J. Eukaryot. Microbiol., 54(1), 2007 pp. 86–92
© 2006 The Author(s)
Journal compilation © 2006 by the International Society of Protistologists
DOI: 10.1111/j.1550-7408.2006.00233.x

Phylogeny of Phagotrophic Euglenids (Euglenozoa) as Inferred from Hsp90 Gene Sequences

SUSANA A. BREGLIA, CLAUDIO H. SLAMOVITS and BRIAN S. LEANDER

Program in Evolutionary Biology, Departments of Botany and Zoology, Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

ABSTRACT. Molecular phylogenies of euglenids are usually based on ribosomal RNA genes that do not resolve the branching order among the deeper lineages. We addressed deep euglenid phylogeny using the cytosolic form of the heat-shock protein 90 gene (hsp90), which has already been employed with some success in other groups of euglenozoans and eukaryotes in general. Hsp90 sequences were generated from three taxa of euglenids representing different degrees of ultrastructural complexity, namely *Petalomonas cantuscygni* and wild isolates of *Entosiphon sulcatum*, and *Peranema trichophorum*. The hsp90 gene sequence of *P. trichophorum* contained three short introns (ranging from 27 to 31 bp), two of which had non-canonical borders GG-GG and GG-TG and two 10-bp inverted repeats, suggesting a structure similar to that of the non-canonical introns described in *Euglena gracilis*. Phylogenetic analyses confirmed a closer relationship between kinetoplastids and diplonemids than to euglenids, and supported previous views regarding the branching order among primarily bacteriovorous, primarily eukaryovorous, and photosynthetic euglenids. The position of *P. cantuscygni* within Euglenozoa, as well as the relative support for the nodes including it were strongly dependent on outgroup selection. The results were most consistent when the jakobid *Reclinomonas americana* was used as the outgroup. The most robust phylogenies place *P. cantuscygni* as the most basal branch within the euglenid clade. However, the presence of a kinetoplast-like mitochondrial inclusion in *P. cantuscygni* deviates from the currently accepted apomorphy-based definition of the kinetoplastid clade and highlights the necessity of detailed studies addressing the molecular nature of the euglenid and diplonemid mitochondrial genome.

Key Words. *Entosiphon*, Euglenida, Euglenozoa, heat-shock protein 90, introns, kDNA, kinetoplast, *Peranema*, *Petalomonas*, phylogeny.

THE Euglenozoa constitute a monophyletic group of flagellates that includes three major clades: euglenids, kinetoplastids, and diplonemids. Although, this monophyly has been confirmed by studies using nuclear small subunit (SSU) recombinant DNA (rDNA) and protein sequences, the relationships between the three major groups varies according to the methods and the markers employed (Simpson and Roger 2004; Talke and Preisfeld 2002). The most compelling molecular phylogenetic data indicate that diplonemids and kinetoplastids are more closely related to each other than to euglenids (Maslov, Yasuhira, and Simpson 1999; Moreira, López-García, and Vickerman 2004; Simpson, Lukěs, and Roger 2002; Simpson and Roger 2004; Simpson, Stevens, and Lukěs 2006b; Simpson et al. 2004). Morphologically, these three groups share a distinctive flagellar apparatus in which the flagella, inserting within an apical or subapical pocket, are reinforced by paraxonemal rods, a proteinaceous scaffolding adjacent to the usual 9+2 axoneme (Marande, Lukěs, and Burger 2005; Simpson 1997; Willey, Walne, and Kivic 1988).

Kinetoplastids include both trypanosomatids (parasites) and bodonids (either free living or parasites) and are apomorphically defined by a novel arrangement of mitochondrial DNA (mtDNA) forming conspicuous inclusions called ''kinetoplasts'' (Lukěs et al. 2002). In the majority of kinetoplastids, the kinetoplast DNA (kDNA) is arranged in two kinds of circular molecules known as ''maxicircles'' and ''minicircles.'' Maxicircles are present in a few dozen copies, and encode both mitochondrial protein genes and some ''guide rRNA'' (gRNA) genes. Minicircles are present in thousands of copies and encode gRNAs (Morris et al. 2001). Transcripts of some protein-coding genes undergo editing in which the addition or removal of uridine residues produce the final mRNA molecules (Lukěs, Hashimi, and Zíková 2005). Nonetheless, the general configuration of kDNA is variable, and the main states described so far include pro-kDNA, poly-kDNA, pan-kDNA, and mega-kDNA (see reviews by Lukěs et al. 2002; Marande et al. 2005).

Diplonemids comprise a small group of mostly free-living phagotrophs (Kivic and Walne 1984; Marande et al. 2005) and some facultative parasites (Kent et al. 1987). There is no evidence of a kDNA-like mitochondrial inclusion in diplonemids, but the structure of its single mitochondrion has been shown to be unusual (e.g. highly branched, with few flattened cristae). The mtDNA is arranged in circular chromosomes of two slightly different sizes and distributed in a pan-kDNA-like fashion (Marande et al. 2005; Maslov et al. 1999).

Euglenids form a very diverse group of free-living flagellates that inhabit a variety of aquatic environments. Members of this group have very different modes of nutrition, such as osmotrophy, phagotrophy, and phototrophy. The most distinctive structural character of euglenids is a pellicle composed of proteinaceous strips that run lengthwise over the cell (Leander and Farmer 2001a). Some euglenids are completely rigid, whereas others are highly plastic. Cell plasticity depends on the number of pellicle strips in the cell, and how they articulate with each other along the lateral margins. Longitudinally arranged strips are usually associated with cells that cannot change their shape, whereas helically arranged strips are associated with a type of peristaltic movement called ''euglenoid movement'' (Leander and Farmer 2001b; Leander, Witek, and Farmer 2001b; Suzaki and Williamson 1985, 1986a, b). The diversity of pellicle surface patterns provides a significant amount of information useful for discriminating among the euglenid lineages (Brosnan et al. 2003; Esson and Leander 2006; Leander 2004; Leander and Farmer 2001a; Leander et al. 2001b).

