

# The Chaperonin Genes of Jakobid and Jakobid-Like Flagellates: Implications for Eukaryotic Evolution

John M. Archibald,<sup>\*1</sup> Charles J. O'Kelly,<sup>†</sup> and W. Ford Doolittle<sup>\*</sup>

<sup>\*</sup>Department of Biochemistry and Molecular Biology, Dalhousie University; and <sup>†</sup>Bigelow Laboratory for Ocean Sciences

The jakobids are free-living mitochondriate protists that share ultrastructural features with certain amitochondriate groups and possess the most bacterial-like mitochondrial genomes described thus far. Jakobids belong to a diverse group of mitochondriate and amitochondriate eukaryotes, the excavate taxa. The relationships among the various excavate taxa and their relationships to other putative deep-branching protist groups are largely unknown. With the hope of clarifying these issues, we have isolated the cytosolic chaperonin CCTalpha gene from the jakobid *Reclinomonas americana* (strains 50394 and 50283), the jakobid-like malawimonad *Malawimonas jakobiformis*, two heteroloboseans (*Acrasis rosea* and *Naegleria gruberi*), a euglenozoan (*Trypanosoma brucei*), and a parabasalid (*Monocercomonas* sp.). We also amplified the CCTdelta gene from *M. jakobiformis*. The *Reclinomonas* and *Malawimonas* sequences presented here are among the first nuclear protein-coding genes to be described from these organisms. Unlike other putative early diverging protist lineages, a high density of spliceosomal introns was found in the jakobid and malawimonad CCTs—similar to that observed in vertebrate protein-coding genes. An analysis of intron positions in CCT genes from protists, plants, animals, and fungi suggests that many of the intron-sparse or intron-lacking protist lineages may not be primitively so but have lost spliceosomal introns during their evolutionary history. In phylogenetic trees constructed from CCTalpha protein sequences, *R. americana* (but not *M. jakobiformis*) shows a weak but consistent affinity for the Heterolobosea and Euglenozoa.

## Introduction

The jakobid flagellates are free-living, mitochondriate, heterotrophic protists (O'Kelly 1993, 1997). They include the families Jakobidae, genus *Jakoba* Patterson, 1990 (Patterson 1990) and Histonidae, genera *Histon* Voigt, 1901, and *Reclinomonas* Flavin and Nerad, 1993 (Flavin and Nerad 1993). The jakobids have figured prominently in hypotheses about the origin and early evolution of eukaryotes, originally because of the presence of ultrastructural features shared with certain amitochondriate lineages (O'Kelly 1993). Small subunit ribosomal RNA (SSUrRNA) and protein phylogenies often place the amitochondriates at or near the base of the eukaryotic tree (e.g., Sogin et al. 1989; Leipe et al. 1993; Hashimoto et al. 1994; Stiller et al. 1998); O'Kelly (1993) hypothesized that if amitochondriate protists were the earliest diverging eukaryotic lineages, then the earliest diverging mitochondriate eukaryotes should share ultrastructural features with these groups.

The sequencing of jakobid mitochondrial genomes has provided startling insight into mitochondrial evolution and has further suggested a pivotal role for the jakobids in our understanding of the evolution of eukaryotes. These organisms possess the most bacterial-like mitochondrial genomes characterized thus far (Lang et al. 1997; Gray et al. 1998; Gray, Burger, and Lang 1999). Most striking is the mitochondrial DNA (mtDNA) of *Reclinomonas americana*, which encodes 97 genes, more than are present in any other mitochondrial genome. Several of these have never before been

found encoded in mtDNA, including a gene for a bacterial translation factor (*tufA*), a putative cytochrome oxidase assembly protein (*cox11*), a secretion pathway protein (*secY*), and genes for four subunits of a bacterial-type RNA polymerase (*rpoA-D*); single subunit phage-type RNA polymerases are thought to function in all other known mitochondria (Lang et al. 1997; Gray et al. 1998). Operon-like ribosomal protein gene clusters similar to those found in bacteria are also present.

More recently, the jakobids have been considered members of a much larger assemblage of protists, the excavate taxa. The excavates are a diverse group of amitochondriate, mitochondriate, and hydrogenosomal lineages that share as their uniting feature the presence of a ventral feeding groove (Patterson, Simpson, and Weerakoon 1999; Simpson and Patterson 1999). In addition to the jakobids (*Jakoba*, *Reclinomonas*, and *Histon*) and the jakobid-like nanoflagellate *Malawimonas* (O'Kelly and Nerad 1999), the excavates include the heteroloboseans, diplomonads, retortamonads, *Trimastix*, and *Carpediemonas* (Simpson and Patterson 1999). Whereas ultrastructural data suggest that these organisms share a common excavate (i.e., feeding groove bearing) ancestor, there is no consensus view on the relationships amongst the various excavate taxa or their relationship to other mitochondriate, amitochondriate, and hydrogenosome-containing groups. Indeed, current views on the origin and evolution of eukaryotes are in a state of flux. Many of the putatively deep-branching and primitively amitochondriate protist lineages (including some of the excavate taxa) are now thought to be derived from mitochondrion-bearing ancestors (see Roger 1999 for recent review). Further, the ability of current phylogenetic methods to accurately reconstruct the deepest branches of phylogenetic trees has come into question (Hirt et al. 1999; Stiller and Hall 1999; Philippe and Germot 2000, and references therein). Philippe and Adoutte (1998) suggested that a big bang occurred at

<sup>1</sup> Present address: Department of Botany, University of British Columbia, Vancouver, Canada.

Key words: jakobids, eukaryotic evolution, introns, chaperonins, CCT.

Address for correspondence and reprints: John M. Archibald, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4. E-mail: jarch@interchange.ubc.ca.

*Mol. Biol. Evol.* 19(4):422–431. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

the base of eukaryotes and that the major cladogenetic events in eukaryotic evolution occurred in quick succession. Thus, there is at present no clear picture as to which protist groups—if any—actually represent early diverging lineages.

We are studying the evolution of chaperonins in diverse amitochondriate and mitochondriate protists. Group II chaperonins, a class of molecular chaperone found in Archaea and the eukaryotic cytosol, are double-ring protein complexes that mediate the proper folding of nascent or denatured proteins (for review see Willison and Horwich 1996). The eukaryotic cytosolic chaperonin complex (CCT or TriC) is best known for its role in mediating the proper folding of the cytoskeletal proteins actin and tubulin (Willison and Kubota 1994; Kubota, Hynes, and Willison 1995). Recent experiments suggest that a large number of newly translated proteins may also interact with CCT (Melki et al. 1997; Thulasiraman, Yang, and Frydman 1999). Unlike bacterial and organellar chaperonins, which are usually homo-oligomeric, the eukaryotic CCT complex contains eight different homologous subunits, alpha, beta, gamma, delta, epsilon, zeta, eta, and theta (Willison and Horwich 1996). The duplications producing the different CCT genes are known to have occurred prior to the divergence of the parabasalids and diplomonads from other eukaryotes (Archibald, Logsdon, and Doolittle 2000). Individual CCT subunits are over 500 amino acids in length and are highly conserved, both desirable properties for a phylogenetic marker. The presence of eight distinct CCT paralogs makes it possible to perform multiple reciprocal rootings of the phylogenetic tree of eukaryotes. Unlike SSUrRNA, EF-1 $\alpha$ , actin, and tubulin, the CCTs are relatively poorly sampled. Focusing on the alpha subunit of CCT, we therefore sought to broaden the diversity of taxonomic representation and add to the *Trichomonas vaginalis* and *Giardia lamblia* CCTalpha's sequenced in a recent study (Archibald, Logsdon, and Doolittle 2000).

