PHYLOGENY OF DEEP-LEVEL RELATIONSHIPS WITHIN EUGLENOZOA BASED ON COMBINED SMALL SUBUNIT AND LARGE SUBUNIT RIBOSOMAL DNA SEQUENCES

by

BING MA

(Under the Direction of Mark A. Farmer)

ABSTRACT

My master's research addressed questions about the evolutionary histories of Euglenozoa, with special attention given to deep-level relationships among taxa. The Euglenozoa are a putative early-branching assemblage of flagellated Eukaryotes, comprised primarily by three subgroups: euglenids, kinetoplastids and diplonemids. The goals of my research were to 1) evaluate the phylogenetic potential of large subunit ribosomal DNA (LSU rDNA) gene sequences as a molecular marker; 2) construct a phylogeny for the Euglenozoa to address their deep-level relationships; 3) provide morphological data of the flagellate *Petalomonas cantuscygni* to infer its evolutionary position in Euglenozoa.

A dataset based on LSU rDNA sequences combined with SSU rDNA from thirty-nine taxa representing every subgroup of Euglenozoa and outgroup species was used for testing relationships within the Euglenozoa. Our results indicate that a) LSU rDNA is a useful marker to infer phylogeny, b) euglenids and diplonemdis are more closely related to one another than either is to the kinetoplatids and c) *Petalomonas cantuscygni* is closely related to the diplonemids.

INDEX WORDS: Euglenozoa, phylogenetic inference, molecular evolution, 28S rDNA, LSU rDNA, 18S rDNA, SSU rDNA, euglenids, diplonemids, kinetoplastids, conserved core regions, expansion segments, TOL

PHYLOGENY OF DEEP-LEVEL RELATIONSHIPS WITHIN EUGLENOZOA BASED ON COMBINED SMALL SUBUNIT AND LARGE SUBUNIT RIBOSOMAL DNA SEQUENCES

by

BING MA

B. Med., Zhengzhou University, P. R. China, 2002

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2005

PHYLOGENY OF DEEP-LEVEL RELATIONSHIPS WITHIN EUGLENOZOA BASED ON COMBINED SMALL SUBUNIT AND LARGE SUBUNIT RIBOSOMAL DNA SEQUENCES

by

BING MA

Major Professor: Mark A. Farmer

Committee: Robert K. Kuzoff
Kojo Mensa-Wilmot

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2005

DEDICATION

To Qingfen Kong and Xiguo Ma

ACKNOWLEDGEMENTS

First, I deeply appreciate that Mark Farmer leads me into the intriguing field of evolutionary biology that has great impact to my future career and whole life. He is so kind and considerate to offer me this opportunity to pursue my master's degree, help me to figure out my future areas of interests in research, and set up my work and handle the encountered difficulties. Moreover, I benefited a great deal from financial support obtained from the National Foundation PEET (Partnerships for Enhancing Expertise in Taxonomy, grant no. DEB 4-21348) by Mark and from the Department of Cellular Biology in University of Georgia.

I own a huge debt of gratitude to Bob Kuzoff. I never imagine that a chance know him in the class of molecular systematics would result in the acquisition of a fabulous mentor, professional collaborator, and a fantastic friend. I never imagine that a student from a foreign country without any background in molecular evolution could turn out to have a professional publication in this field with in-depth understanding of the related techniques and analyses. He was uncommonly generous with his time and genuinely concerned for both my welfare as an honest person with good characters, a graduate student with good skills in all aspects, and a professionals with a good scope of knowledge. Very few students are fortunate enough like me to have such a fantastic mentor so accessible and so willing to help. I really want to say that my life become better because of him, for now and in future.

There have been many people who have helped me out along the path of my graduate career, beginning with another committee member Kojo Mensa-Wilmot, who is also a great instructor to give me generous help. I also thank for other faculties and staffs in Department of

Cellular Biology and Plant Biology. I was so lucky to have so many great teachers and friends here to help me. I also own thanks to John Shields and other EM center professionals and staffs. I thank Wenli Li, Sarah Jardeleza and other fellow students for helpful discussions and technical assistance.

Most importantly, this accomplishment would not have been possible without the constant encouragement and emotional support from my dear mom and dad, Qingfen Kong and Xiguo Ma. Also Wenzhi Chen, who give so much support to me, both mentally and physically. Words simply cannot describe how indebted I am to them. I am trying to be their proud by presenting my best always. I love you all always.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

THE EUGLENOZOA

The Euglenozoa are an assemblage of flagellated Eukaryotes that are purported to have diverged early during the evolutionary history of eukaryotes (Sogin, Elwood et al. 1986; Van de Peer, Neefs et al. 1993). Collectively, the Euglenozoa number around 1,600 species including both free-living and parasitic organisms with some of the latter being important animal pathogens. They were originally proposed as a monophyletic group based on the presence of discoidal mitochondrial cristae, non-tubular flagellar hairs, paraxial rods, and a closed pattern of mitosis with an intranuclear spindle (Cavalier-Smith 1981).

Simpson (1997) extended and modified the criteria for circumscription of the Euglenozoa. He placed an emphasis on three apparent synapormorphies for the group. First is a flagellar root pattern in Euglenozoa: the three microtubular roots of the flagellar apparatus are asymmetrically arranged with two basal bodies, one root lines a portion of the feeding apparatus. The second synapomorphy is the presence of a paraxial rod in one or both of the flagella. This paraxial rod is primarily composed of two polypeptides arranged in either an amorphous or nearly crystalline array that typically spans the entire length of the flagellum but does not extend into the flagellar transition zone. The third synapomorphy is the presence of tubular, thick walled extrusive organelles (e.g. trichocysts, mucocysts) that may have been lost or reduced in many derived Euglenozoa.

THREE MAIN SUBGROUPS OF EUGLENOZOA

Based on morphological features, the three subgroups of Euglenozoa proposed are the euglenids (Butschli 1884), the kinetoplastids (Honigberg 1963) and the diplonemids (Cavalier-Smith 1993). Support for the three main subgroups of Euglenozoa was later recovered through analysis of small subunit ribosomal DNA (SSU rDNA) sequences (Maslov, Yasuhira et al. 1999; Preisfeld, Berger et al. 2000; Moreira, Lopez-Garcia et al. 2001; Preisfeld, Busse et al. 2001; Busse and Preisfeld 2003).

Euglenids are distinguished by the presence of a series of pellicle strips that are organized by supporting microtubules underneath the cell membrane. An additional glycoprotein layer is appressed to the inner surface of the plasma membrane generating the ribbon-shaped pellicle strips, which extend the length of the cell, providing structural support for a variety of unique cell shapes. Euglenids are typically larger than 30µm in length and may reach 500µm. They are one of the best-known groups of free-living flagellates and may be green, photosynthetic autotrophs or colorless osmotrophs and heterotrophs. They are commonly found in freshwater, while some, like species of *Eutreptiella*, are marine organisms.

Kinetoplastids are distinguished from other protists by the presence of a *kinetoplast*, a large structure formed by an accumulation of mitochondrial DNA, which is found in close association with flagellar bases. Like euglenids, the cytoskeletal microtubules underlying the plasma membrane form a continuous or discontinuous supporting corset, but there is no additional glycoprotein layer present, an apparent requirement for pellicle strip formation. Kinetoplastids are typically smaller than euglenids with most species being less than 30µm long. Kinetoplastid taxa include a number of parasites responsible for diseases in humans (e.g.

sleeping sickness, Chagas' disease), animals, and plants, while others are free-living organisms (e.g. Bodonids).

The third subgroups, the diplonemids, also have a corset of supporting microtubules under their cortical membranes, but unlike the euglenids these microtubules are not organized into pellicle strips. Diplonemids are also smaller than euglenids having a length of less than 30μm. Most diplonemids are thought to be strictly free-living organisms, but recently von der Heyden and co-workers (von der Heyden, Chao et al. 2004) proposed a group of pathogenic diplonemids that may target a range of hosts. Finally there are a number of Euglenozoan taxa (e.g., *Postgaardi*, *Calkinsia* and *Petalomonas cantuscygni*) that have uncertain phylogenetic affinities and do not clearly group with any of the three defined subgroups.

PETALOMONAS CANTUSCYGNI

Petalomonas cantuscygni is a small, colorless, relatively rigid protist found in marine environments. Many of the morphological features of *P. cantuscygni* are thought to resemble the ancestral form of Euglenozoa. It has a single emergent flagellum, relatively few longitudinally arranged pellicular strips, mitochondrial inclusions and is phagotrophic which is considered to be the state of the ancestral form of the Euglenozoa (Cann and Pennick 1986). It does not have plastids, an eyespot or a paraflagellar body, which are believed to be derived features associated with the photosynthetic euglenids. *Petalomonas cantuscygni* has a Type I feeding apparatus supported by a few microtubules (Triemer and Farmer 1991). There are in total four types of Euglenid feeding apparati: Type I (MTR/pocket), Type II (plicate type), type III (short extensive type) and type IV (siphon type) (Triemer and Farmer, 1991). The type I feeding apparatus consists of a cytoplasmic pocket that arises adjacent to the flagellar opening and extends towards

the posterior of the cell. It is believed to be the simplest and most ancestral feeding type among the four (Triemer 1986; Triemer and Ott 1990; Triemer and Farmer 1991). Besides the Type I feeding apparatus, it also has two basal bodies with three asymmetrically distributed microtubular rootlets, and a mitotic spindle that forms within a closed nuclear envelope (Triemer and Farmer, 1991).

Previous studies on *Petalomonas cantuscygni* have focused on its anatomical structures and used SSU rDNA to refer its phylogenetic position. However, the morphological data conflict with the molecular phylogenetic analyses, which make the position of *P. cantuscygni* even more complicated. Homologous structures in flagellar apparatus, cytoskeleton, feeding apparatus and mitotic apparatus show that the euglenids, even the photosynthetic ones, and kinetoplastids may share a close common ancestor and are more closely related to each other than either is to other groups (Triemer and Farmer, 1991). Interestingly, *Petalomonas cantuscygni* has a euglenid-like pellicle composed of longitudinal strips, as well as a number of fibrous mitochondrial inclusions near the flagellar apparatus that resemble a kinetoplast. In many ways *Petalomonas cantuscygni* bridges the gap between the euglenids and kinetoplastids and can be thought of as a "missing-link" between the two groups (Triemer and Farmer 1991; Leander and Farmer 2001). However, recent molecular systematic studies suggest that euglenids and diplonemids are more closely related groups than either is to kinetoplastids (Busse and Preisfeld 2003; von der Heyden, Chao et al. 2004), which contradicts the notion that *Petalomonas cantuscygni* is a "missing-link" between euglenids and kinetoplastids.

GENERAL OUTLINE AND OBJECTIVES

Phylogenetic relationships among the three main subgroups of Euglenozoa and sundry orphan taxa have been explored through several lines of evidence, but no clear picture has emerged. In this study we: (1) explore and take advantage of the merits of LSU rDNA; (2) evaluate LSU rDNA's phylogenetic potential to re-construct a phylogeny for Euglenozoa; and (3) focus on *Petalomonas cantuscygni* to provide additional morphological data to infer its evolutionary position within Euglenozoa.

The objective of this study is to address the deep-level relationship within Euglenozoa. Chapter 2 focused on the question using LSU rDNA combined with SSU rDNA as genetic marker; and chapter 3 is specifically focussed on *Petalomonas cantuscygni* and looking for evidence to address its evolutionary position. Based on our phylogeny and other data, we make the conclusions in chapter 4.

REFERENCES

Busse, I. and A. Preisfeld (2003). "Discovery of a group I intron in the SSU rDNA of Ploeotia costata (Euglenozoa)." Protist **154**(1): 57-69.

Busse, I. and A. Preisfeld (2003). "Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa)." Int J Syst Evol Microbiol **53**(Pt 2): 617-24.

Butschli, O. (1884). "Protozoa in Bronn's Klassen und Ordnung des." Tierreiches. **2**: 617-1097.

Cann, J. P. and N. C. Pennick (1986). "Observations on Petalomonas-Cantuscygni, N-Sp, a New Halo-Tolerant Strain." Archiv Fur Protistenkunde **132**(1-2): 63-71.

Cavalier-Smith, T. (1981). "Eukaryote kingdoms: seven or nine?" Biosystems **14**(3-4): 461-81.

Cavalier-Smith, T. (1993). "Kingdom protozoa and its 18 phyla." Microbiol Rev **57**(4): 953-94.

Honigberg, B. M., Balamuth, W. (1963). "Subphylum Sarcomastigophora nom. Nov. to embrace the flagellate and amoiboid assemblages of protozoans." Journal of Protozoology **10 (suppl.) 27**.

Leander, B. S. and M. A. Farmer (2001). "Evolution of Phacus (Euglenophyceae) as inferred from pellicle morphology and SSU rDNA." Journal of Phycology **37**(1): 143-159.

Maslov, D. A., S. Yasuhira, et al. (1999). "Phylogenetic affinities of Diplonema within the euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences." Protist **150**(1): 33-42.

Moreira, D., P. Lopez-Garcia, et al. (2001). "New insights into the phylogenetic position of diplonemids: G+C content bias, differences of evolutionary rate and a new environmental sequence." International Journal of Systematic and Evolutionary Microbiology **51**: 2211-2219.

Preisfeld, A., S. Berger, et al. (2000). "Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data." Journal of Phycology **36**(1): 220-226.

Preisfeld, A., I. Busse, et al. (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**(Pt 3): 751-8.

Simpson, A. G. B. (1997). "The identity and composition of the Euglenozoa." Archiv Fur Protistenkunde **148**(3): 318-328.

Sogin, M. L., H. J. Elwood, et al. (1986). "Evolutionary Diversity of Eukaryotic Small-Subunit Ribosomal-Rna Genes." Proceedings of the National Academy of Sciences of the United States of America **83**(5): 1383-1387.

Triemer, R. E. (1986). "Light and Electron-Microscopic Description of a Colorless Euglenoid, Serpenomonas-Costata N-G, N-Sp." Journal of Protozoology **33**(3): 412-415.

Triemer, R. E. and M. A. Farmer (1991). "An Ultrastructural Comparison of the Mitotic Apparatus, Feeding Apparatus, Flagellar Apparatus and Cytoskeleton in Euglenoids and Kinetoplastids." Protoplasma **164**(1-3): 91-104.

Triemer, R. E. and D. W. Ott (1990). "Ultrastructure of Diplonema-Ambulator Larsen and Patterson (Euglenozoa) and Its Relationship to Isonema." European Journal of Protistology **25**(4): 316-320.