Both morphology-based and molecular-based phylogenetic analyses suggest that rigid phagotrophs form the earliest diverging branches within euglenids (Leander 2004; Leander, Triemer, and Farmer 2001a). As such, an important taxon for reconstructing early events in the evolutionary history of euglenozoans is the phagotrophic euglenid *Petalomonas cantuscygni* (Busse, Patterson, and Preisfeld 2003; Leander et al. 2001a). *Petalomonas cantuscygni* is usually considered the earliest branching euglenid because it possesses an unusual combination of ultrastructural characters, including a simple feeding apparatus, pellicle strips (synapomorphic for euglenids), and a kDNA-like inclusion within each mitochondrion (Leander et al. 2001a). Unfortunately, the SSU rDNA sequence from *P. cantuscygni* is highly divergent, which precludes resolution of its phylogenetic position within the

Corresponding Author: B. Leander, Canadian Institute for Advanced Research, Program in Evolutionary Biology, Departments of Botany and Zoology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4—Telephone number: +604-822-2474; FAX number: +604-822-6089; e-mail: bleander@interchange.ubc.ca

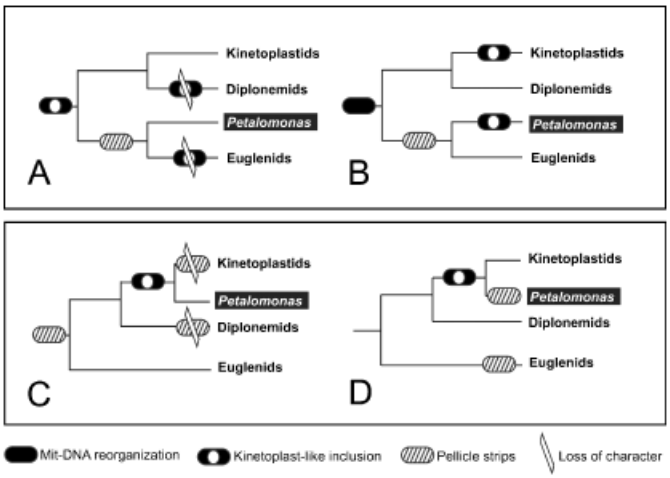

Fig. **1.** Hypothetical frameworks for inferring character evolution within the Euglenozoa, considering two different phylogenetic positions for *Petalomonas*: scenarios (**A**, **B**), *P. cantuscygni* is the sister lineage to all other euglenids; scenarios (**C**, **D**), *P. cantuscygni* is the sister lineage to kinetoplastids. The putative synapomorphies for the Kinetoplastida and the Euglenida are ''kinetoplasts'' and ''pellicle strips,'' respectively. For each possible topology, two alternative scenarios for the evolutionary origin and losses of pellicle strips and (kinetoplast-like) mitochondrial inclusions are phylogenetically mapped. In scenarios A, C, and D, the reorganization of mitochondrial DNA (mtDNA) is inferred to predate or occur concurrently with the presence of the inclusion (as indicated by the black oval). Scenario A invokes the independent loss of a kinetoplast-like inclusion, but not the reorganized mtDNA, in euglenids and diplonemids. Scenario B specifically postulates a reorganization of the mtDNA in all euglenozoans before to the independent acquisition of inclusions in *Petalomonas* and kinetoplastids. Scenario C invokes the independent loss of pellicle strips in kinetoplastids and diplonemids. Scenario D invokes convergent evolution of pellicle strips in *Petalomonas* and euglenids.

Euglenozoa (Busse and Preisfeld 2003; Busse et al. 2003; Leander and Farmer 2001a, b; Leander et al. 2001a, b; Linton et al. 2000; Marin et al. 2003; von der Heyden et al. 2004). This uncertainty about its phylogenetic position coupled with the unusual suite of characters found in *P. cantuscygni* led us to consider two main hypothetical scenarios for understanding character evolution within the Euglenozoa (Fig. 1).

If *P. cantuscygni* is the earliest branching euglenid, then the presence of a kDNA-like inclusion might not be a synapomorphy for the kinetoplastid clade but a plesiomorphic condition present in the ancestor of all euglenozoans (Fig. 1A, B). This scenario requires that kDNA-like inclusions were secondarily lost in the last common ancestor of most euglenids, as well as in diplonemids (Fig. 1A). Alternatively, kDNA-like inclusions might have been independently acquired by kinetoplastids and *P. cantuscygni* (Fig. 1B). If *P. cantuscygni* is more closely allied with kinetoplastids, then the most parsimonious inference would be that kDNA-like inclusions arose once in the last common ancestor of these two lineages (Fig. 1C, D). This scenario would also suggest that the feature inferred to be synapomorphic for euglenids, namely pellicle strips, is actually an ancestral state for all euglenozoans, which were subsequently and independently lost in kinetoplastids and diplonemids (Fig. 1C). Alternatively, pellicle strips might have evolved convergently in euglenids and *P. cantuscygni* (Fig. 1D). It should be noted that two other hypothetical scenarios are formally possible: (1) *P. cantuscygni* is the sister lineage to a clade consisting of euglenids, diplonemids, and kinetoplastids; and (2) *P. cantuscygni* is the sister lineage to a clade consisting of diplonemids and kinetoplastids.

Resolving the branching order of *P. cantuscygni* and other phagotrophic euglenids is necessary to evaluate the hypotheses outlined above. However, they represent several distinctive and divergent lineages that cannot be satisfactorily resolved with SSU rDNA data (Busse and Preisfeld 2002, 2003; Busse et al. 2003; Simpson and Roger 2004). Comparison of nucleus-encoded protein sequences, such as the heat-shock protein 90 (hsp90) gene, provides a viable alternative (Leander and Keeling 2003; Shalchian-Tabrizi et al. 2006; Simpson and Roger 2004; Simpson et al. 2002, 2006a; Stechmann and Cavalier-Smith 2003). Among euglenids, hsp90 sequences are known only from the phototroph *Euglena gracilis*, precluding any attempt to analyze relationships within this group. With the aim of increasing the dataset of available hsp90 sequences from euglenids and addressing our phylogenetic hypotheses (Fig. 1), we sequenced the cytosolic hsp90 gene from three taxa, namely the phagotrophs *Peranema trichophorum*, *Entosiphon sulcatum*, and *P. cantuscygni*.

MATERIALS AND METHODS

Collection of organisms. *Entosiphon sulcatum* was isolated from sediment samples collected from a pond at the Queen Elizabeth Park, Vancouver, BC, in May 2004 (Fig. 2, 3). Cells were manually isolated using drawn-out glass Pasteur micropipettes and temporarily grown at 16 °C in an 802 Sonnerborn's *Paramecium* medium (rye grass Cerophyl of 250 mg/L with *Klebsiella* sp. as food source). *Peranema trichophorum* was manually isolated from a duck pond at Granville Island, Vancouver, BC, in June 2005, and grown at room temperature in a fresh-water KNOP medium with egg yolk as food source (Saito et al. 2003) (Fig. 4). Differential interference contrast light micrographs were captured with a Zeiss Axioplan 2 imaging microscope Gottingen, Germany connected to a Leica DC500 digital color camera (Wetzlar, Germany). Genomic DNA from *P. cantuscygni* (CCAP 1259/1) was a gift from S. Jardeleza and M. Farmer (University of Georgia, Athens, GA).

