J. Phycol. **42,** 721–730 (2006)
© 2006 by the Phycological Society of America
DOI: 10.1111/j.1529-8817.2006.00216.x

PHYLOGENY OF PHOTOSYNTHETIC EUGLENOPHYTES BASED ON COMBINED CHLOROPLAST AND CYTOPLASMIC SSU rDNA SEQUENCE ANALYSIS[1]

Rafał Milanowski, Sylwia Kosmala, Bożena Zakryś[2] *and Jan Kwiatowski*

Department of Plant Systematics and Geography, Warsaw University, Al. Ujazdowskie 4, PL-00-478 Warszawa, Poland

Eighteen new 16S rDNA and 16 new 18S rDNA sequences from 24 strains, representing 23 species of photoautotrophic euglenoids, were obtained in nearly their entire length. Maximum parsimony, maximum likelihood, and Bayesian phylogenetic analyses were performed on separate data (39 sequences of 16S rDNA and 58 sequences of 18S rDNA), as well as on combined data sets (37 sequences). All methods of sequence analysis gave similar results in those cases in which the clades received substantial support. However, the combined data set produced several additional well-supported clades, not encountered before in the analyses of green euglenoids. There are three main well-defined clades (A, B/C/D, and G) on trees from the combined data set. Clade G diverges first, while clades A and B/C/D form sister groups. Clade A consists of *Euglena* species *sensu stricto* and is divided into three sub-clades (A1, A2, and A3). Clade A3 (composed of *E. deses* and *E. mutabilis*) branches off first; then, two sister clades emerge: A1 (composed of *E. viridis*-like species) and A2 (consisting of *E. agilis* and *E. gracilis* species). Clade B/C/D consists of the *Strombomonas*, *Trachelomonas*, *Cryptoglena*, *Monomorphina*, and *Colacium* genera. Clade G comprises *Phacus* and *Lepocinclis*, as well as the *Discoglena* species of *Euglena*, with *Discoglena* branching off first, and then *Phacus* and *Lepocinclis* emerging as sister groups.

Key index words: ***Colacium***; ***Cryptoglena***; **Endosymbiosis;** ***Euglena***; **Euglenozoa;** ***Eutreptia***; ***Eutreptiella***; ***Lepocinclis***; **molecular phylogeny;** ***Monomorphina***; ***Phacus***; **ribosomal RNA; SSU rDNA;** ***Strombomonas***; ***Trachelomonas***

Abbreviations: **BA, Bayesian analysis; bs, non-parametric bootstrap; di, decay index; ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor-joining; nt, nucleotide; pp, posterior probability**

In the last several years, vigorous and continuous attempts have been made to redraw the picture of euglenoid phylogeny based mainly on molecular data, such as: cytoplasmic small subunit (SSU) rDNA (Montegut-Felkner and Triemer 1997, Linton et al. 1999, 2000, Preisfeld et al. 2000, 2001, Leander and Farmer 2001, Leander et al. 2001, Moreira et al. 2001, Müllner et al. 2001, Marin et al. 2003, Nudelman et al. 2003, Von der Heyden et al. 2004) and LSU (Brosnan et al. 2003), chloroplast SSU rDNA (Milanowski et al. 2001, Zakryś et al. 2002), or *rbc*L (Thompson et al. 1995), and *par 1* and *par2* genes (Talke and Preisfeld 2002). In the analyses where a wide range of strains were used, the dominating molecular marker was cytoplasmic SSU rDNA (Marin et al. 2003).

Generally accepted conclusions emerging from the above works may be summarized as follows: (1) euglenoids are a monophyletic assemblage belonging, together with Diplonemida and Kinetoplastida, to a broader group of Euglenozoa (Cavalier-Smith 1981); (2) colorless euglenoids could be divided into two categories: primary and secondary heterotrophs, the latter, such as *Khawkinea* or *Astasia,* having lost their chloroplasts recently; and (3) green and secondarily colorless euglenoids form a monophyletic group, conforming to a view (Gibbs 1978) that an euglenoid chloroplast emerged as a consequence of a single endosymbiotic episode between a heterotrophic euglenoid and a chlorophycean alga. The recent discovery of photosynthesis-related genes in the genomes of some Kinetoplastids (Hannaert et al. 2003) challenged that view. However, recent analyses suggest that there is no evidence that Kinetoplastids had a plastid-bearing ancestor (Leander 2004, Rogers and Keeling 2004, El-Sayed et al. 2005).

Estimations of phylogenetic relationships within photosynthetic euglenoids are far from satisfactory. They pointedly argue that the existing classification, based on morphology, requires substantial revisions. This need was particularly urgent in the case of the genus *Euglena,* which seemed to be at least paraphyletic or even polyphyletic. Within the group of green euglenoids, several well-established clades formed an unresolved assemblage.

In the work reported here, we combined analyses of chloroplast and cytoplasmic SSU rDNA, to further our understanding of the relationships between main clades of green euglenoids. Our new phylogenetic trees allow us to draw an outline of changes in morphological features within these euglenoids.

MATERIALS AND METHODS

Euglenoid strains and culture conditions. The new 16S rDNA and 18S rDNA sequences from 24 strains, representing 1723

[1]Received 26 October 2005. Accepted 9 February 2006.
[2]Author for correspondence: e-mail zakrys@biol.uw.edu.pl.

TABLE 1. Strains and accession numbers for cytoplasmic and chloroplast small subunit rDNA sequences.