Van de Peer, Y., J. M. Neefs, et al. (1993). "Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock." J Mol Evol **37**(2): 221-32.

von der Heyden, S., E. E. Chao, et al. (2004). "Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa." Journal of Eukaryotic Microbiology **51**(4): 402-416.

CHAPTER 2

PHYLOGENY OF DEEP-LEVEL RELATIONSHIPS WITHIN EUGLENOZOA BASED ON COMBINED SMALL SUBUNIT AND LARGE SUBUNIT RIBOSOMAL DNA SEQUENCES[1]

[1] Bing Ma, Mark A. Farmer, Robert K. Kuzoff. To be submitted to *Molecular Biology and Evolution*.

ABSTRACT

The deep-level relationships within the Euglenozoa remain poorly resolved and controversial, despite a number of previous studies on the morphological structures and some recent systematic studies based on SSU ribosomal DNA. SSU rDNA is the most widely used molecular marker for phylogenetic reconstruction at higher taxonomic levels, but it has not led to conclusive results in unraveling relationships among the Euglenozoa. Previous studies in plant, animal, fungi and other organisms using full length LSU rDNA have suggested the potential of this gene for phylogeny retrieval at taxonomic levels comparable to those investigated with SSU rDNA. We present a protocol for PCR amplification and sequencing of approximately eighty percent of the full length of LSU rDNA sequences. Comparisons of sequence dissimilarity indicate that full length LSU rDNA sequences of Euglenozoa exhibit a level of base substitution 1.16 to 2.67 times as high as that of entire SSU rDNA sequences, while the combined data set provides 3.08 times as many phylogenetically informative characters as SSU rDNA alone. We have targeted deep-level relationships in the Euglenozoa using LSU rDNA combined with SSU rDNA phylogenetic markers. We examine sequences of thirty-nine taxa from the major subgroups of Euglenozoa as Euglenids, Diplonemids, and Kinetoplatids, and from the Euglenozoa's closest relative sister taxon, the Percolozoa. All of our analysis, including NJ, MP, ML and Bayesian approaches, strongly supports a close relationship between Diplonemids and Euglenids to the exclusion of Kinetoplastids.

INTRODUCTION

An effort among contemporary biologists to construct a global phylogeny for all lineages of living organisms has been underway for the past 20 years. The rooting of the tree of

life (TOL) and determining the relationships among its major lineages, especially the widely divergent groups of protists, remains controversial (Brown and Doolittle 1997; Pace 1997; Baldauf 2003). Lying between the two prokaryotic lineages, (Bacteria and Archaea), and the multicellular organisms (plant, fungi and animals) are a great many groups of intriguing protistan clades, most of which are comprised of a variety of mainly unicellular eukaryotic species (Baldauf, Roger et al. 2000). Resolving the branching order within and among these protistan clades and identifying suitable place-holder taxa for higher-level analyses are essential for recovering an accurate TOL (Stechmann and Cavalier-Smith 2003). The primary reason that the relationships between protsitan taxa are still pooly known is due to the face that enormous stretches of time separate extant protists from their last common ancestor. One group of protists shares a common ancestor with the metazoans, and fungi and another very different clade of protists gave rise to the plants. Additionally, efforts to reconstruct a phylogeny for eukaryotes using molecular data may be hampered by lateral transfer of individual genes or larger genetic regions, presumably through symbiotic interactions or other events (Doolittle 2000).

The Euglenozoa are a monophyletic assemblage of flagellates that diverged early during the evolutionary history of eukaryotes and share a suite of ultrastructural features that unite them (Cavalier-Smith 1981; Sogin, Elwood et al. 1986; Van de Peer, Neefs et al. 1993). Three apparent synapomorphies of the Euglenozoa are: 1) a flagellar base with two basal bodies and three asymmetrically arranged microtubular roots; 2) a paraxial rod in one or both of the flagella composed primarily of two polypeptides arranged in either an amorphous or highly crystalline array; and 3) tubular, thick walled extrusive organelles (e.g. trichocysts, mucocysts) that may have been lost or reduced in many derived taxa (Simpson 1997). The Euglenozoa are thought to be comprised of around 1,600 species including both free-living and parasitic

organisms with some of the later being important human and animal pathogens. Based on morphological features, three major subgroups in the Euglenozoa have been proposed. These are the as Euglenids (Butschli 1884), the Kinetoplastids (Honigberg, B. M., Balamuth 1963) and the Diplonemids (Cavalier-Smith 1993). Support for the validity of these three unresolved subgroups of Euglenozoa was later recovered through analysis of small subunit ribosomal DNA (SSU rDNA) sequences (Maslov, Yasuhira et al. 1999; Preisfeld, Berger et al. 2000; Moreira, Lopez-Garcia et al. 2001; Preisfeld, Busse et al. 2001; Busse, Patterson et al. 2003; Busse and Preisfeld 2003).

Euglenids are distinguished by the presence of a series of pellicle strips that are organized by supporting microtubules underneath the cell membrane. An additional glycoprotein layer is appressed to the inner surface of the plasma membrane generating the ribbon-shaped pellicle strips, which extend the length of the cell, providing structural support for a variety of unique cell shapes. Euglenids are typically larger than 30µm in length and may reach 500µm. They are one of the best-known groups of free-living flagellates and may be green, photosynthetic autotrophs or colorless osmotrophs and heterotrophs. They are commonly found in freshwater, while some, like species of *Eutreptiella*, are marine organisms.

The Kinetoplastids are distinguished from other protists by the presence of a *kinetoplast*, a mitochondrial-derived structure formed by an accumulation of interlocking loops of DNA (kDNA), which is found in close association with flagellar bases. As in Euglenids, the cytoskeletal microtubules underlying the plasma membrane and form a continuous or discontinuous supporting corset, but they lack the additional glycoprotein layer found in Euglenids, which is an apparent requirement for pellicle strip formation. Kinetoplastids are typically smaller than Euglenids and are less than 30µm in length. This group contains a number

of parasites responsible for diseases in humans (e.g. sleeping sickness, Chagas' disease), other animals, and plants, as well as many other free-living bacterivores and micro-predators (e.g. Bodonids).

Diplonemids also have a corset of supporting microtubules under their cortical membranes, but unlike the Euglenids these microtubules are not organized into pellicle strips. They are tiny flagellates that are less than 30μm in length, and have a feeding and flagellar apparatus very similar to Bodonids and some Euglenids (Porter 1973; Willey 1988; Montegutfelkner and Triemer 1994). Most Diplonemids are presently thought to be strictly free-living marine and freshwater organisms, but recently von der Heyden et al. (2004) proposed a group of pathogenic Diplonemids that may target a range of hosts. Outside of the three major subgroups are a number of protists (e.g., *Postgaardi, Calkinsia* and *Petalomonas cantuscygni*) that are clearly members of the Euglenozoa but which have unclear phylogenetic affinities.

Phylogenetic relationships among the three main subgroups of Euglenozoa and various sundry orphan taxa have been explored using several lines of evidence, but no clear picture has emerged. Based on morphological similarities including the shape of the feeding and flagellar apparatus, patterns of mitosis, and cytoskeletal composition, Euglenids and Kinetoplastids were proposed to be more closely related to each other than either is to Diplonemids (Triemer and Farmer 1991). In contrast, recent neighbor-joining (Busse, Patterson et al. 2003) and distance-based (von der Heyden, Chao et al. 2004) phylogenetic analyses of SSU rDNA suggested that the Euglenids and Diplonemids were most closely related. Other recent studies using heat shock protein marker Hsp 70 and 90 (Simpson and Roger 2004) suggested that Kinetoplastids and Diplonemids are sister taxa. Most radically, it has recently been proposed that Kinetoplastids are actually descended from a photosynthetic Euglenids ancestor (Hannaert,

Saavedra et al. 2003; Martin and Borst 2003) an opinion that has been challenged by (Leander 2004). Because of the evolutionary, ecological and medical significance of this group of protists and the conflicting results obtained to date, relationships among the Euglenozoa are an area of intensive ongoing research.

The selection of appropriate molecular markers is essential to creating an accurate reconstruction of Euglenozoan phylogeny. A variety of molecular markers have previously been used for systematic studies in this group. Studies based on paraxonemal rod proteins 1 (PAR1) and 2 (PAR2) (Talke and Preisfeld 2002), heat shock protein 70 and 90 (Baldauf, Roger et al. 2000; Simpson, Lukes et al. 2002; Simpson and Roger 2004), *Cox*1 (Tessier, vanderSpeck et al. 1997); chloroplast-based *rbcL* (ribulose-1,5-bisphosphate carboxylase large subunit; (Thompson, Copertino et al. 1995), and mitochondria-based COI (cytochrome oxidase subunit I; (Maslov, Yasuhira et al. 1999) have yielded conflicting results. Additionally, systematic studies of the Euglenozoa using SSU rDNA have not provided adequate resolution or confident support for inferred clades, especially with deep-level relationships. Although SSU rDNA has several advantages over many protein coding genes including being effectively a single-copied gene, highly conserved, it is not translated and it is easy to amplify, these merits are partly offset by its other properties. SSU rDNA is relatively short and contains fewer phylogenetically informative characters, and these informative sites are prone to base substitution saturation (Busse and Preisfeld 2002; Busse, Patterson et al. 2003; Busse and Preisfeld 2003).

For a variety of reasons, large subunit ribosomal DNA (LSU rDNA) is an attractive molecular marker to sample for further testing evolutionary relationships in the Euglenozoa. LSU rDNA comprises a mosaic of slowly-evolving conserved core (CC) regions and fast-evolving expansion segments (ES), facilitating studies of relationships at both higher and

lower taxonomic levels (Ware, Tague et al. 1983; Clark, Tague et al. 1984; Hassouna, Michot et al. 1984). In previous studies, LSU rDNA has been successfully used in plants (Kuzoff, Sweere et al. 1998; Korall and Kenrick 2004), animals (Passamaneck, Schander et al. 2004; Winchell, Martin et al. 2004), insects (Suh, Noda et al. 2001), fungi (Cavalier-Smith 2002; Inderbitzin, Lim et al. 2004), yeast (Kerrigan, Smith et al. 2003), and other protists, including dinoflagellates (Hansen and Daugbjerg, 2004) and ciliates (Snoeyenbos-West, Cole et al. 2004). As is the case with SSU rDNA, LSU rDNA is effectively a single-copy nuclear gene present in all Euglenozoa, and it is easy to manipulate experimentally. Also, LSU rDNA is about four to five thousand base pairs and contains a significant number of phylogenetically informative characters. Additionally, LSU rDNA is not likely to be transferred horizontally among species (Daubin, Lerat et al. 2003). It promises to be an additional source of evidence from the nuclear genome that can be analyzed in combination with SSU rDNA due to concerted evolution. Recently, partial LSU rDNA sequences were sampled from a subset of the photosynthetic Euglenozoa (Brosnan, Shin et al. 2003). However, in their study, only around a third of the full length of LSU rDNA sequence was sequenced and analyzed.

In this study we explore and take advantage of the merits of LSU rDNA to evaluate its phylogenetic potential and to reconstruct a phylogeny for the Euglenozoa. In this study, we sample nearly full length LSU rDNA (4000 bp) from a wide range of Euglenozoan taxa and outgroup species. In combination with SSU rDNA, these sequences provide abundant phylogenetic signals. We have designed primers of LSU rDNA that are effective for its amplification of LSU rDNA and sequencing, and we also suggest feasible strategies for aligning the different portions of LSU rDNA (e.g., expansion segments and conserved core regions). We evaluate the phylogenetic potential of LSU rDNA using partitioned data matrices and comparing

their sequence dissimilarity and phylogenetic utility. Phylogenetic relationships in the Euglenozoa were estimated using uncorrected and corrected minimum evolution, unweighted and weighted parsimony, maximum likelihood, and Bayesian analyses. Tree topologies were evaluated using Shimodaira-Hasegawa test (SH test) (Shimodaira and Hasegawa 1999; Shimodaira 2001). Finally, we suggest suitable place-holder taxa to represent the Euglenozoa in broader studies of the TOL.

MATERIALS AND METHODS

Strains and culture conditions

The species sampled in this study are listed in Table 1 including culture collection information and GenBank accession numbers for both SSU rDNA and LSU rDNA. Among the thirty-nine taxa, twenty-eight LSU rDNA and five SSU rDNA sequences were newly sequenced in this study. *Colacium vesiculosum, Entosiphon sulcatum, Euglena laciniata, Euglena terricola, Euglena stellata, Eutreptiella gymnastica, Leishmania major, Lepocinclis ovata, Strombomonas acuminate, Trachelomonas hisida var. crenulatocolis, Trachelomonas volvocina, Trypanoplasma borreli, Trypanosoma brucei,* and *Trypanosoma cruzi* were sequenced from different live cultures for LSU and SSU rDNA (table 1). *Naegleria gruberi* and *Naegleria fowleri* were sequenced as outgroups to the Euglenozoa, and the alveolates *Euplotes aediculatus, Perkinsus andrewsi,* and *Plasmodium falciparum* were used as outgroup taxa in addition to *Naegleria*.

Eutreptiella gymnastica, Eutreptiella marina, and Eutreptiella sp. were grown in K medium (Keller M. D. 1987) in 50 ml tissue culture flasks at 15ºC on a 12:12 light dark cycle, while the other organisms were cultured in ESSEX medium (Brown 2003) grown at 20ºC on a

12:12 light dark cycle, except that *Petalomonas cantuscygni* was grown in ESNW medium (recipe below) and the others.

ESNW medium (Enriched Soil extract Natural Seawater medium) is a variation of natural seawater medium to which soil extract (1:250) and ES vitamins (1:250) (Harrison, Waters et al. 1980) are added. The recipe for soil extract is as follows: to 1L distilled water added 50g Garden soil, 0.2g $NH_4MgPO_4 \cdot 6H_2O$, 0.2g $CaCO_3$, 0.2g crushed barley, and 10 pieces of dry split peas; heat to 70˚C and maintain for 5 hours; remove from heat and cover with cheesecloth; let stand for two days at room temperature; decant the supernatant; filter through a 0.2 μm filter; autoclave for 30 min; and cool to 20˚C for use. The active components of ES vitamin solution include: Thiamine 0.1 $g \cdot L^{-1}$, Cyanocobalamin 2 $mg \cdot L^{-1}$, and Biotin 1 $mg \cdot L^{-1}$.

