2005), we found 12 clusters with a recognizable match in *Prototheca* but no significant similarity to a sequence in any other organism to date. These genes may be interesting cases to study the origin of parasitism in the ancestors of these genera.

Phylogeny of *Helicosporidium* Genes

The phylogenetic history of Helicosporidia has not been obvious from the initial observations of these parasites. They have, at various times, been allied with the Cnidospora (apicomplexa, microsporidia and other parasites), or lower fungi (Kellen and Lindegren 1973; Kudo 1966; Weiser 1970). The first molecular data from *Helicosporidium* were something of a surprise, therefore, as phylogenies of these genes showed a relationship to green algae. This conclusion has been supported by nuclear and plastid SSU rRNA genes, as well as actin and beta-tubulin, EFL and plastid elongation factor Tu . Moreover, most analyses with relevant data (the exception being EFL) also suggest that the closest relative of *Helicosporidium* is the opportunistic parasite *Prototheca* (de Koning and Keeling 2004; Tartar et al. 2002, 2003). Overall, these results and a few characteristics of vegetative *Helicosporidium* cells that also resemble *Prototheca* (Boucias et al. 2001) lend strong support to the conclusion that *Helicosporidium* is a green alga related to *Prototheca*. We have therefore examined the EST data to see if there is uniform support from a large number of genes for this conclusion, which has so far been based on strong evidence from a few genes. At the broadest level, the top blast hit to virtually every EST cluster is either a green alga or a plant, so we selected a class of genes to examine as a whole.

Ribosomal proteins are generally highly conserved and also highly- expressed and therefore abundant in EST samples, so we identified all clusters encoding putative ribosomal proteins from *Helicosporidium*, resulting in a set of 69 proteins. We then identified homologues for 65 of

these in *Prototheca* EST data (Borza et al. 2005), and conducted ML phylogenetic and bootstrap analyses on all 69 proteins. Representative trees from both SSU and LSU are shown in Figure 3, and the summary of the results is shown in Tables 2 and 3. The four genes shown in Figure 3 all support the sister relationship between *Helicosporidium* and *Prototheca* relatively strongly (bootstrap support ranging from 81 to 97) and some support the relationship of these to other green algae (e.g., S23, and L31), while others only support the relationship of greens to land plants (e.g., S3). The trees in Figure 3 are only intended to serve as examples; the important information comes from the overall view of all 69 phylogenies. Of the 27 small subunit proteins identified (Table 2), 19 placed *Helicosporidium* and *Prototheca* as sisters (14 with support over 80%). In one case no *Prototheca* data exist and in five others they were not sisters, but in every one of these six cases, *Helicosporidium* branched with the other green algae. Similarly, 42 large subunit proteins were analyzed (Table 3) and 33 of these placed *Helicosporidium* as sister to *Prototheca* (13 with support over 80%). Again, of the remaining nine genes, no *Prototheca* data were available for three, but in only a single case was the *Helicosporidium* gene not related to green algal homologues (this case is rpl6, where *Helicosporidium* was related to red algal homologues). Overall, 80% of the relevant phylogenies (52 out of 65 genes) showed a specific relationship between *Helicosporidium* and *Prototheca*, and 98% (68 out of 69) showed a relationship of *Helicosporidium* to either *Prototheca* or other green algae.

Concluding Remarks

In recent years, our understanding of Helicosporidia has been transformed by the application of molecular methods to the group, which in turn was only possible by establishing methods for cultivation. What were not long ago regarded as enigmatic parasites of unknown origin, are now interpreted as highly derived trebouxiophyte green algae. This is a remarkable evolutionary transformation supported by virtually all data, as exemplified by the large proportion of ribosomal proteins that show such a relationship. Harder questions have not been addressed, however. In particular, those relating to the molecular mechanisms of their unique form of infection and how such a system evolved from a free living, photosynthetic green alga (probably through a constitutively parasitic form like modern *Prototheca*) to the form we see today remain to be answered. The answers to these questions likely lay in the most difficult proteins to study, those with no readily identifiable homologue, and many such candidates have now been identified.

