#/149

Protist, Vol. 150, 33–42, March 1999 © Urban & Fischer Verlag
http://www.urbanfischer.de/journals/protist

Protist

ORIGINAL PAPER

Phylogenetic Affinities of *Diplonema* within the Euglenozoa as Inferred from the SSU rRNA Gene and Partial COI Protein Sequences

Dmitri A. Maslov[a,1], Shinji Yasuhira[b,2], and Larry Simpson[b]

[a]Department of Biology, University of California, Riverside, CA 92521, USA
[b]Howard Hughes Medical Institute, Department of Molecular, Cell and Developmental Biology and Department of Medical Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095-1662, USA

Submitted November 4, 1998; Accepted February 6, 1999
Monitoring Editor: Mitchell L. Sogin

In order to shed light on the phylogenetic position of diplonemids within the phylum Euglenozoa, we have sequenced small subunit rRNA (SSU rRNA) genes from *Diplonema* (syn. *Isonema*) *papillatum* and *Diplonema* sp. We have also analyzed a partial sequence of the mitochondrial gene for cytochrome c oxidase subunit I from *D. papillatum*. With both markers, the maximum likelihood method favored a closer grouping of diplonemids with kinetoplastids, while the parsimony and distance suggested a closer relationship of diplonemids with euglenoids. In each case, the differences between the best tree and the alternative trees were small. The frequency of codon usage in the partial *D. papillatum* COI was different from both related groups; however, as is the case in kinetoplastids but not in *Euglena*, both the non-canonical UGA codon and the canonical UGG codon were used to encode tryptophan in *Diplonema*.

Introduction

The phylum, Euglenozoa, includes four major groups: euglenoids, kinetoplastids (consisting of bodonids and trypanosomatids), diplonemids (including the genera *Diplonema* (syn. *Isonema*) and *Rhynchopus*) and *Postgaardi* (incertae sedis) (Cavalier-Smith 1993; Corliss 1994; Simpson 1997). This grouping is supported by a large amount of morphological and ultrastructural data; however, the relationship among these groups is not resolved. In particular, the phylogenetic affinity of *Diplonema* is uncertain, with these organisms sharing a variety of features with both euglenoids and kinetoplastids, but clearly differing from either group (Triemer 1992; Triemer and Farmer 1991; Triemer and Ott 1990). Initially it was proposed that *Diplonema* represents an evolutionary link between bodonids, which were considered ancestral to the entire group, and euglenoids (Kivic and Walne 1984). Subsequent scenarios considered an early divergence of all three lineages from the common ancestor, and in one case *Diplonema* was proposed to be a close relative of euglenoids (Willey et al. 1988), while in another it represented a separate branch (Triemer 1992; Triemer and Farmer 1991). Although morphological data

[1]Corresponding author;
fax 1-909-787-4286
e-mail maslov@ucrac1.ucr.edu
[2]Present address: Research Reactor Institute, Kyoto University, Kumatori-cho, Noda, Sennan-gun, Osaka 590-04, Japan

1434-4610/99/150/01-33 $ 12.00/0

proved to be insufficient to corroborate or refute any of these hypotheses, no molecular data on diplonemids has been obtained until now. An usual base (J) was recently found in the nuclear DNA of *Diplonema* and kinetoplastids, but no search was performed for this base in euglenoids (Van Leeuwen et al. 1998). In the present work, we have attempted to resolve the relationships among the major euglenozoan groups by a molecular phylogenetic analysis of the small subunit (SSU) ribosomal RNA and cytochrome c oxidase subunit I (COI) polypeptide sequences.

Results

Phylogenetic Position of *Diplonema* Based on the SSU Data

The SSU rRNA gene sequences from *D. papillatum* and *Diplonema* sp. were amplified from the total cell DNA with the 5' and 3' conserved sequence oligonucleotides, S762 and S763, respectively, and aligned with the sequences from kinetoplastids and euglenoids. The alignment contained 17 taxa and 1349 alignable characters including gaps. The kinetoplastids included representatives of the major trypanosomatid clades (*Leishmania tarentolae, Herpetomonas muscarum, Phytomonas serpens, Blastocrithidia culicis* and *Trypanosoma brucei*) (Hollar et al. 1998) and four currently available bodonids (*Dimastigella trypaniformis, Rhynchobodo* sp., *Trypanoplasma borreli* and *Bodo caudatus*). Four euglenoid sequences (*Euglena gracilis, Khawkinea quartana, Lepocinclis ovata* and *Petalomonas cantuscygni*), were retrieved from the GenBank™ database. The alignment contained two slowly evolving outgroup sequences – *Physarum polycephalum* and *Saccharomyces cerevisiae*.

Analysis of the data with the maximum likelihood method using the PAUP* program produced the tree shown in Figure 1. Three major groups of Euglenozoa are represented by the corresponding monophyletic clades, and the bootstrap support for their monophyly is high. The clade of *Diplonema* is shown as a sister-group of kinetoplastids. However this affinity is not supported by the bootstrap analysis. Indeed, the maximum likelihood consensus majority tree (not shown) shows the clade of diplonemids as a sister-group of euglenoids with a 77% bootstrap level. The branching order which united kinetoplas-