No.	Taxon name	Strain origin	GenBank accession numbers	
			16S rDNA	18S rDNA
1.	*Colacium mucronatum* Bourrelly et Chadefaud	UTEX 2524		AF326232
2.	*Colacium vesiculosum* Ehr.	UW-Łazienki (L)	AF289238	**DQ249874**
3.		UTEX 1315		AF081592
4.	*Cryptoglena pigra* Ehr.	AICB 350	**AY626052**	**DQ249875**
5.		CCAP 1212/1		AJ532437
6.	*Euglena agilis* Carter	UTEX 1605		AF115279
7.		UW-Nowy Targ (NT)	AY158150	
8.		UW-Pruszków (P)	AF289239	
9.		UW-Wąwocko (W)	AY158151	**DQ249876**
10.	*Euglena anabaena* Mainx	SAG 1224-15b	AF289240	AJ532430
11.	*Euglena cantabrica* Pringsheim	SAG 1224-25	**AY626047**	AJ532412
12.	*Euglena carterae* (Pringsh.) Marin et Melkonian	SAG 1224-22		AJ542406
13.	*Euglena chadefaudii* Bourrelly	CCAP 1224/17g (as *E. viridis*)	**AY626049**	**AY626063**
14.		ACOI 2951 (as *E. viridis*)	**AY626046**	**AY626064**
15.	*Euglena cuneata* Pringsheim	SAG 26.93	**AY626048**	**AY626065**
16.	*Euglena deses* Ehr.	SAG 1224-19b	**AY626043**	AJ532409
17.	*Euglena geniculata* Dujardin	SAG 1224-4b	AF289241	AY070249
18.		SAG 1224-4f	AY070252	AY070248
19.	*Euglena gracilis* Klebs	Unknown	X12890	M12677
20.	*Euglena granulata* (Klebs) Schmitz	SAG 1224-8c		AJ532421
21.	*Euglena laciniata* Pringsheim	SAG 1224-31		AJ532420
22.	*Euglena limnophila*	ACOI 1026	**AY626056**	**DQ249877**
23.	Lemmermann	ASW 08039		AJ532453
24.	*Euglena mutabilis* Schmitz	SAG 1224-9b	**AY626044**	AJ532405
25.	*Euglena polymorpha* Dangeard	ACOI 921	AF289242	**AY626062**
26.	*Euglena proxima* Dangeard	SAG 1224-11b	**AY626050**	DQ249878
27.	*Euglena sanguinea* Ehr.	UTEX 2345		AJ532422
28.	*Euglena spathirhyncha* Skuja	SAG 1224-42	**AY626060**	AJ532454
29.	*Euglena stellata* Mainx	SAG 1224-14	AF289244	AJ532419
30.	*Euglena tristella* Chu	SAG 1224-35	AF289246	AY070247
31.	*Euglena viridis* Ehr.	SAG 1224-17d	AF289248	AY070246
32.	*Eutreptia pertyi* Pringsheim	UTEX 1290		AF081589
33.	*Eutreptia viridis* Perty	SAG1226-1c	AF289247	AJ532395
34.	*Eutreptiella braarudii* Throndsen	CCMP-1594 (as *E. gymnastica*)	**DQ249873**	**DQ249879**
35.	*Eutreptiella gymnastica* Throndsen	SCCAP K-0333		AJ532400
36.	*Lepocinclis büetschlii* Lemmermann	Isolated in Korea		AF096993
37.	*Lepocinclis fusiformis* (Carter) Lemmermann	ACOI 1025	AF289249	AY935697
38.	*Lepocinclis ovum* (Ehr.) Minkiewicz	SAG 1244-8		AF110419
39.	*Lepocinclis spirogyroides* (Ehr.) Marin et Melkonian	SAG 1224-13b	AF289243	AJ532462
40.	*Lepocinclis tripteris*	UW-OB	AF289245	AY935696
41.	(Dujardin) Marin et Melkonian	UTEX 1311		AF445459
42.	*Monomorphina pyrum* (Ehr.)	ACOI 2778	**AY626053**	DQ116996
43.	Mereschkowsky	UTEX 2354		AF112874
44.		SAG 1244-5 (as *L. ovata*)		AF061338
45.	*Monomorphina striata* (Francé) Marin et Melkonian	CCAP 1261/9	**AY626054**	AJ532432
46.	*Phacus granum* Dreżepolski	AICB 349	**AY626058**	**DQ249880**
47.	*Phacus orbicularis* Hübner	AICB 502	**AY626057**	AY935698
48.		AICB 525	AF289250	AY935699
49.		ASW 08045		AF283315
50.	*Phacus pleuronectes* (Ehr.) Dujardin	SAG 1261-3b	AF289251	AJ532475
51.	*Phacus similis* Christen	ACOI 1226	**AY626059**	**DQ249881**
52.		SAG 58.81		AF119118
53.	*Phacus skujae* Skvortzov	AICB 323	AF289252	**DQ249882**
54.	*Strombomonas acuminata*	ACOI 2476	**AY626051**	**DQ249883**
55.	(Schmarda) Deflandre	Isolated in New Jersey (USA)		AF445461
56.	*Strombomonas acuminata* var. *conspersa* (Schmarda) Deflandre	SAG 1280-1		AY015000
57.	*Strombomonas costata* Deflandre	ACOI 2992	AF289253	**DQ249884**
58.	*Trachelomonas similis* Stokes	SAG 1283-14	**AY626055**	**DQ249885**
59.	*Trachelomonas volvocina* Ehr.	AICB524	AF289254	**DQ249886**
60.		SAG 1283-4		AJ532451

Those obtained in this study are in boldface.

ACOI, Culture Collection of Algae at the Department of Botany, University of Coimbra, Portugal; AICB, Culture Collection of Algae at the Institute of Biological Research Cluj-Napoca, Romania; ASW, Culture Collection of Algae at the University of Vienna, Austria; CCAP, Culture Collection of Algae and Protozoa at Center for Ecology and Hydrology, Cumbria, UK; CCMP, The Provasoli-Guillard National Center of Culture of Marine Phytoplancton, Bigelow Laboratory for Ocean Sciences, USA; SAG, Sammlung von Algenkulturen Pflanzenphysiologisches Institut der Universität Göttingen, Germany; SCCAP, Scandinavian Culture Center for Algae and Protozoa, University of Copenhagen, Denmark; UTEX, Culture Collection of Algae at the University of Texas at Austin, TX, USA; UW, Department of Plant Systematics and Geography of Warsaw University, Poland.

TABLE 2. Models of nucleotide substitutions chosen by Modeltest 3.7 program (Posada and Crandall 1998) and MrModeltest 2.2 (Nylander 2004).

Alignment	Modeltest 3.7 for PAUP		MrModeltest 2.2 for MrBayes	
	LRT test	AIC test	LRT test	AIC test
16S rDNA	GTR + I + G	GTR + I + G	GTR + I + G	GTR + I + G
18S rDNA	SYM + I + G	GTR + I + G	SYM + I + G	GTR + I + G
16S–18S rDNA	TrN + I + G	GTR + I + G	GTR + I + G	GTR + I + G

GTR, general time-reversible (Lanave et al. 1984, Tavare 1986, Rodriguez et al. 1990); SYM, symetrical (Zharkikh 1994); TrN, Tamura and Nei 1993. AIC, Akaike information criterion; LRT, likelihood ratio test.

photoautotrophic taxa, are shown in Table 1, together with all other strains, whose sequences were used in this study. All strains were cultivated as described earlier (Milanowski et al. 2001).

DNA isolation, amplification, and sequencing. Isolation of total DNA, amplification of 16S rDNA and 18S rDNA regions, purification and sequencing of PCR products were performed as described previously (Milanowski et al. 2001, Zakryś et al. 2002). Two additional PCR primers were used for sequencing of the *E. cantabrica* 16S rDNA 750 nucleotides (nt) long group II intron, located on the rRNA helix 20: 16SecabIntL: 5′-CTTTGACATTACTGCGTG-3′ and 16SecabIntR, 5′-CTACTCATACGGCTACTC-3′.

