International Journal of Systematic and Evolutionary Microbiology (2001), **51**, 773–781 Printed in Great Britain

Phylogenetic analysis of chloroplast small-subunit rRNA genes of the genus *Euglena* Ehrenberg

Rafał Milanowski, Bożena Zakryś and Jan Kwiatowski

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

Author for correspondence: Bożena Zakryś. Tel: +48 22 628 2595. Fax: +48 22 622 6646. e-mail: zakrys@mercury.ci.uw.edu.pl

Almost complete sequences of plastid SSU rDNA (16S rDNA) from 17 species belonging to the order Euglenales (*sensu* Németh, 1997; Shi *et al.*, 1999) were determined and used to infer phylogenetic relationships between 10 species of *Euglena*, three of *Phacus*, and one of each of *Colacium*, *Lepocinclis*, *Strombomonas*, *Trachelomonas* and *Eutreptia*. The maximum-parsimony (MP), maximum-likelihood (ML) and distance analyses of the unambiguously aligned sequence fragments imply that the genus *Euglena* is not monophyletic. Parsimony and distance methods divide Euglenaceae into two sister groups. One comprises of representatives from the subgenera *Phacus*, *Lepocinclis* and *Discoglena* (*sensu* Zakryś, 1986), whereas the other includes members of *Euglena* and *Calliglena* subgenera (*sensu* Zakryś, 1986), intermixed with representatives of *Colacium*, *Strombomonas* and *Trachelomonas*. In all analyses subgenera *Euglena* – together with *Euglena polymorpha* (representative of the subgenus *Calliglena*) – and *Discoglena* – together with *Phacus* and *Lepocinclis* – form two well-defined clades. The data clearly indicate that a substantial revision of euglenoid systematics is very much required, nevertheless it must await while more information can be gathered, allowing resolution of outstanding relationships.

Keywords: chloroplast SSU rDNA, *Euglena*, *Calliglena*, *Discoglena*, molecular phylogeny

INTRODUCTION

The euglenoid flagellates are an ancient, distinct group of protists related to kinetoplastids (Triemer & Farmer, 1991; Cavalier-Smith, 1981, 1993; Corliss, 1994; Dawson & Walne, 1994; Kivic & Walne, 1983; Montegut-Felkner & Triemer, 1997; Linton *et al.*, 1999) comprising green and colourless forms. The origin of the euglenoid chloroplast is not entirely certain. Most likely, it is monophyletic in nature and obtained by secondary symbiosis from green algae (Gibbs, 1978, 1981; Morden *et al.*, 1992; Delwiche & Palmer, 1997).

This paper was presented at the XIIIth meeting of the International Society for Evolutionary Protistology in České Budějovice, Czech Republic, 31 July–4 August 2000.

Abbreviations: ML, maximum-likelihood; MP, maximum-parsimony; 16S rDNA, chloroplast SSU rDNA.

The GenBank accession numbers for chloroplast SSU rDNA (16S rDNA) sequences reported in this papers are given in Methods.

The genus *Euglena* consists of organisms highly diversified with respect to cell architecture. Species included within the genus represent almost entire morphological monad diversity encountered within green euglenoids. Such an immense variety provoked many authors to construct numerous intrageneric classifications. Klebs (1883), for example, described five 'types' of *Euglena* (viridis, deses, oxyuris, spirogyra and acus); whereas Hansgirg (1892) described six 'sections' (Auteuglena, Platyglena, Oxyglena, Spiroglena, Acuglena and Pseudophacus). Other scholars have designated distinct 'groups' [Lemmermann (1913) – four; Gojdics (1953) – eight; Pringsheim (1956) – five], 'categories' [Chu (1946) – four] or 'subgenera' [Zakryś (1986) – three] within the genus *Euglena*, with respect to the chloroplast structure. Other propositions have included the division of the genus into separate 'groups', according to the presence or absence of flagella and/or the type of cell movement. Thus, Elenkin (1924) proposed three 'groups', whereas Popova (1966) designated two 'groups' with 11 'sub-

groups'. Such classifications, based mainly on morphological characters, may reflect morphological convergence, rather then phylogenetic affiliations of considered taxa, as recent molecular studies suggest (Montegut-Felkner & Triemer, 1997; Linton *et al*., 1999, 2000; Preisfeld *et al*., 2000; Thompson *et al*., 1995).

We present here the phylogeny of 18 species of phototrophic euglenoids from the order Euglenales (Németh, 1977; Shi *et al*., 1999), which are classified into several genera: *Colacium* Ehrenberg, *Euglena* Ehrenberg, *Eutreptia* Perty, *Lepocinclis* Perty, *Phacus* Dujardin, *Strombomonas* Deflandre and *Trachelomonas* Ehrenberg. We have conducted this work to investigate the validity of the concept of three subgenera (*Euglena*, *Calliglena* and *Discoglena*) within the genus *Euglena* (Zakryś, 1986). Therefore, the main emphasis is placed on the genus *Euglena* (10 species), whereas the remaining genera are represented by three (*Phacus*), or only single species (the rest of the taxa). This is the first communication on the phylogeny of this group based on the chloroplast SSU rDNA (16S rDNA), commonly used to construct phylogenies of algae (Buchheim *et al*., 1996; Turner *et al*., 1989). Molecular phylogenies of euglenoids published so far were based on 18S rRNA (Montegut-Felkner & Triemer, 1997; Linton *et al*., 1999, 2000; Preisfeld *et al*., 2000) or *rbcL* genes (Thompson *et al*., 1995).

