Phylogeny of the photosynthetic euglenophytes inferred from the nuclear SSU and partial LSU rDNA

Stacy Brosnan,¹ Woongghi Shin,² Karl M. Kjer³ and Richard E. Triemer²

Correspondence

Stacy Brosnan

szimmer@rci.rutgers.edu

¹Division of Life Sciences, Rutgers University, Nelson Hall, 604 Allison Road, Piscataway, NJ 08854, USA

²Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

³Department of Entomology, Rutgers University, Cook College, Blake Hall, 93 Lipman Drive, New Brunswick, NJ 08901, USA

Previous studies using the nuclear SSU rDNA have indicated that the photosynthetic euglenoids are a monophyletic group; however, some of the genera within the photosynthetic lineage are not monophyletic. To test these results further, evolutionary relationships among the photosynthetic genera were investigated by obtaining partial LSU nuclear rDNA sequences. Taxa from each of the external clades of the SSU rDNA-based phylogeny were chosen to create a combined dataset and to compare the individual LSU and SSU rDNA datasets. Conserved areas of the aligned sequences for both the LSU and SSU rDNA were used to generate parsimony, log-det, maximum-likelihood and Bayesian trees. The SSU and LSU rDNA consistently generated the same seven terminal clades; however, the relationship among those clades varied depending on the type of analysis and the dataset used. The combined dataset generated a more robust phylogeny, but the relationships among clades still varied. The addition of the LSU rDNA dataset to the euglenophyte phylogeny supports the view that the genera *Euglena*, *Lepocinclis* and *Phacus* are not monophyletic and substantiates the existence of several well-supported clades. A secondary structural model for the D2 region of the LSU rDNA was proposed on the basis of compensatory base changes found in the alignment.

INTRODUCTION

The euglenophytes are a distinct group of protists consisting of green phototrophic species and colourless phagotrophic and osmotrophic species. These organisms are characterized by a proteinaceous pellicle composed of individual strips lined by microtubules, a β -1,3-glucan storage product known as paramylon, an intranuclear mitotic spindle with a persistent nucleolus, condensed chromosomes throughout the cell cycle and paraxial rods associated with the flagella (Leedale, 1967). The evolutionary relationships among these

taxa are unclear from the available data, although various classification schemes have been proposed (Stein, 1878; Klebs, 1883; Senn, 1900; Hollande, 1952; Huber-Pestalozzi, 1955; Leedale, 1967; Bourrelly, 1970).

The taxonomic system in common use today has its roots in that proposed by Hollande (1952), who created a scheme that was based largely on body shape (radial symmetry versus asymmetry), nutritional mode (phototrophic, osmotrophic or phagotrophic), flagellar structure and degree of metaboly (rigid versus metabolic). His classification consisted of three major groups, each of which included both pigmented and colourless genera. The presence or absence of chloroplasts was considered to be of minor importance. Later, Leedale (1967) incorporated new physiological and electron microscopic information into the Hollande scheme and created a modified scheme that recognized six separate orders based on similar morphological features. Like Hollande, he grouped photosynthetic and colourless forms together.

Recently, a re-evaluation of the phylogeny of the euglenophytes has been conducted using molecular data (Thompson *et al.*, 1995; Montegut-Felkner & Triemer, 1997; Linton *et al.*,

Published online ahead of print on 8 November 2002 as DOI 10.1099/ijs.0.02518-0.

Abbreviation: AIC, Akaike information criterion.

The GenBank/EMBL/DDBJ accession numbers for the LSU rDNA sequences determined in this study are AY130223–AY130242 and AY130814–AY130826, as listed in Table 1.

Nexus files for three datasets generated for phylogenetic analysis, the SSU rDNA, the LSU rDNA and a combined dataset, a complete alignment of all 34 taxa and a proposed secondary structure of the portion of the LSU rDNA sequenced in this analysis from *Euglena gracilis* are available as supplementary material in IJSEM Online.

1999, 2000; Preisfeld *et al.*, 2000, 2001; Moreira *et al.*, 2001; Müllner *et al.*, 2001; Milanowski *et al.*, 2001; Nudelman *et al.*, 2003) in combination with morphological characteristics (Leander & Farmer, 2000, 2001a, b; Leander *et al.*, 2001).

The earliest molecular study (Thompson *et al.*, 1995) was conducted using the chloroplast gene encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcl*). Their primary objective, however, was to analyse *Euglena* chloroplast intron evolution, and they therefore focused entirely on the genus *Euglena*. A high degree of homoplasy in *rbcl* and misidentification of some taxa prevented Thompson *et al.* (1995) from strongly supporting their phylogeny. Milanowski *et al.* (2001) later sequenced the plastid SSU (16S) rDNA for 17 species belonging to the Euglenales. Their analysis also led to the conclusion that the genus *Euglena* is not monophyletic; however, this gene provided only weak resolution of the internal nodes of the phylogeny.

More recent studies focusing on the nuclear SSU rDNA have confirmed previous speculations, derived from morphological data, that phagotrophy arose early in euglenoid evolution, with the subsequent acquisition of phototrophy, osmotrophy and parasitism (Montegut-Felkner & Triemer, 1997; Linton *et al.*, 1999; Preisfeld *et al.*, 2000). The osmotrophic taxa were polyphyletic, and osmotrophy was shown to have evolved many times. In some instances, osmotrophy arose secondarily due to the loss of photosynthesis (Müllner *et al.*, 2001). In the *Euglena* clade, for example, the presence of the osmotrophic taxa *Astasia longa* and *Khawkinia quartana* indicated that they had lost their chloroplasts (Linton *et al.*, 1999, 2000; Müllner *et al.*, 2001), which correlated with previous data demonstrating the presence of a partial chloroplast genome in *A. longa* (Seimeister & Hachtel, 1990; Gockel *et al.*, 1994).

The phototrophic taxa formed a monophyletic assemblage at the apex of the tree consisting of two clades, one with taxa having two emergent flagella (*Eutreptiales sensu* Leedale) and the other with taxa having one emergent flagellum (*Euglenales sensu* Leedale) (Preisfeld *et al.*, 2001; Moreira *et al.*, 2001). The *Eutreptiales* diverged prior to the Euglenales and were a paraphyletic group (Preisfeld *et al.*, 2000, 2001; Müllner *et al.*, 2001; Moreira *et al.*, 2001). Within the Euglenales, the phototrophic genera *Euglena*, *Phacus* and *Lepocinclis* were also paraphyletic (Linton *et al.*, 1999, 2000; Müllner *et al.*, 2001; Leander & Farmer, 2001b; Nudelman *et al.*, 2003).

The need for a revision of euglenoid taxonomy is apparent. To accomplish this, however, a clear understanding of the relationships among species is necessary. The chloroplast genes examined to date are unable to resolve internal nodes of the photosynthetic taxa and are not applicable to the colourless genera. Because mitochondrial genes evolve more rapidly (Woese, 1996), they are inappropriate for examination of more distant evolutionary relationships. Therefore, a portion of the nuclear LSU rDNA was chosen to resolve

further the phylogeny of the photosynthetic euglenoids. Organisms were included from each of the external clades of the SSU rDNA-based phylogeny to create a combined dataset and compare the LSU and SSU rDNA phylogenies.