Sequence alignment and phylogenetic analysis. The sequences used for phylogenetic analyses are shown in Table 1. Sequence alignments were obtained using the ClustalX program (Thompson et al. 1997) set to default, manually checked, and edited according to the secondary structure of *Euglena gracilis* rRNA (Wuyts et al. 2002). Several regions of uncertain homology, which could not be unambiguously aligned (for 16S rRNA: V1, 2, 5, 6, 7, and 9; for 18S rRNA: V2, helix 16, V3, 4, 5, 7, and 8) were omitted from the analyses. The alignments used for analyses are available in EMBL [16S: ALIGN_000948; 18S: ALIGN_000949; 16S/18S: ALIGN_000950]. All nucleotides were treated as independent and unordered, multi-state characters of equal weight. The alignments were edited using the program GenDoc 2.6 (Nicholas and Nicholas 1997).

Phylogenetic analyses were performed separately for 16S rDNA (39 sequences), 18S rDNA (58 sequences), and for combined data (37 sequences). Maximum likelihood (ML) and maximum parsimony (MP) analyses, as well as homogeneity test (χ^2) of nucleotide distribution and partition homogeneity test were performed by PAUP*, Version 4.0b10 for Microsoft Windows (Swofford 1998). To find the best tree, the heuristic search option was used with MULPARS, tree-bisection-reconnection (TBR) branch swapping, ACCTRAN optimization and random addition with the number of replicates depending on the method used (1000 for MP and 10 for ML). Bootstrap support for specific nodes (Felsenstein 1985) was estimated with the default options using 1000 and 100 replications, for MP and ML analyses, respectively, as implemented in PAUP*. Models of sequence evolution (Table 2) and their parameters (Table 3) for the ML method were chosen, separately for each of the three alignments, by Modeltest 3.7 program (Posada and Crandall 1998). Auto decay indices (Bremer 1994) were calculated by AutoDecay 4.02 program (Eriksson 1998) for MP analyses. Bayesian analyses (BA) were performed by MrBayes 3.1 (Huelsenbeck and Ronquist 2001). Models of sequence evolution for Bayesian analyses (Table 2) were chosen separately for 16S rDNA and 18S rDNA as well as for combined data by MrModeltest 2.2 (Nylander 2004). For each chosen model, two types of analyses were performed where the covarion model of evolution was used or not. In case of combined sequences, two partitions were specified for 18S and 16S rDNA, respectively; however analyses without partitioning were also performed. The sequence of *Eutreptia pertyi*, *Eutreptia viridis*, *Eutreptiella gymnastica*, and *Eutreptiella braarudii* were used to root the trees, which were drawn by Tree View, Version 1.6.1, for Microsoft Windows (Page 1996).

RESULTS

Alignments and phylogenetic analyses. Three sequence alignments were prepared: 16S rDNA

TABLE 3. Parameters proposed by the Modeltest 3.7 program (Posada and Crandall 1998) for chosen models of sequence evolution.

Alignment	Model	A	C	G	T	I	G (α)
16S	GTR + I + G	0.3007	0.1526	0.2458	0.3009	0.2784	0.5659
18S	GTR + I + G	0.2317	0.2583	0.2705	0.2395	0.3301	0.5921
18S	SYM + I + G	0.2500	0.2500	0.2500	0.2500	0.3227	0.5755
16S–18S	GTR + I + G	0.2637	0.2081	0.2654	0.2627	0.2706	0.5276
16S–18S	TrN + I + G	0.2838	0.1964	0.2463	0.2734	0.2692	0.5186
Alignment	**Model**	**A–C**	**A–G**	**A–T**	**C–G**	**C–T**	**G–T**
16S	GTR + I + G	1.3641	5.5420	1.6637	0.5486	9.8830	1.0000
18S	GTR + I + G	1.1675	2.9619	1.0015	0.5343	4.2133	1.0000
18S	SYM + I + G	1.0764	2.8412	0.8466	0.5829	4.0199	1.0000
16S–18S	GTR + I + G	1.2047	3.9983	1.3499	0.5534	5.8575	1.0000
16S–18S	TrN + I + G	1.0000	3.8417	1.0000	1,0000	5.7744	1.0000

A, C, G, and T, frequency of nucleotides; I, fraction of unchangeable nucleotides; G(α), shape parameter (α) of gamma (G) distribution of nucleotide substitution rates; A–C, A–G, A–T, C–G, C–T, and G–T, rates of reversible nucleotide substitutions. GTR, general time-reversible (Lanave et al. 1984, Tavare 1986, Rodriguez et al. 1990); SYM, symetrical (Zharkikh 1994); TrN, Tamura and Nei (Tamura and Nei 1993).

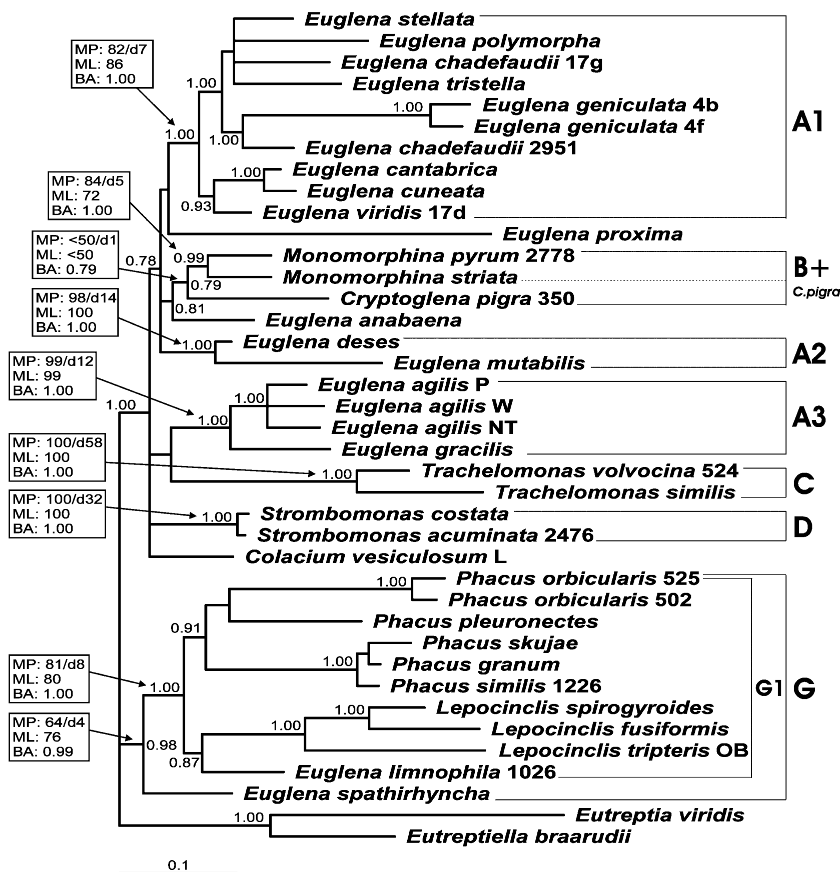

FIG. 1. Phylogenetic tree of the 16S rDNA sequence obtained by Bayesian inference (model GTR + G + I, covarion = yes). Numbers at the nodes show posterior probabilities of the tree bipartitions. Probabilities of less than 75% are not shown. Numbers in rectangles show bootstrap values/decay indices obtained for the main clades by maximum parsimony, bootstrap values obtained by maximum likelihood analysis (model GTR + I + G), and posterior probabilities obtained by Bayesian inference (model GTR + I + G, covarion = no).