METHODS

Strains and culture conditions. The following strains were used in this study: *Colacium vesiculosum* (UW Łazienki), *Euglena agilis* (UW Pruszków-1), *Euglena anabaena* (SAG 1224-15b), *Euglena geniculata* (SAG 1224-4b), *Euglena polymorpha* (SAG 1224-32), *Euglena spirogyra* (SAG 1224-13b), *Euglena stellata* (SAG 1224-14), *Euglena tripteris* (UW OB.), *Euglena tristella* (SAG 1224-35), *Euglena viridis* (SAG 1224-17d), *Lepocinclis fusiformis* (ACOI 1025), *Phacus orbicularis* (AICB 525), *Phacus pleuronectes* (SAG 1261-3b), *Phacus skujai* (AICB 323), *Strombomonas costata* (ACOI 2992), *Trachelomonas volvocina* (AICB 524) and *Eutreptia viridis* (SAG 1226-1c). They were obtained from the following collections: 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; SAG, Sammlung von Algenkulturen Pflanzenphysiologisches Institut der Universität Göttingen, Germany; UW, Culture Collection of Algae at Department of Plant Systematics and Geography of Warsaw University, Poland. Clones derived from each strain were cultivated in liquid soil-water medium, enriched with a small piece of garden pea (medium 3c, SAG Göttingen; Schlösser, 1994) under identical conditions, in a growth chamber maintained at 17 °C and 16:8 h light/dark, ca. 27 μmol photons m^{-2} s^{-1} provided by cool-white fluorescent tubes.

DNA isolation, amplification and sequencing. The total DNA was isolated from 20–30 mg centrifuged cells, using DNeasy Tissue Kit (Qiagen), according to the manufacturer's protocol (with proteinase K addition). PCR cycle conditions were individually tailored to amplify overlapping fragments of the 16S rRNA gene from a single species, using different combinations of forward and reverse primers listed in Table 1. A 50 μl reaction mixture contained 1 U Taq polymerase (MBI Fermentas), 0·2 mM dNTPs, 2·5 mM $MgCl_2$, 10 pmol each primer, reaction buffer (MBI Fermentas) and 10–50 ng DNA. The PCR protocol consisted of 5 min of denaturation at 95 °C, followed by seven initial cycles comprising 1 min at 95 °C, 2 min at 44–58 °C and 0·5–1 min at 72 °C, then by 30 cycles comprising 0·5 min at 95 °C, 0·5 min at 54–64 °C and 0·5–1 min at 72 °C. The final extension step was performed for 7 min at 72 °C. PCR products were purified either by precipitation with 98 % ethanol and 0·3 M potassium acetate, or by electrophoresis and excision of a predominant band, followed by purification with QIAEXII Gel Extraction Kit (Qiagen). PCR products were sequenced from both strands by cycle sequencing using BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) and primers from Table 1. The readings from ABI Prism 377 DNA sequenator, after removal of primer sequences, were assembled into 'contigs' by the SeqMan program of the LASERGENE package (DnaStar) and checked manually for consistency.

Sequence accession numbers. The GenBank accession numbers for chloroplast SSU rDNA (16S rDNA) sequences reported in this paper are: AF289238, *Colacium vesiculosum*; AF289239, *Euglena agilis*; AF289240, *Euglena anabaena*; AF289241, *Euglena geniculata*; AF289242, *Euglena poly-*

Table 1. Primers used for PCR amplification and sequencing of euglenoid 16S rRNA genes

F denotes forward primers, R–reverse. Position of 3′ end refers to 16S rDNA of *Euglena gracilis* (GenBank no. X12890).

Primer	Position of 3′ end	Sequence (5′–3′)
16SF	39	TTGATCCTGGCTCAGGATGAACGCT
16S223F	242	ATGAGCTTGCATCTGATTAG
16S379R	381	CACGCGGCATTGCTCCGTCA
16S647F	664	ATTTCCAGTGTAGCGGTG
16S781R	781	ACTTAGTATCCATAGTTTACG
16S834R	828	AGGCGGGACACTTAACGCGTT
16S1180R	1170	TGTAGCACGTGTGTCGCCCAG
16SR	1470	CAAGGAGGTGATCCAGCCGCACCTT

morpha; AF289243, *Euglena spirogyra*; AF289244, *Euglena stellata*; AF289245, *Euglena tripteris*; AF289246, *Euglena tristella*; AF289248, *Euglena viridis*; AF289249, *Lepocynclis fusiformis*; AF289250, *Phacus orbicularis*; AF289251, *Phacus pleuronectes*; AF289252, *Phacus skujai*; AF289253, *Strombomonas costata*; AF289254, *Trachelomonas volvocina*; AF289247, *Eutreptia viridis*.

Sequence alignment and phylogenetic analysis. The additional sequence of *Euglena gracilis* 'Z' (X12890) was retrieved from GenBank. Alignment of sequences was obtained using the CLUSTAL W 1.8 program (Thompson *et al*., 1994) with default options, manually checked and edited according to the secondary structure of *Euglena gracilis* (Van de Peer *et al*., 1999). Several regions, which could not be unambiguously aligned, were omitted from analyses. All nucleotides were treated as independent and unordered, multistate characters of equal weight. The alignment of 18 euglenoid sequences used for analysis is available on-line at http://ulmus.bot.uw.edu.pl/~jmkwiato/aln.html, or upon request from the corresponding author. Neighbour-joining trees (Saitou & Nei, 1987) were obtained with the MEGA program, version 1.0 (Kumar *et al*., 1993). Jukes–Cantor (Jukes & Cantor, 1969), Kimura two-parameter (Kimura, 1980), and Tamura–Nei (Tamura & Nei, 1993) models with equal or varying rates among sites were used to calculate distances between sequences. Gamma distribution shape parameters from 0·5 to 1·0 were tested. Maximum-parsimony (MP) and maximum-likelihood (ML) analyses and the nucleotide homogeneity test were performed by PAUP*, version 4.0b4a for Microsoft Windows (Swofford, 1998). Heuristic Search option with MULPARS, tree-bisection-reconnection (TBR) branch swapping, ACCTRAN optimization, and random addition with 100 replicates was used to find the best tree. For ML analyses, the model of nucleotide substitution of Rodriguez *et al*. (1990) was applied, in addition to the models used for distance analyses. Bootstrap support of specific nodes (Felsenstein, 1985) was estimated with 1000 replications (100 for ML analysis) and default options, as implemented in MEGA and PAUP*. Decay indices (Bremer, 1988, 1994) were calculated by the SEPAL program (Salisbury, 2000). The sequence of *Eutreptia viridis*, member of Eutreptiaceae, was used to root the trees. There is evidence, both molecular (Linton *et al*., 1999, 2000; Preisfeld *et al*., 2000; Thompson *et al*., 1995) and morphological (Bourelly, 1970; Németh, 1997; Shi *et al*., 1999), that Eutreptiaceae are indeed an outgroup with respect to Euglenaceae. Trees were drawn by TreeView, version 1.6.1 for Microsoft Windows (Page, 1996).