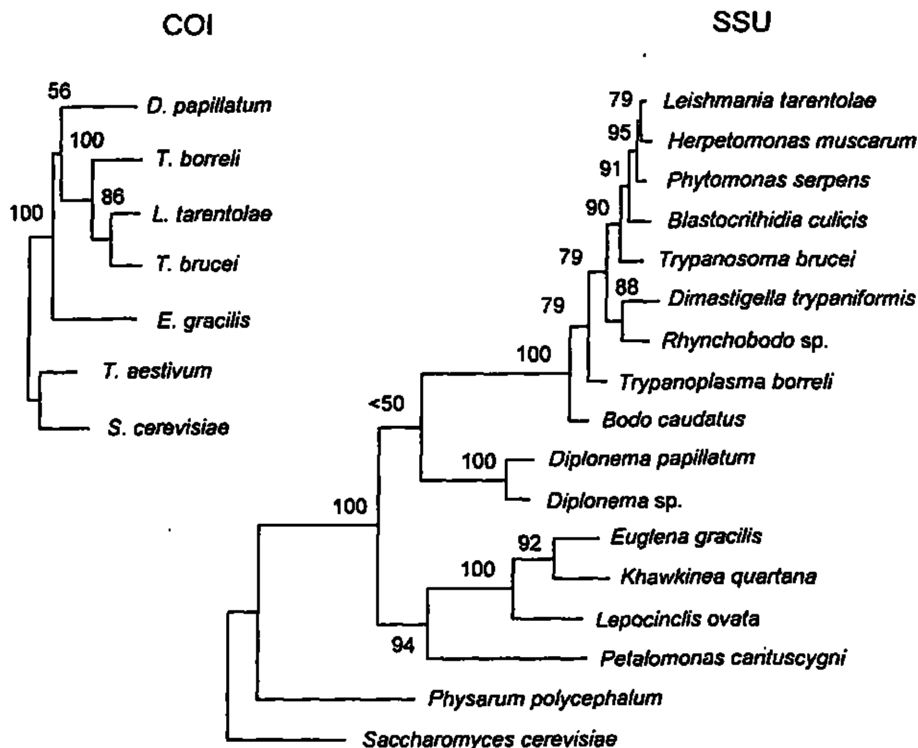

Figure 1. Maximum likelihood SSU rRNA and COI protein phylogenetic trees of Euglenozoa. Ln likelihood of the SSU tree is –9936.69451. Bootstrap values were obtained for 100 pseudoreplicates. The sequences have the following Genbank™ accession numbers: *L. tarentolae* – M84225, *H. muscarum* – L18872, *P. serpens* – AF016323, *B. culicis* – U05679, *T. brucei* – M12676, *D. trypaniformis* – X76494, *Rhynchobodo* sp. – U67183, *T. borreli* – L14840, *B. caudatus* – X53910, *D. papillatum* – AF119811 (this work), *Diplonema* sp. – AF119812 (this work), *E. gracilis* – M12677, *K. quartana* – U84732, *L. ovata* – AF061338, *P. cantuscygni* – U84731, *P. polycephalum* – X13160, *S. cerevisiae* – J01353. Ln likelihood of the COI tree is –3572.64 +/– 97.57. Bootstrap values were obtained for 1000 pseudoreplicates. The corresponding Genbank™ accession numbers are: *L. tarentolae* - M10126, *T. brucei* – M94286, *T. borreli* - U11683, *D. papillatum* – AF119813 (this work), *E. gracilis* – U49052, *T. aestivum* – Y00417, *S. cerevisiae* – M97514.

tids and diplonemids as shown in Fig. 1 occurred in only 19% of the bootstrap replicates. This value may be an underestimate of the actual support level: in order to ease the computations, the bootstrap analysis, unlike the search for the best tree, was performed assuming that all sites evolve with an equal rate, and this simplified condition favors the affinity of diplonemids with euglenoids.

In order to additionally address possible effects of unequal rates of sequence evolution in different lineages (Felsenstein 1978) on the support of the diplonemid-kinetoplastid clade, we also performed a bootstrap analysis after omitting the long branch of *P. cantuscygni* and the short trypanosomatid branches. Although the support level increased to 47% (not shown), it remained statistically irrelevant.

With three monophyletic euglenozoan clades – diplonemids (D), kinetoplastids (K) and euglenoids (E) – and the clade of an outgroup, there can be only three alternative trees. Each of these alternatives was evaluated using the likelihood, parsimony and distance algorithms by comparing trees without topological constraints with trees constructed under the corresponding user-defined constraints. The results are presented in Table 1. The best unconstrained maximum likelihood tree shows diplonemids and kinetoplastids as sister-groups – ((K,D),E) (see also Fig. 1), while the shortest unconstrained parsimony and distance trees contain diplonemids as the closest relatives of euglenoids – (K,(D,E)). However, the trees with enforced alternative topologies – (K,(D,E)) for likelihood and ((K,D),E) for parsimony and distance – are not significantly different from the corresponding unconstrained trees. With each of these methods, the topology which shows diplonemids as an earliest branch of Euglenozoa – (D,(K,E)) – was the least favorable.

It should also be noticed that using the SSU rRNA sequences of a diplomonad (*Giardia lamblia*) and a microsporidian (*Vairimorpha necatrix*) as outgroups substantially increases a bootstrap support for the diplonemid-kinetoplastid clade (up to 90% with some alignments, data not shown). However, such a high level of support may be an artefact associated with a biased nucleotide composition and fast substitution rates observed in the outgroup sequences.