Methods

Strains, cultivation, and library construction: The *Helicosporidium* sp. (ATCC 50920) isolated from the black fly *Simulium jonesii* (Boucias et al. 2001) was propagated on artificial media (TC-100 insect medium supplemented by 5% fetal calf serum) containing gentamycin (50 μg/ml) and incubated at 26 °C. Vegetative cells were inspected for purity by light microscopy and collected by low-speed centrifugation, re-suspended into 10 ml of TriReagent (Sigma) plus glass beads (0.45 mm), and broken using a Braun MSK homogenizer. Following cell breakage, total RNA was extracted using the TriReagent manufacturer protocol. An aliquot of this total RNA was used to isolate polyA mRNA, using the Oligotex mRNA purification kit (Qiagen). The cDNA library was prepared in the Uni-ZAP XR plasmid using the ZAP-cDNA synthesis kit (Stratagene). Following the manufacturer's protocol, the cDNAs were ligated directionally into the Uni-ZAP XR vector, and the ligation reaction products were packaged using the Gigapack III Gold packaging extract. The library was titered, amplified, and mass excised converting phage into the pBluescript phagemid.

Expressed sequence tag sequencing and analysis: Colonies from the mass excised library were selected at random and plasmid DNA was isolated from 1,536 clones. All clones were sequenced from the 5' end using dye-terminator chemistry. Trace files were vector- and quality-trimmed, then sequences greater than 50 bp were clustered using Protist EST Program database, PEPdb (http://megasun.bch.umontreal.ca/pepdb/pepdb.html). All clusters were used to search GenBank using tBlastx and internally using Blastn. Clusters were also subjected to the automatic annotation using AutoFACT (http://megasun.bch.umontreal.ca/Software/AutoFACT.htm). All EST sequences with annotation have been deposited in the public database PEPdbPUB (http://amoebidia.bcm.umontreal.ca/public/pepdb/agrm.php), and NCBI dbEST (accessions CX128248-CX129443).

Ribosomal proteins were identified based on the automatic annotation results. *Prototheca* homologues were identified by the same procedure or