DNA preparation

Genomic DNA was isolated from cell pellets using the DNeasy plant mini-kit (Qiagen Incorporated, Valencia, CA) according to manufacturer's instructions with an extended cell lysis time of 10 min to 30 min. Centrifuged cell pellets were collected at 8,000 rpm from cultured log-phase growth cells and the starting weight ranged from 50μg to 100μg. The extracted DNA was quantified by spectrophotometer (SmartSpec 3000, BIO-RAD Laboratories, Hercules, CA) and diluted to a concentration of 10μg/μl to 30μg/μl to be ready for the following PCR amplification steps.

LSU rDNA sequences were amplified via PCR using isolated genomic DNA from twenty-eight species (listed in Table 1). A 100μl PCR reaction contained the following: 42.63μl water; 5μl DMSO; 10μl 10X PCR buffer (0.2M $MgCl_2$, 0.1M KCl, 0.1M Tris PH 8.3 and 1% Tween-20); 0.2mM dNTP dilution mix; 1mM forward primer and 1mM reverse primer solutions; 1.67μl JumpStart Taq DNA polymerase (SIGMA-ALDRICH Corporate, St. Louis, MO) at the

concentration of 2.5U/μl; 0.7μl Taq Extender PCR Additive (Stratagene Corporate, La Jolla, CA), and 10μl diluted DNA template. The PCR reaction mixtures were covered with mineral oil and PCR amplifications were carried out in MJ Research PCT-200 Peltier Thermal Cycler as following steps: (1) a hot start at 94ºC for 3 min; (2) 30 amplification cycles of 94ºC for 0.5 min, 45ºC-59ºC for 0.5 min, 72ºC for 3.5 min; (3) a terminal extension phase at 72ºC for 3 min. The annealing temperatures depend on different primer combinations, which were optimized using a gradient PCR program. The primers used in amplification are listed in Table 2.

The quality of PCR products was checked in a 1.2% agarose TBE gel first. The remaining PCR products were subsequently purified via precipitation with equal volumes of 20% Polyethylene Glyco-8, 000 (PEG; (Soltis and Soltis 1997; Kuzoff, Sweere et al. 1998) at 37ºC for 15 min, centrifuged for 15 min at 14,000 rpm at 4ºC, washed with 400μl 80% and 95% ethanol for 5 min respectively, and dried in a Savant Speed-Vac (Global Medical Instrumentation Incorporated, Ramsey, MN) for 10 min. The resulting dry pellets were resuspended in 25μl dH2O and quantitated by spectrophotometer, and DNA template for sequencing was prepared at a concentration of 2ng/μl per 100 base pairs of PCR product length. Low-yield PCR products from problematic taxa were re-amplified from TAE gel purified bands (QIAquick gel extraction kit, Qiagen Incorporated, Valencia, CA) or TA-cloned (TOPO-TA cloning kit, Qiagen Incorporated, Valencia, CA) prior to sequencing.

DNA sequencing

The primary amplification primers were designed based on an alignment of the full-length LSU rDNA sequences of *Euglena gracilis, Trypanosoma brucei* and *Leishmania major*, which were available in GenBank. Using these primary primers (Operon Biotechnologies Incorporated, Huntsville, AL), a pilot group including *Euglena gracilis, Eutreptiella marina,*

Phacus pusillus, Petalomonas cantuscygni, and *Strombomonas acuminata* was used to optimize PCR and sequencing conditions. Primer positions and relative locations are provided in Table 2 and Fig 1. Nearly complete sequences were amplified using 22F and 5465R, or 22F and 3510R together with 2300F and 5465R. For those taxa that could not be amplified in this way, internal primers were used to produce pieces of LSU rDNA sequences and taxon-specific primers were designed to amply the recalcitrant regions using the available sequences in conjunction with SSU rDNA sequences.

The DNA sample for cycle sequencing was prepared with the recipe of 4μl ABI BigDye Terminator Cycle Sequencing Ready Reaction Mix (Applied Biosystems, Foster City, CA), 0.4μl 10mM primer and 5.6μl dH2O. The single-primer cycle sequencing program was as follows: 96ºC for 10 seconds, 50ºC for 5 seconds and 60ºC for 4 min. To precipitate the cycle sequencing products, 1μl 3M sodium acetate (pH 4.6) and 25μl 95% ethanol were added and incubated at room temperature for 15 min, spun in a microcentrifuge for 20 min at 14,000 rpm, rinsed the pellet with 250μl 70% ethanol twice, spun for another 5 min at maximum speed, and dried the pellet in a Savant Speed-Vac (Global Medical Instrumentation Incorporated, Ramsey, MN) for 10 min. Automated sequencing of LSU rDNA was conducted on an ABI3700 automated sequencer (facilities in the Department of Plant Biology at the University of Georgia) or ABI3730 automated sequencer (Davis Sequencing Incorporated, Davis, CA). Sequence chromatogram output files were initially aligned and edited base by base using Sequencher version 4.1 (1994; Gene Codes Corporation, Ann Arbor, Michigan).

Sequence alignment and location of expansion segments and conserved core regions

Our definitions of the locations of expansions segments and conserved core regions were based on two alignments: one comprising LSU sequences from species across the TOL and

the other comprising Euglenozoa and five outgroup taxa. There are 14 representatives from the three major clades in the TOL alignment: *Escherichia coli* (V00331), *Halobacterium salinarum* (NC002607), *Haemophilus influenzae* (NC000907), *Uncultured crenarchaeote* (AJ627422), *Giardia muris* (X65063), *Saccharomyces cerevisiae* (Z73326), *Schizosaccharomyces pombe* (Z19136), *Oryza sativa* (M11585), *Arabidopsis thaliana* (X52322), *Euglena gracilis* (M12677), *Crithidia fasciculata* (Y00055), *Tetrahymena pyriformis* (X54004), *Xenopus laevis* (X02995), and *Caenorhabditis elegans* (X03680).

The alignment strategies for this set of taxa were as follows: all the eukaryotes taxa except *Euglena gracilis* were aligned in ClustalX 1.8 (Thompson, Gibson et al. 1997) in slow-accurate mode with pairwise alignment gap opening penalty of 10 and a gap extension penalty of 0.1, and multiple alignment gap extension penalty of 0.2. The alignment produced was imported into MacClade version 4.05 PPC (Maddison 1992), and series of local pairwise alignments were performed till the conserved core regions from all the species were aligned up. A difficult to align raw sequence from *Escherichia coli* was added to the alignment by making a pairwise alignment (Needlema.Sb and Wunsch 1970) with *Giardia muris*, and other prokaryotes taxa were added by making a pairwise alignment with their closest relatives. A difficult to align raw sequence from *Euglena gracilis* was added to the TOL alignment in a similar manner, but initially aligned with *Oryza sativa.* In this way, we aligned all the conserved core regions. We subsequently made adjustments to the expansion segments portions of the alignment as follows: for a poorly aligned region between two conserved core regions, we perform a local realignment using ClustalX 1.8. For alignment of these regions, optimized gap opening and gap extension values for pairwise alignment ranged from 10-25 and 0.1-6.6, respectively, and ranged from

10-25 and 0.2-6.6, respectively for multiple alignments. The alignment for the Euglenozoa and outgroup sequences was produced using the same general strategies.

Conserved core regions and expansion segments were based on the previous definition (Ware, Tague et al. 1983; Clark, Tague et al. 1984; Hassouna, Michot et al. 1984) and our new alignments. Conserved core regions are aligned portions of rDNA sequences that have more than fourteen consecutive base pairs of less than six base pairs across the TOL. Expansion segments are defined into four classes based on their extent of conservation. Class I expansion segments are found across investigated Eukaryotes. Class II expansion segments are unique to the Euglenozoa and could not be found in conserved core regions and Class I expansion segments. Class III expansion segments are regions that are conserved only within subgroups of Euglenozoa including Euglenids, Kinetoplastids, and Diplonemids. Class IV expansion segments are regions sharing pronounced length variation within subgroups of Euglenozoa (for detailed contents, please refer to discussion section).

The positions and lengths (drawn to scale) of fourteen conserved core regions and different classes of expansion segments in the Euglenozoa LSU rDNA sequence alignment are illustrated in Figure 2a Figure 2b illustrates how these regions were defined. Consensus motifs for the 10 bases at the beginning and end of each CCs in the Euglenozoa alignment including outgroup taxa are listed in Table 3. Standard ambiguity codes were used to designate positions within each motif exhibiting nucleotide variation among the aligned sequences. Sizes and exact locations of each CC regions were ascertained through inspection of the aligned sequences.

Sequence dissimilarity estimation

The sequences used in this analysis contain thirty-nine Euglenozoan and outgroup taxa that represent roughly eighty percent of the full gene length LSU rDNA genes. This corresponds

to the alignable portion of LSU rDNA and contains most of the phylogenetic signal. The following five partitions of the data matrix were used to estimate the sequence dissimilarities: (1) conserved core regions of LSU rDNA alone; (2) expansion segments class I, II and III of LSU rDNA; (3) expansion segments class IV of LSU rDNA; (4) entire LSU rDNA sequence; and (5) entire SSU rDNA sequences. The most suitable model of sequence evolution for each partition was determined using Modeltest 3.6 (Posada and Crandall 1998; Posada and Buckley 2004), and corresponding parameter values were estimated in PAUP* 4.0b10 (Swofford 1993; Swofford 2002). The taxa used to compare sequence dissimilarity were representatives from the four major subgroups of Euglenozoa (Simpson 1997; von der Heyden, Chao et al. 2004): (1) Euglenids; (2) Core Bodonids; (3) Trypanosomids; and (4) Diplonemids. Each suite of five representative sequences was selected from the first three subgroups in terms of their phylogenetic distances (marked as * in Table 1), while there were only two taxa available for Diplonemids. The same suite of representative taxa was used for each subgroup. The average sequence dissimilarity for each suite of sequences was calculated among these five partitions by comparing levels of base substitution under the TrN+I+G as model of sequence evolution and corresponding and parameter values estimated by PAUP* 4.0b10. We also added the enigmatic taxon *Petalomonas cantuscygni* into this analysis to compare pairwise distances between it and representatives from the above four subgroups. Additionally, the number of steps per twelve consecutive bases was generated by MacClade 4.05 PPC in order to assess variation in levels of base substitution among sites using the most parsimonious tree.

Phylogentic informativeness of Euglenozoa LSU rDNA sequences

The following six data sets were used to estimate the phylogenetic utility of Euglenozoan LSU rDNA sequences: (1) conserved core regions; (2) expansion segments class I,

II and III; (3) expansion segments class IV; (4) entire LSU rDNA sequence; (5) entire SSU rDNA sequences; and (6) combined SSU and LSU rDNA sequences. To assess its phylogenetic utility, five indicators were used including skewness values, rescaled consistency index (RC), the shape of gamma distribution (α), the proportion of invariant sites (pi) and B_{50} values on inferred trees. Skewness values provide an indication of the presence of phylogenetic signal in a data set (Hillis and Huelsenbeck 1992), and Alpha (α) is thought to be a good indicator of data set quality (Yang 1998; Lin and Danforth 2004). Alpha shows a significant negative correlation with pi and a positive correlation with RC, which means that data sets with lower values of alpha show less homoplasy (Swofford, Kasckow et al. 1996; Lin and Danforth 2004). In order to compare the relative supportive values after phylogenetic inference based on each of the above six data sets, B_{50} values were calculated as the percentage of the nodes that have bootstrap values above 50 on the bootstrap 50% majority-rule consensus tree.

Most parsimonious trees were produced using PAUP* 4.0b10, and the number of islands were recorded, which was subsequently used to decide the number of random taxa addition (RTA) needed for each bootstrap pseudo-replicate. Heuristic searches were performed using RTA with 100 replications, TBR branch swapping, MULPARS, accelerated transformation (ACCTRAN) character state optimization, gaps equivalent to missing data and multistate taxa coded as uncertain. Bootstrap analysis was performed using 100 pseudo-replicates and sampling was limited to parsimony informative characters. Skewness tests were conducted using 10,000 randomly selected trees. RC, alpha and pi for each gene and each data set were also calculated in PAUP* 4.0b10.

Phylogenetic analysis

Phylogenetic analyses of thirty-nine taxa including twenty-eight newly generated sequences of Euglenozoan LSU rDNA (listed in Table 1) were conducted in PAUP* 4.0b10 using a combined data set of concatenated SSU and LSU rDNA sequences. Modeltest 3.6 selected TrN (Tamura and Nei 1993) with a gamma correction for amount site rate variation and invariant sites as the most suitable model of sequence evolution from the hierarchical likelihood ratio rests (hLRTs). Based on this model, the corresponding likelihood parameters were estimated and then applied in PAUP* 4.0b10. Several phylogenies were produced using different methods, including: uncorrected and maximum likelihood corrected neighbor joining; unweighted parsimony and parsimony with substitutions weighted according to the instantaneous rate matrix or characters weighted according to their RC values. Resultant phylogenies were used as starting trees for maximum likelihood estimation under the estimated likelihood parameters. Bootstrap analysis was performed using 100 replicates with sampling limited to parsimony-informative characters. The same strategies were used on a data set including SSU rDNA only reference. MrBayes v. 3.0b4 (Huelsenbeck and Bollback 2001; Huelsenbeck and Ronquist 2001; Huelsenbeck, Ronquist et al. 2001) was also used with three datasets partitions as LSU rDNA conserved core regions and SSU rDNA, expansion segments class I to III, and expansion segment IV of LSU rDNA. Each run composed four chains starting from random trees, and 500,000 generations with trees sampled every one hundred generations. The majority rule consensus tree was calculated after the removal of first 200 trees corresponding to a burn-in period.

Topologies resulting from each of the above phylogenetic analyses were compared through a one-tailed SH test (Shimodaira and Hasegawa 1999) implemented by PAUP* 4.0b10.

Test distributions were generated using ten thousand resampling estimated log-likelihood (RELL) bootstrap replicates (Goldman, Anderson et al. 2000).

RESULTS

Phylogenetic properties of Euglenozoa LSU rDNA sequences

LSU rDNA in the Euglenozoa has a mosaic structure of conserved core regions (CC) and expansion segments (ES), with the later exhibiting greater variation in length and patterns of nucleotide replacement. Site variability is greatest in ES class IV, where the number of steps per twelve sites across the tree appreciably surpasses 100; while CC regions, at the other extreme, typically have fewer than 50 (figure 3, Table 4a). The overall base substitution rate of LSU rDNA is between that of expansion segments class IV and class I-III, while the conserved core regions of LSU rDNA has the slowest rate of evolution. Furthermore, the lengths of all the Euglenozoa species show great variation with independently derived length expansion in the Trypanosomids, *Euglena* and *Eutreptiella* clades. These results underscore the distinctions among LSU rDNA partitions and suggest that: 1) they will be useful for phylogeny reconstruction at a range of taxonomic levels; and 2) suitable phylogenetic methods will be required when these heterogeneous regions are analyzed simultaneously.