Cloning and Analysis of the COI mRNA Sequence

In order to obtain additional information for resolving the phylogenetic position of *Diplonema*, we analyzed the sequence of the mitochondrial COI protein. The usefulness of this marker is validated by previous work which showed that mitochondria are monophyletic and that nuclear and mitochondrial phylogenies are congruent (Inagaki et al. 1997; Sicheritz-Pontén et al. 1998; Tessier et al. 1997; Viale and Arakaki 1994; Yasuhira and Simpson 1997).

Total cell RNA was isolated from *D. papillatum*. cDNA to polyadenylated mRNA was synthesized by reverse transcription using $dT_{20}N$ primer, and a partial COI mRNA was amplified by PCR using the oligonucleotides C112 and $dT_{20}N$. The conserved sequence-specific oligonucleotide C112 anneals to a site within the central region of the kinetoplastid COI mRNA approximately 700 nt from the 5' end or 800-900 nt from the 3' end. The PCR yielded several

Table 1. Parameters of the phylogenetic trees of Euglenozoa with or without specific topological constraints.

Tree topology	Marker	Likelihood[1]	Parsimony[2]	Distance[3]
((K, D), E)	SSU	<u>–9936.69451</u>	1912	1.48893 (2 trees)
	COI	<u>–3572.64</u>	536	
(K, (D, E))	SSU	–9936.72057	<u>1900</u>	<u>1.48734</u> (2 trees)
	COI	–3575.73	<u>534</u>	
((K, E), D)	SSU	–9938.29060	1914	1.52350 (2 trees)
	COI	–3574.26	537	

[1] Ln likelihood value found by a heuristic search using empirical nucleotide frequencies, assuming two substitution types, and estimating a proportion of invariable sites and a transition/transversion ratio via maximum likelihood.
[2] Performed by a branch-and-bound search. The score indicates a number of steps.
[3] Minimal evolution score found by a heuristic search with Kimura 3-parameter distances and allowing for among site variation; starting tree obtained through neighbor-joining followed by global rearrangements. The two trees in each case differ only by arrangement of *L. tarentolae*, *H. muscarum* and *P. serpens* branches. The similar differences in tree scores were also obtained for sums of unweighted least squares.
For each optimality criterion, the underlined tree corresponds to the constrained tree which coincides with the best unconstrained tree.

DNA fragments (data not shown), including among them a prominent 900 bp fragment, which was then gel-isolated, cloned and sequenced. The sequence contained an ORF for 287 amino acids beginning at the very 5' end and ending at a UAG codon one nucleotide upstream of the poly-A tail (Fig. 2). A database homology search with the BLAST program showed that this sequence was homologous to the 3' regions of known COI gene sequences. Several attempts were made to obtain the missing 5' end of the COI mRNA by 5' RACE PCR, but they were unsuccessful except for the amplification of a noncontiguous region which encodes additional 53 amino acids (data not shown).

In order to ascertain the authenticity of the amplified sequence, we isolated mtDNA and hybridized it to the amplified COI sequence. Total cell DNA of *D. papillatum* was fractionated by equilibrium density centrifugation in CsCl-Hoechst 33258 and CsCl-EtBr gradients, which separate DNA according to its buoyant density or topological state, respectively. Two bands were seen in each case (Fig. 3). DNA from each band was recovered, digested with restriction enzymes and analyzed in agarose gels. A set of discrete fragments apparently produced by a low complexity component was seen in the covalently closed material from the lower band of a CsCl-EtBr gradient (Fig. 3A). In the topologically relaxed DNA from the upper band, a smear produced by a high complexity component was additionally seen. We conclude that the lower band contains a covalently closed mitochondrial DNA, while the upper band contains a nicked or linearized fraction of mtDNA and the nuclear DNA. The bands in the CsCl-Hoechst gradient were too close to avoid mixing of the viscous material from the upper band with DNA from the lower band (Fig. 3B). However, as seen in the restriction digests, DNA recovered from the lower band was enriched with the putative mitochondrial component, while DNA from the upper band also contained a smear of the nuclear DNA. We thus concluded that, unlike in the majority of eukaryotes, the mtDNA in *D. papillatum* is more G+C rich than the nuclear DNA.

```
...GGUCUACGUGAUCAUACUCCCGGUGUUCGGGCUGGUAUCCCAUUCCCUACAUCGAGGAGGACUUUUUUCGCUCUACAACAUGCUGGGCAUGGUGUACGCCAUGAUAGCCAUAGCCGUGGUC 121
...V Y V I I L P V F G L V S H S L H R G G L F S L Y N M L G M V Y A M I A I A V V

GGGUACUUCGUAUGAGCACACCACAUGUUCACUGUGGGUCUGGAUGUAGACACCAGGGUGUACUUCUCCAGUGCCACGUUGCUCAUAGCGCUACCUACAUCCAUCAAGGUGUUCUCCUGG 241
G Y F V W A H H M F T V G L D V D T R V Y F S S A T L L I A L P T S I K V F S W

AUGGUGGGACUGCGUAGGAUGGCACUAGCCAGCAGCGCUGCCUGGUACGUCGUGGCCUUCCUCCUCAUGUUCCUCCUAGGUGGUGUCACAGGGCUGGUGCUAGCCAACAGUGAGGUGGAC 361
M V G L R R M A L A S S A A W Y V V A F L L M F L L G G V T G L V L A N S E V D

CUCGUAAUGCAUGACAGCUAUUACGUCGUGGCCCACUUCCACUAUGUGCUCUCAUUAGGAGCGGUCUUCGGUCUGCUCAACGGUGUCCUCUCAUGCCACGAGCUGUGCUCUGGGUAUAGG 481
L V M H D S Y Y V V A H F H Y V L S L G A V F G L L N G V L S C H E L C S G Y R

AGCGCUGCAUGGUUACUCCGUGUACAGGUGGUCCUCUUGGUCUGAGGAACCACGUGCAUCUUCUGAGGCAUGCACCUGAGCGGGACCCUGGGACUACCACGUAGAGUACCAGAUGCACCU 601
S A A W L L R V Q V V L L V W G T T C I F W G M H L S G T L G L P R R V P D A P

GAUGGAUACCUAGGUACAGUGGUAUCCACCACCUGUGGUAUACUGGUAGUCCUACUAGCAGUAGCGCUGUUGCUGUGUGCAUCCUUGGAGGCAUCCUUGUGGGAUACACAGCAGCUGCGU 721
D G Y L G T V V S T T C G I L V V L L A V A L L L C A S L E A S L W D T Q Q L R

GCCACCACUAGGAGCACCAACCCUGGGAUGCACAACCACAUGCUCAGUGGCAUGCACAGCCUUGACCACGCUACUCGGGGUCAGCUGAUGCUACAUACAUGUGUGUCUACGUCCCACAGA 841
A T T R S T N P G M H N H M L S G M H S L D H A T R G Q L M L H T C V S T S H R

GCAUCCAUGCAUCUGGAGGAGUAGU-poly(A)
A S M H L E E Ter
```