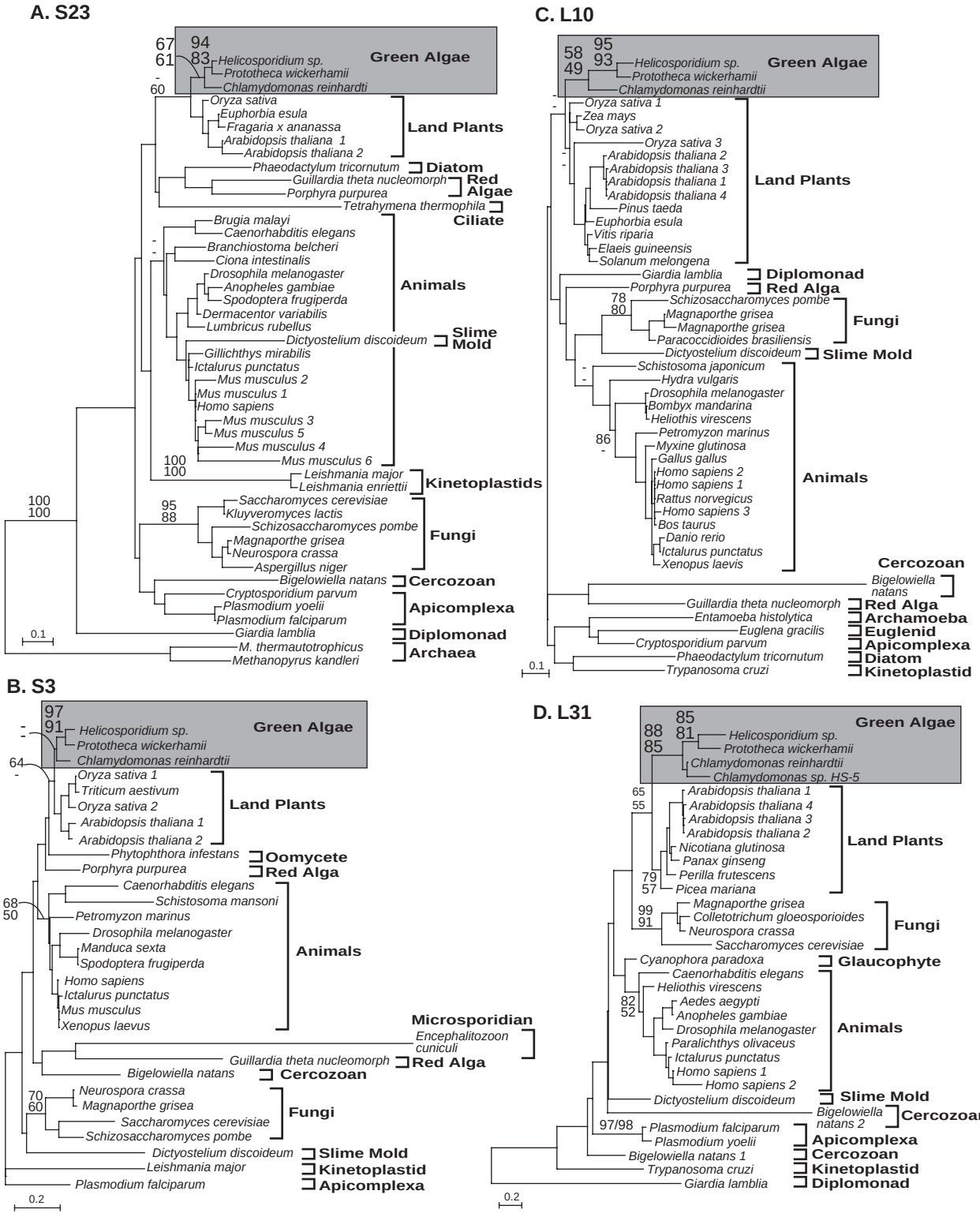

Figure 3. Protein maximum likelihood phylogenies of ribosomal SSU proteins (A) S23 and (B) S3 and LSU proteins (C) L10 and (D) L31. In all cases, green algae are shaded while other major groups are named and bracketed to the right. Numbers at nodes correspond to bootstrap support from ML (top or left) and distance (bottom or right). Dashes indicate support less than 50% and numbers are only shown for nodes relevant to *Helicosporidium* or supporting other major groups.

Table 2. Summary of phylogenetic analyses of SSU ribosomal proteins.

SSU protein	Taxa/ characters	BS% *Helicosporidium*+green algae[1]	BS% *Helicosporidium+Prototheca*[2]
S2	39/159	36[3]	84
S3	30/203	64[3]	97
S3A	40/230	29	58
S5	38/138	36	96
S6	33/136	82	—
S8	35/168	41	70
S9	34/92	29	69
S10	39/90	—	9
S11	37/138	89	91
S13	40/139	—	85
S14	39/118	—	89
S15	45/74	20[3]	86
S15A	53/123	70	98
S16	34/123	—	99
S19	39/137	38	99
S20	29/101	—	92
S21	37/71	43	54
S23	46/137	67	94
S24	34/122	46[3]	NA
S25	28/87	53	—
S26	31/115	70	—
S27	37/78	27[3]	61
S28	28/61	83[3]	—
S29	28/54	—	6
S30	25/56	47	80
SA	37/103	58	—

[1]Dashes (-) indicate this relationship was not observed in the bootstrap tree.
[2]NA indicates that no *Prototheca* data were available for comparison.
[3]In these cases the green algae were not monophyletic, but the green algae plus plants were monophyletic, so the support for the green algae and plants collectively is reported.

by similarity to *Helicosporidium* homologues. All *Helicosporidium* clusters and *Prototheca* sequences (Borza et al. 2005) (available at: amoebidia.bcm.umontreal.ca/public/pepdb/welcome.php) corresponding to putative ribosomal proteins were translated and added to amino acid multiple sequence alignments that included a broad diversity of eukaryotic 60S and 40S proteins (generally between 30 and 50 taxa in total). In all cases, alignments included representative animals, fungi and land plants, and whatever protists and algae that were available in public databases. In cases where no green algal sequence existed in GenBank, the *Chlamydomonas reinhardtii* sequence was assembled from EST or genomic data and added to the alignment so that all alignments included representatives of both land plants and green algae. Phylogenetic trees and 100 bootstrap replicates were inferred from all alignments using PhyML 2.3 (Guindon and Gascuel 2003) using the JTT substitution matrix with site-to-site rate variation modeled on a discrete gamma distribution with 4 variable rate categories and the shape parameter alpha estimated from the data. For the four genes where the phylogenies are shown, distance trees and bootstraps were also performed by TREE-PUZZLE 5.2 (Schmidt et al. 2002) using the WAG substitution frequency matrix and a gamma distribution with 8 rate categories and invariable sites with the alpha parameter and proportion of invariable sites estimated from the data. Trees were constructed using WEIGHBOR 1.2 (Bruno et al. 2000). Distance bootstraps were conducted in the same way using the shell script PUZZLEBOOT (A. Roger and M. Holder; www.tree-puzzle.de).