## Materials and Methods

### Genomic DNAs

The *R. americana* (ATCC number 50394 and 50283) and *M. jakobiformis* (ATCC number 50310) genomic DNAs (gDNAs) used in this study were kindly provided by Dr. B. F. Lang (Université de Montréal, Canada). Dr. S. L. Hajduk and M. Oli (University of Alabama) provided gDNA from *Trypanosoma brucei* (strain MiTat 1.2 427/221), Dr. A. J. Roger (Dalhousie University, Canada) provided *Acrasis rosea* gDNA, and *Naegleria gruberi* gDNA was a gift from Dr. R. J. Redfield (University of British Columbia, Canada), Dr. J. M. Logsdon Jr. (Emory University), and J. Dacks (Dalhousie University, Canada).

### Cloning and Sequencing of CCT Genes

Degenerate PCR primers were used to amplify CCT genes from gDNA. A combination of universal chaperonin primers (Archibald, Logsdon, and Doolittle 2000) and CCTalpha-specific primers were used to amplify the CCTalpha gene (forward primers: CCT-2-for [5'-AACG

ACGGTGCNACNATHYT-3'], CCT-9-for [5'-CCAGTCGGTCTNGAYAARATG-3']; reverse primers: CCT-4-rev [5'-CTCTACAGCNCCNSCNCC-3'], CCT-10-rev [5'-TGATCAGRTCRTCDATNC-3'], CCT-11-rev [5'-AGGTCGTCGATGCKNARDAT-3'], TF-9-rev [GCAGCTATCARRTCRTCDAT-3']). With most primer combinations, 90%–95% of the gene was amplified. The CCTdelta gene was amplified with CCT-9-for (above) and CCT-7-rev [5'-ACGATGCACATNGHRTCRTG-3'].

PCR reactions were carried out on an MJ Research Inc. PTC-100 thermal cycler using GIBCO-BRL *Taq* polymerase, buffer, and dNTP. The PCR reactions performed with *R. americana* and *M. jakobiformis* DNAs contained 5% acetamide (final concentration). Following an initial denaturation of 3 min at 92°C, reactions were performed with 40–45 cycles of 92°C for 15 s, 50–54°C for 30 s, and 72°C for 3 min. PCR products of the expected size were cloned directly from low-melt agarose with the TOPO-TA cloning kit (Invitrogen). Clones were screened for the presence of inserts by *EcoRI* restriction of isolated plasmid DNAs or by PCR-screening directly from *E. coli* cells with the M13 universal forward and reverse primers. Sequencing was performed manually (T7 sequencing kit, Pharmacia) and with LiCor and ABI automated sequencers.

CCT sequences from *Plasmodium falciparum* were identified by BLAST (Altschul et al. 1990) at the PlasmoDB website (<http://www.plasmodb.org/>). CCT-alpha was obtained from chromosome 11 sequence, whereas CCTdelta was found in genomic sequence from chromosome 13. Preliminary sequence data for *P. falciparum* chromosomes 10 and 11 was obtained from The Institute for Genomic Research website ([www.tigr.org](http://www.tigr.org)). Sequencing of chromosomes 10 and 11 was part of the International Malaria Genome Sequencing Project and was supported by award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. Sequence data for *P. falciparum* chromosome 13 was obtained from The Sanger Centre website at <http://www.sanger.ac.uk/Projects/P.falciparum/>. Sequencing of *P. falciparum* chromosome 13 was accomplished as part of the Malaria Genome Project with support by The Wellcome Trust.

For *T. brucei*, a small fragment of the CCTalpha gene was found by searching genomic data from unfinished microbial genomes. Preliminary sequence data was obtained from The Institute for Genomic Research website at <http://www.tigr.org>. Based on this sequence an exact-match reverse primer (Tbru.CCTa.R-1 [5'-GTAAAGGCGAACACAGTCAT-3']) was designed and used in combination with degenerate forward primers (CCT-2-for, CCT-9-for; above) to amplify most of the CCTalpha coding sequence from *T. brucei* gDNA. From the sequence of five independent clones, two different CCTalpha clones were apparent, one of which matched the original fragment present in the *T. brucei* genome data. The ambiguities between the two clone types were almost all synonymous substitutions, suggesting the presence of multiple copies of the CCTalpha gene in *T. brucei* (see *Results*). The sequences presented in this

study have been deposited in GenBank under the following accession numbers: AF322043–AF322050.

### Southern Hybridization

To confirm the source and copy number of the CCT genes, PCR products obtained from *R. americana* (ATCC number 50394) and *M. jakobiformis* (ATCC number 50310) were isolated (BIORAD, Prep-a-gene), labeled with  $\alpha^{32}\text{P}$  (Prime-It II random primer labeling kit, Stratagene), and hybridized to restricted gDNA.

### Sequence Characterization and Phylogeny

A combination of BLAST and DNA Strider (Douglas 1995) was used to determine the positions of introns in the new CCT genes and in CCTs retrieved from the public databases. Inferred amino acid sequences from the *Reclinomonas*, *Malawimonas*, heterolobosean, *Trypanosoma* and *Monocercomonas* CCTs were added manually, based on globally conserved regions, to an alignment constructed previously (Archibald, Logsdon, and Doolittle 2000). From this master data set, smaller alignments containing subsets of the data were constructed. Each alignment was assessed individually and regions of ambiguity were removed to ensure that only confidently aligned amino acid positions were used for phylogenetic reconstruction. Gaps for missing data (e.g., the extreme N- and C-termini missing from PCR-generated sequences) were also removed. The complete CCTalpha-only alignment contained 26 taxa and 424 unambiguously aligned amino acid positions. The data sets for rooted analyses contained 23 CCTalpha sequences (three highly similar mammalian CCTs were removed) and between five and eight outgroup sequences chosen for maximal taxonomic diversity. These data sets contained the following number of sites: CCTalpha-beta—381 sites, CCTalpha-gamma—360 sites, CCTalpha-delta—378 sites, CCTalpha-eta—369 sites, CCTalpha-epsilon—351 sites, CCTalpha-theta—334 sites, and CCTalpha-zeta—357 sites. All alignments are available from J.M.A. upon request (jarch@interchange.ubc.ca).