(1543 nt), 18S rDNA (3037 nt), and the combined 16S and 18S rDNA (4580 nt). After the removal of sites of an uncertain homology, which could not be unambiguously aligned, 1402 positions were left in the 16S rDNA alignment (of which 700 were constant and 528 were MP informative). In the 18S rDNA and the combined sequence alignments, the numbers were 1270, 637, 495 and 2672, 1383, 916, respectively. χ^2 tests for all three alignments showed the homogenous nucleotide distributions (16S: $P = 0.99$; 18S and 16S-18S: $P = 1.00$), permitting reliable phylogenetic analyses.

16S rDNA. Trees obtained by the three different methods (MP, ML, and BA) are congruent with respect to the nodes that have substantial support (bs, bootstrap; di, decay index; or pp, posterior probabilities). Figure 1 shows the tree obtained by the Bayesian approach, on which seven well-defined clades were identified (A1-3, B, C, D, G).

Clades A1–3 represent species of *Euglena*. A1 consists of ten strains of seven *Euglena* species classified under the subgenus *Euglena sensu* Zakryś and of one representative of the subgenus *Calliglena* (Zakryś 1986). This clade is well-supported on trees obtained by all the methods (bs—82/86, di—7, pp—1.00), although the relationships within it remain unresolved. *E. deses* and *E. mutabilis* forming clade A2 (bs—98/100, di—14, pp—1.00) as well as *E. gracilis* and *E. agilis* forming clade A3 (bs—99, di—12, pp—1.00), all belong to the subgenus *Calliglena sensu* Zakryś.

Clades B, C, and D are represented by pairs of strains from different genera. Clade B (bs—72/84, di 5, pp—0.99/1.00) consists of the *Monomorphina* species, clade C (bs—100, di—58, pp—1.00) of *Trachelomonas*, and clade D (bs—100, di—32, pp—1.00) of *Strombomonas*. *Cryptoglena pigra* is located at the base of clade B on all the trees, although with a weak support (bs—<50, di—1, pp—0.79). The mutual phylogenetic affinities of the clades A1–3, B (with *C. pigra*), C, D as well as species *E. proxima*, *E. anabaena*, and *Colacium vesiculosum* are unresolved; however, on the ML and BA trees they form a sister group to the clade G. This topology is especially strongly supported on a BA tree, where the covarion model was used (pp – 1.00). Clade G (bs – 64/76, di – 4, pp – 0.98/0.99) consists of five species classified as *Phacus*: three as *Lepocinclis* and two as

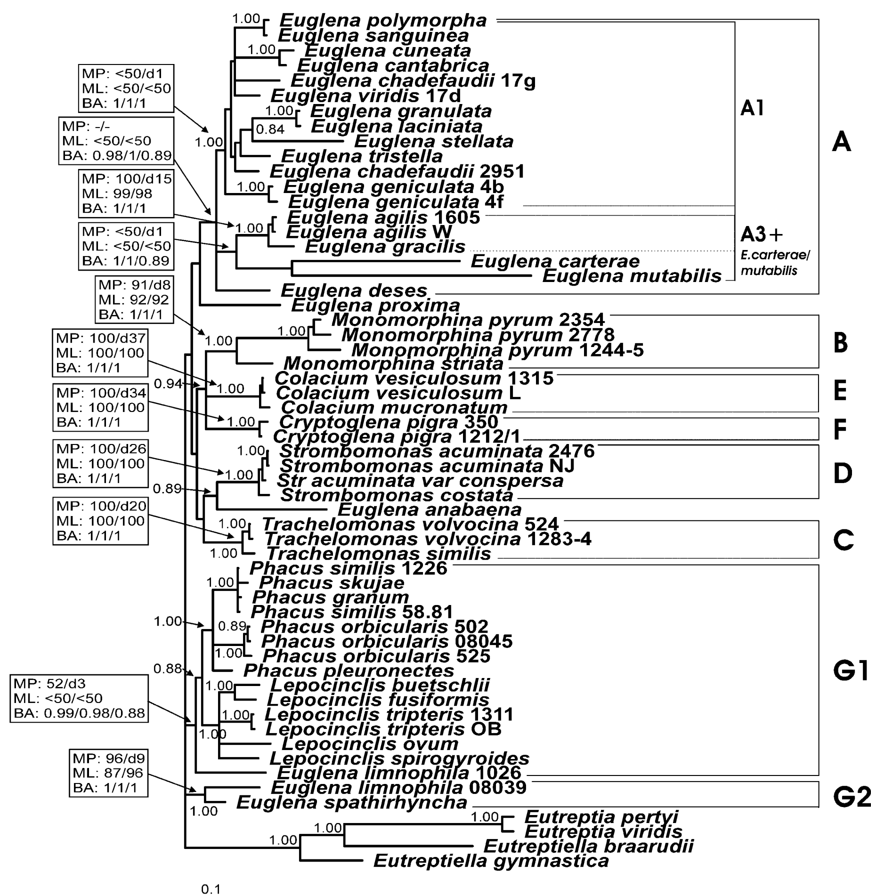

FIG. 2. Phylogenetic tree of the 18S rDNA sequence obtained by Bayesian inference (model GTR + G + I, covarion = yes). Numbers at the nodes show posterior probabilities of the tree bipartitions. Probabilities of less than 75% are not shown. Numbers in rectangles show bootstrap values/decay indices obtained for the main clades by MP (−/− indicates that clade does not exist in MP analysis), two bootstrap values obtained by ML analysis (models GTR + I + G/SYM + I + G), and three posterior probabilities obtained by Bayesian inference (models GTR + I + G, covarion = no/SYM + I + G, cov = no/SYM + I + G, cov = yes).

Euglena (subgenus *Discoglena sensu* Zakryś). On all trees the *Phacus* and *Lepocinclis* genera are both monophyletic.