Table 3. Summary of phylogenetic analyses of LSU ribosomal proteins.

LSU protein	Taxa/ characters	BS% *Helicosporidium*+green algae[1]	BS% *Helicosporidium*+*Prototheca*[2]
L4B	36/183	77	95
L5	37/210	89	99
L6	39/98	—	—
L7	35/180	55[3]	70
L7A	33/188	70	54
L8 (L2)	39/176	73	64
L9	40/146	32	39
L10	47/119	58	95
L11	41/163	71	92
L13	27/191	89	NA
L14	33/116	81	—
L15	39/153	—	86
L17	37/134	52[3]	—
L18	37/164	71	47
L18A	36/151	77	47
L19	37/146	95	62
L21	36/148	26[3]	NA
L22	36/95	58[3]	42
L23	43/132	54	95
L23A	37/128	58[3]	80
L24	40/99	78	99
L26	33/119	21	34
L27	40/113	86	75
L27A	40/102	54	49
L29	27/53	—	32
L30	39/107	31	66
L31	33/84	85	88
L32	35/123	67	79
L34	36/107	57[3]	91
L35	38/105	—	52
L35A	34/94	—	62
L36	36/76	78[3]	70
L37	34/81	88	—
L37A	38/83	50	—
L38	30/68	24	41
L39	19/51	51	NA
L40 (CEP52)	42/121	19	47
L44	39/96	91	79
P0	41/86	44	—
P1	38/78	46	88
P2	45/66	—	94

[1]Dashes (-) indicate this relationship was not observed in the bootstrap tree.
[2]NA indicates that no *Prototheca* data were available for comparison.
[3]In these cases the green algae were not monophyletic, but the green algae plus plants were monophyletic, so the support for the green algae and plants collectively is reported.

Acknowledgements

This work was supported by grants from the Canadian Institutes for Health Research (MOP-42517 to P.J.K.) and the National Science Foundation (0131017 to D.G.B.). We thank R.W. Lee for permission to use *Prototheca* EST data, PEPdb for clustering and annotating *Helicosporidium* data, and M.B. Rogers and A. Fischer for providing many of the alignments. A.P.dK. is supported by

scholarships from CIHR and the Michael Smith Foundation for Health Research. P.J.K. is a scholar of the CIAR and a new investigator of the CIHR and MSFHR.

References

Bläske V-U, Boucias DG (2004) Influence of *Helicosporidium* sp. (Chlorophyta: Trebouxiophyceae) infection on development and survival of three noctuid species. Environ Entomol **33**: 54—61

Borza T, Popescu CE, Lee RW (2005) Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wicherhamii*. Eukaryot Cell 4: in press

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

Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol **17**: 189—197

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

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**: 696—704

Keeling PJ, Inagaki Y (2004) A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1{alpha}. Proc Natl Acad Sci USA **101**: 15380—15385

Keilin D (1921) On the life history of *Helicosporidium parasiticum* n. g., n. sp., a new species of protist parastie in the larvae of *Dashelaea obscura* Winn (Diptera:Ceratopogonidae) and in some arthropods. Parasitology **13**: 97—113

Kellen WR, Lindegren JE (1973) New host records for *Helicosporidium parasiticum*. J Invertebr Pathol **22**: 296—297

Kudo RR (1966) Protozoology. Thomas Publishing, Springfield, IL

Lindegren JE, Hoffmann DF (1976) Ultrastructure of some developmental stages of *Helicosporidium* sp. in the navel orangeworm, *Paramyelois transitella*. J Invertebr Pathol **27**: 105—113

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**: 502—504

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

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

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

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**: 41

Weiser J (1970) Helicosporidium parasiticum Keilin infection in the caterpillar of a hepialid moth in Argentina. J Protozool **17**: 436—440