Conserved core regions in the Euglenozoa display distinct phylogenetic properties due to their highly conserved mode of evolution. SSU rDNA have a rate of base substitution rate that is several times higher than that of CC regions of Euglenozoa LSU rDNA (Table 4). Noticeably, Trypanosomatids have an especially slow rate of base substitution in their CC regions that is approximately one twelfth as fast as either SSU rDNA or ES I-III. This suggests that purifying selection is especially strong in these genetic regions. Conserved core regions of Euglenozoa

have a GC content that is near 50% and the highest RC values of all LSU partitions, suggesting that they may be attractive for phylogenetic analysis. However, because of their conservative pattern of evolution, CC regions also have the fewest phylogenetically informative characters and provide little resolution among lineages of Euglenozoa.

Expansion segments, especially class IV of LSU rDNA show markedly different features from conserved core regions (Table 5). ES class IV show the highest rate of base substitution, greatest variation in length and the largest deviation from equal base frequencies. They also have the lowest B_{50} and lowest RC values, which suggests that they contain comparatively less phylogenetic signal, and have greater conflict among their characters producing a higher ratio of phylogenetic noise. However, ES class IV have the highest alpha value among all partitions, a relatively even distribution of rates among sites and perhaps appreciably phylogenetic potential. Skewness values also indicate that all partitions contain significant nonrandom structure that likely reflects phylogenetic signal ($P<0.01$). When all ES, and CC regions are analyzed separately, they produce highly concordant topologies, suggesting that despite their distinctive properties, all LSU partitions retain useful phylogenetic signal. B_{50} values derived from separate analyses of each partition indicate that all classes of ES, and the conserved core regions provide for robust inferences. The GC content of CC regions and all ES do not differ substantially from 50%, suggesting that they are less likely to violate assumptions of equal base frequencies in phylogenetic analyses and, therefore, engender inaccurate results.

Phylogenetic inferences of Euglenozoa

Deep-level relationships in the Euglenozoa are recovered with high resolution and strong support in both maximum likelihood (Fig 4) and Bayesian majority-rule consensus topologies (Fig 5). The Euglenozoa form a strongly supported monophyletic group, that is sister

to the heterolobosians *Naegleria gruberi* and *Naegleria fowleri*. The three major subgroups of Euglenozoa form strongly supported clades. *Petalomonas cantuscygni*, an orphan taxon in the Euglenozoa, is sister to Diplonemids. Pairwise distance comparison in all partitions between *Petalomonas cantuscygni* and representatives from diverse subgroups of Euglenozoa also suggest that *Petalomonas* is most closely related to the Diplonemids (Table 4b). This clade is in turn sister to Euglenids with high bootstrap support. Kinetoplastids form a monophyletic group with 100 bootstrap support that is the sister clade to all other Euglenozoa. Some classically recognized genera, including *Bodo*, *Phacus* and *Euglena* are paraphyletic. In Euglenids, the two monophyletic group *Strombomonas* and *Trachelomonas* together form a clade, and *Eutreptiella* is a monophyletic clade at the base of Euglenids. In Kinetoplastids, three parasitic genera *Leishmania, Crithidia* and *Trypansomatids* form a monophyletic group.

The topology derived from maximum likelihood analysis of SSU and LSU rDNA combined data set shows significant improvement over the phylogeny based on SSU rDNA alone. Topologies based solely on the analysis of SSU rDNA were in complete topological agreement with those based on combined data set, but provided reduced support for deep level relationships. For example, the support value for *Petalomonas cantuscygni* to group with Diplonemids is 77 in SSU rDNA while the number is increased to 100 in combined data set.

Maximum likelihood phylogeny based on the combined data set was selected to be the best tree among all the other phylogenies recovered through different computational algorithms. An SH test was performed on phylogenies obtained from uncorrected and maximum-likelihood-corrected neighbor joining, unweighted parsimony and parsimony with substitutions weighted according to the instantaneous rate matrix or characters weighted according to their RC values, together with the last two hundred trees produced by the Bayesian

analysis. The maximum likelihood tree was indicated as the best tree among all the phylogenies. The last two hundred trees produced in Bayesian are not significantly worse than the maximum likelihood tree. However, trees based on all other methods are significantly worse ($P<0.05$).

DISCUSSION

Phylogenetic properties of Euglenozoa LSU rDNA data partitions

One of the classic issues in using ribosomal DNA sequences for phylogentic research is to distinguish regions that are conserved across the TOL from those that have not. Definitions of conserved core regions set forth in earlier studies vary, sometimes substantially. Previous circumscriptions fall into two traditions: one is based on an alignment of sampled LSU rDNA sequences from a suite of species across the TOL (Ware, Tague et al. 1983; Clark, Tague et al. 1984; Hassouna, Michot et al. 1984; Hancock and Dover 1988; Kuzoff, Sweere et al. 1998); and the other, based on a prior study of Euglenozoa, characterizes conserved segments of ribosomal RNA as "discrete RNA species" (White, Rudenko et al. 1986; Spencer, Collings et al. 1987; Schnare, Cook et al. 1990; Schnare and Gray 1990). In the latter, 5.8S is called "LSU rRNA species 1"; moving in the 3' direction, conserved regions are dubbed rDNA species 2 through rDNA species 14, which is misleading because it suggests that these are rDNA genes. Our findings are closer to the former tradition, but we propose a modified definition that is more tightly centered on the idea that conserved core regions are present across the whole TOL. Our revised definition is based on analysis of more recently sampled LSU rDNA sequences from across the TOL and is motivated by concerns for accuracy in both phylogenetic and functional studies.

Similarly, we recognize a suite of expansion segments that are distinguished on the basis of their phylogenetic distributions. Segments of the ES are highly conserved within large clades in the TOL and important components of LSU rDNA. Portions of them likely play a functional role in mature LSU rDNA, although their exact functions remain to be determined (Wuyts, De Rijk et al. 2000; Wuyts, Van de Peer et al. 2001). Expansion segments were once described as "inserts" of foreign DNA sequences that have been integrated into rDNA (Ware, Tague et al. 1983). The term expansion segments was later introduced to describe "the portions of RNA transcribed from DNA inserted into a stretch of evolutionarily non-conserved sequence [that] apparently does not destroy ribosome function" (Clark, Tague et al. 1984). For practical reasons, the definition was later simplified to the portion of DNA outside the conserved regions (Sofia, Chen et al. 2001). Expansion segments occur in a pattern of alternation with conserved core regions in all eukaryotic nuclear rDNA sequences including yeast (Veldman, Klootwijk et al. 1981), plants (Kuzoff, Sweere et al. 1998), Metazoa (Ware, Tague et al. 1983; Clark, Tague et al. 1984; Hassouna, Michot et al. 1984) and other single-celled eukaryotes. We classify expansion segments based on their phylogenetic distributions both across the Eukaryotes and among Euglenozoa, permitting distinct models of sequence evolution to be applied to each class when analyzed phylogenetically. Most importantly, our definitions reveal deep-level synapomorphies; for example, expansion segments class I and II are synapomorphies for eukaryotes and Euglenozoa, respectively, while ES class III contains synapomorphies for the three main subgroups of Euglenozoa; Euglenids, Kinetoplastids and Diplonemids. A similar approach could be applied to other eukaryotes groups, enabling greater accuracy in ribosomal DNA-based studies of phylogenetic relationships.

Rapidly evolving regions, such as expansion segments class IV, are often excluded from phylogenetic analyses, but there are several reasons why we chose to include these regions in ours. First, previous theoretical studies have underscored the value of retaining rapidly evolving characters in phylogenetic studies (Yang 1998). Second, excellent methods are available to diminish the phylogenetic noise that can be engendered through analyses of such characters (Yang 1996; Pupko, Huchon et al. 2002). Third, several preliminary analyses and metrics generated here indicate that these regions contain appreciable phylogenetic signal that is concordant with that of SSU rDNA and other LSU rDNA partitions. For example, a relatively high value for alpha (α) suggests that ES IV are well suited to phylogenetic study. Phylogenetic analysis of ES IV alone yields a topology that is highly concordant with that based on the combined data set. The B_{50} value for LSU rDNA when ES IV is excluded is lower than that for total LSU rDNA. Finally, branch support metrics for shallow nodes are especially enhanced when ES III are included.

Optimizing phylogeny of the Euglenozoa

In systematic studies of Euglenozoa, the choice of molecular marker and taxon sampling are major factors affecting the topology. Resolving deep-level relationships within the Euglenozoa using SSU rDNA is limited both by the relatively small number of informative characters (Busse and Preisfeld 2002) and the high level of saturation in rapidly evolving sites (Moreira, Lopez-Garcia et al. 2001; Busse and Preisfeld 2003). Additionally, unbalanced taxonomic sampling has also been problematic. For example, species of Euglenids and Kinetoplastids are well sampled in some recent studies (Marin, Palm et al. 2003; von der Heyden, Chao et al. 2004), but are represented by only a handful of sequences in others studies (Maslov, Yasuhira et al. 1999; Preisfeld, Busse et al. 2001; Simpson and Roger 2004). For Diplonemids,

there are only two genera, so taxon sampling is typically adequate. In our phylogeny, we sampled a range of taxa in each subgroup as well as outgroups that were selected on the basis of previous studies to span the phylogenetic breadth of Euglenozoa.

In addition to expanded sampling of characters and taxa, our study also benefits from the use of reliable phylogenetic algorithms and a proper rooting strategy. In our study, we infer the phylogeny of the Euglenozoa using maximum likelihood estimation and Bayesian analysis. As indicated by an SH test, these methods produced superior topologies to those recovered by other computational algorithms (e.g., weighted parsimony and maximum-likelihood corrected neighbour-joining). Previous phylogenetic studies in the Euglenozoa relied primarily on unweighted parsimony or uncorrected neighbor-joining and did not assess results using any topology test (Marin, Palm et al. 2003; von der Heyden, Chao et al. 2004). In addition, we select Percolozoa (including *Naegleria gruberi* and *Naegleria flowleri*) as the outgroup taxa to the Euglenozoa because previous studies of ultrastructural features indicated that they have both discoid mitochondrial cristae (Cavalier-Smith 2000; Cavalier-Smith 2002; Simpson 2003; Stechmann and Cavalier-Smith 2003) and molecular data based on a variety of molecular markers including SSU rDNA, actin, beta-tubulin, hsp70, hsp90 and other genes (Fast, Xue et al. 2002; Baldauf 2003; von der Heyden, Chao et al. 2004) support the notion that Percolozoa are the sister group to Euglenozoa.

Phylogeny of deep-level relationships within Euglenozoa

We specifically focused on the branching order of the three subgroups of Euglenozoa and the orphan taxon *Petalomonas cantuscygni*. Our results indicate that the Diplonemids and *Petalomonas cantuscygni* are sister to one another and together they form a clade with the Euglenids. This topology is in agreement with a circumscription of Diplonemids and Euglenids

into subphylum Plicostoma, which was previously proposed on the basis of shared morphological features of a type II feeding apparatus including plicate vanes and two supporting rods (Kivic and Walne 1984; Larsen and Patterson 1990; Triemer and Farmer 1991; Cavalier-Smith 1998). The most striking morphological difference between Euglenids and Diplonemids is the presence of pellicle strips. Euglenids have pellicle strips formed from additional glycoprotein layers with a supported microtubule corset underneath. Diplonemids, in contrast, do not possess discoid mitochondrial cristae and lack axonemal rods within their flagella (Triemer and Farmer 1991). Our study provides a robust framework for future comparative studies (e.g., whether there is any vestigial structure of pellicle strips in Diplonemids). This phylogeny also agrees with previous studies based on SSU rDNA sequences (von der Heyden, Chao et al. 2004) in that both place the Kinetoplastid clade at the base of Euglenozoa, but our study provides much higher bootstrap support for this placement. However, our phylogeny contradicts with the notion that *Petalomonas cantuscygni* is a basal Euglenid as has been suggested by previous studies (Preisfeld, Berger et al. 2000; Leander and Farmer 2001; Mullner, Angeler et al. 2001; Busse, Patterson et al. 2003). Sampling additional orphan taxa for SSU and LSU rDNA will likely be helpful to resolve their phylogenetic affinities as well.

Other reports had supported the notion that Kinetoplastids and Diplonemids are closer to each other than either is to Euglenids. The most persuasive evidence for this hypothesis is that TGA codons encodes for tryptophan in Diplonemids and Kinetoplastids, whereas Euglenids use the universal genetic code (Inagaki, HayashiIshimaru et al. 1997; Yasuhira and Simpson 1997; Maslov, Yasuhira et al. 1999). Our topology suggests that this condition is either plesiomorphic for the Euglenozoa and was lost in the Euglenids or it was independently derived in Diplonemids and Kinetoplastids. Importantly we notice that there are few shared morphological features

between Diplonemids and Kinetoplastids (Elbrachter, Schnepf et al. 1996; Simpson 1997). Recently a maximum likelihood tree based on hsp 70 and 90 protein markers also supported the notion that Diplonemids and Kinetoplastids are more closely related to one another than either is to Euglenids (Simpson and Roger 2004). However, the taxon sampling in this study is extremely unbalanced among the three subgroups. For example, only one taxon was used to represent the Euglenids. Additionally, hsp 70 and 90 are part of a gene family that is known for frequent duplications (Kapoor, Curle et al. 1995; Atkinson, Bolitho et al. 1998; Sung, Vierling et al. 2001), therefore, it is entirely possible that paralogous genes were analyzed in this study. A maximum likelihood tree based on COI protein sequences (Maslov, Yasuhira et al. 1999) also supported the notion that Diplonemids were closer to Kinetoplastids. However, the taxon sample in this study was also very sparse and the bootstrap support for inferred clades was only moderate. Additionally, conflicting results are recovered in this study when parsimony and distance based methods are used to analyze the data. Those approaches suggest that Diplonemids and Euglenids are most closely related.