Figure 2. Partial mRNA and deduced protein sequence of the COI from *D. papillatum*. All tryptophan codons are boxed, UGA codons are additionally shown in bold face and the UGA codon conserved among kinetoplastids is underlined.

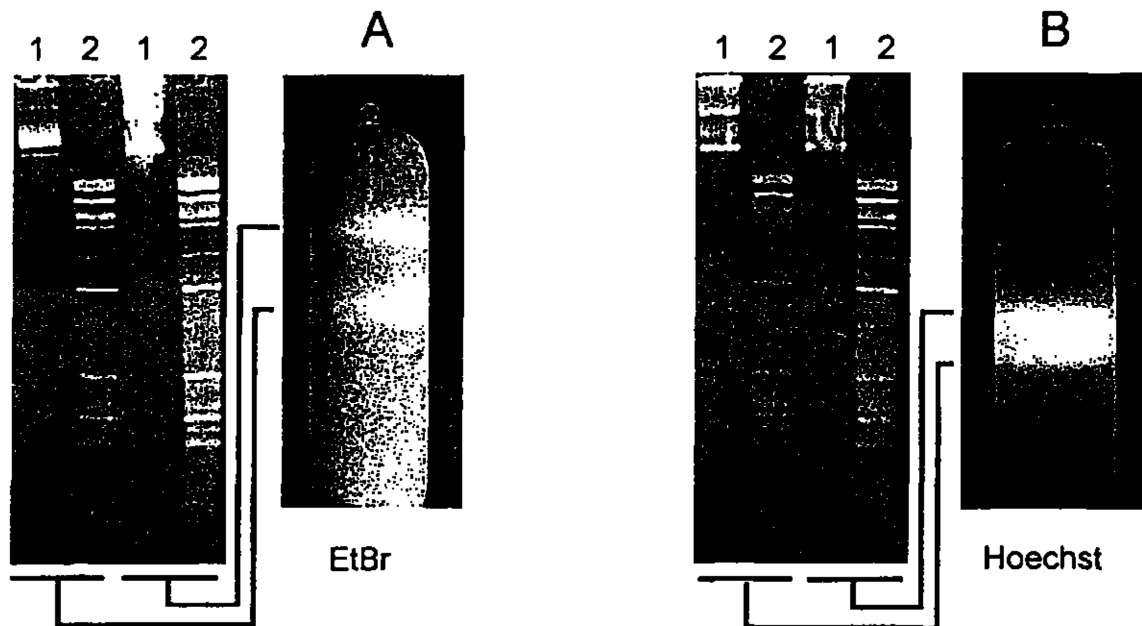

Figure 3. Fractionation of the total cell DNA from *D. papillatum* by equilibrium density centrifugation in CsCl gradients with ethidium bromide (A) and Hoechst 33258 (B). The gradient-purified DNA from the upper and lower band of each gradient, either undigested (lane 1) or digested with MspI (lane 2) was analyzed in a 1% agarose gel shown to the left of the corresponding gradient.

Results of the restriction digests shown in Figure 4A and 4C indicate that each digest contains a discrete set of nonstoichiometric fragments with the sizes not exceeding several kilobases. Some hexanucleotide recognizing enzymes (BglII, HindIII, EcoRI and others) apparently do not digest this DNA, while most tetranucleotide cutters do. There is no evidence for the presence of uniform length minicircle-like molecules as those seen in digests of kinetoplast DNA from trypanosomatids. Therefore, this mtDNA is probably composed of several types of circles. The exact size, number of types and topological arrangement of these molecules remain to be investigated.

Probing of the mtDNA digests with the cloned COI cDNA revealed hybridization with one or a few restriction fragments or hybridization with undigested molecules, depending on the enzyme used (Fig. 4B). Specific restriction fragments producing hybridization signals could be identified in most cases. Cloning and sequence analysis of the mitochondrial genomic COI sequence are in progress.