METHODS

Strains. Thirty-four taxa, previously sequenced for the SSU rDNA, were sequenced for the LSU rDNA in this analysis. The names, culture-collection information and GenBank accession numbers of the 34 taxa sequenced for both SSU and LSU rDNA can be found in Table 1. Currently, two taxa included for analysis are incorrectly identified in the UTEX culture collection (Culture Center of Algae, Austin, TX, USA): *Euglena agilis* UTEX 1605 is misidentified as *Euglena pisciformis* var. *obtusa*, its synonym (Linton *et al.*, 1999), and *Phacus oscillans* UTEX 1285 is misidentified as *Phacus caudata* (see Linton *et al.*, 2000). Twenty-nine of the 34 taxa sequenced for the SSU rDNA came from the same cultures as those from which the LSU rDNA was sequenced. *Colacium vesiculosum*, *Euglena oxyuris*, *Hyalophacus ocellatus*, *Lepocinclis bütschlii*, *Phacus acuminatus* and *Phacus pleuronectes* were sequenced from different cultures for the LSU and SSU rDNA. Four of these six organisms were sequenced from Korean isolates for the SSU rDNA and were unavailable for further study. *H. ocellatus* and *Phacus acuminatus* were sequenced from the cultures SAG 228.80 and UTEX 1288, respectively, for the SSU rDNA. *Peranema trichophorum*, a colourless phagotrophic species, and *Eutreptiella* sp., a metabolic photosynthetic species with two emergent flagella, were sequenced as outgroups.

DNA isolation, amplification and sequencing. Genomic DNA was isolated from a centrifuged pellet of cultured cells using either the Chelex procedure (Goff & Moon, 1993) with Chelex 100 resin (Bio-Rad) or the DNeasy Tissue kit (Qiagen) using the animal-tissues protocol. A 1.3–1.6 kb region of the LSU rDNA starting at stem 11 (Larsen, 1992) and ending at stem 35a was amplified using 15–20 ng total genomic DNA or 10 µl DNA-containing solution from the Chelex extraction. A 100 µl reaction contained 10 µl DNA (20 ng), 59.5 µl distilled, deionized water, 10 µl 10 × reaction buffer with 6 µl 25 mM MgCl₂, 4 µl 10 mM dNTPs (Applied Biosystems), 5 µl each primer (10 mM) and 0.5 µl DNA polymerase (5 U µl⁻¹; Sigma). Amplification was carried out using the following program: a hot start at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 2 min, annealing at 37–45 °C (depending on strain) for 2 min and extension at 72 °C for 4 min, finishing with an extension at 72 °C for 11 min and an indefinite hold at 4 °C. Primers specific to the LSU rDNA in euglenoids were created by comparing the LSU rDNA of *Euglena gracilis*, which had been sequenced previously by M. N. Schnare and others (GenBank accession no. X53361), with both published LSU rDNA primers (Freshwater & Bailey, 1998; Palumbi, 1996) and the LSU rDNA of other distant taxa, such as *Mus*. Three published LSU rDNA primers were modified slightly and an additional 17 primers were created (Table 2). The entire LSU rDNA fragment used in this analysis was amplified for each organism using primers 1F and C1R. The PCR product was then sized, cut from an agarose gel and purified using either a QIAEX II gel extraction kit (Qiagen) or the MinElute gel extraction kit according to manufacturer's protocol. The purified DNA was sequenced using an ABI 377 dye-terminator cycle sequencer (Perkin Elmer Applied Biosystems). The entire 1.3–1.6 kb fragment was sequenced for most organisms using three forward primers and three reverse primers: 1F, 2F, 1XF, 2R, 1XR and C1R. The additional primers were created for regions that were not adequately conserved for all taxa sequenced. Forward and reverse primers were used to generate the final LSU rDNA sequence for each organism. Any discrepancies among sequences from different primers were resolved by reading the ABI chromatogram for those primers.

Table 1. Sequenced strains

Abbreviations: ASW, Algenkulture-Sammlung an der Universität Wien, Vienna, Austria; ATCC, American Type Culture Collection, Manassas, VA, USA; CBS, Carolina Biological Supply Company, Burlington, NC, USA; CCMP, Provasoli–Guillard National Center for Culture of Marine Phytoplankton, West Boothbay Harbor, ME, USA; SAG, Sammlung von Algenkulturen Pflanzenphysiologisches Institut der Universität Göttingen, Germany; UTEX, Culture Center of Algae, Austin, TX, USA; UW, Culture Collection of Algae at the Department of Plant Systematics and Geography of Warsaw University, Poland.

Organism	Culture used for sequencing		GenBank accession numbers	
	LSU rDNA	SSU rDNA	LSU rDNA	SSU rDNA
<i>Astasia longa</i>	UTEX 512	Same	AY130223	AF112871
<i>Colacium mucronatum</i>	UTEX LB 2524	Same	AY130224	AF326232
<i>Colacium vesiculosum</i>	UW Łazienki	Korean isolate	AY130225	AF081592
<i>Euglena acus</i> var. <i>major</i>	UTEX 1316	Same	AY130226	AF152104
<i>Euglena agilis</i> (<i>Euglena pisciformis</i>)	UTEX 1605	Same	AY130227	AF115279
<i>Euglena gracilis</i>	–	–	X53361	M12677
<i>Euglena oxyuris</i>	SAG 1224-10b	Korean isolate	AY130814	AF090869
<i>Euglena spirogyra</i>	UTEX 1307	Same	AY130228	AF150935
<i>Euglena stellata</i>	UTEX 372	Same	AY130229	AF150936
<i>Euglena tripteris</i>	UTEX 1311	Same	AY130230	AF445459
<i>Euglena viridis</i>	UTEX 85	Same	AY130231	AF112872
<i>Eutreptiella</i> sp.	CCMP 389	Same	AY130232	AF112875
<i>Hyalophacus ocellatus</i>	SAG 1237-1	SAG 228.80	AY130822	AF445458
<i>Khawkinea quartana</i>	ATCC 30895	Same	AY130233	U84732
<i>Lepocinclis bütschlii</i>	UTEX 523	Korean isolate	AY130815	AF096993
<i>Lepocinclis ovum</i>	SAG 1244-8	Same	AY130235	AF110419
<i>Peranema trichophorum</i>	CBS CE 13-1838	Same	AY130826	U84733, U84734
<i>Phacus acuminatus</i>	SAG 1261-1	UTEX 1288	AY130821	AF283311
<i>Phacus acuminatus</i> (<i>Phacus brachykentron</i>)	UTEX 1317	Same	AY130820	AF286209
<i>Phacus agilis</i>	SAG 10.88	Same	AY130236	AY014998
<i>Phacus oscillans</i> (misnamed as <i>Phacus caudata</i>)	UTEX 1285	Same	AY130823	AF181968
<i>Phacus pleuronectes</i>	SAG 1261-3b	Korean isolate	AY130824	AF081591
<i>Phacus pleuronectes</i> var. <i>triquetra</i> (<i>Phacus alatus</i>)	SAG 1261-2b	Same	AY130825	AY014999
<i>Phacus pusillus</i>	UTEX 1282	Same	AY130237	AF190815
<i>Phacus pyrum</i>	UTEX 2354	Same	AY130238	AF112874
<i>Phacus pyrum</i> var. <i>ovata</i> (<i>Lepocinclis ovata</i>)	SAG 1244-5	Same	AY130234	AF061338
<i>Phacus similis</i>	SAG 58.81	Same	AY130239	AF119118
<i>Phacus striatus</i> (<i>Phacus aenigmaticus</i>)	ASW 08012	Same	AY130819	AF283313
<i>Strombomonas acuminata</i> (<i>Strombomonas verrucosa</i>)	NJ (US) isolate	Same	AY130240	AF445461
<i>Strombomonas acuminata</i> (<i>Strombomonas verrucosa</i> var. <i>conspersa</i> , <i>Trachelomonas conspersa</i>)	SAG 1280-1	Same	AY130241	AY015000
<i>Trachelomonas echinata</i>	SAG 1283-22	Same	AY130242	AY015001
<i>Trachelomonas hispida</i>	UTEX 1326	Same	AY130817	AF090377
<i>Trachelomonas hispida</i> (<i>Trachelomonas oblonga</i> var. <i>punctata</i>)	UTEX 1325	Same	AY130818	AF445462
<i>Trachelomonas volvocinopsis</i> var. <i>spiralis</i>	UTEX 1313	Same	AY130816	AY015004