Euglenids and Kinetoplastids were once associated together by a suite of morphological features (Triemer and Farmer 1991). However, ultrastructural data for Diplonemids were not compared in this study. Recently, it has been shown that some Trypanosomes contain several genes that seem to be homologous with proteins found in plants or algae (Andersson and Roger 2002; Hannaert, Saavedra et al. 2003). To some, this suggested that a chloroplast had been acquired early in the history of Kinetoplastids and was subsequently lost. This hypothesis suggests that: 1) the Euglenids are not a monophyletic group; and 2) the Kinetoplstids are descendant from a photosynthetic Euglenid ancestor. In our phylogeny, Kinetoplastids are located at the base of the Euglenozoa, and photosynthetic Euglenids are

derived in a monophyletic clade with osmotrophic, phagotrophic or heterotrophic Euglenids forming its basal branches. Since these aspects of our topology are rather well supported, we suggest that the hypothesis of Anderson et al. (2002) and others is unsupported and that there might be alternative explanations for the algal-like genes in the Kinetoplastids.

Place-holder Euglenozoa in the TOL

Based on our phylogeny, we select representative taxa from the Euglenozoa to be used in future analyses of the TOL. We selected place-holders on the basis of three criteria: 1) the breadth of their phylogenetic distribution; 2) shorter branch length; and 3) shorter sequence length. In order to decrease the chance of long branch attraction (LBA), taxa clustered in one derived clade with relatively long branch lengths were avoided in our selection process. As our phylogeny indicated, *Eutreptiella marina, Diplonema papillatum* and *Trypanoplasma borreli* have relatively short branch lengths and also have shorter LSU rDNA sequences. *Naegleria gruberi*, from the sister clade to Euglenozoa, was also included to maximize phylogenetic breadth and reduce the potential for long branch attraction. This set of representative taxa is different from what has been used previously to represent the Euglenozoa in TOL phylogeny reconstruction. *Euglena gracilis*, a well-known photosynthetic protist, is most frequently used to represent the Euglenids, and *Trypanosoma brucei* or *Trypanosoma cruzi* are often used as place-holder taxa for Kinetoplastids. However, these taxa are all derived and have relatively long branchs and relatively long LSU rDNA sequences, Bodonids and Diplonemids are generally not represented at all. We suggest that future studies of Eukaryotes groups utilize more appropriate place-holder taxa in those clades, to maximize the chance of recovering an accurate estimate of the TOL.

CONCLUSIONS

We conclude that Euglenozoa LSU rDNA sequences contain significant phylogenetic signal in both conserved core regions and all different classes of expansion segments. LSU rDNA sequences evolve 1.16 to 2.67 times faster than SSU rDNA, and yield 2.41 times as many characters. Different partitions of LSU rDNA sequences presumably are useful for phylogeny retrieval at different taxonomic levels with suitably estimated parameters for LSU rDNA. Expansion segments class I-III have similar evolutionary rates with SSU rDNA. Therefore, they should be appropriate for phylogeny retrieval at taxonomic levels similar to those investigated with SSU rDNA. Expansion segment IV evolves at a rate of 11.89 to 90.49 times as fast as the conserved core regions of LSU rDNA. They need to be appropriately weighted, or perhaps excluded, from phylogenetic analyses at much higher taxonomic levels, but they appear to be informative at lower-level relationship (e.g. species level) in Euglenozoa. In general, LSU rDNA have great phylogenetic potential to be applied to studies at different taxonomic levels with appropriate partitions and weighting strategies.

The conclusions on the deep-level relationships are drawn based on our phylogeny which generally has high resolution and robust support for inferred clades. The Euglenozoa is a monophyletic group and includes Euglenids, Diplonemids and Kinetoplastids and a range of orphan taxa. One taxon of previous phylogenetic uncertainty, *Petalomonas cantuscygni*, is here confidently grouped together with diplonemids, to form a clade that is a closer relative to the Euglenids than to Kinetoplastids. Bodonids, *Phacus* and *Euglena* are paraphyletic. In Euglenids, the genera *Strombomonas* and *Trachelomonas* together form a clade, and *Eutreptiella* is a monophyletic clade at the base of photosynthetic Euglenids. In Kinetoplastids, three parasitic genera *Leishmania, Crithidia* and *Trypansomatids* form a monophyletic group.

Based on our phylogeny, *Naegleria gruberi* was selected as an outgroup, and *Eutreptiella marina, Diplonema papillatum, Trypanoplasma borreli* are selected as place-holder taxa to represent Euglenozoa because their short branch lengths, sequence lengths and relative phylogenetic breadth within the clade.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Alastair G. B. Simpson (Dalhousie University) for so generous providing the genomic DNA of *Bodo saliens, Bodo uncinatus, Cryptobia helices, Rhynchopus sp, Trypanoplasma borreli,* and *Rhynchobodo sp.,* Dr. Francine Marciano-Cabral (Virginia Commonwealth University) for providing the genomic DNA of *Naegleria gruberi,* and *Naegleria fowleri,* and Dr. David Porter (Department of Plant Biology, Univeristy of Georgia) for providing living culture of *Diplonema papillatum.* This research was supported by the U.S. National Science Foundation PEET (Partnerships for Enhancing Expertise in Taxonomy) grant no. DEB 4-21348.

REFERENCES

Andersson, J. O. and A. J. Roger (2002). "A cyanobacterial gene in nonphotosynthetic protists - An early chloroplast acquisition in eukaryotes?" Current Biology **12**(2): 115-119.

Atkinson, R. G., K. M. Bolitho, et al. (1998). "Apple ACC-oxidase and polygalacturonase: ripening-specific gene expression and promoter analysis in transgenic tomato." Plant Molecular Biology **38**(3): 449-460.

Baldauf, S. L. (2003). "The deep roots of eukaryotes." Science **300**(5626): 1703-6.

Baldauf, S. L., A. J. Roger, et al. (2000). "A kingdom-level phylogeny of eukaryotes based on combined protein data." Science **290**(5493): 972-7.

Brosnan, S., W. Shin, et al. (2003). "Phylogeny of the photosynthetic euglenophytes inferred from the nuclear SSU and partial LSU rDNA." Int J Syst Evol Microbiol **53**(Pt 4): 1175-86.

Brown, J. R. and W. F. Doolittle (1997). "Archaea and the prokaryote-to-eukaryote transition." Microbiol Mol Biol Rev **61**(4): 456-502.

Brown, P., Zakry's B., and Farmer, M. A. (2003). "Plastid morphology, ultrastructure, and development in Colacium and the loricate euglenophytes (Euglenophyceae)." J. Phycol. **39**: 115-121.

Busse, I., D. J. Patterson, et al. (2003). "Phylogeny of phagotrophic euglenids (Euglenozoa): A molecular approach based on culture material and environmental samples." Journal of Phycology **39**(4): 828-836.

Busse, I. and A. Preisfeld (2002). "Phylogenetic position of Rhynchopus sp. and Diplonema ambulator as indicated by analyses of euglenozoan small subunit ribosomal DNA." Gene **284**(1-2): 83-91.

Busse, I. and A. Preisfeld (2002). "Unusually expanded SSU ribosomal DNA of primary osmotrophic euglenids: molecular evolution and phylogenetic inference." J Mol Evol **55**(6): 757-67.

Busse, I. and A. Preisfeld (2003). "Application of spectral analysis to examine phylogenetic signal among euglenid SSU rDNA data sets (Euglenozoa)." Organisms Diversity & Evolution **3**(1): 1-12.

Busse, I. and A. Preisfeld (2003). "Discovery of a group I intron in the SSU rDNA of Ploeotia costata (Euglenozoa)." Protist **154**(1): 57-69.

Busse, I. and A. Preisfeld (2003). "Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa)." International Journal of Systematic and Evolutionary Microbiology **53**: 617-624.

Butschli, O. (1884). "Protozoa in Bronn's Klassen und Ordnung des." Tierreiches. **2**: 617-1097.

Cavalier-Smith, T. (1981). "Eukaryote kingdoms: seven or nine?" Biosystems **14**(3-4): 461-81.

Cavalier-Smith, T. (1993). "Kingdom protozoa and its 18 phyla." Microbiol Rev **57**(4): 953-94.

Cavalier-Smith, T. (1998). "A revised six-kingdom system of life." Biological Reviews **73**(3): 203-266.

Cavalier-Smith, T. (2000). "Flagellate megaevolution: the basis for eukaryote diversification." 361-390.

Cavalier-Smith, T. (2002). "The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa." Int J Syst Evol Microbiol **52**(Pt 2): 297-354.

Clark, C. G., B. W. Tague, et al. (1984). "Xenopus-Laevis-28S Ribosomal-Rna - a Secondary Structure Model and Its Evolutionary and Functional Implications." Nucleic Acids Research **12**(15): 6197-6220.

Daubin, V., E. Lerat, et al. (2003). "The source of laterally transferred genes in bacterial genomes." Genome Biology **4**(9): -.

Doolittle, W. F. (2000). "Uprooting the tree of life." Scientific American **282**(2): 90-95.

Elbrachter, M., E. Schnepf, et al. (1996). "Hemistasia phaeocysticola (SCHERFFEL) comb nov, redescription of a free-living, marine, phagotrophic kinetoplastid flagellate." Archiv Fur Protistenkunde **147**(2): 125-136.

Fast, N. M., L. Xue, et al. (2002). "Re-examining alveolate evolution using multiple protein molecular phylogenies." J Eukaryot Microbiol **49**(1): 30-7.

Goldman, N., J. P. Anderson, et al. (2000). "Likelihood-based tests of topologies in phylogenetics." Syst Biol **49**(4): 652-70.

Hancock, J. M. and G. A. Dover (1988). "Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs." Mol Biol Evol **5**(4): 377-91.

Hannaert, V., E. Saavedra, et al. (2003). "Plant-like traits associated with metabolism of Trypanosoma parasites." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 1067-1071.

Harrison, P. J., R. E. Waters, et al. (1980). "A Broad-Spectrum Artificial Seawater Medium for Coastal and Open Ocean Phytoplankton." Journal of Phycology **16**(1): 28-35.

Hassouna, N., B. Michot, et al. (1984). "The Complete Nucleotide-Sequence of Mouse 28S Ribosomal-Rna Gene - Implications for the Process of Size Increase of the Large Subunit Ribosomal-Rna in Higher Eukaryotes." Nucleic Acids Research **12**(8): 3563-3583.

Hillis, D. M. and J. P. Huelsenbeck (1992). "Signal, noise, and reliability in molecular phylogenetic analyses." J Hered **83**(3): 189-95.

Honigberg, B. M., Balamuth, W. (1963). "Subphylum Sarcomastigophora nom. Nov. to embrace the flagellate and amoiboid assemblages of protozoans." Journal of Protozoology **10 (suppl.) 27**.

Huelsenbeck, J. P. and J. P. Bollback (2001). "Empirical and hierarchical Bayesian estimation of ancestral states." Systematic Biology **50**(3): 351-366.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.

Huelsenbeck, J. P., F. Ronquist, et al. (2001). "Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology." Science **294**(5550): 2310-2314.

Inagaki, Y., Y. HayashiIshimaru, et al. (1997). "Algae or protozoa: Phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences." Journal of Molecular Evolution **45**(3): 295-300.

Inderbitzin, P., S. R. Lim, et al. (2004). "The phylogenetic position of Spathulospora based on DNA sequences from dried herbarium material." Mycological Research **108**: 737-748.

Kapoor, M., C. A. Curle, et al. (1995). "The Hsp70 Gene Family of Neurospora-Crassa - Cloning, Sequence-Analysis, Expression, and Genetic-Mapping of the Major Stress-Inducible Member." Journal of Bacteriology **177**(1): 212-221.

Keller M. D., S., R. C., Claus, W., and Guillard, R. R. L. (1987). "Media for the culture of oceanic ultraphytoplankton." J. Phycol. **23**: 633-638.

Kerrigan, J., M. T. Smith, et al. (2003). "Ascobotryozyma cognata sp. nov., a new ascomycetous yeast associated with nematodes from wood-boring beetle galleries." Mycological Research **107**: 1110-1120.

Kivic, P. A. and P. L. Walne (1984). "An Evaluation of a Possible Phylogenetic Relationship between the Euglenophyta and Kinetoplastida." Origins of Life and Evolution of the Biosphere **13**(3-4): 269-288.

Korall, P. and P. Kenrick (2004). "The phylogenetic history of Selaginellaceae based on DNA sequences from the plastid and nucleus: extreme substitution rates and rate heterogeneity." Molecular Phylogenetics and Evolution **31**(3): 852-864.

Kuzoff, R. K., J. A. Sweere, et al. (1998). "The phylogenetic potential of entire 26S rDNA sequences in plants." Molecular Biology and Evolution **15**(3): 251-263.

Larsen, J. and D. J. Patterson (1990). "Some Flagellates (Protista) from Tropical Marine-Sediments." Journal of Natural History **24**(4): 801-937.

Leander, B. S. (2004). "Did trypanosomatid parasites have photosynthetic ancestors?" Trends in Microbiology **12**(6): 251-258.

Leander, B. S. and M. A. Farmer (2001). "Evolution of Phacus (Euglenophyceae) as inferred from pellicle morphology and SSU rDNA." Journal of Phycology **37**(1): 143-159.

Lin, C. P. and B. N. Danforth (2004). "How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets." Molecular Phylogenetics and Evolution **30**(3): 686-702.

Maddison, W. P. a. D. R. M. (1992). "MacClade. Version 3.0."

Marin, B., A. Palm, et al. (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**(1): 99-145.

Martin, W. and P. Borst (2003). "Secondary loss of chloroplasts in trypanosomes." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 765-767.

Maslov, D. A., S. Yasuhira, et al. (1999). "Phylogenetic affinities of Diplonema within the euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences." Protist **150**(1): 33-42.

Montegutfelkner, A. E. and R. E. Triemer (1994). "Phylogeny of Diplonema-Ambulator (Larsen and Patterson).1. Homologies of the Flagellar Apparatus." European Journal of Protistology **30**(2): 227-237.

Moreira, D., P. Lopez-Garcia, et al. (2001). "New insights into the phylogenetic position of diplonemids: G+C content bias, differences of evolutionary rate and a new environmental

sequence." International Journal of Systematic and Evolutionary Microbiology **51**: 2211-2219.

Mullner, A. N., D. G. Angeler, et al. (2001). "Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence." International Journal of Systematic and Evolutionary Microbiology **51**: 783-791.

N., H. G. a. D. (2004). "Ultrastructure of Gyrodinium spirale, the type species of Gyrodinium (Dinophyceae), including a phylogeny of G. dominans, G. rubrum and G. spirale deduced from partial LSU rDNA sequences." Protist **155**: 271-294.

Needlema.Sb and C. D. Wunsch (1970). "A General Method Applicable to Search for Similarities in Amino Acid Sequence of 2 Proteins." Journal of Molecular Biology **48**(3): 443-&.

Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.

Passamaneck, Y. J., C. Schander, et al. (2004). "Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences." Molecular Phylogenetics and Evolution **32**(1): 25-38.

Porter, D. (1973). "Isonema-Papillatum Sp-N, a New Colorless Marine Flagellate - Light-Microscopic and Electron-Microscopic Study." Journal of Protozoology **20**(3): 351-356.

Posada, D. and T. R. Buckley (2004). "Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests." Systematic Biology **53**(5): 793-808.

Posada, D. and K. A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**(9): 817-818.

Preisfeld, A., S. Berger, et al. (2000). "Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data." Journal of Phycology **36**(1): 220-226.

Preisfeld, A., I. Busse, et al. (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**(Pt 3): 751-8.

Pupko, T., D. Huchon, et al. (2002). "Combining multiple data sets in a likelihood analysis: which models are the best?" Mol Biol Evol **19**(12): 2294-307.

Schnare, M. N., J. R. Cook, et al. (1990). "14 Internal Transcribed Spacers in the Circular Ribosomal DNA of Euglena-Gracilis." Journal of Molecular Biology **215**(1): 85-91.

Schnare, M. N. and M. W. Gray (1990). "Sixteen discrete RNA components in the cytoplasmic ribosome of Euglena gracilis." J Mol Biol **215**(1): 73-83.

Shimodaira, H. (2001). "Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection." Communications in Statistics-Theory and Methods **30**(8-9): 1751-1772.

Shimodaira, H. and M. Hasegawa (1999). "Multiple comparisons of log-likelihoods with applications to phylogenetic inference." Molecular Biology and Evolution **16**(8): 1114-1116.

Simpson, A. G., J. Lukes, et al. (2002). "The evolutionary history of kinetoplastids and their kinetoplasts." Mol Biol Evol **19**(12): 2071-83.

Simpson, A. G. B. (1997). "The identity and composition of the Euglenozoa." Archiv Fur Protistenkunde **148**(3): 318-328.

Simpson, A. G. B. (2003). "Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota)." International Journal of Systematic and Evolutionary Microbiology **53**: 1759-1777.

Simpson, A. G. B. and A. J. Roger (2004). "Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa." Molecular Phylogenetics and Evolution **30**(1): 201-212.

Snoeyenbos-West, O. L., J. Cole, et al. (2004). "Molecular phylogeny of phyllopharyngean ciliates and their group I introns." J Eukaryot Microbiol **51**(4): 441-50.

Sofia, H. J., G. Chen, et al. (2001). "Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods." Nucleic Acids Res **29**(5): 1097-106.

Sogin, M. L., H. J. Elwood, et al. (1986). "Evolutionary Diversity of Eukaryotic Small-Subunit Ribosomal-Rna Genes." Proceedings of the National Academy of Sciences of the United States of America **83**(5): 1383-1387.

Soltis, D. E. and P. S. Soltis (1997). "Phylogenetic relationships in Saxifragaceae sensu lato: A comparison of topologies based on 18S rDNA and rbcL sequences." American Journal of Botany **84**(4): 504-522.

Spencer, D. F., J. C. Collings, et al. (1987). "Multiple Spacer Sequences in the Nuclear Large Subunit Ribosomal-Rna Gene of Crithidia-Fasciculata." Embo Journal **6**(4): 1063-1071.

Stechmann, A. and T. Cavalier-Smith (2003). "Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90." Journal of Molecular Evolution **57**(4): 408-419.

Stechmann, A. and T. Cavalier-Smith (2003). "Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90." J Mol Evol **57**(4): 408-19.

Suh, S. O., H. Noda, et al. (2001). "Insect symbiosis: derivation of yeast-like endosymbionts within an entomopathogenic filamentous lineage." Mol Biol Evol **18**(6): 995-1000.

Sung, D. Y., E. Vierling, et al. (2001). "Comprehensive expression profile analysis of the Arabidopsis hsp70 gene family." Plant Physiology **126**(2): 789-800.

Swofford, C. D., J. W. Kasckow, et al. (1996). "Substance use: A powerful predictor of relapse in schizophrenia." Schizophrenia Research **20**(1-2): 145-151.

Swofford, D. L. (1993). Phylogenetic Analysis using Parsimony (PAUP) 3.1.1. Washington, D.C., Smithsonian Institute.

Swofford, D. L. (2002). PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, MA, Sinauer.

Talke, S. and A. Preisfeld (2002). "Molecular evolution of euglenozoan paraxonemal rod genes par1 and par2 coincides with phylogenetic reconstruction based on small subunit rDNA data." Journal of Phycology **38**(5): 995-1003.

Tamura, K. and M. Nei (1993). "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees." Molecular Biology and Evolution **10**(3): 512-526.

Tessier, L. H., H. vanderSpeck, et al. (1997). "The cox1 gene from Euglena gracilis: A protist mitochondrial gene without introns and genetic code modifications." Current Genetics **31**(3): 208-213.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.

Thompson, M. D., D. W. Copertino, et al. (1995). "Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus Euglena." Nucleic Acids Res **23**(23): 4745-52.

Triemer, R. E. and M. A. Farmer (1991). "An Ultrastructural Comparison of the Mitotic Apparatus, Feeding Apparatus, Flagellar Apparatus and Cytoskeleton in Euglenoids and Kinetoplastids." Protoplasma **164**(1-3): 91-104.

Van de Peer, Y., J. M. Neefs, et al. (1993). "Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock." J Mol Evol **37**(2): 221-32.

Veldman, G. M., J. Klootwijk, et al. (1981). "The primary and secondary structure of yeast 26S rRNA." Nucleic Acids Res **9**(24): 6935-52.

von der Heyden, S., E. E. Chao, et al. (2004). "Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa." Journal of Eukaryotic Microbiology **51**(4): 402-416.

Ware, V. C., B. W. Tague, et al. (1983). "Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis." Nucleic Acids Res **11**(22): 7795-817.

White, T. C., G. Rudenko, et al. (1986). "Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs." Nucleic Acids Res **14**(23): 9471-89.

Willey, R. L., Walne, P.L. Kivic, P. (1988). "Phagotrophic and the origins of euglenoid flagellates." CRC Critical Reviews in Plant Sciences **7**: 303-340.

Winchell, C. J., A. P. Martin, et al. (2004). "Phylogeny of elasmobranchs based on LSU and SSU ribosomal RNA genes." Molecular Phylogenetics and Evolution **31**(1): 214-224.

Wuyts, J., P. De Rijk, et al. (2000). "Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA." Nucleic Acids Research **28**(23): 4698-4708.

Wuyts, J., Y. Van de Peer, et al. (2001). "Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA." Nucleic Acids Res **29**(24): 5017-28.

Yang, Z. H. (1996). "Among-site rate variation and its impact on phylogenetic analyses." Trends in Ecology & Evolution **11**(9): 367-372.

Yang, Z. H. (1998). "On the best evolutionary rate for phylogenetic analysis." Systematic Biology **47**(1): 125-133.

Yasuhira, S. and L. Simpson (1997). "Phylogenetic affinity of mitochondria of Euglena gracilis and kinetoplastids using cytochrome oxidase I and hsp60." Journal of Molecular Evolution **44**(3): 341-347.

Table **2.1.** SSU and LSU rDNA sequences analyzed (listed alphabetically)
Abbreviations: UTEX, The Culture Collection of Algae at the University of Texas at Austin, Austin, TX, USA; CCMP, the Provasoli-Guillard National Center for the Culture of Marine Phytoplankton; ACOI, Culture Collection of Algae at the Department of Botany, University of Coimbra, Portugal; SAG, Sammlung von Algenkulteren Gőttingen; UW, Culture Collection of Algae at the Department of Plant Systematics and Geography at Warsaw University, Poland; ATCC, American Type Culture Collection, Manassas, VA, USA; SCCAP, Santa Clara Community Action Program, Santa Clara University, Santa Clara, CA, USA; Tt-JH strain, Jirí Lom, Institution of Parasitology, Ceské Budejovice, Czech Republic; S604, isolate from Lake Nelson, Piscataway, NJ, US.

Taxon	Culture used for sequencing		GenBank accession No.	
	SSU rDNA	LSU rDNA	SSU rDNA	LSU rDNA
*Bodo saliens**	ATCC50358	ATCC50358	AF174379	This study
*Bodo uncinatus**	ATCC30904	ATCC30904	AF208884	This study
Colacium mucronatum	UTEX 2524	UTEX 2524	AF326232	This study
Colacium vesiculosum	UTEX LB 1315	UW Łazienki	AF081592	This study
*Crithidia fasciculate**	MA isolation	MA isolation	Y00055	Y00055
*Cryptobia helices**	Czech Republic isolation	Czech Republic isolation	AF208880	This study
*Diplonema papillatum**	ATCC50162	ATCC50162	AF119811	This study
*Euglena gracilis**	-	-	M12677	X53361
Euglena laciniata	SAG 1224-31	UTEX 1312	AJ532420	This study
Euglena terricola	UTEX LB 1311	UTEX LB 1310	AF445459	This study
Euglena tripteris	UTEX LB 1311	UTEX LB 1311	AF286210	This study
Euglena stellata	SAG 1224-14	UTEX 372	AJ532419	This study
Euplotes aediculatus	-	-	AF164136	AF223571
Eutreptiella gymnastica	SCCAP K-0333	CCMP 1549	AJ532400	This study
*Eutreptiella marina**	CCMP 390	CCMP 390	This study	This study
Eutreptiella sp.	UTEX 2003	UTEX 2003	This study	This study
*Leishmania major**	-	-	X53915	AC098846
*Lepocinclis ovata**	SAG B1244-5	UTEX 1305	AF061338	This study
*Naegleria gruberi**	ATCC 30224	ATCC 30540	M18732	This study
*Naegleria fowleri**	Czech republic isolation	ATCC 30894	AF338423	This study
Perkinsus andrewsi	-	-	AY305326	AY305327
*Petalomonas cantuscygni**	CCAP1259/1	CCAP1259/1	U84731	This study
Phacus pusillus	UTEX LB 1282	UTEX LB 1282	AF190815	This study
Phacus pyrum	UTEX 2354	UTEX 2354	AF112874	This study
*Rhynchobodo sp.**	ATCC 50359	ATCC 50359	U67183	This study
*Rhynchopus sp.**	ATCC 50230	ATCC 50230	This study	This study
Strombomonas	SAG 1280-2	SAG 1280-1	AY015000	This study

*acuminata**				
Strombomonas costata	ACOI 2992	ACOI 2992	This study	This study
Strombomonas triquetra	S604	S604	This study	This study
Tetrahymena pyriformis	-	-	X56171	X54004
Trachelomonas echinata	SAG1283-22	SAG1283-22	AY015001	This study
Trachelomonas hisida var. crenulatocolis	SAG 1283-8	UTEX 539	AJ532442	This study
*Trachelomonas volvocina**	Korea isolation	SAG 1283-4	AF096995	This study
*Trypanoplasma borreli**	ATCC 50433	Tt-JH strain	L14840	This study
*Trypanosoma brucei**	-	-	AJ009141	AE017168
*Trypanosoma cruzi**	-	-	AF228685	L22334
*Trypanosoma grosi**	-	-	AB175624	AB175624
Trypanosoma sp. Pteromys	-	-	AB175626	AB175626
Trypanosoma otospermophili	-	-	AB175625	AB175625

*marked taxa are the taxa that were used in Sequence Dissimilarity Estimation.

Table **2.2.** Primers used for PCR amplification and sequencing of LSU rDNA sequences

Primer	Specifically amplified group	Direction	5' to 3' Sequencing	Position*
LSU_22F	All Euglenozoa	Forward	CCCRCYGAACTTAAGCATATYACTC	25
LSU_33F	All Euglenozoa	Forward	CTTAAGCATATYACTCAGYGGAGG	33
LSU_71F	Euglenids	Forward	CGAYKGYYIYAGTAAIGGCGA	72
LSU_520F	All Euglenozoa	Forward	CCGMYAGIGMASAAGTASWSYGA	440
LSU_535F	Euglenids	Forward	CCGATAGIGIACAAGTASIGTGA	440
LSU_540F	Euglenids	Forward	GAGTAGMRYKGYTTGGGAITG	367
LSU_625F	All Euglenozoa	Forward	RRRACCGATAKYRIACAAGTA	436
LSU_1560F	All Euglenozoa	Forward	AGRCYMATCGARCCAYCTAGTAGC	1322
LSU_2300F	All Euglenozoa	Forward	GCAGATCTTGGTKGTAGTAGCGA	2148
LSU_2900F	Euglenids	Forward	GAGYTYTCKTTTCMYCMTRATSCA	2574
LSU_3510R	All Euglenozoa	Forward	CRKCYAKTTTGCCGACTTCCCTKAG	2987
LSU_4180F	Euglenids	Forward	GTTTGACTCYAGTYTGRYWCTGTGC	3413
LSU_4718F	All Euglenozoa	Forward	GTYTCGARACATCKRYCAGWTGGGG	3808
LSU_5070F	All Euglenozoa	Forward	ATCCTTCGATGTCGGCTCTTCCTA	4103
DIP_1490F	Diplonemids	Forward	GAGAGCTGACTTAAGGTGTCCGAG	1148
DIP_3600F	Diplonemids	Forward	ACGCGGGGACAGGGTTAACTATC	1735
SSU_5650F	All Euglenozoa	Forward	AARRAATTGAYGGAAKGGCACCAC	in SSU rDNA
LSU_628R	All Euglenozoa	Reverse	TCACKSTACTTGTIYRMTATCG	441
LSU_1567R	All Euglenozoa	Reverse	ACTCCTTGGTCCGTGTTTCRAGAC	1032
LSU_2120R	*Eutreptiella*	Reverse	GAGTTGTTACACAYTCCTYAGCGG	1959
LSU_2160R	All Euglenozoa	Reverse	GCRYCATCCATTTTCGGRGCYG	1997
LSU_3883R	All Euglenozoa	Reverse	CAAACKCAACAGGGTCTTCTTTCC	3393
LSU_3885R	All Euglenozoa	Reverse	CAAACTCAACAGGGTCTTCTTTCC	3393
LSU_4060R	All Euglenozoa	Reverse	AAYGAGATTCCYKCTGTCCCKAGT	4202
LSU_4750R	All Euglenozoa	Reverse	GCCACAAGCCAGTTATCCCTGT	4050
LSU_5070R	All Euglenozoa	Reverse	TAGGAAGAGCCGACATCGAAGGA	4103
LSU_5465R	All Euglenozoa	Reverse	RRGRGTTCCTCTCGTACTACC	4398
LSU_5710R	All Euglenozoa	Reverse	GTAAAACCAACCTGTCTCACGACG	4200
DIP_2400R	Diplonemids	Reverse	TCGTCAGTGGCAGCTTTGAGGC	1545

* Relative position in *Euglena gracilis* LSU rDNA (X53361).