Phylogenetic trees were inferred from amino acid sequences using maximum likelihood (ML) and ML-distance methods of tree reconstruction. ML analyses were performed with the following programs: proML in PHYLIP 3.6 (<http://evolution.genetics.washington.edu/phylip.html>), using the Dayhoff amino acid substitution matrix, the global rearrangements option, a randomized sequence input order (10 jumbles), and an among-site rate variation (ASRV) model with eight rate categories plus an invariable rate category (the relative rates for each category were estimated in PUZZLE 4.02 [Strimmer and von Haeseler 1997]); protML (using the JTT-F amino acid substitution matrix) in MOLPHY (Adachi and Hasegawa 1996); quartet puzzling in PUZZLE 4.02, accounting for ASRV with an eight rate-category discrete approximation to the  $\Gamma$  distribution plus an invariable rate category. ML-distance trees were inferred from  $\Gamma$ -corrected distance matrices calculated in PUZZLE 4.02 using FITCH (with global rearrangements) in PHYLIP, version 3.57 (Felsenstein 1993). Support for

proML trees was obtained by bootstrapping (100 replicates). For protML trees, REL values (obtained from quick-add searches of the best 1,000 or 2,000 trees in protML options -q -n 1,000, 2,000; Adachi and Hasegawa 1996) were used as measures of statistical support. Support values for ML-distance trees were obtained by bootstrapping (500 replicates) with PUZZLEBOOT 1.02 (A. Roger and M. Holder; <http://members.tripod.de/korbi/puzzle/>). PUZZLE was used to statistically assess the significance of different tree topologies using the Kishino-Hasegawa test (Kishino and Hasegawa 1989).

## Results

### Jakobid and Malawimonad CCT Genes Possess Numerous Spliceosomal Introns

We used a degenerate PCR approach to amplify the CCTalpha gene from *R. americana*, *M. jakobiformis*, two heteroloboseans (*A. rosea* and *N. gruberi*), the euglenozoan *T. brucei*, and the parabasalid *Monocercomonas* sp. We also amplified a somewhat smaller fragment of the CCTdelta gene from *M. jakobiformis*. Repeated attempts to amplify CCTalpha from *Jakoba libera* (a jakobid) were unsuccessful, despite the use of all possible primer combinations under a wide range of conditions. PCR reactions using additional degenerate primers designed with a strong bias toward G + C at the third codon position (see below) also failed.

The jakobid-malawimonad CCT genes presented here are among the first nuclear protein-coding genes to be characterized from these protists. The most striking feature of the *R. americana* and *M. jakobiformis* CCTs is the presence of multiple spliceosomal introns. The CCTalpha gene from *R. americana* possessed five introns (ranging from 67 to 145 nt in length, with some size heterogeneity between homologous introns in the two strains), whereas the *M. jakobiformis* CCTalpha contained seven introns (58–127 nt long). The *M. jakobiformis* CCTdelta gene possessed seven introns between 61 and 73 nt in length, despite being a shorter fragment of coding sequence than that obtained for CCTalpha (~1 kb of ORF). When the putative introns were removed, the inferred protein sequences were readily alignable with orthologs from a wide range of other eukaryotes, with no size heterogeneity at the intron-exon boundaries. All the introns possessed standard 5'-GT...AG-3' intron boundaries, with the exception of a single intron in one of the two CCTdelta clones from *M. jakobiformis*, which possessed a CT at the 5' intron-exon boundary. It is unclear whether this represents a PCR-generated artifact or a legitimate noncanonical 5' intron boundary. A summary of the sizes and positions of the introns found in the *R. americana* and *M. jakobiformis* CCT genes is shown in figure 1. With the exception of a single intron in the *A. rosea* CCTalpha, none of the nonjakobid-nonmalawimonad CCTalpha genes sequenced in this study contained introns.

The sequencing of independent clones suggested the presence of multiple copies of the CCTalpha gene in *R. americana* (50283 and 50394), *N. gruberi*, *T. brucei*, and *Monocercomonas* sp., as well as the CCTdelta

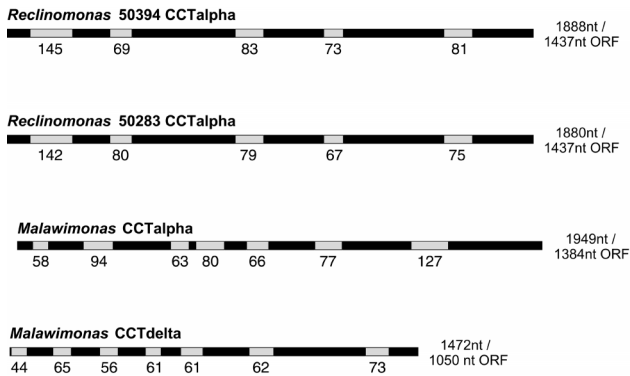


FIG. 1.—Spliceosomal introns in the *Reclinomonas* and *Malawimonas* CCT genes. The schematic shows the positions and sizes of the introns found in the *R. americana* and *M. jakobiformis* CCTAlpha genes and the *M. jakobiformis* CCTDelta gene.

gene in *M. jakobiformis*. Between the different clone types, most of the observed substitutions were synonymous; however, slightly different amino acid sequences were inferred from different clones of the *Monocercomonas* sp. CCTAlpha and *M. jakobiformis* CCTDelta genes. Some size heterogeneity between homologous introns in the different clones was also observed. The presence of multiple copies of CCT genes was subsequently confirmed for CCTAlpha in *R. americana* 50394 and for the *M. jakobiformis* CCTDelta by the southern hybridization of PCR products to restricted gDNAs (data not shown). Overall, a strong G + C bias was present in the *R. americana* and *M. jakobiformis* CCT genes, particularly at the third codon position. Edgcomb et al. (2001) observed a similar bias in jakobid tubulins.

### CCTAlpha Phylogeny

In an attempt to determine the relationships between the jakobid and malawimonad flagellates and

their relationship to other eukaryotes, we performed phylogenetic analyses on the CCTAlpha data set. We also performed multiple rooted analyses using the seven other CCT paralogs to root the CCTAlpha tree. Figure 2 shows a ML-distance phylogeny inferred from CCTAlpha protein sequences. Interestingly, the jakobid *R. americana* and the jakobid-like *M. jakobiformis* showed no affinity for one another in the CCTAlpha tree. Whereas the phylogenetic position of *M. jakobiformis* with respect to the other protists was quite unstable, *R. americana* consistently (but weakly) branched with the two heteroloboseans (as in fig. 2) and only occasionally with *T. brucei* (in proML analyses; data not shown). In many ways, the CCTAlpha topology is similar to those obtained with more widely used phylogenetic markers, such as SSUrRNA, actin, EF-1 $\alpha$ , and alpha- and beta-tubulin. For example, the sisterhood of animals and fungi, strongly supported by an ever-increasing wealth of molecular data (see Baldauf 1999 for recent review) is moderately well supported with all phylogenetic methods. Also, the single mycetozoan representative in the CCTAlpha data set, *Dictyostelium discoideum*, branches weakly but consistently at the base of animals and fungi, in line with suggestions that the Mycetozoa are an outgroup to the animal-fungal clade (Baldauf 1999). The alveolates, represented in our data set by the ciliate *Tetrahymena pyriformis* and the apicomplexan parasite *P. falciparum*, receive reasonable support as a monophyletic grouping, and as expected, the two heterolobosean sequences (*A. rosea* and *Nagleria gruberi*) branch together with strong support and with all phylogenetic methods. The parabasalid CCTAlphas (*Monocercomonas* sp., sequenced here and *T. vaginalis*, sequenced previously; Archibald, Logsdon, and Doolittle 2000) are characterized by extremely long branches, similar to phylogenies constructed with other molecules.