DNA isolation, amplification, and sequencing. Genomic DNA was extracted from cultures of *E. sulcatum* and *P. trichophorum* using a standard trimethylhexadecylammonium bromide (CTAB) extraction protocol: pelleted material was suspended in 400 L CTAB extraction buffer (1.12 g Tris, 8.18 g NaCl, 0.74 g EDTA, 2 g CTAB, 2 g polyvinylpyrolidone, 0.2 ml 2-mercaptoethanol in 100 ml water), homogenized in a glass tissue grinder, incubated at 65 °C for 30 min and separated with chloroform:isoamyl alcohol (24:1). The aqueous phase was then precipitated in 70% ethanol. Sequences of the cytosolic hsp90 gene were amplified using PCR primers designed from different alignments of euglenozoan hsp90 sequences. *Petalomonas cantuscygni* and *E. sulcatum* genes were amplified with the primer pairs F4–R3 for the 5′-portion of the gene and F2eug-970R for the 3′-half. Primer sequences were: F4, 5′-GGAGCCTGATHATHAAYACNTTYTA-3′; R3, 5′-GATGACYT TNARDATYTTRTT-3′; F2eug, 5′-GTNTTCATYATGGACAACT GYGAGGA-3′; and 970R, 5′-TCGAGGGAGAGRCCNARCTT-RATCAT-3′. Attempts to amplify the hsp90 gene from *P. trichophorum* using these primers were unsuccessful, so we designed new primers using the nucleotide sequences obtained from *E. sulcatum* and *P. cantuscygni*. The new primer sequences were: 3Forward, 5′-CTTGGAACGATTGCCAGA-3′; 627Reverse, 5′-CCAATTGTCC TTCAACAGA-3′; 598Forward, 5′-CGATTGGGAGGACCACTT-3′; and Eug2106R, 5′-GAKAGACCAAGYTTRTCAT-3′. PCR amplifications consisted of an initial denaturing period (95 °C for 3 min), 35 cycles of denaturing (93 °C for 45 s), annealing (55 °C for 45 s), extension (72 °C for 2 min), and a final extension period (72 °C for 5 min). PCR products of the expected size were gel-isolated and cloned into the vector pCR2.1 using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Two to four clones from each product were sequenced with the ABI Big-Dye reaction mix using the vector

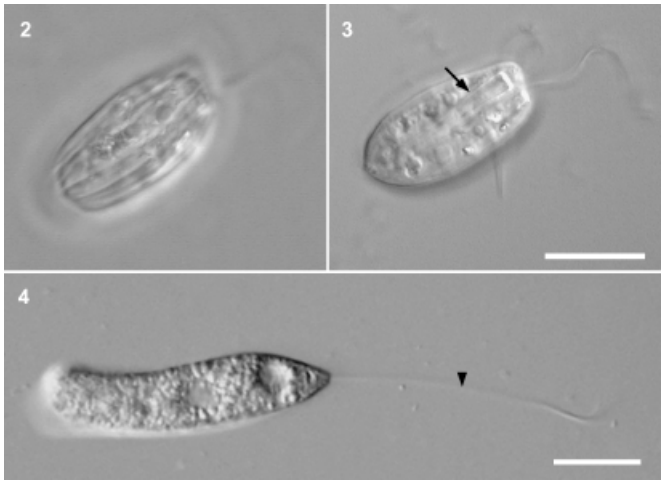

Fig. **2–4.** Light micrographs of wild isolates of two phagotrophic euglenids from which heat-shock protein 90 genes were sequenced. **2.** *Entosiphon sulcatum* showing the longitudinally arranged pellicle strips. **3.** A deeper focal plane of *E. sulcatum* showing the rods of feeding apparatus (arrow). The two heterodynamic flagella are also visible (scale bar = 10 μm). **4.** A gliding *Peranema trichophorum* showing the extended anterior flagellum (arrowhead) and the absence of a conspicuous recurrent flagellum, which is pressed against the cell within a flagellar strip (scale bar = 20 μm).

primers oriented in both directions. New sequences were identified by BLAST and phylogenetic analysis and deposited in GenBank: *P. cantuscygni* (DQ683346), *P. trichophorum* (DQ683345), and *E. sulcatum* (DQ683347).

Alignments. Nucleotide sequences and conceptual amino acids translations were combined with published euglenozoan sequences into two datasets (Table 1). The three new sequences were aligned with 18 representative sequences from kinetoplastids (both trypanosomatids and bodonids), diplonemids, a previous published euglenid (*E. gracilis*), the heterolobosean *Naegleria gruberi*, two plants (*Ipomoea nil* and *Oryza sativa*), and a jakobid (*Reclinomonas americana*). The resultant nucleotide and amino acid datasets were aligned with CLUSTALX (Thompson et al. 1997) and manually adjusted with MacClade. Ambiguously aligned positions, gaps, and the third codon position in the nucleotide dataset were excluded. Final nucleotide and amino acid datasets contained 1,069 and 540 positions for the alignments containing plants and *N. gruberi*, and 925 and 469, when the outgroup was *R. americana*.

Molecular phylogenetic analyses. Phylogenetic relationships were inferred using maximum likelihood (ML), distance, and Bayesian methods with the programs PHYML (Guindon and Gascuel 2003), Weighbor (Bruno, Socci, and Halpern 2000), and MrBayes (Huelsenbeck and Ronquist 2001), respectively. For ML, the alignments of amino acid sequences were analyzed under a WAG model of substitution considering corrections for site-to-site rate variation (γ) with eight categories of rate variation and proportion of invariable sites. Nucleotide datasets were analyzed using a GTR model plus γ correction with eight categories and proportion of invariable sites. For both datasets, 500 bootstrap replicates were performed with the same parameters described above.

Distances were calculated with TREE-PUZZLE 5.0 (Strimmer and Von Haeseler 1996) using the HKY and WAG substitution models for the nucleotide and amino acid datasets, respectively, and among-site rate variation was modeled with a γ correction with eight categories and proportion of invariable sites. One thousand bootstrap datasets were generated with SEQBOOT (Felsenstein 1993). Respective distances were calculated with the shell script ''puzzleboot'' (M. Holder and A. Roger, http://www.tree-puzzle.de) using the parameters estimated from the original datasets (e.g. α shape parameter and nucleotide transition/transversion ratio) and analyzed with Weighbor.