Figure 2 shows the cloned 3' end segment of the COI mRNA along with the deduced protein. The polypeptide sequence is 62% similar (41% identical) to the corresponding portion of the *E. gracilis* COI sequence and 65% similar (37% identical) to the *T. borreli* sequence. There is no similarity at the C-terminus. The amplified COI sequence is likely to be derived from a polyadenylated mRNA and there is only a single nucleotide between the UAG termination codon and the poly(A) tail. This arrangement was identical in all three cDNA clones sequenced. However, there is also a possibility that the 3' UTR was indeed longer and contained an A-rich sequence which acted as an annealing site for the $dT_{20}N$ primer. Future studies should clarify this question.

An important feature of the COI sequence is the use of UGA codons as tryptophan instead of termination. The sequence contains seven deduced tryptophan residues, four of which are encoded by standard UGG codons, while three tryptophans, including one conserved residue, are encoded by noncanonical UGA codons.

The codon usage is summarized in Table 2. The usage is biased towards codons with G or C in the 3rd position (65.3% of all codons). This reflects the 55.3% G+C content of this fragment and correlates with a relatively high G+C-richness of this mtDNA as mentioned above.

Phylogenetic Analysis of COI Sequences

The amplified 3' partial sequence of *D. papillatum* COI and additional 53 residues from the 5' region of

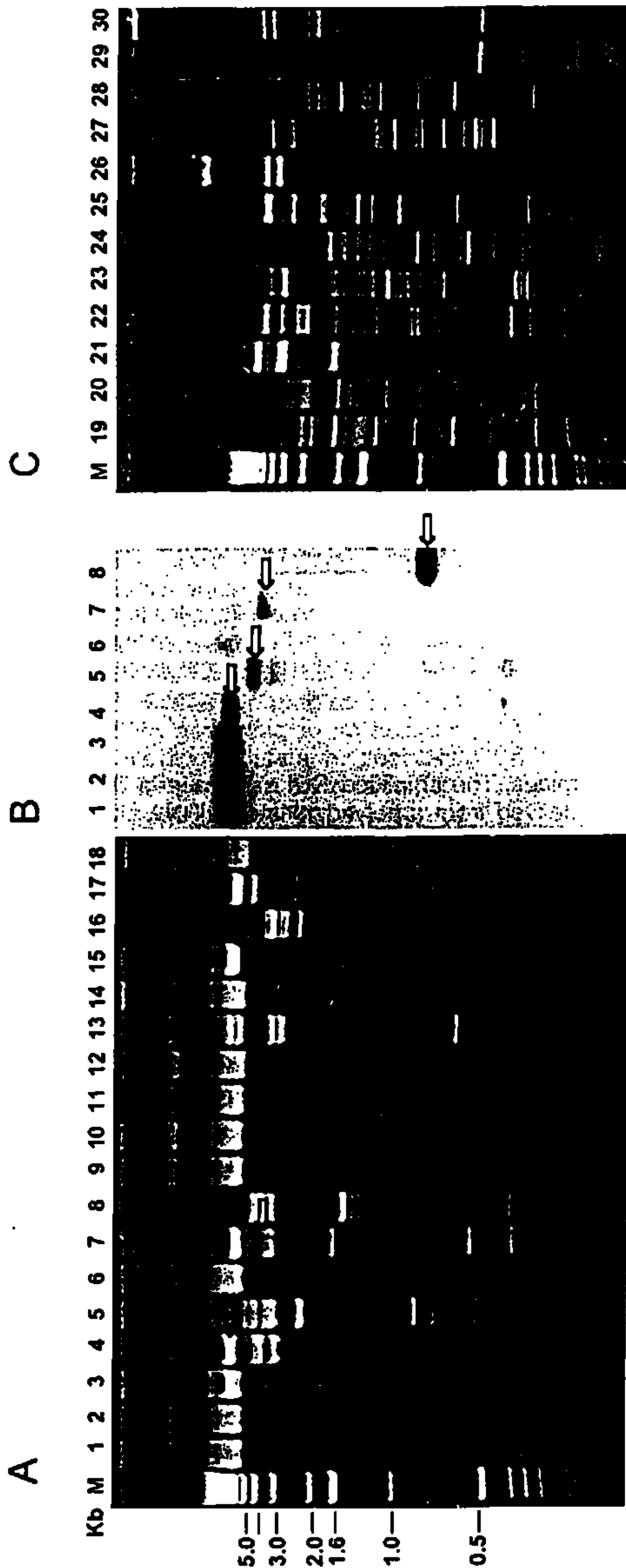

Figure 4. Restriction digest (A and C) and Southern hybridization (B) analyses of the mitochondrial DNA from *D. papillatum*. DNA was digested with BglII (lane 1), HindIII (2), XbaI (3), PstI (4), ScaI (5), EcoRI (6), BamHI (7), SphI (8), ClaI (9), DraI (10), HincII (11), SalGI (12), XhoI (13), NotI (14), NdeI (15), PvuII (16), SmaI (17), MspI (19), Sau3AI (20), HinfI (21), HaeIII (22), MvaI (23), AluI (24), AvaI (25), BstXI (26), HhaI (27), HpaII (28), RsaI (29), TaqI (30) or undigested (18) and fractionated in a 1% agarose gel calibrated with the 1 kb ladder (lane M, Life Technologies). The gel shown in panel A was blotted and hybridized with the labeled COI cDNA probe (panel B). The arrows indicate positions of the hybridizing bands.

the sequence were aligned with the corresponding sequences from *E. gracilis*, *T. borreli*, *T. brucei* and *L. tarentolae*. The yeast and wheat COI sequences were used as outgroups. After elimination of the nonalignable C-terminus and a few short internal regions, the alignment contained 293 characters. As for the SSU genes, analysis of the data with maximum likelihood (PROTML) resulted in a tree which showed *Diplonema* as a sister group to kinetoplastids (Fig. 1). However, two alternative topologies were not significantly different (Table 1). Protein parsimony analysis performed with PHYLIP resulted in two equally short trees, one of them containing *Diplonema* as a sister lineage to kinetoplastids and another one showing it as a sister lineage to *E. gracilis* (data not shown). Protein parsimony analysis performed with PAUP slightly favored the second of these two options (Table 1). Protein distance analysis done with PHYLIP using the Kimura distance formula or the Dayhoff PAM matrix for distance computation combined with two different methods of tree construction (least squares and neighbor-joining), showed *Diplonema* as the closest relative of kinetoplastids (not shown).