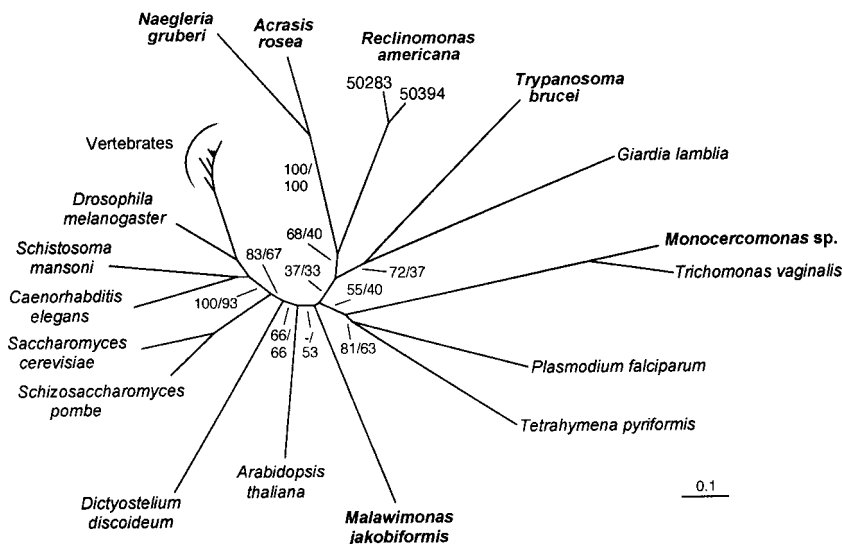


FIG. 2.—Phylogeny of CCT alpha protein sequences. The tree shown is a distance (Fitch-Margoliash) tree inferred from a ML distance matrix calculated with a rate heterogeneity model (JTT-F +  $\Gamma$  + inv, proportion of invariable sites = 0.08,  $\Gamma$  distribution shape parameter alpha = 1.23). The tree was constructed using a data set containing 26 taxa and 424 unambiguously aligned amino acid positions. Sequences determined in this study are highlighted. Support values for important nodes are ML RELL values and ML-distance bootstrap values, respectively. The scale bar indicates the inferred number of substitutions per site.

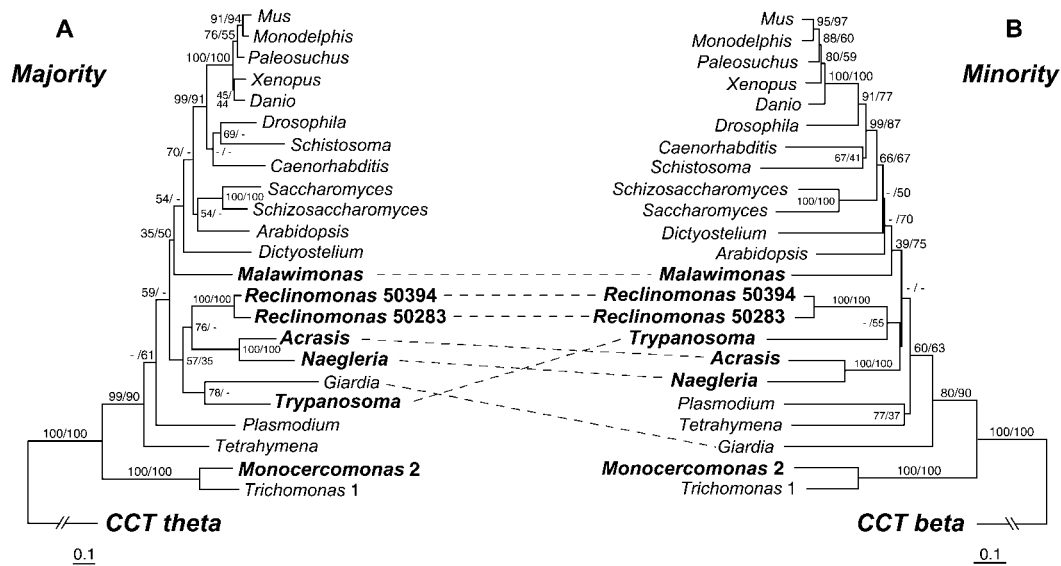


FIG. 3.—Rooted analyses of CCTalpha. *A*, Majority. The CCTalpha-theta tree showing the CCTalpha topology most often observed when rooted with another CCTalpha paralog. The tree shown is the ML tree obtained from a heuristic search of 1,000 trees in protML (lnL = -11253.17), with branch lengths reestimated in PUZZLE under an ASRV model (the proportion of invariable sites was 0.03 and the  $\Gamma$  distribution shape parameter  $\alpha = 1.52$ ). The tree was constructed from an alignment containing 334 unambiguously aligned amino acid positions. *B*, Minority. The CCTalpha-beta tree, an example of an alternate CCTalpha topology. Tree shown is the ML-distance tree inferred with the Fitch-Margoliash algorithm from an ML distance matrix calculated with an ASRV model (JTT-F +  $\Gamma$  + inv, proportion of invariable sites = 0.08,  $\Gamma$  distribution shape parameter  $\alpha = 1.49$ ). The alignment contained 381 sites. Support values are provided if greater than 40% (ML REll values and ML-distance bootstrap values, respectively). Dashed lines between the trees in *A* and *B* highlight differences in the deepest branches of the two CCTalpha topologies. The scale bars indicate the inferred number of amino acid substitutions per site.

The CCTalpha topology differs from other molecular phylogenies of eukaryotes in several interesting ways. The placement of the alveolates near the parabasalids is not seen with most phylogenetic markers but is reminiscent of phylogenies of actin, where the ciliate sequences are paraphyletic and branch near the base of the tree (e.g., Bricheux and Brugerolle 1997). Also, the amitochondriate diplomonad *G. lamblia* branches with the mitochondriate euglenozoan *T. brucei*. Together, these taxa form a weakly supported but consistently observed clade with the heteroloboseans (*N. gruberi* and *A. rosea*) and the two *R. americana* strains. Despite the fact that the *Giardia* CCTalpha is a fairly long branch, this relationship was still observed in the majority of rooted analyses (see below).