For Bayesian analyses, the program MrBayes was set to operate with GTR for nucleotides, with WAG for amino acids, a γ correction with eight categories and proportion of invariable sites, and four Monte–Carlo–Markov chains (MCMC; default temperature = 0.2). In each case, a total of 2,000,000 generations was calculated with trees sampled every 100 generations and with a prior burn-in of 200,000 generations (i.e. 2,000 sampled trees were discarded). A majority rule consensus tree was constructed from 18,000 post-burn-in trees with PAUP* 4.0. Posterior probabilities correspond to the frequency at which a given node is found in the post-burn-in trees.

Five alternative topologies differing in the relative position of *P. cantuscygni* were generated with MacClade with the dataset having *R. americana* as the ourgroup. Approximately unbiased (AU) tests were performed with CONSEL (Shimodaira and Hasegawa 2001) using the likelihoods calculated with PUZZLE 5.2 (Strimmer and Von Haeseler 1996) with the same models and parameters indicated above.

RESULTS

We sequenced the cytosolic hsp90 gene of the phagotrophic euglenids *P. trichophorum*, *E. sulcatum*, and *P. cantuscygni*. The nucleotide sequence recovered from the PCR product isolated from genomic DNA of *P. trichophorum* contained several frame shifts and was longer than anticipated, suggesting that the hsp90 coding sequence may be interrupted by spliceosomal introns. Through close examination of the sequence using BLASTX, we were able to detect three short intervening sequences whose removal resulted in a continuous ORF and maximized the similarity at the amino acid level of this sequence with other euglenozoan hsp90 genes. The nucleotide sequences of these introns are 5′-GTATGTTCACTTTCCTTTTTCTCTTAG-3′ (27 bp), 5′-GGCAACACTACTAAGATTGTGATGCCTGG-3′ (29 bp), and 5′-GGAAATTTTATGAATTAGTAATTTTTCCATG-3′ (31 bp). The 27-bp intron had canonical GT/AG borders, whereas both the 29 and 31-bp introns had non-canonical borders.

For our phylogenetic analyses overall, the level of statistical support throughout the most commonly encountered tree topology was generally consistent between both nucleotide (excluding the third position) and amino acid datasets (Fig. 5). Our analyses recovered a highly supported sister relationship between the kinetoplastid clade and the diplonemid clade to the exclusion of euglenids. Within the kinetoplastid clade, four species of trypanosomatids formed a strongly supported monophyletic group, and the nodes grouping bodonid lineages showed varying levels of support. Within the euglenid clade, the phagotroph *P. trichophorum* and the phototroph *E. gracilis* branched together to the exclusion of *E. sulcatum*. The position of *P. cantuscygni* was less clear. Our analyses using plants (*I. nil* and *O. sativa*) and heterolobosean (*N. gruberi*) sequences as outgroups placed *P. cantuscygni* as the sister lineage to the kinetoplastid-diplonemid clade (Fig. 5B). Support for this relationship was robust in all analyses using the nucleotide dataset, excluding the third codon position (ML bootstrap = 88, Bayesian posterior probability = 0.98, Weighbor bootstrap = 92). Moreover, this relationship was recovered in ML and distance trees using the complete amino acid dataset, but was only weakly supported with bootstrap analyses (data not shown). However, the Bayesian tree derived from the amino acid dataset placed *P. cantuscygni* as the sister lineage to the euglenid clade. In order to address these inconsistencies,

Table **1.** Accession numbers (NCBI nucleotide and protein databases) for the sequences employed in this study.

Taxon	DNA	Protein
Ipomoea nil	M99431	AAA33748
Oryza sativa	XM_483191	XP_483191
Reclinomonas americana	DQ295221	ABC54646
Naegleria gruberi	AY122634	AAM93756
Ichthyobodo necator	AY651251	AAV66335
Rhynchobodo sp.	AY651252	AAV66336
Leishmania amazonensis	M92926	AAA29250
Leishmania infantum	X87770	CAD30506
Trypanosoma brucei	X14176	CAA32377
Trypanosoma cruzi	M15346	AAA30202
Rhynchopus sp.	AY122622	AAM93744
Diplonema papillatum	AY122623	AAM93745
Dimastigella trypaniformis	AY122624	AAM93746
Rhynchomonas nasuta	AY122625	AAM93747
Neobodo saliens	AY122626	AAM93748
Trypanoplasma borreli	AY122628	AAM93750
Cryptobia salmositica	AY122629	AAM93751
Cryptobia helices	AY122631	AAM93753
Bodo saltans	AY122632	AAM93754
Bodo cf. uncinatus	AY122633	AAM93755
Euglena gracilis	AY288510	AAQ24861
Petalomonas cantuscygni	**DQ683346**	**ABG77329**
Peranema trichophorum	**DQ683345**	**ABG77328**
Entosiphon sulcatum	**DQ683347**	**ABG77330**

New sequences are highlighted in bold letters.

we repeated the analyses using different outgroup schemes by excluding the distant taxa *I. nil* and *O. sativa* and replacing a divergent relative (*N. gruberi*) by the jakobid *R. americana*. However, because the published hsp90 sequence from this species is shorter, our alignment was reduced, accordingly. Without having significant effects on the overall topology, outgroup changes had drastic effects on the position of *P. cantuscygni*. With the new scheme of outgroups, *P. cantuscygni* branched as the most basal lineage within the euglenids with high bootstrap support (Fig. 5A). A similar effect occurred with the protein dataset, but in this case the branching order between *P. cantuscygni* and *E. sulcatum* was not clearly resolved (Fig. 5C). Nonetheless, the support for the inclusion of *P. cantuscygni* within the euglenid clade was robust after *N. gruberi* was replaced by *R. americana* as the outgroup (Fig. 5A).

In order to gain additional insight into how well the data supported the phylogenetic position of *P. cantuscygni*, we performed AU tests for comparing the likelihoods of five alternative topologies differing in the relative position of this taxon (Fig. 6). The topologies assayed were similar to the tree shown in Fig. 5A, except for the position of *P. cantuscygni*, which was alternatively placed as a sister branch to the euglenids (topology A in Fig. 6), to the kinetoplastid–diplonemid group (Fig. 6B), to the kinetoplastids (Fig. 6C), to the diplonemids (D, Fig. 4), and at the base of Euglenozoa (Fig. 6D). The tests were performed on both the nucleotide (excluding the third position) and amino acid datasets. The topologies placing *P. cantuscygni* specifically with either diplonemids or kinetoplastids (Fig. 6) were rejected by the AU test at a 5% level, using the nucleotide and amino acids datasets (Table 2). The topology with *P. cantuscygni* placed as a sister lineage of all other euglenozoas (Fig. 6E) had a considerably lower likelihood than the topologies placing *P. cantuscygni* as sister to either euglenids (Fig. 6A) or the kinetoplastid–diplonemid group (Fig. 6B) (i.e. the AU value was very close to the 5% threshold; Table 2). However, topology B (Fig. 6) was not rejected at a 5% level, attesting to the difficulties in resolving this part of the tree.