Discussion

We have analyzed the phylogenetic relationships of *Diplonema* with other Euglenozoa using SSU rRNA gene and mitochondrial COI protein sequences. Confirming results of the previous morphological studies (Triemer 1992; Triemer and Farmer 1991), our analysis showed that diplonemids rep-

Table 2. Codon frequency in the partial COI mRNA of *D. papillatum*.

Gly	GGG	6.00	Arg	AGG	4.00	Trp	UGG	4.00	Arg	CGG	1.00
Gly	GGA	7.00	Arg	AGA	2.00	Trp	UGA	3.00	Arg	CGA	1.00
Gly	GGU	8.00	Ser	AGU	3.00	Cys	UGU	3.00	Arg	CGU	4.00
Gly	GGC	3.00	Ser	AGC	7.00	Cys	UGC	3.00	Arg	CGC	0.00
Glu	GAG	5.00	Lys	AAG	1.00	End	UAG	1.00	Gln	CAG	4.00
Glu	GAA	0.00	Lys	AAA	0.00	End	UAA	0.00	Gln	CAA	0.00
Asp	GAU	4.00	Asn	AAU	0.00	Tyr	UAU	3.00	His	CAU	5.00
Asp	GAC	4.00	Asn	AAC	5.00	Tyr	UAC	8.00	His	CAC	11.00
Val	GUG	16.00	Met	AUG	14.00	Leu	UUG	5.00	Leu	CUG	15.00
Val	GUA	9.00	Ile	AUA	5.00	Leu	UUA	2.00	Leu	CUA	10.00
Val	GUU	0.00	Ile	AUU	0.00	Phe	UUU	1.00	Leu	CUU	2.00
Val	GUC	0.00	Ile	AUC	3.00	Phe	UUC	10.00	Leu	CUC	13.00
Ala	GCG	3.00	Thr	ACG	3.00	Ser	UCG	1.00	Pro	CCG	1.00
Ala	GCA	8.00	Thr	ACA	5.00	Ser	UCA	2.00	Pro	CCA	2.00
Ala	GCU	3.00	Thr	ACU	3.00	Ser	UCU	2.00	Pro	CCU	3.00
Ala	GCC	10.00	Thr	ACC	7.00	Ser	UCC	10.00	Pro	CCC	0.00

resent a third lineage within Euglenozoa in addition to kinetoplastids and euglenoids, and this lineage branches deeply within this group. However, molecular phylogeny, as well as the previous morphological studies, fail to provide an unequivocal answer to the question which of these two clades represents a closest relative of diplonemids. Maximum likelihood, a proven powerful and accurate method of phylogenetic reconstruction, favors a closer relationship of diplonemids with kinetoplastids rather than euglenoids with both markers, while parsimony and distance supported the alternative topology. However, differences between the best tree and suboptimal trees are small with each method.

One interesting feature of the mitochondrial genetic code of *D. papillatum*, the use of UGA and UGG as tryptophan codons in COI mRNA is also shared by kinetoplastids (de la Cruz et al. 1984; Simpson et al. 1987). In COI from *E. gracilis* and the euglenoid, *Eutreptiella gymnastica*, only the UGG codon is used for this purpose (Inagaki et al. 1997; Tessier et al. 1997; Yasuhira and Simpson 1997). In addition, COI mRNA in *E. gracilis* does not contain a poly-A tail (Yasuhira and Simpson 1997), while the *D. papillatum* COI mRNA apparently is polyadenylated, as are mitochondrial messengers in kinetoplastids (Bhat et al. 1992). Although these observations support a closer affinity of diplonemids with kinetoplastids, they are based only on a single mitochondrial gene and need to be verified by analysis of additional genes.

Our preliminary analysis of the mitochondrial DNA from *D. papillatum* indicates that it contains covalently closed molecules. No evidence for a kinetoplast-like network could be seen, although existence of some catenation needs to be investigated. It is unclear if topologically relaxed molecules also present in our preparations are due to the damage caused by handling or whether they represent replication intermediates. In any case, isolated mtDNA of *D. papillatum* is quite dissimilar from the heterogeneous linear molecules isolated from mitochondria of *E. gracilis* (Yasuhira and Simpson 1997), and also from trypanosomatid networks (Simpson 1986; Simpson 1987). Although only limited information is available on the molecular organization of mtDNA of bodonids (Blom et al. 1998; Hajduk et al. 1986; Lukeš et al. 1998; Yasuhira and Simpson 1996), it seems the non-catenated or weakly catenated structure is more similar to that seen in *Diplonema*.

Finally, although the data obtained in this work favor a closer association of diplonemids with kinetoplastids than with euglenoids, it is evident that more work is required in order to fully resolve phylogenetic relationships among these organisms.