RESULTS

Almost complete chloroplast 16S rRNA genes were sequenced from 17 species. The sequences, excluding PCR primers, varied in length from 1386 bp in *Lepocinclis*, to 1430 bp in *Eutreptia*. Several regions, constituting 139 positions, which could not be unambiguously aligned, were omitted from analyses. The resulting matrix of 18 sequences, including *Euglena gracilis* sequence obtained from GenBank, had 1350 of total characters, 745 of which were constant. A total of 203 variable characters were parsimony-uninformative, and 402 were parsimony-informative. When gaps were treated as a fifth base, the number of constant, variable uninformative and informative characters was 730, 215 and 405, respectively. Removal of the ambiguous positions substantially improved the homogeneity of base frequencies across taxa [$\chi^2 = 24{\cdot}80$ (df = 51), $P = 0{\cdot}9993$].

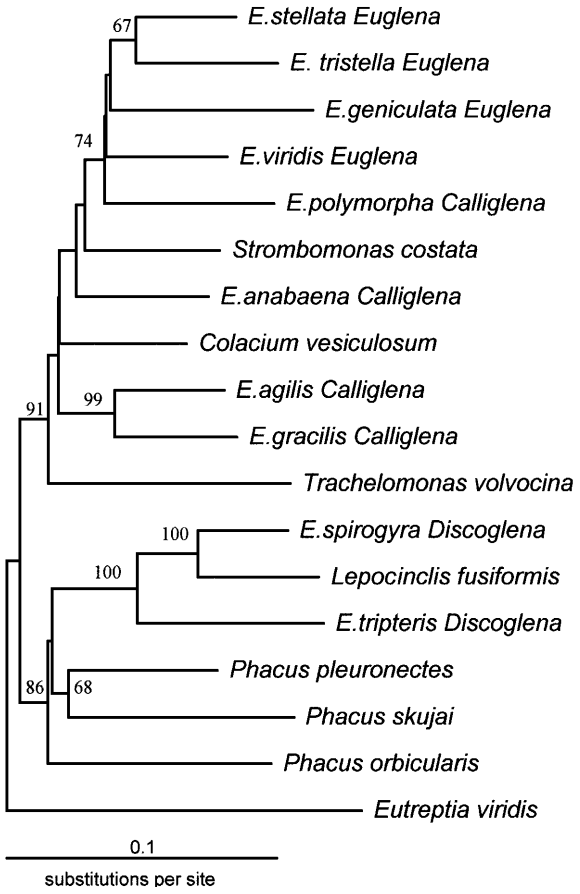

Fig. 1. The neighbour-joining tree of euglenoid 16S rDNA sequences based on Jukes–Cantor distances. Bootstrap values higher than 50% (1000 replications) are shown at the nodes. Subgenera: *Euglena*, *Calliglena*, *Discoglena*.

Fig. 1. shows the distance tree obtained by the Neighbour-joining method using the Jukes–Cantor model of nucleotide substitutions (Jukes & Cantor, 1969). *Eutreptia*, not included in Euglenaceae, and known to diverge first (Bourelly, 1970; Németh, 1997; Linton *et al*., 1999; Shi *et al*., 1999; Preisfeld *et al*., 2000; Thompson *et al*., 1995), was used to root the tree. The main feature of this tree is its division into two well-defined sister groups, both supported by high bootstrap values; 91 and 86%, for the upper and lower clades, respectively. The upper branch consists of representatives of two *Euglena* subgenera (*sensu* Zakryś): *Euglena* (E) and *Calliglena* (C), intermixed with species of three other euglenoid genera: *Colacium*, *Strombomonas* and *Trachelomonas*. None of the affiliations between genera are well supported. Only grouping the species of *Euglena* subgenus together with *Euglena polymorpha* (C) (74%), and that of two other *Calliglena* species (*Euglena gracilis* and *Euglena agilis*) with each other (99%), gain substantial support.