Sequence alignment. SSU and LSU rDNA sequences were manually aligned as suggested by Sogin & Gunderson (1987), using the genetic data environment (GDE 2.2) program (Smith *et al.*, 1994). The secondary structure of *Euglena gracilis* was used as a guide in the alignment of the SSU rDNA, as described by Linton *et al.* (2000). LSU rDNA sequences were aligned with the published *Euglena gracilis* secondary-structure model of Wuyts *et al.* (2001) and Schnare *et al.* (1996), as recommended by Kjer (1995), using Microsoft Word on a PC. Stem-loop structures were numbered according to Larsen (1992). Both of the published secondary-structure models were based on a single species of *Euglena* and were used as a starting point to infer LSU rDNA structure. With our additional sequences, we were able to

design a euglenoid secondary-structure model for the D2 region, using comparative analysis (Gutell *et al.*, 1994; Kjer, 1995), as shown in Fig. 1. The portion of the LSU rDNA sequenced contained three intron regions. These regions were determined for each sequence by comparing them with the published *Euglena gracilis* sequence from which the introns had been spliced out. These same introns were also located within regions of conserved secondary structure and could be determined by flanking stem regions as well. Only conserved regions that could be unambiguously aligned were used in the analyses. Because of length heterogeneity and insertion of introns in the portion of the LSU rDNA sequenced, a large number of nucleotides could not be used in the analyses. Of the 1300–1600 bp sequenced,

Table 2. Primer sequences

Forward primers end in F and reverse primers end in R. Primers containing 'Eutrep' within the name are specific to *Eutreptiella* sp., and those containing 'oxy' are specific for *Euglena oxyuris*.

Primer	Sequence (5'–3')	Secondary-structure location
1F	TTAAGCATATCACTCAGTGGAGG	Stem 11
1VF*	GATTGCTGCAGTAATGGCGAATG	Stem 13
B18R	CCCAAGCCACTCTACTC	Stem 21
B18F	GAGTAGAGTGGCTTGGG	Stem 21
2F	GTAGAGTGATCGAAGGATG	Stem 23
2R	CATCCTTCGATCACTCTAC	Stem 23
B2F	GAAGGATGCAAAGGAC	Between stems 23 and 24
B2R	GTCCTTTGCATCCTTC	Between stems 23 and 24
C2F	CCGATAGYAAACAAGTA	Between stems 4' and 23
C2R	TACTTGTTRCTATCGG	Between stems 4' and 23
Eutrep 2F	GTAGACTGATCGAAACGTG	Stem 23
Eutrep 2R	CAGGCACTTTGAACCTCTTTG	Stem 23
Eutrep M1F	GTTTGACTTTGCTTTGG	Intron 4
MoxyF	TGTGGTGGCAAACCTGCT	D2
MoxyR	CAGGCCCTGGCAAGCAGGGT	D2
1XF*	GAAACACGGACCAAGGAG	Stem 26
1XR*	AGACTCCTTGGTCCGTGTTTCGAGAC	Stem 26
D6R	GTTACCACCTTTCGGG	Stem 32
C1R†	GCTATCCTGAGGGAAACTTCG	Stems 36 and 37
3R	CGGCAGGTGAGTTGTTAC	Stem 44

*Modified versions of primers IV and IX of Palumbi (1996).

†Modified version of primer 28C¹ of Freshwater & Bailey (1998).

only 683 nucleotides were used in the parsimony, distance, maximum-likelihood and Bayesian analyses. Of the 1800–2200 bp sequenced for the SSU rDNA, 1518 were used in the phylogenetic analyses.

Signal analysis. The g1 statistic was used to determine whether phylogenetic signal was significantly non-random (Hillis & Huelsenbeck, 1992). Taxa from highly supported clades were sequentially removed and 100 000 random trees were generated to ensure that signal was not present for only one or two strongly supported clades. Signal was also evaluated for relationships among clades by randomly choosing nine taxa (two outgroups and one taxon from each of the seven clades) for each dataset. Ten sets of nine taxa were generated randomly to ensure that signal was not based on relationships of specific taxa in the analysis (Kjer *et al.*, 2001). The left skew of tree distributions from an exhaustive search for each set of nine taxa was examined and the g1 statistic was compared with the significance levels presented by Hillis & Huelsenbeck (1992). Taxa were then removed from the set of nine until signal became random. One of the outgroup taxa, *Peranema trichophorum*, was removed first, followed by the taxa representing either clade G or clade F (see Fig. 2).

Phylogenetic analysis. Three datasets were generated for phylogenetic analysis: the SSU rDNA, the LSU rDNA and a combined dataset; nexus files are available as supplementary data in IJSEM Online. It was determined that the sequences of three pairs of taxa were identical. One of the identical sequences for each pair was removed from the analysis and, therefore, only 31 taxa are represented in the phylogenies. Parsimony, distance and maximum-likelihood analyses were performed using PAUP* 4.0b10 (Swofford, 2002) and Bayesian analyses were run using MrBayes 2.01 (Huelsenbeck & Ronquist, 2001).

Equally weighted parsimony analyses were conducted on the three datasets. The equally weighted parsimony analysis (heuristic search with random stepwise addition and 10 000 repetitions) carried out to find a minimum-length tree(s) was performed using tree bisection/reconstruction branch swapping, ACCTRAN character state optimization, MULTITREES on, gaps equivalent to missing data and multistate taxa coded as uncertainty. All nucleotides were treated as unordered. Tree support was examined using both the decay index for each node (Bremer, 1994; Donoghue *et al.*, 1992) and non-parametric bootstraps (Felsenstein, 1985). Decay indices were calculated using AutoDecay 4.0 (Eriksson, 1998). For bootstrapping, 1000 pseudoreplicates were performed, each with 10 random addition searches.