DISCUSSION

Non-conventional introns in hsp90 of phagotrophic euglenids. We found that the hsp90 gene of *P. trichophorum* contained at least three putative spliceosomal introns. Although genomic data on euglenids are scarce, some examples of introns in nuclear genes have been reported from *E. gracilis*, and both conventional and non-conventional representatives were found (Breckenridge et al. 1999; Canaday et al. 2001; Henze et al. 1995; Muchhal and Schwartzbach 1994; Russell et al. 2005; Tessier et al. 1995). Introns have not been reported from any hsp90 genes that have been amplified from euglenozoans (Simpson et al. 2002). Nonetheless, like most spliceosomal introns, intron 1 of *P. trichophorum* had typical GT–AG borders and a characteristic pyrimidine-rich tract adjacent to the 3′-border. Introns 2 and 3 had the non-canonical borders GG–GG and GG–TG, respectively. Initially in euglenids, introns with non-canonical borders were found mainly in nuclear genes encoding chloroplast-targeted proteins and, thus, were considered to be of prokaryotic origin (Henze et al. 1995; Muchhal and Schwartzbach 1994; Tessier et al. 1995). However, it was later found that genes encoding for non-plastidic proteins in *E. gracilis* contain both classes of introns (Canaday et al. 2001; Russell et al. 2005). Non-conventional introns are characterized by non-canonical and variable dinucleotides at the borders and by the presence of two inverted repeats that are adjacent to each border of the intron. These repeats would have the capacity for base pairing and for bringing together both splice sites in order to constitute a distinct, and probably spliceosome-free mechanism for splicing (Canaday et al. 2001; Russell et al. 2005). In addition to the non-canonical borders GG–GG and GG–TG, introns 2 and 3 from *P. trichophorum* have two 10-bp inverted repeats arranged in the same fashion as those described in the phototroph *E. gracilis*, showing that non-conventional introns also occur in phagotrophic euglenids. This finding suggests that an endosymbiotic origin of non-conventional introns in *E. gracilis* is unlikely, because eukaryovorous euglenids like *P. trichophorum*, diverged before the endosymbiotic event that gave rise to photosynthesis in euglenids (Leander 2004). Nonetheless, the three introns found in *P. trichophorum* (27, 29, and 31 bp) are smaller than any other intron known from euglenids and are among the smallest introns ever described in eukaryotes (Gilson et al. 2006; Zagulski et al. 2004). Interestingly, the known intron size range in euglenids goes from 27 bp (this study) to 9.2 kbp (Canaday et al. 2001). This wide size variation might explain the difficulties often encountered when amplifying protein-coding genes from genomic DNA in certain species of euglenids.

Phylogeny and character evolution. We have made the first attempt to exploit protein-coding genes to explore the phylogenetic relationships among phagotrophic euglenids. Although still very limited in the number of euglenid taxa, our data suggest that the hsp90 gene constitutes a useful tool for inferring euglenid phylogeny and overcoming some of the limitations exhibited by SSU rDNA sequences. At the same time, our results highlight the importance of appropriate outgroup selection when analyzing euglenozoan phylogeny. Analyses of the nucleotide and amino acid alignments recovered the expected relationships among the three main groups of Euglenozoa (kinetoplastids, diplonemids, and euglenids; Maslov et al. 1999; Moreira et al. 2004; Simpson et al. 2002, 2004, 2006a; Simpson and Roger 2004). The relationships among the phototroph *E. gracilis* and the phagotrophs *P. trichophorum* (with features appropriate for consuming eukaryotic-sized food), and *E. sulcatum* (with features suitable for consuming bacterial-sized food) were similar to those of previous rDNA phylogenies (Busse et al. 2003) and hypothetical scenarios inferred from morphological data (Leander et al. 2001a). Resolving the position of *P. cantuscygni*, however, was more problematic