Given the general lack of resolution among the protists in the CCTalpha phylogeny, we examined the significance of alternate topologies using the Kishino-Hasegawa test (Kishino and Hasegawa 1989), in particular, the relative positions of the *R. americana* and *M. jakobiformis* sequences. Interestingly, trees in which *R. americana* and *M. jakobiformis* were specific sister groups were found to be worse than the topology shown in figure 2 at a 5% level of significance, as were trees in which *M. jakobiformis* was placed as an outgroup to the *R. americana*-Heterolobosea-*Giardia*-*Trypanosoma* clade. We also tested the effect of removing various long-branch taxa on the support for the topology obtained with the full CCTalpha data set. When the parabasalids were removed, the overall topology was the same, except for the placement of *M. jakobiformis*, which moved to a position adjacent to the alveolates

(data not shown). Interestingly, the support for the *R. americana*-Heterolobosea grouping increased with most phylogenetic methods, as did the support for *Giardia* branching with *T. brucei* (data not shown).

When CCTalpha was rooted with each of the seven other CCT paralogs, different topologies were often obtained from different data sets and from the same data set analyzed with different phylogenetic methods. However, the parabasalids were usually the deepest branch in the CCTalpha tree, and the alveolates were often the next deepest branch (sometimes as a paraphyletic group). Figure 3A shows the CCTalpha-theta tree, which represents the most commonly obtained topology with the various methods and data sets. Significantly, *G. lamblia* was not attached to the long branch of the outgroup in most analyses but grouped with *T. brucei*. Overall, the internal topology was very similar to that obtained in the unrooted analyses (fig. 2). Figure 3B shows the CCTalpha-beta tree, where the *G. lamblia* sequence branches near the base of CCTalpha, and *R. americana* branches with *T. brucei* and not with the heteroloboseans.

#### Intron Phylogeny

The density of spliceosomal introns in the *Reclinomonas* and *Malawimonas* CCT genes is extraordinarily high—on par with that observed in vertebrate genes (Logsdon 1998). To compare the diversity of intron positions in the jakobid and malawimonad CCTs to those in animals, fungi, plants, and other protists, we surveyed all available CCTalpha and CCTdelta genomic

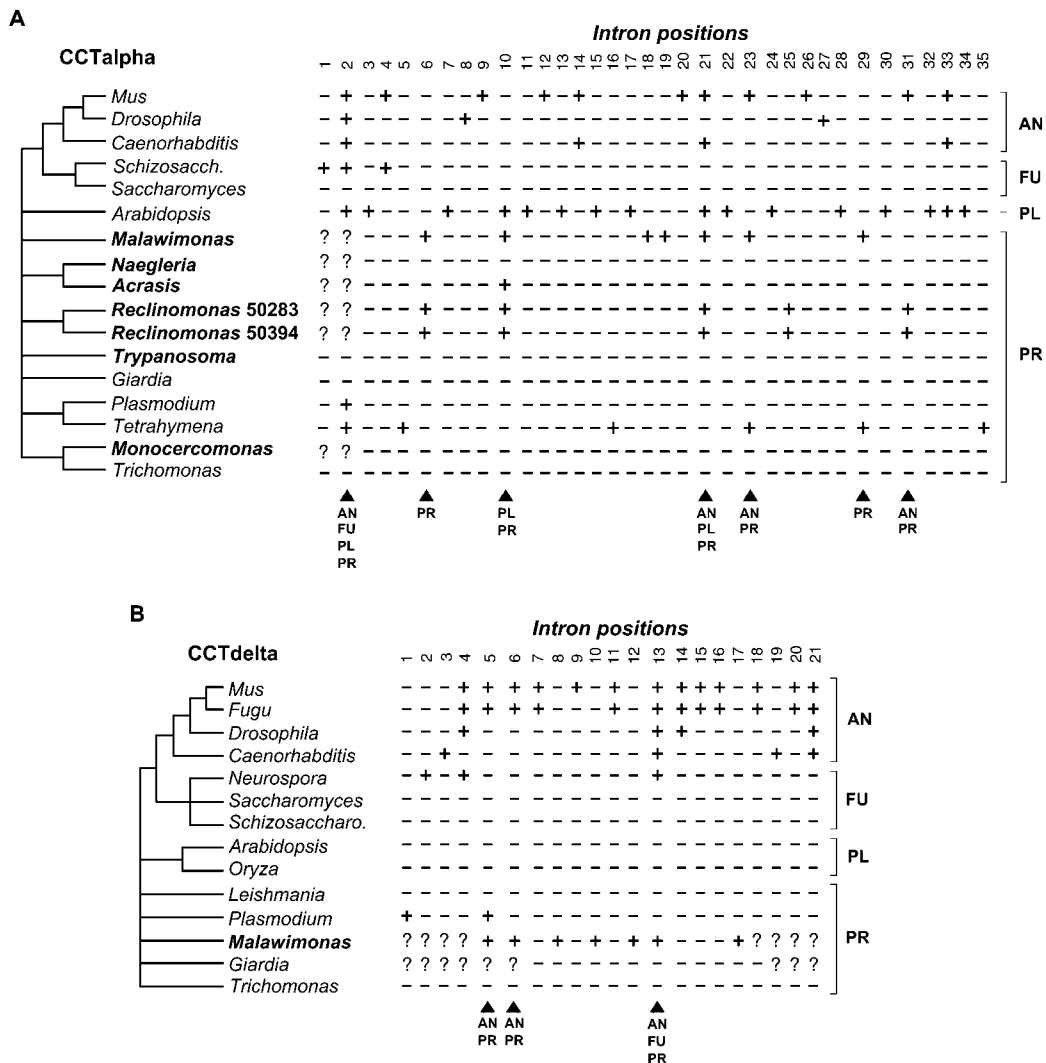


FIG. 4.—CCTalpha and CCTdelta intron phylogenies. A, Phylogenetic distribution of 35 known CCTalpha introns. The tree shown is a consensus topology of rooted and unrooted phylogenies of CCTalpha protein sequences (figs. 2, 3A, and 3B) showing only strongly supported relationships. A plus (+) sign indicates the presence of an intron, a minus indicates intron absence, and a question mark indicates missing data. Intron positions are numbered and potentially conserved intron positions are highlighted. B, Phylogenetic distribution of 21 known CCTdelta introns. Topology shown is a consensus of CCTdelta phylogenetic analyses (data not shown). For both data sets, only introns in exactly the same codon phase were considered to be in the same position. Sequences determined in this study are in bold. AN, animals; FU, fungi; PL, plants; PR, protists.

sequences. The results are presented in figure 4A and B. From the 17 CCTalpha sequences for which genomic sequence data was available, 35 distinct intron positions were observed. For CCTdelta, 14 sequences were available for comparison, and 21 intron positions were found. Interestingly, three out of the five introns present in the *R. americana* CCTalpha gene were shared with *M. jakobiformis*, despite the fact that the two groups showed no affinity for one another in phylogenies of CCTalpha (above), tubulin (Edgcomb et al. 2001), or concatenated mitochondrial proteins (Burger et al. 1999, 2000). Of these three, one was shared with *A. thaliana* and two of three animals (position 21) and another with *A. thaliana* and the heterolobosean *A. rosea* (position 10). One intron (position 2) was conserved between protists (*P. fallciparum* and *T. pyriformis*), plants, one of two fungi, and animals (fig. 4A). For CCTdelta, two introns were

shared between at least one protist and two of four animals (positions 5 and 6) and a third between *M. jakobiformis*, one of three fungi (*Neurospora crassa*), and animals (position 13) (fig. 4B). Despite the general high abundance of introns in plant protein-coding genes (Logsdon 1998)—16 introns were found in the *Arabidopsis* CCTalpha gene—neither of the plant CCTdeltas (*Arabidopsis* or *Oryza*) contained any introns. This likely represents a case of recent retrointegration of a cDNA, although in the absence of additional data, it is difficult to say how recent. A large number of unique intron positions are also present in both data sets. Twenty-four of 35 CCTalpha introns and 9 of 21 CCTdelta introns were present in a single species (considering the two *R. americana* strains as a single group), although for CCTalpha, 12 of the 24 introns were unique to *A. thaliana*. Two of the seven CCTalpha and four of the seven CCTdelta

introns in *M. jakobiformis* were not present in any other taxa.