To examine the influence of possible nucleotide biases, a log-det minimum evolution analysis (Lockhart *et al.*, 1994) was performed on each dataset. Only parsimony-informative characters were included in the analysis, as suggested by Lockhart *et al.* (1994). The outgroup, *Peranema trichophorum*, was removed from the LSU rDNA and combined analyses because some ambiguously aligned regions of the *Peranema* LSU rDNA sequence were coded as missing, which creates undefined distances in log-det. A heuristic search with starting trees obtained via neighbour-joining using a tree bisection/reconstruction branch-swapping algorithm was used. One hundred bootstraps were generated to examine support for the analyses using 100 pseudo-replicates, with one random addition search.

Evolutionary models were examined for each dataset using both maximum-likelihood and Bayesian analyses. The evolutionary model and parameters were selected after running MODELTEST 3.06 (Posada & Crandall, 1998). For both datasets, MODELTEST indicated the need for a general time-reversible model with a gamma correction for among-site

Table 3. Bayesian parameters

Parameter	A–C	A–G	A–T	C–G	C–T	%A	%C	%G	%T	Gamma	%inv.
SSU rDNA											
Mean	1.8308	4.3804	1.6727	0.8779	6.7507	0.2345	0.2406	0.2878	0.2372	0.5379	0.2253
SD	0.0794	0.2854	0.0552	0.0199	0.6247	0.0001	0.0001	0.0001	0.0001	0.0034	0.0013
95 % CI	1.3564– 2.4039	3.3931– 5.4478	1.2462– 2.1497	0.6120– 1.1988	5.2527– 8.3645	0.2179– 0.2515	0.2237– 0.2569	0.2686– 0.3070	0.2372– 0.2544	0.4372– 0.6620	0.1494– 0.2909
LSU rDNA											
Mean	1.5064	3.6235	1.6668	0.4569	8.6853	0.2245	0.2261	0.3331	0.2162	0.5688	0.2027
SD	0.1165	0.3887	0.1191	0.0140	2.0801	0.0001	0.0002	0.0002	0.0001	0.0070	0.0022
95 % CI	0.9189– 2.2464	2.5997– 5.0152	1.0835– 2.4241	0.2631– 0.7443	6.3038– 11.7017	0.2005– 0.2492	0.2047– 0.2497	0.3065– 0.3585	0.1972– 0.2359	0.4242– 0.7413	0.0998– 0.2842
Combined											
Mean	1.4940	3.6867	1.4807	0.6776	6.4580	0.2339	0.2376	0.2972	0.2314	0.5320	0.2139
SD	0.0339	0.1118	0.0259	0.0067	0.3139	0.0000	0.0001	0.0001	0.0000	0.0022	0.0009
95 % CI	1.1682– 1.1917	3.0608– 4.4088	1.1685– 1.8190	0.5077– 0.8567	5.5003– 7.6332	0.2202– 0.2488	0.2242– 0.2531	0.2816– 0.3137	0.2178– 0.2451	0.4435– 0.6295	0.1501– 0.2686

rate variation and invariant sites according to both the likelihood ratio test and the Akaike information criterion (AIC). The only difference between the model selected by the likelihood ratio test and the model selected by the AIC was that the former model assumed equal nucleotide frequencies and only provided values for two rate-matrix substitutions, leaving the others as a value of 1.0. Initial Bayesian analyses were run with a general time-reversible model (nst=6) with rates set to invgamma and nucleotide frequencies set to equal. Each analysis was performed using four Markov chains, with 500 000 cycles for each chain. Trees were saved to a file every 100 cycles, and the first 1000 trees were discarded. Therefore, a majority-rule consensus tree was created from the remaining 4000 trees to examine the posterior probabilities of each clade. The resulting Bayesian evolutionary model parameters were examined for each dataset after completion of the analysis using the 'sump' command. The mean, variance and 95 % confidence interval for the percentage of invariable sites, gamma, the nucleotide frequencies and the general time-reversible R-matrices are given (Table 3). Bayesian analyses were also run using the specified likelihood ratio test and AIC parameters from MODELTEST (Table 4). Since the likelihood ratio test and AIC parameters generated nearly identical tree topologies, only trees based on the AIC parameters are presented. Maximum-likelihood analyses (heuristic search with random stepwise addition and 10 repetitions, tree bisection/reconstruction branch swapping, starting branch length using the Rogers–Swofford approximation method and MULTREES on) were performed using both the parameters specified in the AIC of MODELTEST and the mean value of the parameters produced from the initial Bayesian analyses. One hundred bootstraps were run for each maximum-likelihood analysis using the fast stepwise-addition option. For the Bayesian and maximum-likelihood analysis of the LSU, the outgroup was constrained as monophyletic.

RESULTS

Taxonomic issues

Lepocinclis ovata SAG 1244-5 is morphologically similar to *Phacus pyrum*, the organism with which it forms a clade in SSU rDNA phylogenies (Linton *et al.*, 2000), and shares 95 % sequence similarity. It was originally described as *Phacus pyrum* var. *ovata* Playfair (1921), and was later moved to the genus *Lepocinclis* by Conrad (1934). Because of its similarity in morphology and rDNA sequences, we suggest moving *L. ovata* back to its original placement in the genus *Phacus* as *Phacus pyrum* var. *ovata*. However, it should be noted that a 5 % sequence divergence exists between *Phacus pyrum* and *Phacus pyrum* var. *ovata* (syn. *L. ovata*). Whether this divergence is sufficient to warrant the establishment of a novel species is unknown at present.

The culture *Trachelomonas conspersa* SAG 1280-1 is currently incorrectly named and has a tangled taxonomic history. Originally, *T. conspersa* was renamed *Strombomonas verrucosa* var. *conspersa* by Deflandre (1930) in his monograph establishing the genus *Strombomonas*. However, Popova (1955) questioned the validity of the species *S. verrucosa* Defl. (1930) because of its similarity in morphology to another species of *Strombomonas* described by Deflandre (1930), *Strombomonas acuminata* (Schmarda) Defl. (1930). On the basis of the original material of *S. acuminata* and

Table 4. AIC from MODELTEST

Dataset	A–C	A–G	A–T	C–G	C–T	%A	%C	%G	%T	Gamma	Invar.
SSU rDNA	1.6094	3.9745	1.4901	0.8008	5.9186	0.2357	0.2429	0.2848	0.2366	0.5456	0.2359
LSU rDNA	1.1413	2.9615	1.3830	0.3856	6.5817	0.2300	0.2323	0.3244	0.2132	0.5491	0.1958
Combined	1.4284	3.5259	1.4442	0.6453	6.1324	0.2342	0.2380	0.2989	0.2289	0.5401	0.2216

S. verrucosa, she showed that the cell morphology is very variable for both species, and the shape, size and ornamentation of the lorica depend on the age of the cells and their culture conditions. The name *S. acuminata* takes priority over *S. verrucosa* and, therefore, in the opinion of Popova (1966), only one taxon exists, *S. acuminata* (Schmarda) Defl. (1930), and all other names should be considered as synonyms. Therefore, the morphological forms *S. verrucosa* and *S. verrucosa* var. *conspersa* are currently synonymized with *S. acuminata*. It should be noted that these two taxa have respective sequence divergence of 1.4 and 1.8% for the LSU rDNA and SSU rDNA, and *S. verrucosa* var. *conspersa* is morphologically distinguishable from *S. verrucosa*. However, until additional *Strombomonas* species are sequenced, the amount of divergence necessary to classify a taxon as either a variety or a novel species is unknown. *S. verrucosa* and *S. verrucosa* var. *conspersa* will thus be considered as *S. acuminata* until additional sequence information becomes available.