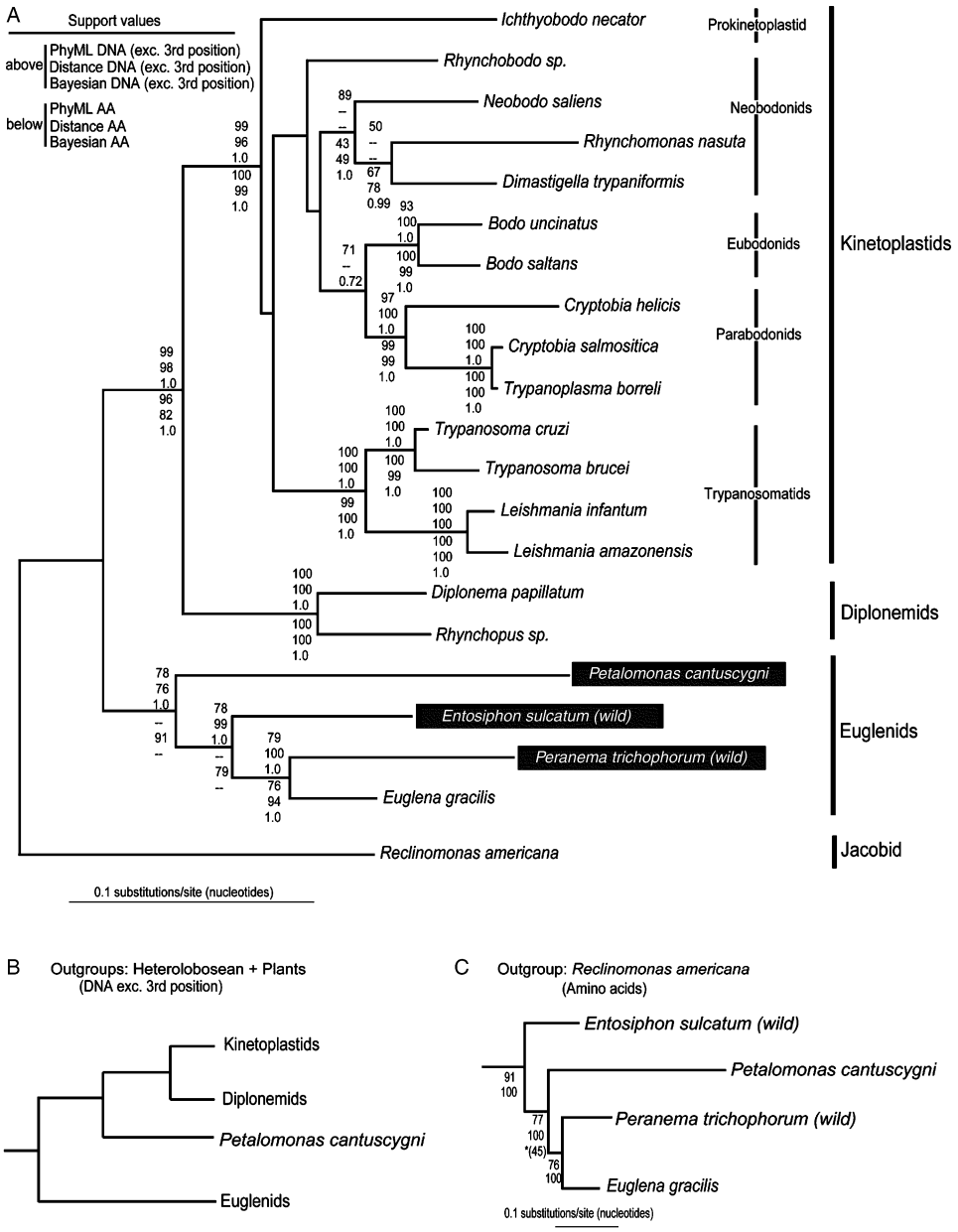

Fig. **5.** Euglenozoan phylogeny as inferred from the heat-shock protein 90 gene. **A.** The topology shown is derived from a+8+I (γ correction with eight categories of rate variation and proportion of invariable sites) maximum likelihood analysis using a GTR model of nucleotide substitutions on 21 unambiguously aligned sequences with the third codon position excluded (925 nucleotide positions). Support values for each node are indicated as bootstrap percentages or Bayesian posterior probabilities as indicated in the upper left-hand table. Numbers above the branches correspond to the nucleotide dataset (third codon position excluded), and numbers below the branches correspond to the amino acid dataset (469 positions). New sequences are indicated with black boxes. **B.** Schematic topology showing the position of *Petalomonas cantuscygni* when the heterolobosean and plant sequences are included as outgroups with the nucleotide dataset (third codon position excluded). This pattern was obtained using maximum likelihood, distance, and Bayesian methods. **C.** Detail of the branching order among euglenids in trees constructed with the amino acids dataset. Numbers at the nodes represent bootstrap percentages using maximum likelihood (above) and Bayesian analysis (below). The bootstrap support of the node indicated (*) drops to 45 when no outgroup is present.

and it was shown to be strongly dependent on outgroup choice. Use of distant taxa (like plants) or a very divergent relative (like *N. gruberi*) apparently distorted the placement of the more divergent taxa in the ingroup, as is the case of *P. cantuscygni*. As the closest relatives to Euglenozoa (Baldauf et al. 2000), a heterolobosean seems to be a proper choice to root the euglenozoan tree, but the only available hsp90 sequence is from *N. gruberi*, an amoeba whose hsp90 sequence is highly divergent and thus potentially problematic. Because jakobids have been shown to branch as sisters to Euglenozoa (Simpson, Inagaki, and Roger 2006a), we decided to include *R. americana* as an outgroup. The addition of *R. americana* (along with *N. gruberi*) increased the support of *P. cantuscygni* branching with the euglenids, and the support for this relationship was significantly higher when we eliminated *N. gruberi* from the dataset. A possible explanation for the different outcome between both datasets is that the highly conserved hsp90 gene lacks sufficient phylogenetic signal at the amino acid level, but contains enough ''hidden'' variation at the nucleotide level

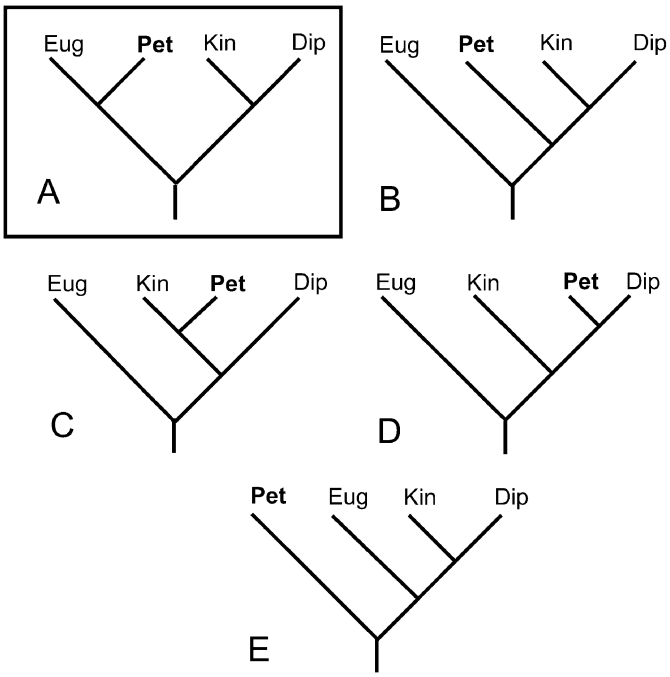

Fig. **6.** Topologies used to evaluate five alternative positions of *Petalomonas cantuscygni* by performing approximately unbiased (AU) likelihood tests with both the nucleotide (third codon position excluded) and the amino acid datasets, with *Reclinomonas americana* as outgroup (see Table 2). Labels at the termini are as follows: Eug, euglenids; Pet, *P. cantuscygni* (bold); Kin, kinetoplastids; and Dip, diplonemids. The topology most favored by the hsp90 phylogenetic analyses and AU likelihood tests is highlighted with a box.

(excluding the ''noise'' created by the third codon position) to robustly resolve the deep phylogenetic position of *P. cantuscygni*. However, whether or not the strong support for the sister relationship between *P. cantuscygni* and the euglenid clade is a consequence of long-branch attraction artifact cannot be definitively ruled out at this stage.