18S rDNA. Figure 2 shows a tree obtained by analyzing the 18S rDNA by the Bayesian method. Eight distinctive clades (A, B, C, D, E, F, G1-2) are visible on the tree. All of them are also present on other 18S rDNA trees (with the exception of clade A in MP analysis), irrespective of the method of tree building. Clade A consists of the strains belonging to the clades A1-3 on the 16S rDNA tree and four additional species, but it is substantially supported only on the BA trees where the covarion model was not used (bs—<50, pp—<0.75–1.00). On the MP tree, *E. anabaena* appears within clade A (weak support: bs—<50, di—1). Of the clades A1–3, only A3 (*E. agilis* + *E. gracilis*) is substantially supported (bs—98–100, di—15, pp—1.00). Clade A1 is significantly supported only on the BA trees (bs—<50. di—1, pp—1.00), and clade A3 does not exist at all. Similar to the 16S rDNA trees, *Monomorphina*, *Trachelomonas*, and *Strombomonas* representatives form clades B (bs—91–92, di—8, pp—1.00), C (bs—100, di—20, pp—1.00), and D (bs—100, di—26, pp—1.00), respectively. Three strains of the *Colacium* genus form clade E (bs—100, di—37, pp—1.00) and two strains of *Cryptoglena pigra* form clade F (bs—100, di—34, pp—1.00). Clade G is not present on all 18S rDNA trees. Instead, its subdivisions, clades G1 and G2 are dispersed on the tree. Clade G1 (bs—<50–52, di—3, pp—<75–0.99) consists of the *Lepocinclis* and *Phacus* species plus one of the two strains of *E. limnophila* (*Discoglena*). Clade G2 (bs—87–96, di—9, pp—1.00) consists of the second strain of *E. limnophila* and *E. spathirhyncha* (*Discoglena*). The relations among main clades as well as the affiliations of *E. anabaena* and *E. proxima* are weakly resolved.

Combined 16S and 18S rDNA. Possible incongruences between data partitions in the combined 16S and 18S rDNA of green euglenoids are most likely caused by rapidly evolving *E. mutabilis* 18S rDNA. The partition homogeneity test produced p values equal to and larger than 0.01, for data containing all taxa and that devoid of *E. mutabilis*, respectively. These results suggest that combining data improved phylogenetic accuracy and that rapidly evolving

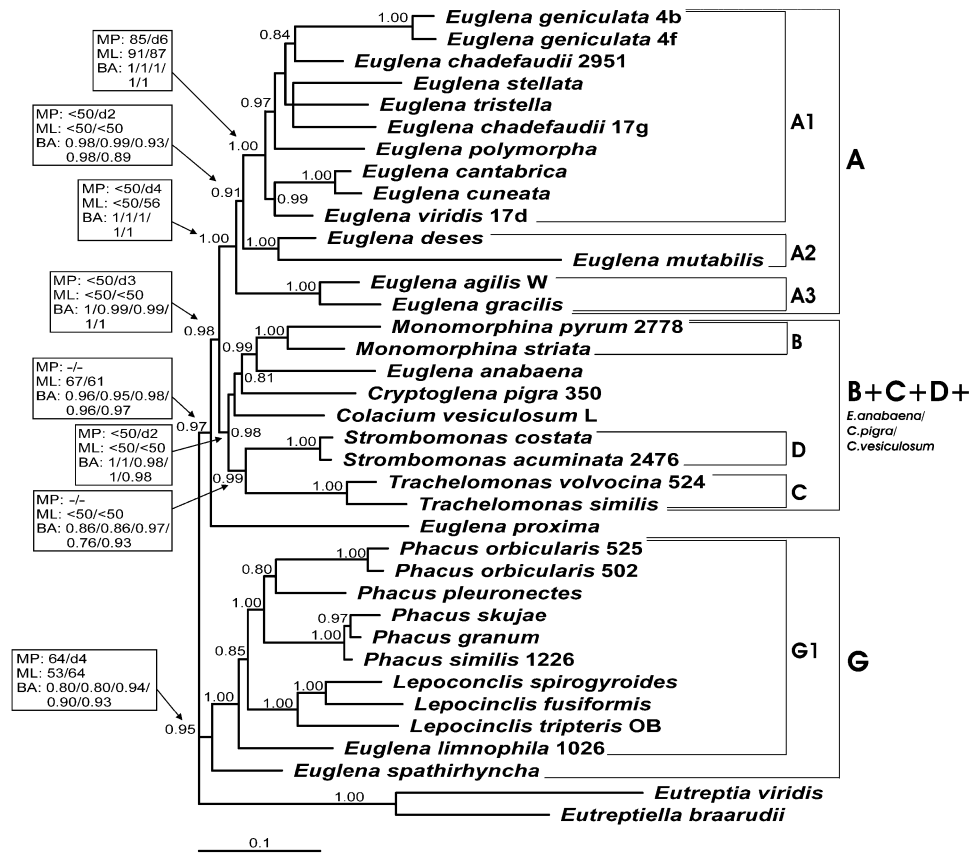

FIG. 3. Pphylogenetic tree of the combined 16S and 18S rDNA obtained by Bayesian inference (partition = yes, model for part 18S and 16S: GTR + G + I, covarion = yes). Numbers at the nodes show posterior probabilities of the tree bipartitions. Probabilities less than 75% are not shown. Numbers in rectangles show bootstrap values/decay indices obtained for the main clades by maximum parsimony (MP) (– / – indicates that the clade does not exist in MP analysis), two bootstrap values obtained by maximum likelihood (ML) analysis (models GTR + I + G/TrN + I + G), and five posterior probabilities obtained by Bayesian inference (part = yes, part 18S and 16S: GTR + I + G, cov = no/part = yes, part 18S: SYM + I + G; part 16S: GTR + I + G, cov = no/part = yes, part 18S: SYM + I + G; part 16S: GTR + I + G, cov = yes/part = no, model GTR + I + G, cov = no/part = no, model GTR + I + G, cov = yes).

E. mutabilis 18S rDNA is responsible for incongruence of the 16S and 18S rDNA trees. Figure 3 shows a tree obtained by the Bayesian method. All trees obtained by the three different methods give almost identical topologies. There are three, well-defined clades: A, B/C/D (plus *E. anabaena*, *C. pigra*, and *C. vesiculosum*) and G. Clade A (bs—<50–56, di—4, pp—1.00) is subdivided into clades A1–3, as on the 16S rDNA tree (Fig. 1), but in contrast to the situation on the latter tree, the relationships between the clades on the combined data tree are better resolved. Clades A1 and A2 are sister groups here (bs—<50, di—2, pp—0.89–0.99), while clade A3 diverges first.