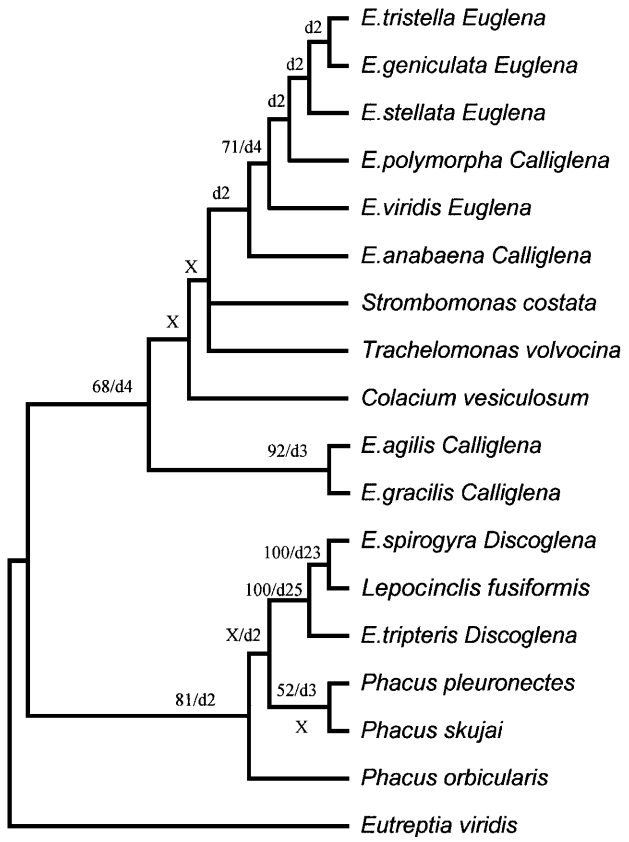

Fig. 2. The strict consensus tree of two most parsimonious trees obtained when gaps are treated as 'missing'. The numbers at nodes indicate the percentage of bootstrap support better than 50% (1000 replications) and decay indices greater than 1 (preceded by d). X denotes the branches collapsed in strict consensus of seven trees obtained when gaps were treated as a fifth character state.

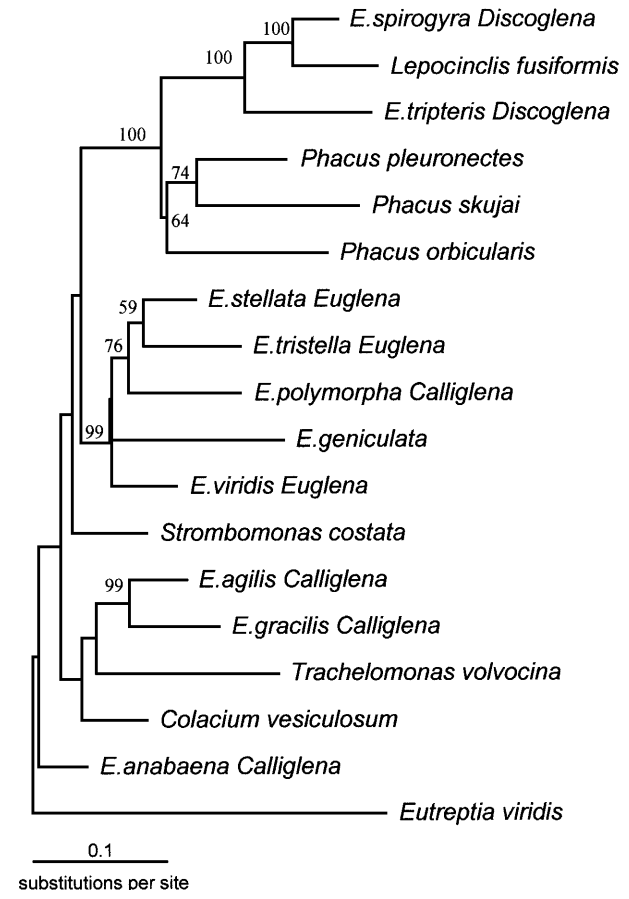

Fig. 3. The maximum-likelihood tree obtained under general-time-reversible model, with unequal rates of substitutions among sites and parameters estimated by ML. Bootstrap values higher than 50% (100 replications) are given at the nodes.

The lower branch consists of early diverging species of *Phacus* and a very well-supported cluster (100%), consisting of *Euglena* subgenus *Discoglena* (D) grouped together with *Lepocinclis*–another genus of Euglenaceae. However, the obtained topology makes *Discoglena* subgenus paraphyletic. Applying Kimura two-parameter (Kimura, 1980) and Tamura-Nei (1993) models of sequence evolution, with rates equal or varying among sites, has not influenced the topology of the tree in any substantial way. The branching order within weakly supported groups was unstable, but the clades with bootstrap values marked in Fig. 1 were always preserved.

Multiple heuristic searches for the most parsimonious topology, with different modes of sequence addition, produced two equally parsimonious trees of 1715 steps, when gaps were treated as 'missing'. Fig. 2 shows a strict consensus of these trees with bootstrap values greater than 50% and decay indices indicating stability of particular nodes. When gaps were treated as a fifth base, seven equally parsimonious trees of 1756 steps were produced. The strict consensus tree obtained from them is different from the one depicted in Fig. 2 in that the internal branches marked by an 'X' have collapsed. Again, two main divisions within euglenoid species, observed on the distance tree of Fig. 1, are visible, although without strong bootstrap and decay support. The same clades of [*Euglena* subgenus + *Euglena polymorpha* (C)] and [*Euglena gracilis* (C) + *Euglena agilis* (C)] in the upper branch; and (*Discoglena* + *Lepocinclis*) in the lower one, are also present. However, all three *Euglena* subgenera are paraphyletic in the parsimony trees.