Table **2.3.a.** Positions and lengths of 14 conserved core regions (CC) in LSU rDNA sequence of *Euglena gracilis* (X53361)

Core Conserved Region	Positions in *Euglena gracilis*	Length	Start CC in *Euglena gracilis*	End CC in *Euglena gracilis*
CC1	18-117	100	GAATGACCCA	CCACGGCTCT
CC2	436-507	72	GACCGATAGT	GTCCCTGAAG
CC3	1028-1068	41	CGTCTCGAAA	AGCTTGCCAG
CC4	1235-1373	139	GGTGAAGCCA	TCAGGATAGC
CC5	1473-1489	17	CGAACTGTGA	TGAATGGGTG
CC6	1904-2016	113	AAAGGATGTT	TGGATGGTGC
CC7	2144-2166	23	GAGCAGATCT	TTGYAGTAGC
CC8	2874-2901	28	AAACTGGCCG	TGGAAGATAC
CC9	3106-3147	42	GGGGAATCTG	TGAAACCCAG
CC10	3182-3198	17	TGACTTCTGC	TGCCCAGTGC
CC11	3226-3465	240	CCAAGCGCGG	TAGGTGGGAG
CC12	3822-3929	108	CAGTTGGGGA	TGGCCAAAGG
CC13	4038-4089	52	AAAATTACCA	CAAAGCGACG
CC14	4105-4211	107	TCGATGTCGG	TGAGACAGGT

*****(a) The listed 10 bases motifs immediately start or close each core conserved regions in the Euglena gracilis LSU rDNA sequence;

Table **2.3.b.** Positions and lengths of 14 conserved core regions (CC) in aligned Euglenozoa LSU rDNA sequence

Core Conserved Region	Positions in alignment	Length	Start CC in alignment	End CC in alignment
CC1	289-406	118	GAAYKWMMSM	CCARKGCTCR
CC2	809-909	101	RACCGATAGY	GTCCCTGAAG
CC3	2013-2055	43	CGTCTYGAAA	AGCTTGCMRG
CC4	2453-2594	142	RRYGAAGCCR	TCAGGATAGC
CC5	2730-2746	17	CGAACTGTGA	TGAATGGGYR
CC6	3434-3554	121	AAAGGATGTY	TGGATGGWGC
CC7	3748-3770	23	GNGCAGATCT	TTGCAGTAGC
CC8	5117-5144	28	AAACTRGCYR	TGGAAGATRY
CC9	5564-5593	30	GGGAATCTGA	TGAAACMCAG
CC10	5643-5660	18	TGACTTCTGC	TGCCCAGTGC
CC11	5703-5945	243	YCAAGCGCGG	YAGGTGGGAG
CC12	6543-6652	110	CAGTYKGGGA	AGGCCAAAGG
CC13	6788-6839	52	AAATTACCAC	CAAAGCGACG
CC14	6856-6964	109	TCGATGTCGG	TGAGACAGGT

*****The listed 10 bases consensus motifs immediately start or close each core conserved regions in the Euglenozoa LSU rDNA sequence alignment.

Table **2.4.a.** Comparison of dissimilarity (a) within LSU rDNA sequences, and (b) between LSU rDNA and SSU rDNA sequences under the model of TrN+I+G

*ES stands for expansion segment and CC strands for conserved core regions; I, II, III, IV stand for the classifications of expansion segments.

Taxonomic Unit	(a) within LSU rDNA sequences				(b) between LSU and SSU rDNA sequences			
	ESI-III/ CC	ESIV/ CC	LSU/ CC	ESIV/ LSU	ESI-III/ SSU	ESIV/ SSU	LSU/ SSU	SSU/ CC
Euglenids	5.3531	24.6810	8.9154	2.7683	1.1095	4.1556	1.5011	5.9392
Trypanosomatids	12.2692	90.4904	31.3846	2.8833	0.9577	7.7013	2.6710	11.7500
Core Bodonids	3.1837	11.8911	4.2228	2.8159	1.1058	3.6767	1.3057	3.2341
Diplonemids	5.3877	17.0399	5.8333	2.9211	0.9328	3.3908	1.1608	5.0254

*The selected model of sequence evolution is TrN+I+G (Tamura and Nei, 1993) selected in Modeltest 3.6.

Table **2.4.b.** Comparison of Dissimilarity Between *Petalomonas cantuscygni* and Representatives from Euglenids, Bodonids, Trypanosomids, and Diplonemids of Euglenozoa

Taxonomic Unit	Taxon	CC	ES I-III	ES IV	LSU rDNA	SSU rDNA
Euglenids	*Euglena gracilis*	0.3498	1.2240	1.2612	1.0079	0.7464
	Lepocinclis ovata	0.3373	1.0263	1.2147	0.0920	0.7102
	Strombomonas acuminate	0.3330	1.0385	0.8403	1.1364	0.7069
	Trachelomonas volvocina	0.3243	1.1713	1.1341	0.9456	0.7776
	Eutreptiella marina	0.3178	1.0639	1.7100	0.9924	0.7131
Diplonemids	*Diplonema papillatum*	0.1901	**0.6810**	**0.7584**	0.5879	0.6038
	Rhynchopus sp.	**0.1874**	0.7030	0.7841	**0.5769**	**0.5721**
Bodonids	*Bodo saliens*	0.2333	1.1241	1.1101	0.8042	0.7761
	Bodo uncinatus	0.2492	1.0609	1.1961	0.8135	0.7304
	Cryptobia helices	0.2790	1.0709	1.3462	0.7761	0.7703
	Rhynchobodo sp.	0.2356	1.2175	1.0687	0.8299	0.8211
	Trypanoplasma borreli	0.2748	1.1506	1.1585	0.8800	0.7274
Trypanosomatids	*Trypanosoma brucei*	0.2361	1.0424	1.1396	0.7983	0.8253
	Trypanosoma cruzi	0.2375	1.0664	1.1978	0.8250	0.7916
	Trypanosoma grosi	0.2354	1.0608	1.0086	0.7689	0.7900
	Crithidia fasciculate	0.2282	1.1174	1.0747	0.7984	0.7473
	Leishmania major	0.2261	1.0945	1.1285	0.8052	0.7856
Percolozoa	*Naegleria gruberi*	0.3053	0.9746	1.0755	0.7980	1.0392
	Naegleria fowleri	0.3139	0.9941	0.9953	0.7911	1.0800

* The underlined bold numbers are the smallest number of pairwise distance in every column.

Table **2.5.** Phylogentic informativeness of Euglenozoa LSU rDNA sequences

Data Sets	g1	B_{50}	Alpha	Pi	RC	GC %	N. islands	N. phyl. info. chara.	N. chara
ES I-III	-0.415261	72.97%	1.223	0.073	0.314	49.47%	7	1394	2139
ES IV	-0.421572	67.57%	3.385	0.001	0.204	54.86%	14	2534	4896
CC Regions	-0.414345	67.57%	0.636	0.244	0.660	49.60%	18	377	1152
LSU rDNA	-0.562530	83.78%	0.876	0.097	0.278	52.85%	3	4305	8187
SSU rDNA	-0.569799	83.78%	0.887	0.081	0.333	50.67%	4	2066	3392
Combined DM	-0.444915	91.89%	0.910	0.094	0.295	52.16%	2	6371	11579

*g1: value resulting from skewness test; alpha: the shape parameter of the gamma distribution; Pi: the proportion of invariant sites; RC: rescaled consistency index.
*N. island indicates number of island where most parsimonious trees hit using 100 Random Taxa Addition (RTA); N. phyl. info. chara. indicates number of phylogenetically informative characters; N. chara. indicates the total number of characters.

Figure **2.1.** Location of primers used in PCR amplification and sequencing of LSU rDNA. This figure was modified and linearized from the circular extrachromosomal rDNA of *Euglena gracilis* (Fig 1 in Spencer J. G et al. 2001; in which the filled boxes indicated the positions of conserved core regions for mature rRNA components and named "LSU 1-10"). Relative locations of the first 10 "LSU" in the original paper were marked as "CC" in this figure. Below this figure are relative positions of 25 primers used to sequence LSU rDNA. 22F, 2300F, 3510R and 5460R were also used in PCR amplification (See table 2 for primer sequences and their exact locations).

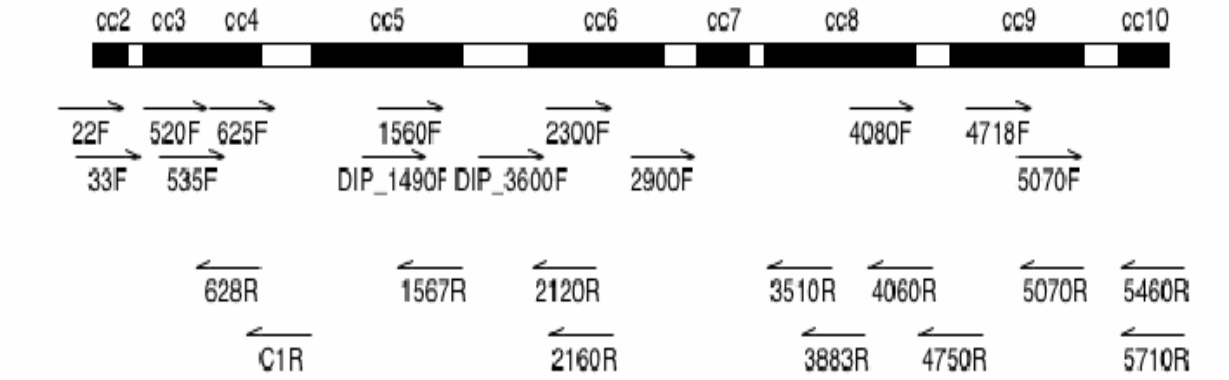


Figure **2.2.** (2a) is the illustration of the positions and lengths of 14 conserved core regions in Euglenozoa LSU rDNA sequence alignment (marked as CC 1-14). (2b) shows of the different classes of expansion segments (ES) and core conserved regions (CC) in the alignment.

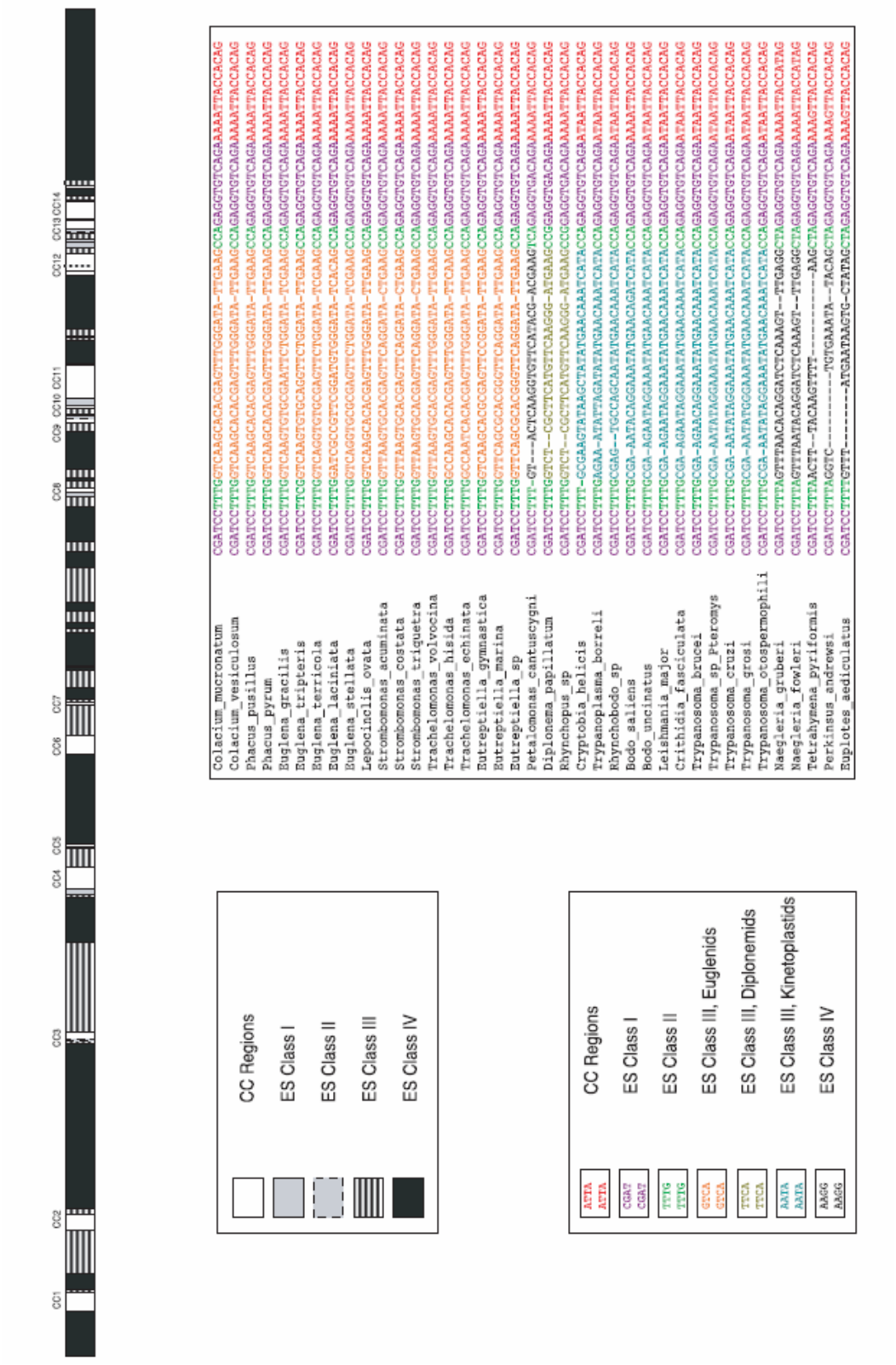


Figure **2.3.** Site variation across LSU rDNA sequence. Above is an illustration of the relative locations of the 14 core conserved regions (CC1-14) of LSU rDNA. Variation in base substitution rates over the length of LSU rDNA for the 39 sequences of LSU rDNA analyzed using in this study were calculated using a window size of twelve consecutive bases.