Clade B/C/D (bs—<50, di—2, pp—0.98–1.00) consists of species forming clades B, C, D on 16S and 18S rDNA trees and *E. anabaena*, *C. pigra*, and *C. vesiculosum*. Internal relations within the group are ambiguous. On the ML and BA trees, *Monomorphina* species, *E. anabaena*, *C. pigra*, and *C. vesiculosum* form a sister group to clades C and D represented by *Trachelomonas* and *Strombomonas* species, respectively. However, this topology is not substantially supported—only the posterior probabilities of group C + D in analysis using the covarion model are significant: 0.93–0.99. On all 16S–18S rDNA trees clades A and B/C/D are sisters group (bs—<50, di—3, pp—0.98–1.00), whereas *E. proxima* is located at the base of A + B/C/D clade on the ML and BA trees (bs—61–67, pp—0.95–0.98). Within clade G (bs—53–64, di—4, pp—0.80–0.95) known from the 16S rDNA tree, clades grouping the *Lepocinclis* and *Phacus* species are sister groups, while *E. limnophila* and *E. spathirhyncha* diverge in a paraphyletic manner.

DISCUSSION

Phylogenetic relationships of green euglenoids. All three data sets produce trees that are congruent with respect to nodes, which are supported substantially. In the case of separate 16S and 18S rDNA analyses, relationships between the well-supported branches are unstable, weakly supported, and dependent on the data set and the method of inference. Strikingly different results are obtained when data are combined. The topology of the trees obtained by

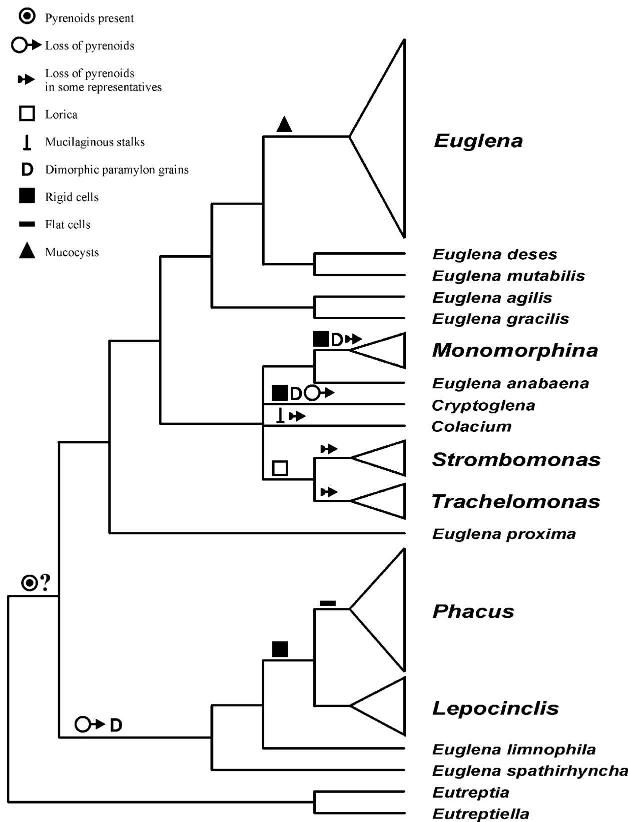

FIG. 4. Character evolution in green euglenoids.

different methods differs only slightly and the support for most of the branches is rather high, particularly with respect to the Bayesian analyses. There is no agreement on how to treat separate partitions of data and the subject remains controversial (Farias et al. 2000). We believe that combining molecular data is beneficial as it leads to the analysis of more characters while it is difficult to establish *a priori* which partition, if any, produces real topology. Moreover, it is possible, at least with respect to BA, to apply different models of sequence evolution to different partitions. Therefore, we place more value on the tree obtained from the combined data, especially when a better resolution is obtained and when the combined analysis does not contradict (weak support does not contradict strong support) the separate analyses. When contradiction is encountered (both different and well-supported topologies), it should be evaluated.

The tree of the combined sequences confirms some of the findings obtained earlier with respect to the relationships within the main groups, mainly A—*Euglena*, (B + E + F)—(*Monomorphina* + *Cryptoglena* + *Colacium*), (C + D)—(*Strombomonas* + *Trachelomonas*), (*Phacus* + *Lepocinclis*), but the relationships between them were not previously resolved. One of the BA trees presented by Brosnan et al. (2003) obtained for the combined SSU and LSU rDNA sequences had similar topology [(A(B/C/D))G], i.e. ((*Euglena*, (*Monomorphina*, (*Strombomonas*, *Trachelomonas*))), (*Phacus*, *Lepocinclis*)), but it had weak support and the topology obtained was different for the ML and MP analyses. In our combined analysis, there are several relationships that are well-resolved for the first time. The most conspicuous and strongly supported involves the grouping together of the *Strombomonas*, *Trachelomonas*, *Cryptoglena*, *Monomorphina*, and *Colacium* genera. This morphologically diversified group of genera shows a relatively low genetic diversity, judging from the lengths of their branches. The group, additionally comprising one strain of *E. anabaena*, is a sister group with respect to *Euglena sensu stricto*, again with strong support. Less conspicuous is the resolution of branching within *Euglena*, which is divided into three sub-clades. The clade composed of *E. deses* and *E. mutabilis* branches off first and then two sister clades emerge: one composed of *Euglena* species, and the second consisting of *E. agilis* and *E. gracilis*. These affiliations were observed by Marin et al. (2003), but with a weak support. The last main and well-resolved clade G consists of the *Phacus* and *Lepocinclis* genera, together with two *Euglena* species, formerly assigned by Zakryś (1986) to the subgenus *Discoglena* (*E. limnophila* and *E. spathirhyncha*). The other three formerly *Discoglena* species (*E. fusca, E. tripteris*, *E. spirogyra*,) were already reclassified by Marin et al. (2003) and Kosmala et al. (2005) into the genus *Lepocinclis*.

Two other *Euglena* species, *E. anabaena* and *E. proxima*, are located outside the *Euglena* clade (A). Neither can be definitely assigned to any one clade by separate analyses, but affiliations of both are resolved the combined analysis. *E. proxima* is a sister group to the (A + B/C/D) clade, while *E. anabaena* lies well within the B/C/D clade. In the system of Zakryś (1986), the lack of pyrenoids is a diagnostic feature, distinguishing *Discoglena* from other subgenera (*Euglena* and *Calliglena*). Even though the system is not consistent with the emerging phylogeny of all euglenoid genera, it does allow distinction between the species from clade A (*Euglena* and *Calliglena*) and G (*Discoglena, Lepocinclis*, and *Phacus*). Pyrenoids have not been observed under the light microscope in *E. proxima*, and therefore it has been assigned to the *Discoglena* subgenus. However, our preliminary results of electron confocal microscopy (data not shown) suggest that *E. proxima* has pyrenoids. Therefore, it is possible that both *E. anabaena* and *E. proxima* are associated with species of the *Calliglena* subgenus, and hence with clade A. The position of *E. proxima* on our combined tree may be the result of long branch attraction, but the position of *E. anabaena* is puzzling, and until confirmation based on more sequence data from closely related species is obtained, both affiliations should be treated with caution.