Topologies of the trees obtained by ML analyses highly depended on the model of nucleotide substitution applied. When the Jukes–Cantor and Felsenstein (Felsenstein, 1981) model was used, the topology obtained was generally consistent with the trees from Figs 1 and 2. The tree with the best score (Ln likelihood = −9175·4), shown in Fig. 3, was produced when the model of Rodriguez *et al.* (Rodriguez *et al.*, 1990; Yang, 1994a, b) was used, with estimated values for the following parameters: base frequencies (A = 0·2826, C = 0·1728, G = 0·2619, T = 0·2827), instantaneous rates of substitutions (AC = 1·288, AG =

Table 2. Kishino–Hasegawa test of different tree topologies under the MP criterion

Tree	−ln Likelihood	−ln L difference	SD	*P*
1	9194·3	18·84	12·64	0·1363
2	9193·8	18·39	14·91	0·2176
3	9200·8	25·36	16·37	0·1215
4	9175·4	(best)		
5	9182·4	6·99	5·42	0·1971
6	9282·2	106·79	24·21	< 0·0001
7	9212·4	36·94	13·70	0·0071
8	9347·7	172·24	30·73	< 0·0001

Table 3. Kishino–Hasegawa test of different tree topologies under the MP criterion

Tree	Length	Length difference	SD	*P*
1	1720	5	9·54196	0·6004
2	1715	(best)		
3	1720	5	4·58225	0·2754
4	1735	20	13·18455	0·1295
5	1741	26	13·77097	0·0592
6	1798	83	19·03064	< 0·0001
7	1748	33	14·29492	0·0211
8	1811	96	19·42815	< 0·0001

6·039, AT = 2·234, CG = 0·4977, CT = 10·19), proportion of invariable sites (0·3095) and gamma-shaped parameter (0·6428). In this tree, *Euglena anabaena* (*Calliglena*) diverges first, followed by a clade composed of two other *Calliglena* species (*Euglena agilis* and *Euglena gracilis*), as well as *Colacium* and *Trachelomonas*. Two sister clades: (*Phacus*, *Lepocinclis*, *Discoglena*) and (*Euglena* subgenus, *Euglena polymorpha*) are supported by very high bootstrap values of 100 and 99%, respectively. Consistently, *Euglena agilis*/*Euglena gracilis* and *Discoglena*/*Lepocinclis* clades gain substantial support, but all other groupings do not.

In order to ascertain how much the ML tree is different from the trees obtained by distance and MP analyses, the Kishino–Hasegawa test (Kishino & Hasegawa, 1989) was performed on the trees from Figs 1, 2 and 3 (trees 1, 2 and 4 in Tables 2 and 3), under the optimality criterion of MP and ML. Additionally, some alternative trees were also tested. One set was identical to the trees from Figs 2 and 3, but the *Euglena* subgenus was made monophyletic by exchanging *Euglena polymorpha* (C) with *Euglena viridis* (E), (trees 3 and 5, respectively). The second set consisted of trees derived from the ML tree in Fig. 3, by interchanging *Strombomonas* and *Euglena viridis* (tree 6) or *Phacus orbicularis* (tree 7), or the latter two (tree 8). Table 2 shows the results of the comparisons under the model of Rodriguez *et al*. (1990), with the above-mentioned parameter values used to obtain the ML tree in Fig. 3. None of the trees were significantly worse than the one in Fig. 3, except the trees 6, 7 and 8, in which either one of the well-established clades (*Phacus*, *Lepocinclis*, *Discoglena*) or (*Euglena* subgenus, *Euglena polymorpha*) was deprived of its early diverging member. Qualitatively identical results were obtained, when other models of the evolutionary change of sequences (Felsenstein, 1981; Hasegawa *et al*., 1985), with or without estimation of various parameter values, were applied under the ML criterion (not shown). Table 3 shows the results of the comparisons under MP criterion. Again, the results are similar, supporting the notion that the (*Phacus*, *Lepocinclis*, *Discoglena*) and (*Euglena* subgenus + *Euglena polymorpha*) clades are well established and that the monophyletic nature of the *Euglena* subgenus could not be excluded. Similar results were obtained when trees from Figs 1 and 2 were altered in the same way as the one from Fig. 3 (not shown).

DISCUSSION

The phylogeny obtained in this as well as other molecular studies (Montegut-Felkner & Triemer, 1997; Linton *et al*., 1999, 2000; Preisfeld *et al*., 2000; Thompson *et al*., 1995) is not consistent with the present classifications based on morphological, particularly non-chloroplast characters such as the shape and rigidity of the cell, the presence of the flagellum, the nature of its movement, or the ability to form palmella.

One of the authors, Zakryś (1986), proposed a system of intrageneric *Euglena* classification in which the following are treated as the main criteria in the classification of the *Euglena* species: (i) the position of chloroplasts in the cell (axial or parietal); (ii) their number, size, shape; (iii) and presence or absence of pyrenoids. We also believe that the paramylon grain features are correlated with the characters mentioned above. In most of the species with small, parietal, disc-shaped and numerous chloroplasts without pyrenoids, the paramylon grains are dimorphic. Those of the one type are scarce (one, two or a few per cell), large rod-like or ring-shaped, while those of the other one are more numerous and small (Zakryś unpublished observations). Two of the chloroplast features (the presence or absence of pyrenoids and the position of chloroplasts in the cell) are assumed to be fundamental for this classification, which has divided the genus *Euglena* into three subgenera: *Euglena*, *Calliglena* and *Discoglena*. All species belonging to the subgenus *Euglena* have axial chloroplasts with pyrenoids. In the subgenus *Calliglena*, chloroplasts are partially or entirely parietal, and also have pyrenoids. In the subgenus *Discoglena*, all chloroplasts are parietal, and do not have pyrenoids. The evidence for the concept of three subgenera (*Euglena*, *Calliglena* and *Discoglena*) comes from the phenetic analysis of the representative

set of 58 taxa by different distance methods, describing 28 characters with 129 states (Batko & Zakryś, 1995).