Three pairs of taxa were found to be identical in both their SSU and LSU rDNA sequences. They were *Trachelomonas oblonga* var. *punctata* Lemm. and *Trachelomonas hispida* (Perty) Stein em. Defl., *Phacus alatus* Klebs and *Phacus pleuronectes* (O.F.M.) Dujardin and *Phacus acuminatus* Stokes and *Phacus brachykentron* Pochm. The strain of *T. oblonga* var. *punctata* obtained from the culture collection was misidentified and should be renamed as *T. hispida*.

In the genus *Phacus*, different taxonomic issues arise. *Phacus alatus* and *Phacus pleuronectes* (the type species for the genus) are morphologically distinct on the basis of their overall body shape and pellicle structure, yet they have identical SSU and LSU rDNA sequences. *Phacus alatus* was originally called *Phacus pleuronectes* var. *triquetra* and we recommend the use of the varietal name. The varietal status recognizes the differences in morphology with respect to the type species, and retention of the species designation reflects the genetic identity of the two gene sequences we have examined.

Similarly, *Phacus acuminatus* and *Phacus brachykentron* are both members of the section *Proterophacus* Pochmann and share many similarities. Both taxa are described as having oval or egg-shaped bodies that terminate in a pointed tail. One to three rounded paramylon grains are present in both. The size ranges for the two species overlap substantially and the two taxa have identical LSU and SSU rDNA sequences. In general, the cell body of *Phacus acuminatus* is wider in the posterior half than in the anterior half. However, the degree of widening varies among the varieties of the species, from very wide in *Phacus acuminatus* var. *javana* (Pochm.) Huber-Pestalozzi (or subsp. *javana* Pochm.) to no widening in *Phacus acuminatus* var. *iowensis* Allorge et Jahn. In summary, the morphological variability between the two species overlaps sufficiently that, when viewed along with the identity of LSU and SSU rDNA sequences, these taxa can be considered to be the same. Since *Phacus acuminatus* Stokes (1885) takes precedence over *Phacus brachykentron*

Pochmann (1942), *Phacus brachykentron* should be considered as *Phacus acuminatus*.

Secondary structure in D2 and intron

Structure had not been proposed for the D2 region of the LSU rDNA by either of the published secondary-structure models (Wuyts *et al.*, 2001; Schnare *et al.*, 1996). By comparing all 34 euglenoid LSU sequences, we found covarying positions throughout this region that conformed to the models of Gutell *et al.* (1994) for other eukaryotes. The secondary-structure alignment for *Euglena gracilis*, based on comparisons with the other euglenoid LSU sequences, is available as supplementary material in IJSEM Online (Supplementary Fig. A), together with a complete alignment of all 34 taxa (Supplementary Fig. B). The proposed secondary-structural model for the D2 region is shown in Fig. 1. Secondary structures for stems 14 and 16 could not be determined using comparative analysis (Gutell *et al.*, 1994; Kjer, 1995) for all 34 euglenoid sequences. Stems 18A, B and C varied in length among taxa and could not be aligned unambiguously. The terminal loop of stem 31A included a large insertion for a number of taxa, causing it to vary in length from 25 to 348 bp. The three introns present in the portion of the LSU rDNA sequenced were located between stems 13 and 18A (intron 2), stems D2_1 and D2_2 (intron 3) and stems D2_3' and D2_6 (intron 4). One taxon, *K. quartana*, lacked intron 2, and two taxa, *Euglena stellata* and *Phacus acuminatus*, lacked intron 3. Within intron 4, we found a stem-loop structure for all but three taxa that was supported by multiple compensatory base changes, shown for *E. gracilis* in Supplementary Figs A and B.

Signal analysis

Both the LSU and SSU rDNA possessed non-random phylogenetic signal for relationships within clades, but did not contain signal for relationships among clades. The g1 statistic for the SSU rDNA remained significant at the 1% level (Hillis & Huelsenbeck, 1992) as taxa were removed from strongly supported clades until only eight taxa remained, at which point signal became insignificantly different from random. At this point, only one taxon from each clade and one outgroup remained. The g1 statistic for the LSU rDNA remained significant at the 1% level until only seven taxa remained (one outgroup and one taxon from six of the seven clades). To examine the lack of signal for relationships among clades further, 10 sets of nine taxa (two outgroups and one taxon from each of the seven clades) were used. Taxa were removed sequentially until signal became insignificantly different from random. For the SSU rDNA, all 10 sets of nine taxa were significant at the 1% level, with g1 statistics ranging from -1.9924 to -1.1713. However, with the removal of the outgroup *Peranema*, all 10 datasets became insignificantly different from random, with g1 statistics ranging from 0.1230 to -0.2304. Since *Peranema* was the last taxon to be removed when we were testing for total signal of the SSU rDNA dataset, it was necessary to check whether all of the signal for the SSU

Fig. 1. Proposed secondary-structure model for the D2 region of the LSU rDNA for *Euglena gracilis* based on comparison of 34 euglenoid sequences.

rDNA was dependent on *Peranema*. Therefore, *Peranema* was removed from the dataset every five taxa to examine whether signal was still significant. If signal was still significant, it was reinserted into the dataset and another five taxa were removed. The result was that the removal of *Peranema* did not affect the total signal until only nine taxa remained. The LSU rDNA dataset also was significant at the 1 % level for all 10 sets of nine taxa, with g1 statistics ranging from -0.5994 to -0.3080 . The removal of *Peranema* from the dataset made the signal for four of the 10 datasets not significantly different from random. Four of the remaining datasets were significant at the 1 % level, and the other two were significant at the 5 % level. Subsequent removal of the taxon representing either clade G or clade F (Fig. 2) left only one dataset with signal significantly non-random at the 5 % level. The removal of the taxon representing clade A (Fig. 2) eliminated the significant signal.

Phylogenetic analyses

All analyses for the SSU rDNA, LSU rDNA and the combined dataset support the same seven external clades (Fig. 2). Although there is strong support for the external clades in most analyses, there is both weak and variable support for the relationship among those clades (Fig. 2; and see Figs 3–5 below).