Taxonomic ramifications. The obtained topology of green euglenoids shows that the genus *Euglena sensu lato* is not polyphyletic as earlier surmised (Milanowski et al. 2001, Marin et al. 2003, Nudelman et al. 2003), but rather paraphyletic due to several *Euglena* species branching off close to the base of the tree. It is

then likely that all of the existing forms of Euglenales evolved from the hypothetical primordial *Euglena* and all contemporary species considered to be *Euglena* are characterized by symplesiomorphies. Therefore, the statement made by Pringsheim that "In fact one can say that the forms described as *Euglena* are those Euglenaceae which do not fit into any of the other genera," (Pringsheim 1956) has a contemporary meaning. Such a hypothetical ancestor of all Euglenales was probably well described by Ehrenberg's words in 1830: "organisms of different shape having a reservoir and having a red eye or not," by which he described green and colorless euglenoids known at that time (*Astasia euchlora*, *Astasia haematodes, E. acus, E. viridis*, and *E. pleuronectes*) and classified in the class *Polygastrica* section (Sectio) *Astasiaea*. This imprecise description still fits *Euglena sensu stricto*, as well as early branching species: *E. spathirhyncha, E. limnophila*, and *E. proxima*. When their phylogenetic position outside clade A (*Euglena sensu stricto*) is confirmed, they should be removed from the genus and either transferred to other genera or assigned a genus of their own. *E. spathirhyncha* and *E. limnophila* clearly do not belong to *Euglena*, whereas the situation with respect to *E. proxima* is far from certain.

We believe that ultimately the genus *Euglena* should be confined to the species of clade A. Clarification is needed with respect to the presence of strains of the same species in more than one clade within this group—*E. chadefaudii* (17G and 2951—this study) and *E. stellata* (Shin and Triemer 2004). As clades A2 and A3 form a paraphyletic assemblage, they should be ultimately assigned to two different subgenera and the concept of the *Calliglena* subgenus should be abandoned or modified.

On the basis of phylogenetic and structural analysis of 18S rDNA sequences, Marin et al. (2003) proposed the partial reclassification of the genus *Euglena*. This included moving *Discoglena* species *sensu* Zakryś (i.e. without pyrenoids), such as *E. spirogyra* and *E. tripteris* to the genus *Lepocinclis*, and leaving the *Euglena* and *Calliglena* species (i.e. with pyrenoids) in the genus *Euglena*. This reclassification did not affect *E. anabaena* (subgenus *Calliglena sensu* Zakryś), *E. limnophila*, and *E. spathirhyncha* (subgenus *Discoglena sensu* Zakryś), because of the uncertain phylogenetic position of these species. Our results are in general agreement with those of Marin et al. (2003) based on 18S rDNA, particularly with respect to *E. spathirhyncha* and *E. limnophila* belonging to clade G, and therefore related to *Phacus* and *Lepocinclis*. On our tree from the combined data, *E. spathirhyncha* is a sister group to the rest of clade G, which consists of *E. limnophila* plus the *Lepocinclis* and *Phacus* genera, whereas on our 18S rDNA tree, the two strains of *E. limnophila* (ASW 08039 and ACOI 1026) do not form a clade. Therefore, both *E. limnophila* and *E. spathirhyncha* should be removed from the genus *Euglena* altogether, i.e. *Discoglena* should be removed from *Euglena*. *E. spathirhyncha* should be given a genus of its own, while the status of *E. limnophila* is not yet certain.

Character evolution. Figure 4 shows a simplified version of the tree from Fig. 3, reflecting topologies obtained by ML and BA analyses with the changes in some morphological characters mapped. It is not clear whether the common ancestor of Euglenales had chloroplasts with pyrenoids. Judging from the presence of pyrenoids in some Eutreptiales (Walne et al. 1986), as well as from their presence in most modern chlorophycean algae, the relatives of the hypothetical precursors of euglenoid chloroplasts, it is likely that they were present in the ancestors of Euglenales. However, pyrenoids are easily lost and acquired in the course of evolution, at least in chlorophycean algae (Nozaki et al. 2002). No matter what the primordial state was, the ancestors of the two main basal clades of the tree differ with respect to the presence of pyrenoids: the ancestor of (A + B/C/D + *E. proxima*) had pyrenoids and the ancestor of G did not. In the first branch, pyrenoids were preserved in clade A (*Euglena sensu stricto*, Marin et Melkonian 2003). In the group B/C/D, the loss of pyrenoids must have occurred after the formation of the genera *Monomorphina*, *Colacium*, *Strombomonas*, *Trachelomonas* (as certain representatives of all of these genera have pyrenoids while others do not), and *Cryptoglena*.

A stiff cell and dimorphic paramylon grains (the presence of both large and small paramylon grains) are present in case of groups B/C/D and G, thus suggesting convergent evolution. However, if the tree is rooted differently and the long branch leading to *E. proxima* is eliminated, the two groups can form a clade and both characteristics may then be considered as emerging only once. The position of the root, based on only one species of *Eutreptia* and one of *Eutreptiella*, both distantly related to the closely related clades of Euglenales, may not be correct. To resolve the issues of the root and the position of *E. proxima*, more sequences from closely related species are needed. In the case of group B/C/D, both a stiff cell and dimorphic paramylon grains are present in *Monomorphina* and *Cryptoglena*, which are separated in the combined analysis by a branch leading to *E. anabaena*. Interestingly, in both of the separate analyses, *Monomorphina* and *Cryptoglena* form a clade, suggesting that both features emerged in the ancestor of the two genera. Dimorphism of paramylon grains is observed in all the members of group G, while the stiffness of the cell appears gradually. The swelling of the pellicle was present in the ancestor of *Phacus* and *Lepocinclis*, leading to the increased stiffness of the cell. In addition, the cells of *Phacus* became flat.

There are clear cases of synapomorphy in Euglenales. Mucocysts are present in all representatives of the monophyletic group A1 in the genus *Euglena*. The genus *Colacium* gained the ability to lead a sedentary life, courtesy of mucilaginous stalks, and a lorica appeared in the ancestor of *Strombomonas* and *Trachelomonas*. It is worth mentioning that monadal forms of *Colacium*,

Strombomonas, and *Trachelomonas* bear resemblance to the monadal forms of *Euglena*.