Nonetheless, the sister relationship between *P. cantuscygni* and the euglenid clade is concordant with previous phylogenetic hypotheses based on ultrastructural data and brings to light some interesting considerations. On one hand, it confirms the synapomorphic character of the pellicle as an exclusive euglenid feature. As per the relationships among euglenids, our results broadly agree with SSU rDNA trees (Busse et al. 2003) and the proposed sequence of events leading to the potential for acquiring eukaryotic endosymbionts, which resulted in the origin and diversification of phototrophic euglenids (Leander 2004; Leander et al. 2001a). On the other hand, *P. cantuscygni* has kDNA-like inclusions within its mitochondria (Leander et al. 2001a). This highly compacted structure strongly resembles the poly-kDNA configuration sensu Lukěs et al. (2002), and has not been observed in other euglenids or diplonemids. Because our most robust phylogenetic hypothesis places *P. cantuscygni* at the base of the euglenid side of the tree, convergent evolution of mitochondrial inclusions might explain their presence in both kinetoplastids and *P. cantuscygni* (Fig. 1B). However, the functional integration of kDNA inclusions at the molecular level suggests that independent acquisition of this sophisticated character is highly unlikely, especially in the absence of data to the contrary. Notwithstanding, it is possible to envision that independent compaction processes starting from an unpacked state gave rise to the pan-kDNA of bodonids and the uncharacterized structure of *P. cantuscygni*. This scenario mirrors a hypothesis on kDNA evolution made by Lukěs et al. (2002), where a plesiomorphic pan-kDNA configuration independently gave rise to (1) the mega-kDNA present in the bodonid *Trypanoplasma*, and (2) the kDNA network of trypanosomatids.

Table **2.** *P* values for approximately unbiased likelihood tests of five alternative positions of *Petalomonas cantuscygni* (topologies shown in Fig. 6).

Topology	A	B	C	D	E
DNA (excluding third position)	0.829	0.309	0.002	0.019	0.082
Amino acids	0.902	0.191	0.003	0.008	0.058

We favor a hypothetical scenario in which the ancestral euglenozoan underwent a novel reorganization of its mitochondrial genome. The mitochondrial genome independently underwent different kinds of rearrangements, some of which acquired further complexity and compaction leading to the inclusions observed in kinetoplastids and *P. cantuscygni* (Fig. 1B). Assessing the plausibility of this scenario requires a significantly improved understanding of the structure and organization of mitochondrial genomes in euglenids and diplonemids, which is still very limited. However, there is evidence showing that the diplonemid *Diplonema papillatum* has a fragmented and reorganized mitochondrial genome (Marande et al. 2005) and *E. gracilis* seems to have an unusual mitochondrial genome as well (Gray, Lang, and Burger 2004; Yasuhira and Simpson 1997). We should be able to address these inferences more concretely once we know more about the molecular organization of the mitochondrial inclusions in *P. cantuscygni* and several of its close relatives (e.g. *Notosolenus* and *Calycimonas*), and the structure of mitochondrial DNA in other euglenids.

ACKNOWLEDGMENTS

This work was supported by grants to B. S. L. from the Natural Sciences and Engineering Research Council of Canada (NSERC 283091-04) and from the Canadian Institute for Advanced Research (CIAR). Genomic DNA from *P. cantuscygni* was kindly provided by S. Jardeleza and M. Farmer (University of Georgia, Athens, GA). This research will be submitted by S. A. B. in partial fulfillment of the requirements for the Ph.D. degree, University of British Columbia, Vancouver, BC Canada.

LITERATURE CITED

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, **290**:972–977.

Breckenridge, D. G., Watanabe, Y., Greenwood, S. J., Gray, M. W. & Schnare, M. N. 1999. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. *Proc. Natl. Acad. Sci. USA*, **96**:852–856.

Brosnan, S., Shin, W., Kjer, K. M. & Triemer, R. E. 2003. Phylogeny of the photosynthetic euglenophytes inferred from the nuclear SSU and partial LSU rDNA. *Int. J. Sys. Evol. Microbiol.*, **53**:1175–1186.

Bruno, W. J., Socci, N. D. & Halpern, A. L. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.*, **17**:189–197.

Busse, I. & Preisfeld, A. 2002. Unusually expanded SSU ribosomal DNA of primary osmotrophic euglenids: molecular evolution and phylogenetic inference. *J. Mol. Evol.*, **55**:757–767.

Busse, I. & Preisfeld, A. 2003. Application of spectral analysis to examine phylogenetic signal among euglenid SSU rDNA data sets (Euglenozoa). *Org. Divers. Evol.*, **3**:1–12.

Busse, I., Patterson, D. J. & Preisfeld, A. 2003. Phylogeny of phagotrophic euglenids (Euglenozoa): a molecular approach based on culture material and environmental samples. *J. Phycol.*, **39**:828–836.

Canaday, J., Tessier, L. H., Imbault, P. & Paulus, F. 2001. Analysis of *Euglena gracilis* alpha-, beta- and gamma-tubulin genes: introns and pre-mRNA maturation. *Mol. Gen. Genomics*, **265**:153–160.

Esson, H. J. & Leander, B. S. 2006. A model for the morphogenesis of strip reduction patterns in phototrophic euglenids: evidence for heterochrony in pellicle evolution. *Evol. Dev.*, **8**:378–388.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). University of Washington, Seattle.

Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J. & McFadden, G. I. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc. Natl. Acad. Sci. USA*, **103**:9566–9571.

Gray, M. W., Lang, B. F. & Burger, G. 2004. Mitochondria of protists. *Annu. Rev. Genet.*, **38**:477–524.

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

Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, W. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. USA*, **92**: 9122–9126.

Huelsenbeck, J. P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**:754–755.

Kent, M. L., Elston, R. A., Neral, T. A. & Sawyer, T. K. 1987. An *Isonema*-like flagellate (Protozoa, Mastigophora) infection in larval geoduck clams, *Panope abrupta*. *J. Invertebr. Pathol.*, **50**:221–229.

Kivic, P. A. & Walne, P. L. 1984. An evaluation of the possible phylogenetic relationship between Euglenophyta and Kinetoplastida. *Orig. Life*, **13**:269–288.

Leander, B. S. 2004. Did trypanosomatid parasites have photosynthetic ancestors? *Trends Microbiol.*, **12**:251–258.

Leander, B. S. & Farmer, M. A. 2001a. Comparative morphology of the euglenid pellicle. II. Diversity of strip substructure. *J. Eukaryot. Microbiol.*, **48**:202–217.

Leander, B. S. & Farmer, M. A. 2001b. Evolution of *Phacus* (Euglenophyceae) as inferred from pellicle morphology and SSU rDNA. *J. Phycol.*, **37**:143–159.

Leander, B. S. & Keeling, P. J. 2003. Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from hsp90 and actin phylogenies. *J. Phycol.*, **40**:341–350.