SSU rDNA. Parsimony (Fig. 3a), log-det (Fig. 3b), Bayesian (Fig. 3c) and maximum-likelihood analyses (Fig. 3d) all generated the same seven external clades, but the relationships among those clades varied. Clades A, B, D and E are strongly supported in all analyses. *Phacus agilis* falls within clade C, though with little support. Bootstrap numbers for clade C represent the support for the clade excluding *Phacus agilis*. *Trachelomonas volvocinopsis* var. *spiralis* falls within clade A in all SSU rDNA analyses, though with little support. Clades F and G are moderately supported, with

Fig. 2. Bayesian phylogeny of the combined dataset generated using an undefined general time-reversible model (nst=6) with rates set to invgamma and nucleotide frequencies set to equal. Numerals at the internodes represent Bayesian probabilities. Probabilities less than 75 % are not shown. Seven strongly supported clades are represented by letters A–G.

Fig. 3. Trees for the SSU rDNA dataset. Letters A–G represent the clades illustrated in Fig. 2. Taxa in parentheses are included in clades but not represented by strong bootstrap support. Broken lines represent taxa that have a different placement in the phylogeny compared with that seen in Fig. 2. (a) Majority-rule consensus tree of two shortest parsimony trees (2670 steps). Numerals above nodes indicate bootstrap support, while numerals under nodes represent decay values. Bootstrap values less than 50% are not shown. (b) Log-det tree. Numerals above nodes indicate bootstrap support. (c) Bayesian trees for both AIC and invgamma models of evolution. Bayesian probabilities for the AIC tree are indicated first, followed by invgamma probabilities if different. Probabilities less than 75% are not shown. (d) Maximum-likelihood trees for both AIC and Bayesian models of evolution. Bootstrap values for the AIC tree are indicated first, followed by bootstrap values for Bayesian parameters if different. TVS, *T. volvocinopsis* var. *spiralis*.

some taxa (*H. ocellatus* and *Phacus acuminatus*) forming long branches within the tree.

The topologies of the SSU rDNA are not significantly different from the combined Bayesian analysis (Fig. 2). Clades F and G are basal to a clade containing A, B, C, D and E. The relationships among clades A, B, C, D and E vary depending on the type of analysis used. This is not unexpected, considering that the signal for relationships among clades, according to the g1 statistic, is insignificantly different from random. Unlike the combined analysis, the

SSU rDNA dataset forms a monophyletic grouping of clade E (the metabolic *Euglena*) and clade B (*Trachelomonas*) in all analyses except log-det, which again forms a monophyletic loricate clade (A and B). However, with the exception of the Bayesian analysis, there is little support for this relationship.

LSU rDNA. Parsimony (Fig. 4a), log-det (Fig. 4b), Bayesian (Fig. 4c) and maximum-likelihood analyses (Fig. 4d) all generated the same seven external clades; however, as in

Fig. 4. Trees for the LSU rDNA dataset. Letters A–G represent clades illustrated in Fig. 2. Broken lines represent taxa that have a different placement in the phylogeny compared with that seen in Fig. 2. (a) Shortest parsimony tree (1507 steps). (b) Log-det tree. (c) Bayesian trees for both the AIC and invgamma models of evolution. (d) Maximum-likelihood trees for both the AIC and Bayesian models of evolution. See legend to Fig. 3 for further details.

Fig. 5. Trees for the combined dataset. Letters A–G represent clades illustrated in Fig. 2. Broken lines represent taxa that have a different placement in the phylogeny compared with that seen in Fig. 2. (a) Majority-rule consensus tree of two shortest parsimony trees (4225 steps). (b) Log-det tree. See legend to Fig. 3 for further details. (c) Bayesian trees for the AIC model of evolution. Bayesian probabilities are indicated above nodes. Probabilities less than 75 % are not shown. (d)–(e) Maximum-likelihood trees with the AIC models of evolution (d) and Bayesian parameters (e). Bootstrap values are indicated by numerals above nodes.

the previous analyses, the relationships among those clades varied. Clades A, B and D were strongly supported in all analyses. Clades C and E were moderately supported in all analyses. There was strong support for some of the relationships within clades G and F for the parsimony, log-det and maximum-likelihood analyses, but there was little support for the clades themselves. Both Bayesian and maximum-likelihood analyses (Fig. 4c, d) generated the same tree topology. Unlike the parsimony, log-det and maximum-likelihood analyses, the Bayesian analyses strongly support six of the seven external clades and relationships among clades A and B and clades D, E, F and G.

The differences in tree topology for the LSU rDNA are largely due to the variability of the clade that is represented as most basal in each tree. Unlike the SSU rDNA and the combined analysis, which place clades F and G at the base of the phylogeny, clades C, D and E are all most basal in one of the trees represented in the analyses of the LSU rDNA. Clades G and F form a clade in all four analyses, but never fall at the base of the tree. The LSU rDNA dataset also differs from the SSU rDNA and combined analysis in its placement of *Euglena tripteris*, which falls as a sister taxon to the larger F and G clade. However, like the combined analysis, clades A and B are monophyletic in all analyses and include *T. volvocinopsis* var. *spiralis* as a sister taxon in the two Bayesian and maximum-likelihood analyses.

Combined dataset. Parsimony (Fig. 5a), log-det (Fig. 5b), Bayesian (Figs 2 and 5c) and maximum-likelihood analyses (Fig. 5d–e) all generated the same seven external clades. Although relationships among clades vary, some general trends in the combined dataset can be observed.

The loricates (clades A and B and *T. volvocinopsis* var. *spiralis*) are monophyletic in all analyses, with the exception of the parsimony analyses, in which *T. volvocinopsis* var.

spiralis is sister to clade C. However, there is no support for the relationships of clade A/B to clades C and D in the parsimony tree. There is strong support for *T. volvocinopsis* var. *spiralis* being sister to the two strongly supported loricate clades *Strombomonas* (A) and *Trachelomonas* (B) in the log-det and Bayesian analyses. The loricates are also sister to clade C in all analyses, with support of 87 % (Fig. 2) and 84 % (Fig. 5c) in the Bayesian trees, but with only weak bootstrap support (Fig. 5a, b, d, e) in the other trees. Clade D is sister to clade C in all analyses except the maximum-likelihood analysis using the same evolutionary model as in the Bayesian analysis (Fig. 5e). However, there is no support for that relationship. Clade E falls either as sister to the A, B, C, D clade (Figs 2 and 5a, b, d) or as sister to a separate F and G clade (Fig. 5c, e). Neither relationship has strong support. Clades F and G remain sister to each other in all analyses and sometimes form a separate clade, though never with strong support.

DISCUSSION

Combining datasets

The SSU and LSU rDNA datasets were analysed separately to examine taxonomic congruence between the datasets. Miyamoto & Fitch (1995) argued that congruence among datasets provides some of the strongest evidence that a phylogenetic estimate is accurate. Interestingly, both the SSU and LSU rDNA datasets support the same seven terminal clades with strong support (Figs 2–5). However, the relationship among the clades varies. This is not surprising, since, according to the *g*₁ statistic, signal for relationships among clades for both the SSU and LSU rDNA is not significantly different from random.

The rate matrices of the combined dataset, the nucleotide

composition and the gamma distribution all fall within the 95 % confidence intervals for these parameters for both the SSU and LSU rDNA datasets (Table 3), indicating that they are not evolving at excessively different rates. The combined analysis stabilized the phylogeny, increasing support for some of the internal nodes of the tree; however, because of the lack of signal for these relationships, only tentative conclusions can be drawn about the relationships among the external clades.