## Discussion

Until recently, attempts to determine the phylogenetic affinities of the jakobid and jakobid-like flagellates to one another and to other protists have relied solely on their ultrastructural features. The analyses presented here are among the first to characterize nuclear protein-coding genes from these organisms and are the first to assess the utility of the cytosolic chaperonin CCTalpha as a marker for eukaryotic phylogeny. Whereas the failure to obtain a CCTalpha sequence from *J. libera* (see *Results*) precludes any direct comparisons between *J. libera*, *R. americana*, and *M. jakobiformis*, the latter two show no detectable phylogenetic affinity for one another in our analyses. This result is consistent with a recent analysis of jakobid alpha- and beta-tubulins (Edgcomb et al. 2001) and with phylogenies constructed from concatenated mitochondrial proteins (Burger et al. 1999, 2000). In the CCTalpha phylogenies *M. jakobiformis* showed no affinity for any particular protist group, whereas *R. americana* branched weakly but consistently with the heteroloboseans and the euglenozoan *T. brucei* (see below). A close relationship between *R. americana* and *M. jakobiformis* was also rejected using the Kishino-Hasegawa test (Kishino and Hasegawa 1989), although this result should be interpreted with caution. The test is known to be biased toward rejection when the best ML tree inferred from the data is included among the topologies tested (Goldman, Anderson, and Rodrigo 2000). A modified version of the test developed by Shimodaira and Hasegawa (1999) aims to eliminate this problem, but at present there is debate over its proper implementation (Shimodaira and Hasegawa 1999; Goldman, Anderson, and Rodrigo 2000).

The weak but consistent affinity of *R. americana* for the heteroloboseans and euglenozoa in CCTalpha phylogenies (figs. 2, 3A, and 3B) is interesting for two reasons. First, it is consistent with beta-tubulin and combined alpha-beta-tubulin trees (Edgcomb et al. 2001). Second, it is consistent with mitochondrial gene-content data. The mtDNA of the heterolobosean *N. gruberi* contains (in addition to many genes shared with plant and protist mtDNAs) two genes thus far only identified in jakobid mitochondrial genomes (*atp3* and *cox11*; M. W. Gray, personal communication; <http://megasun.bch.umontreal.ca/ca.ogmp/projects/ngrub/gen.html>). Whereas the grouping of *R. americana* with the Heterolobosea is consistent with beta-tubulin and combined alpha-beta-tubulin phylogenies (Edgcomb et al. 2001) and mitochondrial gene-content data, it is at odds with mitochondrial cristal morphology. Like *N. gruberi* and the euglenozoan *T. brucei*, *M. jakobiformis* possesses discoidal cristae, yet it is *R. americana*, which possesses tubular mitochondrial cristae, that shows the most affinity for the Heterolobosea and *T. brucei* in the CCTalpha and tubulin trees. Whereas cristal morphology has often been taken to be an evolutionarily stable character, these data suggest that it may not always be reliable for track-

ing the deepest divisions in eukaryotic evolution. More generally, the significance of the retention of ancestral features in jakobid and malawimonad mtDNAs, in terms of their relationship to other eukaryotes, is as yet unclear. The *M. jakobiformis* mtDNA possesses three genes (*rpl18*, 19, 31) that have thus far not been observed in nonjakobid protist mtDNAs (<http://megasun.bch.umontreal.ca/ogmp/projects/mjako/gen.html>), yet all phylogenetic analyses performed thus far on both mitochondrial and nuclear genes suggest that *M. jakobiformis* and the jakobids do not form a monophyletic group.

On the whole, CCTalpha phylogenies show a great deal of congruence with more commonly used phylogenetic markers: animals and fungi appear to be each other's closest relatives, the mycetozoan *Dictyostelium* appears as an immediate outgroup to the animal-fungal clade, and the alveolates and Heterolobosea are monophyletic groups. Most interesting was the placement of the diplomonad *G. lamblia*. Almost without exception, rooted SSUrRNA and protein phylogenies place diplomonads at or near the base of the eukaryotic tree (Roger 1999). In most rooted analyses of CCTalpha, *Giardia* was not positioned at the base of the eukaryotic tree but was nested within mitochondrion-containing groups. This is consistent with the suggestion that diplomonads have lost their mitochondria secondarily (Roger et al. 1998; Roger 1999). The analyses presented here illustrate the problem with deep phylogeny, long-branch attraction (e.g., Germot and Philippe 1999; Stiller and Hall 1999; Philippe and Germot 2000). It is clear that the two parabasalids in the CCTalpha data set have the longest branches of all the taxa—they also emerge consistently at the base of the CCT tree in rooted analyses. For this reason, the deepest branches of the rooted CCTalpha phylogenies presented here should be viewed with caution.

The high density of introns in the jakobid and malawimonad CCT genes raises the question of their origin. Were they recently acquired, or do they represent the retention of old introns? Separate from the issue of the origin of introns themselves (i.e., introns-early vs. introns-late; Logsdon et al. 1995; de Souza et al. 1998; Logsdon 1998) is the question of the diversity and antiquity of spliceosomal introns within eukaryotes. For the most part, protist genes are intron sparse, and many of the lineages that have figured prominently in hypotheses of early events in eukaryotic evolution (e.g., diplomonads and parabasalids) seem to lack introns entirely (Logsdon 1998). Remarkably few intron positions are known to be conserved between animals, fungi, and plants (Fast, Logsdon, and Doolittle 1999; Palmer and Logsdon 1991) and fewer still between animals or fungi (or both), plants, and protists. Interestingly, the CCTalpha data set contains two introns of the latter sort (positions 2 and 21), and CCTdelta contains an intron shared between *M. jakobiformis*, one of three fungi, and animals (position 13). Unfortunately, the CCTalpha data set is somewhat biased against the presence of old introns because of the fact that the sole fungal representatives, *Saccharomyces cerevisiae* and *Schizosaccharo-*

*myces pombe*, are known to be relatively intron sparse compared to other fungi (the intron density in *S. cerevisiae* is only 0.1 per kb; Logsdon 1998). In the CCTdelta data set, three introns were found in the *N. crassa* CCTalpha (*N. crassa* has a higher intron density), and two of these were shared with animals. The CCTdelta data set is similarly biased. No introns were found in the *S. cerevisiae* or *S. pombe* CCTdelta genes, and despite the fact that the intron density in plant protein-coding genes is generally quite high (Logsdon 1998), neither of the plant CCTdeltas (*Arabidopsis* or *Oryza*) contained a single intron. With few protist introns available for comparison, it is difficult to make strong inferences about whether the potentially old jakobid-malawimonad introns are a result of the retention of ancestral intron positions or because of independent parallel insertions. More CCTalpha sequences from each of the major eukaryotic groups may resolve this issue.