Methods

Strains and cultivation conditions: *Diplonema* (syn. *Isonema*) *papillatum* (ATCC50162) and *Diplonema* sp. 2, strain IIIGPC, (ATCC50224) were obtained from the American Type Culture Collection. The strains were cultivated at 27 °C in an enriched *Isonema* medium (ATCC Culture Medium 1728) containing 10% heat-inactivated fetal bovine serum in stationary flasks. For cultures of *D. papillatum*, stationary phase of growth was achieved at 10^7 cells/ml on the fourth day after inoculation with a starting cell density of $0.5–1.0 \times 10^6$ cells/ml. For *Diplonema* sp., only a ten fold lower density of cells could be obtained.

Isolation of total cell DNA and RNA: Cells from stationary cultures were washed with 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 100 mM EDTA, and lysed with 2% Sarcosyl, 0.5 mg/ml pronase at 65 °C for 30 min. The lysate was extracted with buffered (pH 8.0) phenol, phenol-chloroform (1:1) and chloroform. Nucleic acids were precipitated with an equal volume of isopropanol. The pellet was rinsed with ethanol, dissolved and reprecipitated with ethanol. Total cell RNA was isolated by guanidinium thiocyanate/phenol-chloroform extraction procedure using the RNA Isolation Kit (Stratagene).

Isolation of mitochondrial DNA: Total cell DNA isolated from 0.5×10^{10} cells of *D. papillatum* was fractionated using CsCl-Hoechst 33258 or CsCl-ethidium bromide equilibrium density centrifugation in a VTi50 rotor at 45,000 rpm for 20 h. Other conditions were as described previously (Maslov and Simpson 1994; Simpson 1979).

PCR amplification and sequencing of the SSU genes: Small subunit rRNA genes were amplified using the oligonucleotides S762 (GACTTTTGCTTCCTCTAWTG) and S763 (CATATGCTTGTTTCAAGGAC), cloned in the vector pT7Blue (Novagen), and sequenced using the Sequenase kit (version 2.0, Amersham) as described previously (Maslov et al. 1996). The first strand of the SSU gene of *D. papillatum* was sequenced with the following oligonucleotide primers (listed in the order of occurrence):

S847: CATATGCTTGTTTCAAGGACTWAGCCATGCATGCC;
S1404: CTGAGAACGGCTACCACATC;
S713: CCGCGGTAATTCCAGCTCC;
S1381: ACGGTCGACCACCGATGTTA;
S757: TCAGGGGGGAGTACGTTCGC;
S1402: TTGTAGGGGGTGTCTTTTGG;
S828: CAACAGCAGGTCTGTGATGC;

The second strand was sequenced with the primers:

S1401: AGCAACGACGGGCGGTGTGT;
S829: GCATCACAGACCTGCTGTTG;
S1378: CACACAATTCATCGAGAAAG;
S714: CGTCAATTTCTTTAAGTTTC;
S1382: GAAACTCAAAAGAGAACCGC;
S1403: CACCATTACCACCGTTCATA;
S1380: TCAGCAGTGTTGCTATTGGG;

For the first strand of the *Diplonema* sp. SSU gene we used:

S847 (see above); S1645: CCCGCAAGAGTATCTGCCCTATC;
S1736: GGAATTAGGGTTCGATTCCG; S713;
S1644: TTACTGTGAACAAACCAGCGTGT; S757 and S828.

For the second strand of the same gene we used:

S829; S714; S1643: GGTTTGGAGCCTTACCTTAAATTAT;
S1739: CGGGTTTTGATCTTCAACAG;
S755: CTACGAACCCTTTAACAGCA;
S1646: GATGTGGTAGCCGTTTCTCAGGCT.

The SSU rRNA sequences are deposited in the GenBank™ database under the following accession numbers: AF119811 (*D. papillatum*) and AF119812 (*Diplonema* sp.).

Other nucleic acid manipulations: The conditions for RT-PCR and TA-cloning were as described previously (Simpson et al. 1996). Briefly, the cDNA from *D. papillatum* was synthesized using total cell RNA and an oligo-dT primer ($dT_{20}N$). The oligonucleotide C112 (TTYTGRTTYTTYGGNCAYCCNGA) (Lukeš et al. 1994), which corresponds to a highly conserved region of eukaryotic COI sequences, was used together with the oligonucleotide $dT_{20}N$ to PCR amplify a partial COI mRNA. The determined partial COI mRNA sequence is deposited in GenBank™ database under the accession number AF119813. Agarose gel electrophoresis, restriction digestion and Southern blotting were performed by standard protocols. Hybridization with a DNA probe labeled by random priming (Stratagene) was performed at 68 °C in 6 × SSC, 5 × Denhardt's solution, 100 mg/ml salmon sperm DNA and 0.5% SDS.

Phylogenetic analysis: Alignments were generated manually using the program SeqEdit, version 3.1 (Olsen 1990). The SSU and COI alignments are available on request or can be downloaded from the following URL: http://www.lifesci.ucla.edu/RNA/trypanosome/alignments.html. Maximum likelihood, parsimony and distance analyses were performed as described previously (Lukeš et al. 1997) using PAUP* 4.0 beta version (Swofford 1998) (Sinauer Associates, Inc.) and PHYLIP 3.5 (Felsenstein 1995). In addition, maximum likelihood analysis of the COI polypeptide sequences was performed with PROTML, version 2.2, (Adachi et al. 1992) with the JTT amino acid substitution matrix (Jones et al. 1992).