Taxonomic revisions

The seven clades found in all trees do not form groupings consistent with the current taxonomy of the Euglenophyta. The genera *Euglena* and *Phacus* are not monophyletic. Both genera are split into two separate clades, i.e. clades E and F for *Euglena* and clades C and G for *Phacus*. Clade E consists of metabolic *Euglena* species, whereas the *Euglena* species in clade F are more rigid. The *Phacus* species in clade G are the typical flat, leaf-like organisms for which the genus was described. Clade C consists of *Phacus* species that are more rounded in cross-section. Most of these taxa were originally in the genus *Euglena* and were later moved to the genus *Phacus*. A taxonomic revision of *Euglena* and *Phacus* is necessary, but additional taxa must still be sequenced before final amendments can be made. However, on the basis of the data available, we recommend the following: since the type species for the genus *Euglena*, *Euglena viridis*, resides in clade E, this clade will retain the name *Euglena*, and a new name should be proposed for clade F; similarly, *Phacus pleuronectes* (the type species) resides in clade G, and taxa in this clade should retain the name *Phacus*.

Not enough species of the genus *Lepocinclis* have been sequenced to determine whether or not it is monophyletic. Most species of *Lepocinclis* were previously in the genus *Euglena* and were subsequently moved to *Lepocinclis* after the genus was established. The *Lepocinclis* species embedded within the rigid *Euglena* species in clade F will probably acquire the genus name assigned to clade F.

The loricate genera are monophyletic in most analyses, with a separate clade for the genus *Strombomonas* (clade A) and another for the genus *Trachelomonas* (clade B). There is some question as to the placement of *T. volvocinopsis* var. *spiralis*. In the SSU rDNA analyses, it groups with the *Strombomonas* clade; in the LSU rDNA and combined analyses, it sits as a sister taxon to the *Strombomonas* and *Trachelomonas* clade. *T. volvocinopsis* var. *spiralis* differs in size, shape and ornamentation from the organisms placed within the strongly supported *Trachelomonas* clade (B). *T. volvocinopsis* var. *spiralis* is a small, rounded species with a helical ridge pattern ornamenting its lorica. The other *Trachelomonas* species in clade B are oblong with ornamentation of pores and spines. The addition of other *Trachelomonas* species similar to *T. volvocinopsis* var. *spiralis* will help to elucidate the placement of these species in the phylogeny and better establish the subgroups within the

genus *Trachelomonas*. *Strombomonas* will either remain as a separate genus or become a subgroup within *Trachelomonas*.

The genus *Colacium* (clade D) is monophyletic in all analyses. The placement of this genus in relation to the other clades remains uncertain, however.

The colourless, osmotrophic species *A. longa* and *K. quartana* group with the metabolic *Euglena* species (clade E), as shown previously by Linton *et al.* (1999, 2000), indicating that they lost photosynthesis secondarily. Similarly, the colourless *H. ocellatus* groups with *Phacus acuminatus* in the flat *Phacus* clade (G). However, *Hyalophacus* has a large pairwise divergence from the ingroup taxa and may become a long branch (Figs 3–5).

While the taxa presented in this analysis represent a broad range of photosynthetic euglenoids, it is clear that additional taxa, especially from some of the more speciose genera including *Phacus* and *Trachelomonas*, will need to be sampled to resolve the phylogeny in more detail. The presence of the seven strongly supported clades within the euglenoid phylogeny provides a framework for future taxon selection to strengthen our understanding of euglenoid evolution.

Conclusions

Seven strongly supported clades of photosynthetic euglenoids have been identified. Although the relationships among these clades are often weak and may vary, some tentative conclusions can be drawn.

(i) Clades A and B are monophyletic. All taxa that are housed within a lorica are contained in clade A/B, indicating a single origin of the lorica. Taxa within the genus *Strombomonas* were originally assigned to *Trachelomonas*. If the *Strombomonas* clade remains embedded within *Trachelomonas*, the genus *Strombomonas* will need to be dissolved.

(ii) Secondary osmotrophy, due to the loss of photosynthesis, has arisen many times, as evidenced by *A. longa* and *K. quartana* in clade E and *H. ocellatus* in clade G.

(iii) The genus *Euglena* should be limited to those taxa in clade E, which contains the type species, *Euglena viridis*.

(iv) The genus *Phacus* should be limited to those taxa in clade G, which contains the type species, *Phacus pleuronectes*.

(v) The *Phacus* species in clade C will need to be renamed. *Phacus pyrum* Stein was originally described as *Euglena pyrum* by Ehrenberg and later renamed as *Lepocinclis pyrum* Perty before it became *Phacus pyrum* Stein. Similarly, *Phacus pyrum* var. *ovata* is also known as *L. ovata* Conrad. The multiple transfers of these taxa among the various genera demonstrate the uncertainty of the current taxonomy. We are now considering retaining the name *Lepocinclis* and applying it to all taxa in clade C.

(vi) The rigid *Euglena* and *Lepocinclis* species in clade F should be combined under a single name. What that name will be remains uncertain. If clades F and G ultimately form a monophyletic assemblage, clade F may need to be combined with clade G under *Phacus*. Alternatively, a new name will need to be generated for the taxa in clade F. Interestingly, the taxon originally described as *Phacus tripteris* was moved to *Euglena tripteris* by Klebs (1883), moved back to *Phacus tripteris* by Pochmann (1942) and again moved back to *Euglena tripteris* by Huber-Pestalozzi (1955). Perhaps it is time to move it back to its original placement under *Phacus*.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation PEET (Partnership for Enhanced Expertise in Taxonomy) program (grant no. DEB 4-21348). We would like to thank Carole Lewandowski and two anonymous reviewers for their helpful comments and suggestions. We would also like to thank Dr Eric Linton for his help in assisting with some of the phylogenetic analyses. This research was submitted by S. Brosnan in partial fulfilment of the requirements for a PhD degree at Rutgers University, New Brunswick, NJ, USA.