Our analysis of 16S rDNA, as well as that of 18S rDNA (Linton *et al*., 1999, 2000), suggest a clear-cut separation of taxa included in the subgenus *Discoglena* from the rest of the *Euglena* species, which together with the genera *Phacus* and *Lepocinclis* form a well-defined clade. However, the *Discoglena* subgenus represented here by only two species, is clearly paraphyletic. Both *Phacus* and *Lepocinclis* have the same type of photosynthetic apparatus as organisms classified as *Discoglena*. All have numerous, small, disc-shaped, parietal chloroplasts without pyrenoids, often with large paramylon grains. These large paramylon grains distinguish them from the other *Euglena* species, as well as from *Strombomonas*, *Trachelomonas*, *Colacium* and *Eutreptia*. *Eutreptia* is considered to have primitive chloroplast features (Dawson & Walne, 1991): not parietal, small and numerous chloroplasts devoid of pyrenoids – which are present in the cytoplasm – and small paramylon grains. The only character states that differentiate *Phacus* and *Lepocinclis* from the subgenus *Discoglena* are the rigidity and flatness of their cells. However, considering molecular data, such features may not be taxonomically significant. There has been continuous discussion concerning the taxonomic position of *Lepocinclis* and *Phacus*. For example, *Euglena tripteris* was first described by Dujardin (1841) as *Phacus tripteris*, later moved by Klebs (1883) to the genus *Euglena*, but now, according to 16S rDNA data, seem to be associated with *Phacus* again. *Euglena texta* (Duj.) Hübner, is still considered by some authors as *Lepocinclis texta* Lemmermann (Huber-Pestalozzi, 1955; Tell & Conforti, 1986; Compere, 1989; Németh, 1997; Dillard, 2000), but by others as *Euglena texta* (Asaul, 1975; Zakryś, 1986; Zakryś & Walne, 1994; Kim *et al*., 1998; Shi *et al*., 1999). One possibility to alleviate this situation would be to return *Phacus* and *Lepocinclis* to the *Euglena* genus (Linton *et al*., 2000) and classify as *Discoglena* those species which cluster with it, thus making the subgenus monophyletic. Alternatively, if *Discoglena* proves to be nested entirely within a wider *Phacus* or *Lepocinclis* clade it should be renamed accordingly.

When *Eutreptia* is used as an outgroup in distance and parsimony analyses, a sister group to the (*Phacus*, *Lepocinclis*, *Discoglena*) clade consisting of genera: *Colacium*, *Strombomonas* and *Trachelomonas* and subgenera *Calliglena* and *Euglena*, of the genus *Euglena* is present (Figs 1 and 2). Most, if not all, species of these taxa have pyrenoids in chloroplasts (Asaul, 1975; Popova & Safonowa, 1976), which may not always be very visible under light microscope, e.g. in the subgenus *Euglena*, *Colacium vesiculosum*, *Trachelomonas volvocinopsis* and *Euglena proxima*. This uncertainty was resolved by electron microscopic studies (Dragos *et al*., 1979; Péterfi *et al*., 1979; Zakryś & Walne, 1998; Brown, Zakryś & Farmer, unpublished data). Similar findings may be expected in *Strombomonas costata*, since it has relatively large chloroplasts, similar in appearance to those having pyrenoids (Huber-Pestalozzi, 1955; Popova, 1966; Németh, 1997; Shi *et al*., 1999). However, in the cells of the strains of *Colacium* and *Strombomonas* maintained in our laboratory, pyrenoids are not visible under a light microscope.

In some ML analyses, particularly those using more complex models of sequence evolution (Fig. 3) the 'pyrenoid' clade does not exist, but its species diverge successively from the base of the tree. However, the clade consisting of *Euglena* subgenus and *Euglena polymorpha* (C), present on both distance (Fig. 1) and parsimony (Fig. 2) trees, is even much better supported (bootstrap value 99 %). It remains to be seen whether the subgenus *Euglena*, with axially located chloroplasts, constitutes a natural assemblage. Its representatives (*Euglena viridis*, *Euglena stellata*, *Euglena tristella* and *Euglena geniculata*) are always grouped closely together in 16S rDNA, and form a clade in 18S rDNA trees, although only two species (*Euglena viridis*, *Euglena stellata*) are included in the latter study (Linton *et al*., 2000). Even though they are paraphyletic in most analyses of 16S rDNA, the monophyly of the subgenus *Euglena* cannot be ruled out. If they do form a clade, they are not a sister group with respect to the *Calliglena* subgenus, with chloroplasts located differently, and whose one representative, *Euglena polymorpha* is closely associated with them. The rest of the *Calliglena* subgenus, as well as *Colacium*, *Strombomonas* and *Trachelomonas* are outside of the *Euglena* subgenus clade, but their relationships could not be established with great confidence, except for the close affiliation of two *Calliglena* species *Euglena agilis* and *Euglena gracilis*. Our analysis does not seem to agree with the analysis of *rbcL* gene (Thompson *et al*., 1995) in one instance. *Euglena gracilis* is more closely associated with *Euglena geniculata* than with *Euglena pisciformis*, which apparently is synonymous with *Euglena agilis* (Zakryś & Kucharski, 1996; Zakryś *et al*., 1996; Zakryś, 1997a, b). All other relationships between *Calliglena* and subgenus *Euglena* species are not well resolved in the *rbcL* gene analysis (Thompson *et al*., 1995).

Subgenus *Calliglena* is the most genetically diversified of the three *Euglena* subgenera, at least with respect to species sampled here and in the study of 18S rDNA (Linton *et al*., 2000), where *Euglena anabaena* is a sister group to the rest of Euglenales. In some of our ML analyses, *Euglena anabaena* also constitutes a sister lineage to the rest of Euglenales, comprising additionally of the *Colacium*, *Strombomonas* and *Trachelomonas* species, whose chloroplasts are similar to those of *Calliglena*. It is therefore likely, that the chloroplast features of all these taxa represent plesiomorphic character states. They differ from *Eutreptia* only in that they have pyrenoids in chloroplasts, not in the cytoplasm. On the other hand, some of the features of the photosynthetic apparatus of the two remaining subgenera are likely to be autapomorphies. In the case

of *Discoglena*, as well as *Phacus* and *Lepocinclis*, they are the complete lack of pyrenoids in the cell, the disc-like shape of the parietaly located chloroplasts, and the dimorphic appearance of the paramylon grains (see above). In the case of the *Euglena* subgenus, the few large, axially located chloroplasts (only one, two or three per cell), seem to be an autapomorphy.