Acknowledgements

We would like to thank Dr. J. Lukeš for the oligonucleotide C112. This work was supported in part by the Burroughs Wellcome New Investigator Award in Molecular Parasitology to D.M. and the grant AI09102 from National Institutes of Health to L.S.

References

Adachi J, Hasegawa M (1992) MOLPHY: Programs for molecular phylogenetics I – PROTML: maximum likeli-

hood inference of protein phylogeny. Computer science monographs, No. 27. Institute of Statistical Mathematics, Tokyo

Bhat GJ, Souza AE, Feagin JE, Stuart K (1992) Transcript-specific developmental regulation of polyadenylation in *Trypanosoma brucei* mitochondria. Mol Biochem Parasitol **52**: 231–240

Blom D, De Haan A, Van den Berg M, Sloof P, Jirků M, Lukeš J, Benne R (1998) RNA editing in the free-living bodonid *Bodo saltans*. Nucl Acids Res **26**: 1205–1213

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

Corliss JO (1994) An interim utilitarian ("user-friendly") hierarchical classification and characterization of the protists. Acta Protozool **33**: 1–51

de la Cruz V, Neckelmann N, Simpson L (1984) Sequences of six structural genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. J Biol Chem **259**: 15136–15147

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool **27**: 401–410

Felsenstein J (1995) PHYLIP – Phylogeny Inference Package, version 3.5. University of Washington, Seattle. Distributed by the Author

Hajduk S, Siqueira A, Vickerman K (1986) Kinetoplast DNA of *Bodo caudatus*: a noncatenated structure. Mol Cell Biol **6**: 4372–4378

Hollar L, Lukeš J, Maslov DA (1998) Monophyly of endosymbiont containing trypanosomatids: Phylogeny versus taxonomy. J Euk Microbiol **45**: 293–297

Inagaki Y, Hayashi-Ishimaru Y, Ehara R, Igarashi I, Ohama T (1997) Algae or protozoa: Phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences. J Mol Evol **45**: 295–300

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from **protein sequences.** Comput Appl Biosci **8**: 275–282

Kivic PA, Walne PL (1984) An evaluation of a possible phylogenetic relationship between the euglenophyta and kinetoplastida. Origins of Life **13**: 269–288

Lukeš J, Jirků M, Avliyakulov N, Benada O (1998) Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. EMBO J **17**: 838–846

Lukeš J, Jirků M, Dolezel D, Kral'ová I, Hollar L, Maslov DA (1997) Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol **44**: 521–527

Lukeš J, Arts GJ, Van den Burg J, De Haan A, Opperdoes F, Sloof P, Benne R (1994) Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. EMBO J **13**: 5086–5098

Maslov DA, Simpson L (1994) RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan, *Trypanoplasma borreli*. Mol Cell Biol **14**: 8174–8182

Maslov DA, Lukeš J, Jirků M, Simpson L (1996) Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. Mol Biochem Parasitol **75**: 197–205

Olsen GJ (1990) Sequence editor and analysis program. University of Illinois, Urbana, computer program distributed by the author

Sicheritz-Pontén T, Kurland CG, Andersson SGE (1998) A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. Biochim Biophys Acta: Bio-Energetics **1365**: 545–551

Simpson AGB (1997) The identity and composition of the Euglenozoa. Arch Protistenkd **148**: 318–328

Simpson L (1979) Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. Proc Natl Acad Sci USA **76**: 1585–1588

Simpson L (1986) Kinetoplast DNA in trypanosomid flagellates. Int Rev Cytol **99**: 119–179

Simpson L (1987) The mitochondrial genome of kinetoplastid protozoa: Genomic organization, transcription, replication and evolution. Annu Rev Microbiol **41**: 363–382

Simpson L, Frech GC, Maslov DA (1996) RNA editing in trypanosomatid mitochondria. Methods Enzymol **264**: 99–121

Simpson L, Neckelmann N, de la Cruz V, Simpson A, Feagin J, Jasmer D, Stuart K (1987) Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. J Biol Chem **262**: 6182–6196

Swofford DL (1998) PAUP* 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods), beta version. Sinauer Associates, Sunderland, Massachusetts

Tessier LH, Van der Speck H, Gualberto JM, Grienenberger JM (1997) The *cox1* gene from *Euglena gracilis*: A protist mitochondrial gene without introns and genetic code modifications. Current Genet **31**: 208–213

Triemer RE (1992) Ultrastructure of mitosis in *Diplonema ambulator* Larsen and Patterson (Euglenozoa). Europ J Protistol **28**: 398–404

Triemer RE, Farmer MA (1991) An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. Protoplasma **164**: 91–104

Triemer RE, Ott DW (1990) Ultrastructure of *Diplonema ambulator* Larsen & Petersen (Euglenozoa) and its relationship to *Isonema*. Europ J Protistol **25**: 316–320

Van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, Kieft R, Borst P (1998) β-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. Proc Natl Acad Sci USA **95**: 2366–2371

Viale AM, Arakaki AK (1994) The chaperone connection to the origins of the eukaryotic organelles. FEBS Lett **341**: 146–151

Willey RL, Walne PL, Kivic P (1988) Phagotrophy and the origins of the euglenoid flagellates CRC Crit Rev Plant Sci **7**: 303–340

Yasuhira S, Simpson L (1996) Guide RNAs and guide RNA genes in the cryptobiid kinetoplastid protozoan, *Trypanoplasma borreli*. RNA **2**: 1153–1160

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