It is certainly clear from both data sets that recent intron gain has occurred. The *Arabidopsis thaliana* CCTalpha gene alone contained 12 introns in positions not found in any other taxa, and the CCTalpha from the ciliate *T. pyriformis* contained another three (fig. 4A, see Results). For the jakobid-malawimonad CCTalpha's, one of the *R. americana* introns and two of the *M. jakobiformis* introns were located in unique positions, as were four of seven *M. jakobiformis* CCTdelta introns. Data from the jakobid tubulin genes also support this notion. As predicted from ultrastructure and mitochondrial genome content data, *R. americana* and *J. libera* branched strongly together in beta- and alpha-beta-tubulin phylogenies. Yet despite this apparent close relationship, their respective beta-tubulins each possessed a single intron located in different positions, neither of which is present in known beta-tubulin genes (Edgcomb et al. 2001).

As for intron loss, the data are more ambiguous. In general, this uncertainty stems from the lack of a robust phylogeny of the major protist groups upon which intron gain-loss scenarios can be evaluated. Nevertheless, a case for intron loss can be inferred for the Heterolobosea. Consistent with the low intron density observed thus far for heterolobosean protein-coding genes (Logsdon 1998), only one intron was found in the CCTalpha gene from *A. rosea*, and none were present in the *N. gruberi* sequence. However, the intron in the *A. rosea* CCTalpha is shared with the two *R. americana* strains, *M. jakobiformis* and *Arabidopsis* (intron position 10; fig. 4A), suggesting that it may predate the divergence of plants and these protists.

It seems significant that in cases where only one or a few introns are present in the CCT genes (e.g., *Acrasis*, *Plasmodium*, *Schizosaccharomyces*; fig. 4A and B) they are located near the 5' end (in CCTalpha, intron positions 1–10 are located in the first 20% of the coding sequence, whereas introns 1–5 in CCTdelta are present in the first 11% of the gene). Such a bias is also observed for the vast majority of introns in yeast (Fink 1987; Spingola et al. 1999), and for the few remaining introns in the highly reduced nucleomorph genome of the cryptomonad alga *Guillardia theta* (Douglas et al.

2001). In yeast, a process of incomplete reverse transcriptase-mediated intron loss has been suggested to account for such a nonrandom distribution (Fink 1987). However, it is known that at least some of these 5'-end introns play regulatory roles in gene expression (Spingola et al. 1999), suggesting that they may persist not because of incomplete reverse transcription but because of selection against intron loss. Either way, the clustering of introns near the 5'-ends of genes is consistent with loss from a more intron-rich ancestral state. On the other hand, it is also possible that such a pattern simply reflects a bias in the process of intron insertion.

Although no introns have been described in genes from the amitochondriate diplomonads and parabasalids, indirect evidence for their existence in the parabasalid *T. vaginalis* has come from the discovery of a homolog of PRP8, a highly conserved protein component of the spliceosome (Fast and Doolittle 1999). A gene whose product has significant similarity to PRP8 is also present in the near complete genome of the diplomonad *G. lamblia* (Smith et al. 1998) (identified by searching the high-throughput genome sequence database at NCBI). Taken as a whole, the data are consistent with the possibility that the intron-sparse or intron-free nature of many protist genes is a derived feature.

How intron-rich will jakobid and malawimonad nuclear genes turn out to be? At present, such a small sample size makes it impossible to predict. Nevertheless, the data presented here suggest a different picture of spliceosomal intron evolution than is often assumed, one in which intron loss has been a significant factor in shaping eukaryotic nuclear genomes.

## Acknowledgments

We wish to thank B. F. Lang, S. L. Hajduk, M. Oli, A. J. Roger, R. J. Redfield, J. M. Logsdon Jr., and J. Dacks for generously providing the protist gDNAs used in this study. We also thank A. Stoltzfus for mol2con.pl, members of the Doolittle lab for discussion and comments on the manuscript, two anonymous reviewers for helpful comments, and A. Roger for assistance with ML analyses and critical review of the manuscript. For *T. brucei*, preliminary sequence data was obtained from The Institute for Genomic Research website at <http://www.tigr.org>. Preliminary sequence data for *P. falciparum* chromosomes 10, 11, and 13 was obtained from The Institute for Genomic Research website ([www.tigr.org](http://www.tigr.org)). Sequencing of chromosomes 10, 11, and 13 was part of the International Malaria Genome Sequencing Project and was supported by award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. This work was supported by a grant awarded to W.F.D. by the Canadian Institutes for Health Research (CIHR). J.M.A. was supported by a Doctoral Research Award from CIHR.

## LITERATURE CITED

- ADACHI, J. M., and M. HASEGAWA. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* 28:1–150.

- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- ARCHIBALD, J. M., J. M. LOGSDON JR., and W. F. DOOLITTLE. 2000. Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol. Biol. Evol.* **17**:1456–1466.
- BALDAUF, S. L. 1999. A search for the origins of Animals and Fungi: comparing and combining molecular data. *Am. Nat.* **154**(Suppl.):S178–S188.
- BRICHEUX, G., and G. BRUGEROLLE. 1997. Molecular cloning of actin genes in *Trichomonas vaginalis* and phylogeny inferred from actin sequences. *FEMS Microbiol. Lett.* **153**: 205–213.
- BURGER, G., D. SAINT-LOUIS, M. W. GRAY, and B. F. LANG. 1999. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* **11**:1675–1694.
- BURGER, G., Y. ZU, T. G. LITTLEJOHN, S. J. GREENWOOD, M. N. SCHNARE, B. F. LANG, and M. W. GRAY. 2000. Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J. Mol. Biol.* **297**:365–380.
- DE SOUZA, S. J., M. LONG, R. J. KLEIN, S. ROY, S. LIN, and W. GILBERT. 1998. Towards a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. USA* **95**:5094–5099.
- DOUGLAS, S. E. 1995. DNA strider. An inexpensive sequence analysis package for the Macintosh. *Mol. Biotechnol.* **3**:37–45.
- DOUGLAS, S., S. ZAUNER, M. FRAUNHOLZ, M. BEATON, S. PENNY, L. DENG, X. WU, M. REITH, T. CAVALIER-SMITH, and W.-G. MAIER. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* **410**:1091–1096.
- EDGCOMB, V. P., A. J. ROGER, A. G. B. SIMPSON, D. T. KYSELA, and M. L. SOGIN. 2001. Evolutionary relationships among “jakobid” flagellates as indicated by alpha- and beta-tubulin phylogenies. *Mol. Biol. Evol.* **18**:514–522.
- FAST, N. M., and W. F. DOOLITTLE. 1999. *Trichomonas vaginalis* possesses a gene encoding the essential spliceosomal component, PRP8. *Mol. Biochem. Parasitol.* **99**:275–278.
- FAST, N. M., J. M. LOGSDON JR., and W. F. DOOLITTLE. 1999. Phylogenetic analysis of the TATA box binding protein (TBP) gene from *Nosema locustae*: evidence for a microsporidia-fungi relationship and spliceosomal intron loss. *Mol. Biol. Evol.* **16**:1415–1419.
- FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FINK, G. R. 1987. Pseudogenes in yeast? *Cell* **49**:5–6.
- FLAVIN, M., and T. A. NERAD. 1993. *Reclinomonas americana* N. G., N. Sp., a new freshwater heterotrophic flagellate. *J. Eukaryot. Microbiol.* **40**:172–179.
- GERMOT, A., and H. PHILIPPE. 1999. Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *J. Eukaryot. Microbiol.* **46**:116–124.
- GRAY, M. W., G. BURGER, and B. F. LANG. 1999. Mitochondrial evolution. *Science* **283**:1476–1481.
- GRAY, M. W., B. F. LANG, R. CEDERGREN et al. (15 co-authors). 1998. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26**:865–878.
- GOLDMAN, N., J. P. ANDERSON, and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**:652–670.
- HASHIMOTO, T., Y. NAKAMURA, F. NAKAMURA, T. SHIRAKURA, J. ADACHI, N. GOTO, K. OKAMOTO, and M. HASEGAWA. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.* **11**:65–71.
- HIRT, R. P., J. M. LOGSDON JR., B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLEY. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. USA* **96**:580–585.
- KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidia. *J. Mol. Evol.* **29**:170–179.
- KUBOTA, H., G. HYNES, and K. WILLISON. 1995. The chaperonin containing t-complex polypeptide 1 (TCP-1). Multi-subunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230**:3–16.
- LANG, B. F., G. BURGER, C. J. O’KELLY, R. CEDERGREN, G. B. GOLDING, C. LEMIEUX, D. SANKOFF, M. TURMEL, and M. W. GRAY. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**: 493–497.
- LEIPE, D. D., J. H. GUNDERSON, T. A. NERAD, and M. L. SOGIN. 1993. Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* **59**:41–48.
- LOGSDON, J. M. JR. 1998. The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **8**:637–648.
- LOGSDON, J. M. JR., M. G. TYSHENKO, C. DIXON, D. J. JAFARI, V. K. WALKER, and J. D. PALMER. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. USA* **92**:8507–8511.
- MELKI, R., G. BATELIER, S. SOULIE, and R. C. WILLIAMS JR. 1997. Cytoplasmic chaperonin containing TCP-1: structural and functional characterization. *Biochemistry* **36**:5817–5826.
- O’KELLY, C. J. 1993. The jakobid flagellates: structural features of *Jakoba*, *Reclinomonas* and *Histonida* and implications for the early diversification of eukaryotes. *J. Eukaryot. Microbiol.* **40**:627–636.
- . 1997. Ultrastructure of trophozoites, zoospores and cysts of *Reclinomonas americana* Flavin & Nerad, 1993 (*Protista incertae sedis*: Histonidae). *Eur. J. Protistol.* **33**: 337–348.
- O’KELLY, C. J., and T. A. NERAD. 1999. *Malawimonas jakobiformis* n. gen., n. sp. (Malawimonadidae n. fam.): a *Jakoba*-like heterotrophic nanoflagellate with discoidal mitochondrial cristae. *J. Eukaryot. Microbiol.* **46**:522–531.
- PALMER, J. D., and J. M. LOGSDON JR. 1991. The recent origins of introns. *Curr. Opin. Genet. Dev.* **1**:470–477.
- PATTERSON, D. J. 1990. *Jakoba libera* (Ruinen, 1938), a heterotrophic flagellate from deep oceanic sediments. *J. Mar. Biol. Assoc. U K* **70**:381–393.
- PATTERSON, D. J., A. G. B. SIMPSON, and N. WEERAKOON. 1999. Free-living flagellates from anoxic habitats and the assembly of the eukaryotic cell. *Biol. Bull.* **196**:381–384.
- PHILIPPE, H., and A. ADOUTTE. 1998. The molecular phylogeny of protozoa: solid facts and uncertainties. Pp. 25–52 in G. H. Coombs, K. Vickerman, M. A. Sleigh, and A. Warren, eds. *Evolutionary relationships among protozoa*. Kluwer Academic Publishers, London.
- PHILIPPE, H., and A. GERMOT. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.* **17**:830–834.

- ROGER, A. J. 1999. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**(Suppl.):S146–S163.
- ROGER, A. J., S. G. SVARD, J. TOVAR, C. G. CLARK, M. W. SMITH, F. D. GILLIN, and M. L. SOGIN. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. USA* **95**:229–234.
- SHIMODAIRA, H., and M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- SIMPSON, A. G. B., and D. J. PATTERSON. 1999. The ultrastructure of *Carpodemonas membranifera* (Eukaryota) with reference to the “excavate hypothesis.” *Eur. J. Protistol.* **35**:353–370.
- SMITH, M. W., S. B. ALEY, M. SOGIN, F. D. GILLIN, and G. A. EVANS. 1998. Sequence survey of the *Giardia lamblia* genome. *Mol. Biochem. Parasitol.* **95**:267–280.
- SOGIN, M. L., J. H. GUNDERSON, H. J. ELWOOD, R. A. ALONSO, and D. A. PEATTIE. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**:75–77.
- SPINGOLA, M., L. GRATE, D. HAUSSLER, and M. ARES JR. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**:221–234.
- STILLER, J. W., E. C. DUFFIELD, and B. D. HALL. 1998. Amictochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **95**:11,769–11,774.
- STILLER, J. W., and B. D. HALL. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16**:1270–1279.
- STRIMMER, K., and A. VON HAESELER. 1997. PUZZLE. Zoologisches Institut, Universität München, Germany.
- THULASIRAMAN, V., C. F. YANG, and J. FRYDMAN. 1999. *In vivo* newly translated polypeptides are sequestered in a protected folding environment. *EMBO J.* **18**:85–95.
- WILLISON, K. R., and A. L. HORWICH. 1996. Structure and function of chaperonins in archaeobacteria and eukaryotic cytosol. Pp. 107–136 in R. J. Ellis, ed. *The chaperonins*. Academic Press, San Diego.
- WILLISON, K. R., and H. KUBOTA. 1994. The structure, function, and genetics of the chaperonin containing TCP-1 (CCT) in eukaryotic cytosol. Pp. 299–312 in R. I. Morimoto, A. Tissieres, and C. Georgopoulos, eds. *The biology of heat shock proteins and molecular chaperones*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

BRANDON GAUT, reviewing editor

Accepted October 26, 2001