REFERENCES

- Bourrelly, P. (1970). 1. Ordre des Euglénales. In *Les Algues d'Eau Douce*, pp. 123–159. Paris: Boubee (in French).
- Bremer, K. (1994). Branch support and tree stability. *Cladistics* 10, 295–304.
- Conrad, W. (1934). Matériaux pour une monographie due genre *Lepocinclis* Perty. *Arch Protistenkd* 82, 203–249 (in French).
- Deflandre, G. (1930). *Strombomonas* nouveau genre de l'Euglenacees. *Arch Protistenkd* 69, 551–614 (in French).
- Donoghue, M. J., Olmstead, R. G., Smith, J. F. & Palmer, J. D. (1992). Phylogenetic relationships of Dipsacales based on rbcL sequences. *Ann MO Bot Gard* 79, 333–345.
- Eriksson, T. (1998). AutoDecay version 4.0. Program distributed by the author. Department of Botany, Stockholm University, Sweden.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Freshwater, D. W. & Bailey, J. C. (1998). A multigene phylogeny of the Gelidiales including nuclear large-subunit rRNA sequence data. *J Appl Phycol* 10, 229–236.
- Gockel, G., Hachtel, W., Baier, S., Fliss, C. & Henke, M. (1994). Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Curr Genet* 26, 256–262.
- Goff, L. & Moon, D. (1993). PCR amplification of nuclear and plastid genes from algal herbarium specimens and algal spores. *J Phycol* 29, 381–384.
- Gutell, R. R., Larsen, N. & Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* 58, 10–26.
- Hillis, D. M. & Huelsenbeck, J. P. (1992). Signal, noise, and reliability in molecular phylogenetic analyses. *J Hered* 83, 189–195.
- Hollande, A. (1952). Classe des Eugléniens (Euglenoidina Butschli, 1884). In *Traité de Zoologie*, vol. 1, pp. 238–284. Edited by P. P. Grassé. Paris: Mason (in French).
- Huber-Pestalozzi, G. (1955). *Das Phytoplankton des Süßwassers*. IV. *Euglenophyceen*. Edited by A. Thienemann. Stuttgart: E. Schweizerbart'sche (in German).
- Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.
- Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol Phylogenet Evol* 4, 314–330.
- Kjer, K. M., Blahnik, R. J. & Holzenthal, R. W. (2001). Phylogeny of Trichoptera (caddisflies): characterization of signal and noise within multiple datasets. *Syst Biol* 50, 781–816.
- Klebs, G. (1883). Über die Organization einiger Flagellatengruppen und ihre Beziehungen zu Algen und Infusorien. *Untersuchungen Bot Inst Tübingen* 1, 233–362 (in German).
- Larsen, N. (1992). Higher order interactions of 23S rRNA. *Proc Natl Acad Sci U S A* 89, 5044–5048.
- Leander, B. S. & Farmer, M. A. (2000). Comparative morphology of the euglenid pellicle. I. Patterns of strips and pores. *J Eukaryot Microbiol* 47, 469–479.
- 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., Triemer, R. E. & Farmer, M. A. (2001). Character evolution in heterotrophic euglenids. *Eur J Protistol* 37, 337–356.
- Leedale, G. F. (1967). *Euglenoid Flagellates*. Englewood Cliffs, NJ: Prentice Hall.
- Linton, E. W., Hittner, D., Lewandowski, C., Auld, T. & Triemer, R. E. (1999). A molecular study of euglenoid phylogeny using small subunit rDNA. *J Eukaryot Microbiol* 46, 217–223.
- 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.
- Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* 11, 605–612.
- Milanowski, R., Zakryś, B. & Kwiatowski, J. (2001). Phylogenetic analysis of chloroplast small-subunit rRNA genes of the genus *Euglena* Ehrenberg. *Int J Syst Evol Microbiol* 51, 773–781.
- Miyamoto, M. M. & Walter, M. F. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst Biol* 44, 64–76.
- Montegut-Felkner, A. E. & Triemer, R. E. (1997). Phylogenetic relationships of selected euglenoid genera based on morphological and molecular data. *J Phycol* 33, 512–519.
- Moreira, D., López-García, P. & Rodríguez-Valera, F. (2001). New insights into the phylogenetic position of diplomonads: G+C content bias, differences of evolutionary rate and a new environmental sequence. *Int J Syst Evol Microbiol* 51, 2211–2219.
- Müllner, A. N., Angeler, D. G., Samuel, R., Linton, E. W. & Triemer, R. E. (2001). Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *Int J Syst Evol Microbiol* 51, 783–791.
- Nudelman, M. A., Rossi, M. S., Conforti, V. & Triemer, R. E. (2003). Phylogeny of Euglenophyceae based on small subunit rDNA sequences: taxonomic implications. *J Phycol* 39, 226–235.
- Palumbi, S. R. (1996). Nucleic acids. II. The polymerase chain reaction. In *Molecular Systematics*, 2nd edn, pp. 205–247. Edited by D. M. Hillis, C. Moritz & B. K. Mable. Sunderland, MA: Sinauer.

- Playfair, G. J. (1921).** Australian freshwater flagellates. *Proc Linnear Soc NSW* **46**, 99–146.
- Pochmann, A. (1942).** Synopsis der Gattung *Phacus*. *Arch Protistenkd* **95**, 81–252 (in German).
- Popova, T. G. (1955).** *Opredelitel Presnovodnykh Vodoroslej SSSR*, vol. 7, *Euglenovyye Vodorosli*. Moscow & Leningrad: Sovetskaya Nauka (in Russian).
- Popova, T. G. (1966).** *Flora Sporovyky Rastenij SSSR*, vol. 8, *Euglenophyta* [Gen. *Trachelomonas*, *Strombomonas*, *Eutreptia*, *Euglena*]. Moscow & Leningrad: Izd. Akademii Nauka (in Russian).
- Posada, D. & Crandall, K. A. (1998).** MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Preisfeld, A., Berger, S., Busse, I., Liller, S. & Ruppel, H. G. (2000).** Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data. *J Phycol* **36**, 220–226.
- Preisfeld, A., Busse, I., Klingberg, M., Talke, S. & Ruppel, H. G. (2001).** Phylogenetic position and inter-relationships of the osmotrophic euglenoids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int J Syst Evol Microbiol* **51**, 751–758.
- Schnare, M. N., Damberger, S. H., Gray, M. W. & Gutell, R. R. (1996).** Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23S-like) ribosomal RNA. *J Mol Biol* **256**, 701–719.
- Seimeister, G. & Hachtel, W. (1990).** Structure and expression of a gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbcL*) in the colourless euglenoid flagellate *Astasia longa*. *Plant Mol Biol* **14**, 825–833.
- Senn, G. (1900).** Euglenineae. In *Die natürlichen Pflanzenfamilien*, pp. 173–185. Edited by A. Engler & K. Prantl (in German).
- Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. (1994).** The genetic data environment: an expandable GUI for multiple sequence analysis. *Comput Appl Biosci* **10**, 671–675.
- Sogin, M. L. & Gunderson, J. H. (1987).** Structural diversity of eukaryotic small subunit ribosomal RNAs: evolutionary implications. *Ann N Y Acad Sci* **573**, 125–139.
- Stein, F. R. (1878).** *Der Organismus der Infusionsthiere*. III Abt. *Der Organismus der Flagellaten*. Leipzig: W. Engelmann (in German).
- Stokes, A. C. (1885).** Some apparently undescribed Infusoria from fresh water. *Am Nat* **19**, 18–27.
- Swofford, D. L. (2002).** PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, MA: Sinauer.
- Thompson, M. D., Copertino, D. W., Thompson, E., Favreau, M. R. & Hallick, R. B. (1995).** Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Res* **23**, 4745–4752.
- Woese, C. R. (1996).** The world of ribosomal RNA. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*, pp. 23–47. Edited by A. Zimmermann & A. E. Dahlberg. Boca Raton, FL: CRC Press.
- Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T. & De Wachter, R. (2001).** The European large subunit ribosomal RNA database. *Nucleic Acids Res* **29**, 175–177.