Leander, B. S., Triemer, R. E. & Farmer, M. A. 2001a. Character evolution in heterotrophic euglenids. *Eur. J. Protistol.*, **37**:337–356.

Leander, B. S., Witek, R. P. & Farmer, M. A. 2001b. Trends in the evolution of the euglenid pellicle. *Evolution*, **55**:2115–2135.

Linton, E. W., Nudelman, M. A., Conforti, V. & Triemer, R. E. 2000. A molecular analysis of the euglenophytes using SSU rDNA. *J. Phycol.*, **36**:740–746.

Lukěs, J., Hashimi, H. & Zikova, A. 2005. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr. Genet.*, **48**:277–299.

Lukěs, J., Guilbride, D. L., Votypka, J., Zikova, A., Benne, R. & Englund, P. T. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell*, **1**:495–502.

Marande, W., Lukěs, J. & Burger, G. 2005. Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. *Eukaryot. Cell*, **4**:1137–1146.

Marin, B., Palm, A., Klingberg, M. & Melkonian, M. 2003. Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, **154**:99–145.

Maslov, D. A., Yasuhira, S. & Simpson, L. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. *Protist*, **150**:33–42.

Moreira, D., Lopez-Garcia, P. & Vickerman, K. 2004. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int. J. Syst. Evol. Microbiol.*, **54**:1861–1875.

Morris, J. C., Drew, M. E., Klingbeil, M. M., Motyka, S. A., Saxowsky, T. T., Wang, Z. & Englund, P. T. 2001. Replication of kinetoplast DNA: an update for the new millennium. *Int. J. Parasitol.*, **31**:453–458.

Muchhal, U. S. & Schwartzbach, S. D. 1994. Characterization of the unique intron–exon junctions of *Euglena* gene(s) encoding the polyprotein precursor to the light-harvesting chlorophyll a/b binding protein of photosystem II. *Nucl. Acids Res.*, **22**:5737–5744.

Russell, A. G., Watanabe, Y., Charette, J. M. & Gray, M. W. 2005. Unusual features of fibrillarin cDNA and gene structure in *Euglena gracilis*: evolutionary conservation of core proteins and structural predictions for methylation-guide box C/D snoRNPs throughout the domain Eukarya. *Nucleic Acids Res.*, **33**:2781–2791.

Saito, A., Suetomo, Y., Arikawa, M., Omura, G., Khan, S. M. M. K., Kakuta, S., Suzaki, E., Kataoka, K. & Suzaki, T. 2003. Gliding movement in *Peranema trichophorum* is powered by flagellar surface motility. *Cell Motil. Cytoskel.*, **55**:244–253.

Shalchian-Tabrizi, K., Minge, M. A., Cavalier-Smith, T., Nedreklepp, J. M., Klaveness, D. & Jakobsen, K. S. 2006. Combined heat-shock protein 90 and ribosomal RNA sequence phylogeny supports multiple replacements of dinoflagellate plastids. *J. Eukaryot. Microbiol.*, **53**: 217–224.

Shimodaira, H. & Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, **17**:1246–1247.

Simpson, A. G. B. 1997. The identity and composition of the Euglenozoa. *Arch. Protistenkunde*, **148**:318–328.

Simpson, A. G. B. & Roger, A. J. 2004. Protein phylogenies robustly resolve deep-level relationships within Euglenozoa. *Mol. Phyl. Evol.*, **30**:201–212.

Simpson, A. G. B., Lukěs, J. & Roger, A. J. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. *Mol. Biol. Evol.*, **19**: 2071–2083.

Simpson, A. G. B., Inagaki, Y. & Roger, A. J. 2006a. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of ''primitive'' eukaryotes. *Mol. Biol. Evol.*, **23**:615–625.

Simpson, A. G. B., Stevens, J. R. & Lukěs, J. 2006b. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.*, **22**:168–174.

Simpson, A. G. B., Gill, E. E., Callahan, H. A., Litaker, R. W. & Roger, A. J. 2004. Early evolution within kinetoplastids (Euglenozoa), and the late emergence of trypanosomatids. *Protist*, **155**:407–422.

Stechmann, A. & Cavalier-Smith, T. 2003. Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J. Mol. Evol.*, **57**:408–419.

Strimmer, K. & Von Haeseler, A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**:964–969.

Suzaki, T. & Williamson, R. E. 1985. Euglenoid movement in *Euglena fusca*: evidence for sliding between pellicular strips. *Protoplasma*, **124**: 137–146.

Suzaki, T. & Williamson, R. E. 1986a. Pellicular ultrastructure and euglenoid movements in *Euglena ehrenbergii* Klebs and *Euglena oxyuris* Schmarda. *J. Protozool.*, **33**:165–171.

Suzaki, T. & Williamson, R. E. 1986b. Ultrastructure and sliding of pellicular structures during euglenoid movement in *Astasia longa* Pringsheim (Sarcomastigophora, Euglenida). *J. Protozool.*, **33**:179–184.

Talke, S. & Preisfeld, A. 2002. Molecular evolution of euglenozoan paraxonemal rod genes *par1* and *par2* coincides with phylogenetic reconstruction based on small subunit rdna data. *J. Phycol.*, **38**:995–1003.

Tessier, L. H., Paulus, F., Keller, M., Vial, C. & Imbault, P. 1995. Structure and expression of *Euglena gracilis* nuclear rbcs genes encoding the small subunits of the ribulose 1,5-bisphosphate carboxylase/oxygenase: a novel splicing process for unusual intervening sequences? *J. Mol. Biol.*, **245**:22–33.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997. The clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.*, **24**:4876–4882.

von der Heyden, S., Chao, E. E., Vickerman, K. & Cavalier-Smith, T. 2004. Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of Euglenozoa. *J. Eukaryot. Microbiol.*, **51**:402–416.

Willey, R. L., Walne, P. L. & Kivic, P. A. 1988. Phagotrophy and the origins of the euglenoid flagellates. *CRC Crit. Rev. Plant Sci.*, **7**:303–340.

Yasuhira, S. & Simpson, L. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. *J. Mol. Evol.*, **44**:341–347.

Zagulski, M., Nowak, J. K., Le Mouel, A., Nowacki, M., Migdalski, A., Gromadka, R., Noel, B., Blanc, I., Dessen, P., Wincker, P., Keller, A. M., Cohen, J., Meyer, E. & Sperling, L. 2004. High coding density on the largest *Paramecium tetraurelia* somatic chromosome. *Curr. Biol.*, **14**:1397–1404.

Received: 06/13/06, 09/18/06, 10/17/06; accepted: 10/17/06