To summarize, three types of characters can be encountered and were used in the classification of Euglenales: synapomorphies (stalks, loricas, flat cells, mucocysts), which were used to describe and successfully distinguish many taxa, such as *Colacium*, *Strombomonas*, *Trachelomonas*, and *Phacus*; and symplesiomorphies (pyrenoids, metaboly, free-floating cells), which were used erroneously in classification. The third type encompasses features such as the lack of pyrenoids, stiffness of the cell, and the dimorphism of paramylon grains, which may be the result of convergent evolution, and should not be used in classification either. In the history of taxonomic studies of Euglenales, there are many examples of erroneous classification as studies were based on symplesiomorphies of convergent characters (*E. spirogyra*, *E. fusca*, *E. tripteris*, *E. limnophila*). In the era of rapid progress in microscopic techniques and in the building of reliable molecular trees, we can expect new discoveries of morphological synapomorphies, such as the already known mucocysts defining clade A1 (Shin and Triemer 2004), or the presence of a single, spherical (closed) chloroplast in *Monomorphina* (submitted for publication). Nonetheless, it is becoming clear that morphological characteristics are not enough for the classification of organisms as simple as Euglenoids, and that synapomorphies of a different kind are needed, be they of a molecular, physiological, or behavioral nature.

Financial support was provided by the State Committee for Scientific Research (KBN) Grant no. 6PO4CO1617 and 3PO4C08225. We thank Dr. M. F. Santos from the University of Coimbra, Portugal, for providing euglenoid strains, and Dr. Mark Farmer from the University of Georgia, Athens, USA, for providing DNA of *Eutreptiella braarudii*.

Bremer, K. 1994. Branch support and tree stability. *Cladistics* 10:295–304.

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. Syst. Evol. Microbiol.* 53:1175–86.

Cavalier-Smith, T. 1981. Eukaryote kingdoms: seven or nine? *BioSystems* 14:461–81.

Ehrenberg, C. G. 1830. Neue Beobachtungen über blutartige Erscheinungen in Ägypten, Arabien und Sibirien, nebst einer Übersicht und Kritik der früher bekannten. *Pogg. Ann. Physik. Chem.* 94:477–514.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D. & Hall, N. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309:404–9.

Eriksson, T. 1998. AutoDecay 4.02. Program Distributed by Author. Department of Botany, Stockholm University.

Farias, I. P., Orti, G. & Meyer, A. 2000. Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. *J. Exp. Zool.* 288:76–92.

Felsenstein, J. 1985. Confidence limits on phylogenies: approach using the bootstrap. *Evolution* 39:783–91.

Gibbs, S. P. 1978. The chloroplast of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* 56:2883–9.

Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A. M. & Opperdoes, F. R. 2003. Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl. Acad. Sci. USA* 100:1067–71.

Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.

Kosmala, S., Karnkowska, A., Milanowski, R., Kwiatowski, J. & Zakryś, B. (2005). The phylogenetic and taxonomic position of *Lepocinclis fusca* comb. nova (= *Euglena fusca*) (Euglenaceae). Morphological and molecular justification. *J. Phycol.* 41:1258–67.

Lanave, C., Preparata, G., Saccone, C. & Serio, G. 1984. A new method of calculating evolutionary substitution rate. *J. Mol. Evol.* 20:86–93.

Leander, B. S. (2004). Did trypanosomatid parasites have photosynthetic ancestors? *Trends Microbiol.* 12:251–8.

Leander, B. S. & Farmer, M. A. 2001. Comparative morphology of the euglenid pellicle. II. Diversity of strip substructure. *J. Eukaryot. Microbiol.* 48:202–17.

Leander, B. S., Witek, R. P. & Farmer, M. A. 2001. Trends in the evolution of the euglenid pellicle. *Evolution* 55:2215–35.

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.* 46:217–23.

Linton, E. W., Nudelman, M. A., Conforti, V. & Triemer, R. E. 2000. A molecular analysis of the Euglenophytes using SSU rDNA. *J. Phycol.* 36:740–6.

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.

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.* 51:773–81.

Montegut-Felkner, A. E. & Triemer, R. E. 1997. Phylogenetic relationships of selected euglenoid genera based on morphological and molecular data. *J. Phycol.* 33:512–9.

Moreira, D., López-García, P. & Rodriguez-Valera, F. 2001. New insights into the phylogenetic position of diplonemids: G + C content bias, differences of evolutionary rate and a new environmental sequence. *Int. J. Syst. Evol. Microbiol.* 51:2211–9.

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.* 51:783–91.

Nicholas, K. B. & Nicholas, H. B. 1997. GeneDoc: analysis and visualization of genetic variation. Program distributed by authors. http://www.psc.edu/biomed/genedoc/.

Nozaki, H., Onishi, K. & Morita, E. 2002. Differences in pyrenoid morphology are correlated with differences in the rbcL genes of members of the *Chloromonas* lineage (volvocales, chlorophyceae). *J. Mol. Evol.* 55:414–30.

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.* 39:226–35.

Nylander, J. A. A. 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University

Page, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357–8.

Posada, D. & Crandall, K. A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–8.

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.* 36:220–6.

Preisfeld, A., Busse, I., Klingberg, M., Talke, S. & Ruppel, H. G. 2001. Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int. J. Syst. Evol. Microbiol.* 51:751–8.

Pringsheim, E. G. 1956. Contributions towards a monograph of the genus *Euglena*. *Nova Acta Leopold.* 18:1–168.

Rodriguez, F., Oliver, J. L., Marin, A. & Medina, J. R. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142:485–501.

Rogers, M. & Keeling, P. J. 2004. Lateral transfer and re-compartmentalization of calvin cycle enzymes of plants and algae. *J. Mol. Evol.* 58:367–75.

Shin, W. & Triemer, R. E. 2004. Phylogenetic analysis of the genus *Euglena* (Euglenophyceae) with particular reference to the type species *Euglena viridis*. *J. Phycol.* 40:226–35.

Swofford, D. L. 1998. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sinauer Associates, Sunderland, MA.

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.

Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–26.

Tavare, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* 17:57–86.

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.* 23:745–52.

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. *Nucleic Acids Res.* 24:4876–82.

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.* 51:402–16.

Walne, P., Moestrup, O., Norris, R. E. & Ettl, H. 1986. Light and electron microscopical studies of *Eutreptiella eupharyngea* sp. nov. (Euglenophyceae) from Danish and American waters. *Phycologia* 25:109–26.

Wuyts, J., Van de Peer, Y., Winkelmans, T. & De Wachter, R. 2002. The European database on small subunit ribosomal RNA. *Nucleic Acids Res.* 30:183–5.

Zakryś, B. 1986. Contribution to the monograph of polish members of the genus *Euglena* Ehrenberg 1830. *Nova Hedwigia* 42:494–540.

Zakryś, B., Milanowski, R., Empel, J., Borsuk, P., Gromadka, R. & Kwiatowski, J. 2002. Two different species of *Euglena, E. geniculata* and *E. myxocylindracea* (Euglenophyceae), are virtually genetically and morphologically identical. *J. Phycol.* 38:1190–9.

Zharkikh, A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* 39:315–29.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.