Several interesting questions arise from, or are left unresolved by, this study. One is whether lorica (the cell envelope composed of mineral-impregnated mucilage) is a shared-derived character in *Trachelomonas* and *Strombomonas*. On the16S rDNA trees, these species are not closer together than they are to any other member of the 'pyrenoid' clade. Therefore, before resolving the pattern of their relative divergence, no conclusion can be reached about the emergence (and possible loss) of this feature. *Colacium* is a sedentary species, but it is clustered together with free swimming members of *Euglena*, *Strombomonas* and *Trachelomonas* in parsimony and distance trees. Some authors have included the genus *Colacium* in the family Euglenaceae (Pringsheim, 1956; Németh, 1997; Popova, 1966; Shi *et al*., 1999). Others, however, have regarded it as a member of the family Colaciaceae (Popova & Safonova, 1976; Compere, 1989), or even have placed it in a separate order Colaciales (Bourrelly, 1970; Tell & Conforti, 1986), opposite the order Euglenales, consisting of all the rest of the green and colourless euglenoids. Genera *Colacium*, *Lepocinclis*, *Strombomonas* and *Trachelomonas* are represented here by single species. They may be polyphyletic and/or paraphyletic, as suggested by 18S rDNA data, with respect to *Lepocinclis* and *Phacus* (Linton *et al*., 2000). Resolving all of these issues requires greater amount of sequence data on more taxa.

Conclusions

Analysis of 16S rDNA confirms that the genus *Euglena* is not monophyletic.

Euglena species classified as *Discoglena* are firmly associated with the genera *Lepocinclis* and *Phacus*, with which they share essential features of chloroplasts. Different schemes are possible to render these three taxa monophyletic.

The question of whether the subgenus *Euglena*, with unique, axially located chloroplasts forms a clade, could not be convincingly resolved, but they are closely related and their monophyly is not contradicted by 16S rDNA data.

Species of the *Calliglena* subgenus, which share chloroplast features with *Colacium*, *Strombomonas* and *Trachelomonas*, are the most genetically diversified among the *Euglena* species and may be among the first diverging lineages of Euglenales. Therefore, they are not likely to be a natural group.

The molecular data obtained so far, beg for the taxonomic revision of the order Euglenales, and of all the genera included therein. Yet, due to many still unresolved relations, more sequence data are needed before the gruelling task of constructing a natural and consistent system of taxonomical hierarchy can be achieved. We believe that, for the time being, even minor changes in classification may cause more harm than good.

ACKNOWLEDGEMENTS

This work was supported by the State Committee for Scientific Research (KBN) Grant no. 6PO4C01617. We thank Professor Dr U. G. Schlösser, Sammlung von Algenkulturen (SAG), Göttingen, Germany; Professor L. S. Péterfi and Dr C. Popescu from Cluj-Napoca University, Romania; and Dr M. F. Santos from University of Coimbra, Portugal, for providing euglenoid strains. Technical suggestions of Dr P. Borsuk and A. Katana are also appreciated. The input of two anonymous reviewers has greatly helped in improving the manuscript.

REFERENCES

Asaul, Z. I. (1975). *Survey of the Euglenophytes of the Ukrainian SSR*, pp. 1–407. Kiev, Naukova Dumka (in Russian).

Batko, A. & Zakryś, B. (1995). Numerical proof of the subgeneric classification of *Euglena*. *Algol Stud* **79**, 1–18.

Bourelly, P. (1970). Les algues d'eau douce. III: Les algues bleues et rouges. Les Eugléniens, Peridiniens et Cryptomonadines, pp. 1–512. Paris.

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**, 795–803.

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

Buchheim, M. A., Lemieux, C., Otis, C., Gutell, R. R., Chapman, R. L. & Turmel, M. (1996). Phylogeny of the Chlamydomonadales (*Chlorophyceae*): a comparison of ribosomal RNA gene sequences from the nucleus and the chloroplast. *Mol Phylogenet Evol* **5**, 391–402.

Cavalier-Smith, T. (1981). Eukaryote kingdoms: seven or nine?, *Biosystems* **14**, 461–481.

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiol Rev* **57**, 953–994.

Chu, S. P. (1946). Contributions to our knowledge of the genus *Euglena*. *Sinensia* **17**, 75–134.

Compere, P. (1989). Flora pratique des algues d'eau douce de Belgique. I: *Cryptophyceae*, *Dinophyceae*, *Raphidophyceae*, *Euglenophyceae*, pp. 1–208. Meise: Jardin Botanique National de Belgique.

Corliss, J. O. (1994). An interim utilitarian user-friendly hierarchical classification of the protists. *Acta Protozool* **33**, 1–51.

Dawson, N. S. & Walne, P. L. (1991). Structural characterization of *Eutreptia pertyi* (Euglenophyta). I. General description. *Phycologia* **30**, 287–302.

Dawson, N. S. & Walne, P. L. (1994). Evolutionary trends in euglenoids. *Arch Protistenkd* **144**, 221–225.

Delwiche, C. F. & Palmer, J. D. (1997). The origin of plastids and their spread via secondary symbiosis. *Plant Syst Evol* (suppl.) **11**, 53–86.

Dillard, G. E. (2000). Freshwater algae of the Southeastern United States. Part 7: Pigmented *Euglenophyceae*. *Bibl Phycol* **106**, 1–135.

Dragos, N., Péterfi, S. & Craciun, C. (1979). Fine structure of

Euglena. II. *Euglena stellata* Mainx and *Euglena viridis* Ehrenberg. *Nova Hedwigia* **31**, 223–246.

Dujardin, F. (1841). Historie naturalle des Zoophytes-Infusorires, pp. 1–684. Paris.

Elenkin, A. A. (1924). Die Euglenarum sine flagello, sectione nova. *Not Syst Inst Cryptogam Hortic Bot Republ Ross* **3**, 9–11.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368–376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Gibbs, S. P. (1978). The chloroplasts of Euglena may have evolved from symbiotic green algae. *Can J Bot* **56**, 2883–2889.

Gibbs, S. P. (1981). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann NY Acad Sci* **361**, 193–208.

Gojdics, M. (1953). The genus *Euglena*, pp. 1–268. Madison: The University of Wisconsin Press.

Hansgirg, A. (1892). Prodromus der Algenflora von Böhmen. Zweiter Teil. *Arch Naturwiss Durchforsch von Böhmen* **8**, 1–266.

Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating the human-ape split by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160–174.

Huber-Pestalozzi, P. (1955). Das Phytoplankton des Süsswassers, 4. Teil: *Euglenophyceen*, pp. 1–606. Stuttgart: Die Binnengewässer.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kim, J. T. S., Boo, S. M. & Zakryś, B. (1998). Floristic and taxonomic accounts of the genus *Euglena* (*Euglenophyceae*) from Korean fresh waters. *Algae* **13**, 173–197.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the ML estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol* **29**, 170–179.

Kivic, P. A. & Walne, P. L. (1983). Algal photosensory apparatus probably represent multiple parallel evolutions. *Biosystems* **16**, 31–38.

Klebs, G. (1883). Über die Organization einiger Flagellatengruppen und ihre Beziehungen zu Algen und Infusorien. *Unters Bot Inst Tübingen* **1**, 233–362.

Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular evolutionary genetics analysis, version 1.0. University Park, PA: Pennsylvania State University.

Lemmermann, E. (1913). *Eugleninae*. In *Die Süsswasserflora Deutschlands, Österreichs, und der Schweiz*, pp. 215–274. Jena: Gustav-Fisher.

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.

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.

Morden, C. W., Delwiche, C. F., Kuhsel, M. & Palmer, J. D. (1992). Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems* **28**, 75–90.

Németh, J. (1997). A guide for the identification of *Euglenophyta* occurring in Hungary, I, pp. 1–319. Vizi Természet-És Környezetvédelem (in Hungarian).

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Péterfi, L. S., Dragos, N. & Craciun, C. (1979). Fine structure of *Euglena*. I. *Euglena tristella* Chu. *Nova Hedwigia* **31**, 197–221.

Popova, T. G. (1966). *Euglenophyta*. Part 1. Flora sporovych rastenij SSSR **8**, pp. 1–411. Moskva-Leningrad, Izd. Nauka (in Russian).

Popova, T. G. & Safonova, T. A. (1976). Flora plantarum cryptogamarum URSS **9**. Euglenophyta. **2**, pp. 286. Leningrad: Izd. Nauka (in Russian).

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.

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

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

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Salisbury, B. A. (2000). SEPAL: strongest evidence and parsimony analyzer. Version 1.2. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.

Schlösser, U. G. (1994). SAG–Sammlung von Algenkulturen at the University of Göttingen. Cataloge of strains 1994. *Bot Acta* **107**, 113–187.

Shi, Z., Wang, Q., Xie, S., Dai, J. & Chen, L. (1999). Flora algarum sinicsrum aquae dulcis. Tomus VI: *Euglenophyta*, pp. 297–414. Edited by Academiae Sinicae: Science Press (in Chinese).

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

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**, 512–526.

Tell, G. & Conforti, V. (1986). Euglenophyta pigmentadas de la Argentina. *Bibl Phycol* **75**, 1–201.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

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**, 745–752.

Triemer, R. E. & Farmer, M. A. (1991). An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. *Protoplasma* **164**, 91–104.

Turner, S. T., Burger-Wiersma, S. J., Giovannoni, L., Mur, R. & Pace, N. R. (1989). The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature* **337**, 380–382.

Van de Peer, Y., Robbrecht, E., de Hoog, S., Caers, A., De Rijk, P. & De Wachter, R. (1999). Database on the structure of small subunit ribosomal RNA. *Nucleic Acids Res* **27**, 179–182.

Yang, Z. (1994a). Estimating the pattern of nucleotide substitution. *J Mol Evol* **39**, 105–111.

Yang, Z. (1994b). ML phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306–314.

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

Zakryś, B. (1997a). On the identity and variation of *Euglena agilis* Carter (= *E. pisciformis* Klebs). *Algol Studies* **86**, 81–90.

Zakryś, B. (1997b). The taxonomic consequences of morphological and genetic variability in Euglena agilis Carter (Euglenophyta): species or clones in Euglena? *Acta Protozool* **36**, 157–169.

Zakryś, B. & Kucharski, R. (1996). Microevolutionary processes in *Euglena pisciformis*. Genetic drift or adaptation? *Algol Stud* **81**, 23–37.

Zakryś, B. & Walne, P. L. (1994). Floristic, taxonomic and phytogeographic studies of green *Euglenophyta* from the Southeastern United States, with emphasis on new and rare species. *Algol Studies* **72**, 71–114.

Zakryś, B. & Walne, P. L. (1998). Comparative ultrastructure of chloroplasts in the subgenus *Euglena* (Euglenophyta): taxonomic significance. *Cryptogam Algol* **19**, 3–18.

Zakryś, B., Kucharski, R. & Moraczewski, J. (1996). Genetic and morphological variability among clones of *Euglena pisciformis* based on RAPD and biometric analysis. *Algol Stud* **81**, 1–21.