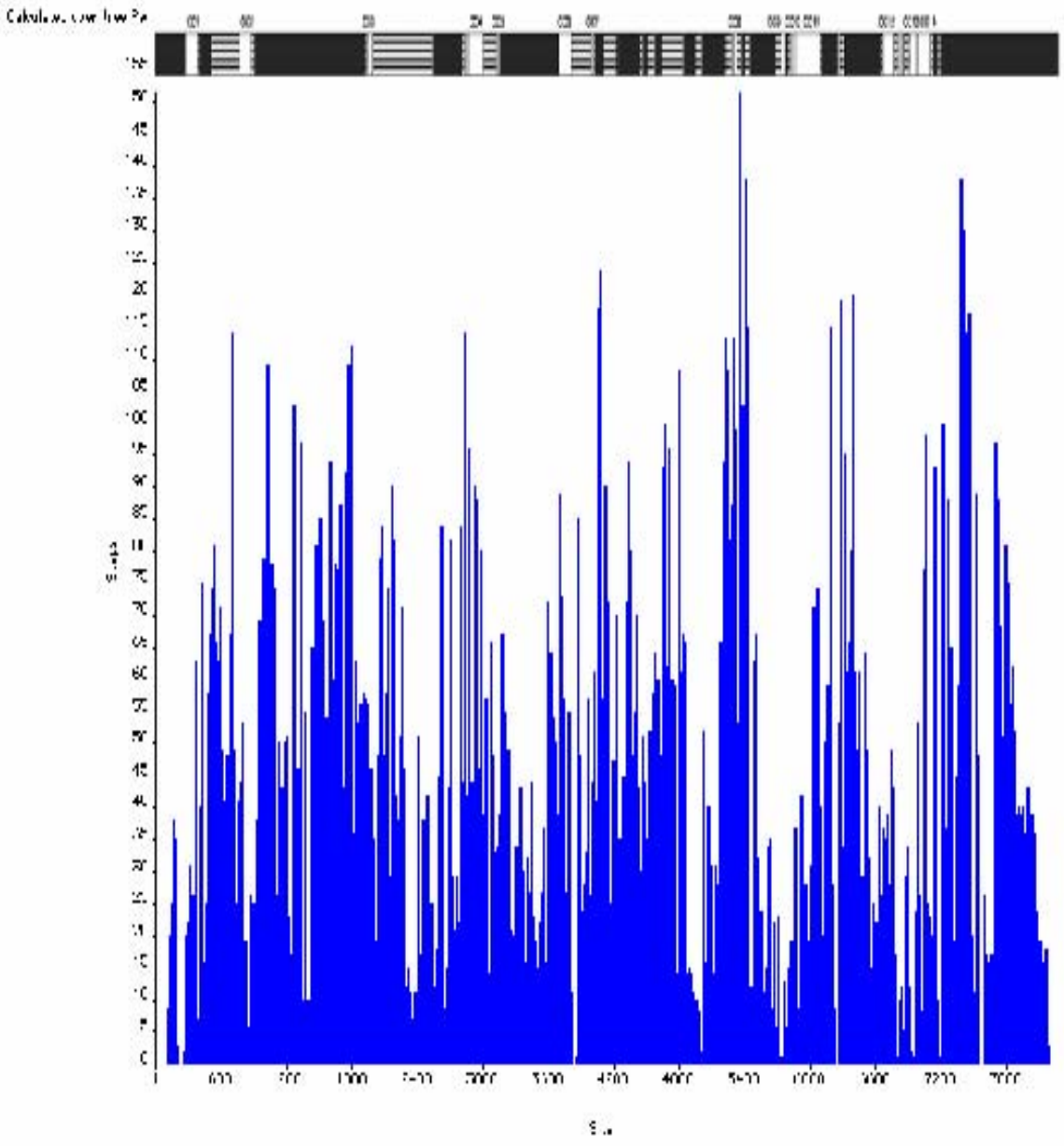


Figure **2.4.a.** Maximum likelihood tree (-LnL=140977.15) of 39 Euglenozoa and outgroup taxa using combined SSU and LSU rDNA sequences (11579 bases). The number at each node is bootstrap support based 1000 replicates using maximum-likelihood corrected neighbour-joining search (before slash), and 100 heuristic research based on parsimony with substitutions weighted according to the instantaneous rate matrix (after slash). Bootstrap supports less than 50% were not shown. The text corresponds to a certain node as indicated by straight lines. Filled right arrows indicate the bootstrap supports for the three subgroups of Euglenozoa as Euglenids, Diplonemids and Kinetoplastids to be monoplyletic clades. Filled arrowhead points to the bootstrap support for that *Petalomonas cantuscygni* and Diplonemids are the closest relatives.

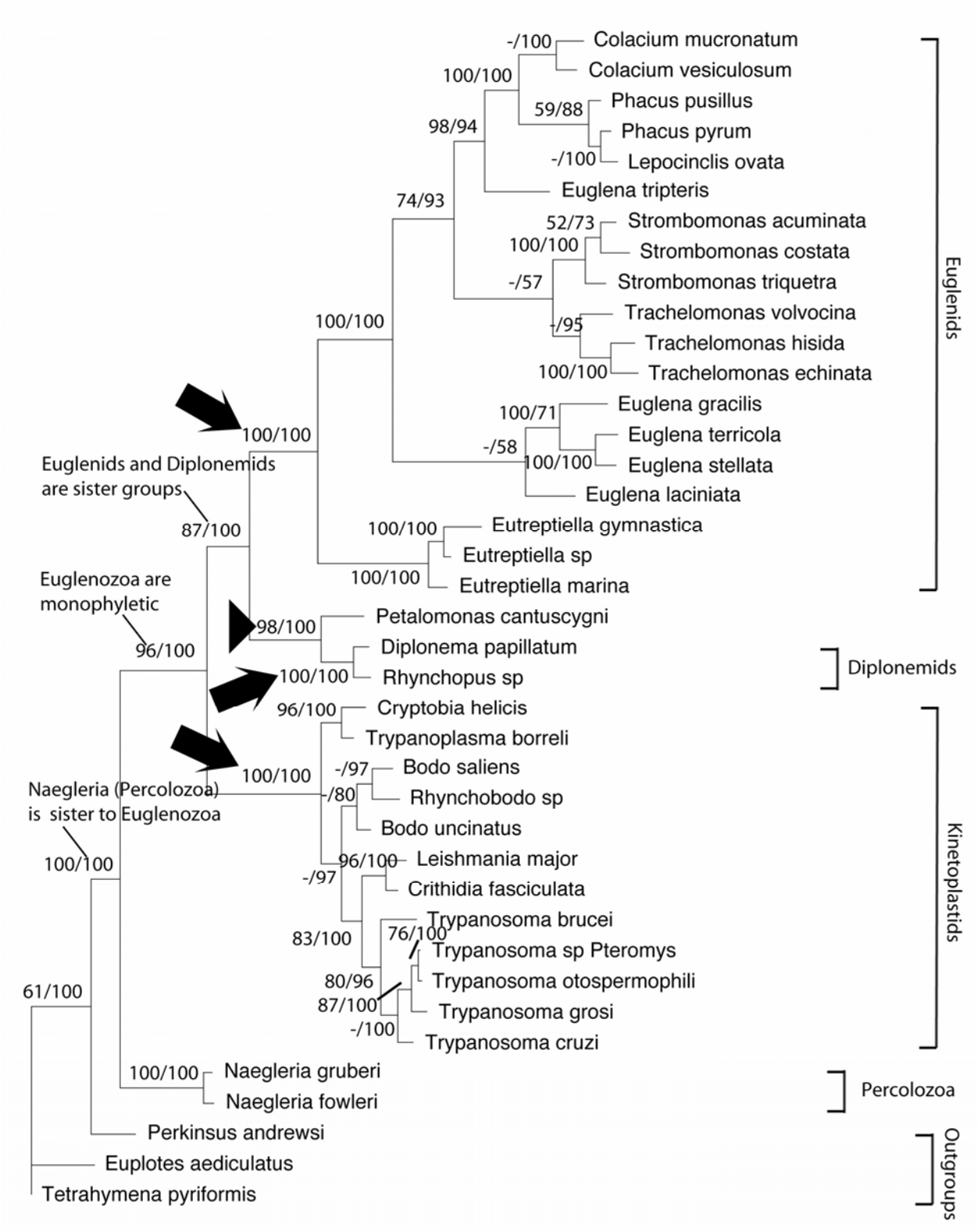

-/100
Colacium mucronatum
Colacium vesiculosum
100/100
59/88
Phacus pusillus
Phacus pyrum
98/94
-/100
Lepocinclis ovata
Euglena tripteris
74/93
52/73
Strombomonas acuminata
100/100
Strombomonas costata
-/57
Strombomonas triquetra
Trachelomonas volvocina
100/100
-/95
Trachelomonas hisida
100/100
Trachelomonas echinata
Euglenids
100/71
Euglena gracilis
100/100
Euglenids and Diplonemids
are sister groups
-/58
Euglena terricola
100/100
Euglena stellata
Euglena laciniata
87/100
100/100
Eutreptiella gymnastica
Eutreptiella sp
Euglenozoa are
monophyletic
100/100
Eutreptiella marina
98/100
Petalomonas cantuscygni
96/100
Diplonema papillatum
Diplonemids
100/100
Rhynchopus sp
96/100
Cryptobia helicis
Trypanoplasma borreli
100/100
-/97
Bodo saliens
Naegleria (Percolozoa)
is sister to Euglenozoa
-/80
Rhynchobodo sp
Bodo uncinatus
100/100
96/100
Leishmania major
-/97
Crithidia fasciculata
Trypanosoma brucei
83/100
76/100
Trypanosoma sp Pteromys
Kinetoplastids
80/96
Trypanosoma otospermophili
61/100
87/100
Trypanosoma grosi
-/100
Trypanosoma cruzi
100/100
Naegleria gruberi
Percolozoa
Naegleria fowleri
Perkinsus andrewsi
Outgroups
Euplotes aediculatus
Tetrahymena pyriformis

Figure **2.4.b.** Maximum likelihood tree (-LnL=43901.783) of 39 Euglenozoa and outgroup taxa using SSU rDNA sequences (3392 bases). The number at each node is bootstrap support based 1000 replicates using maximum-likelihood corrected neighbour-joining search (before slash), and 100 heuristic research based on parsimony with substitutions weighted according to the instantaneous rate matrix (after slash). Bootstrap supports less than 50% were not shown. The text corresponds to a certain node as indicated by straight lines. Filled right arrows indicate the bootstrap supports for the three subgroups of Euglenozoa as Euglenids, Diplonemids and Kinetoplastids to be monoplyletic clades. Filled arrowhead points to the bootstrap support for that *Petalomonas cantuscygni* and Diplonemids are the closest relatives.

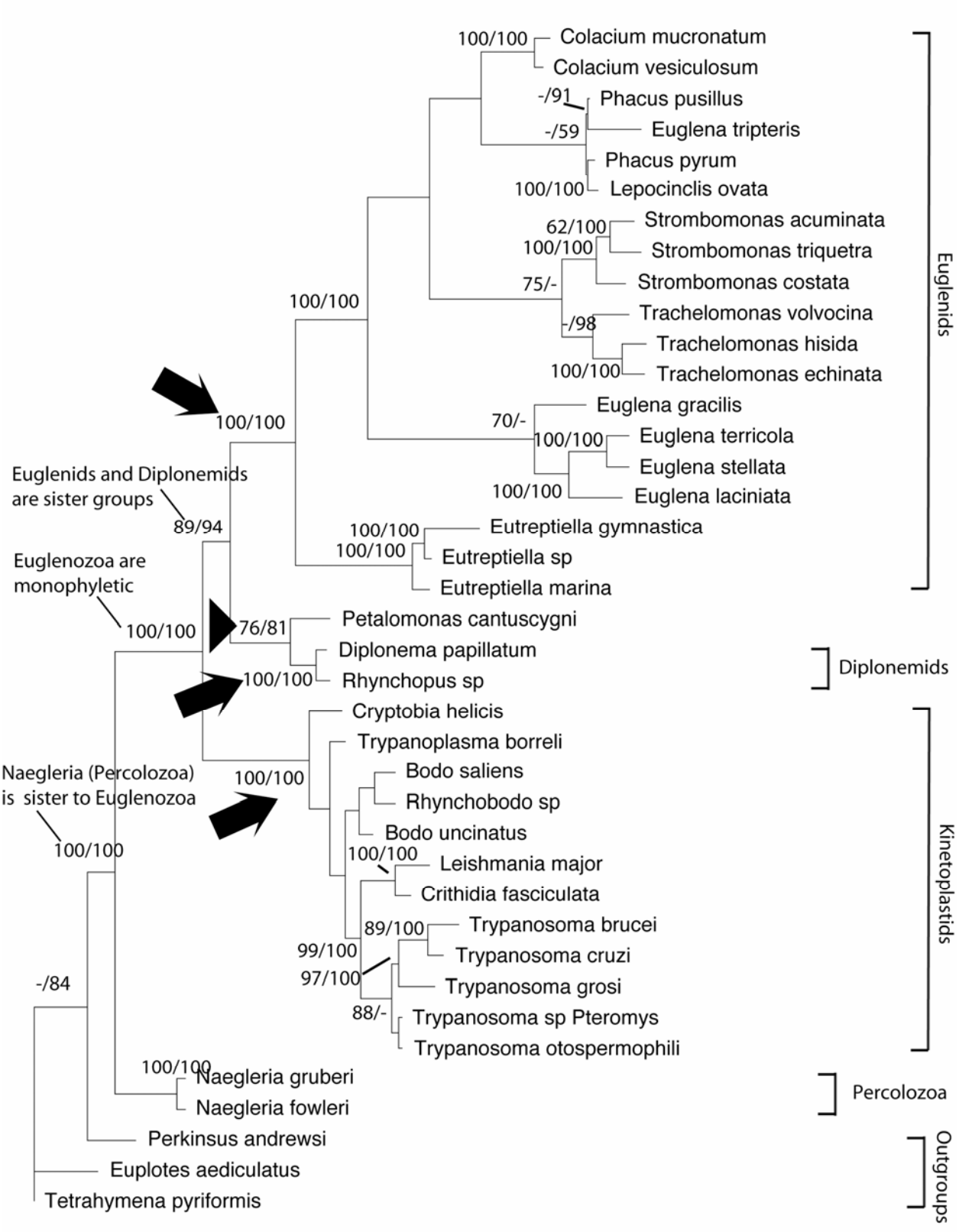

100/100
Colacium mucronatum
Colacium vesiculosum
-/91
Phacus pusillus
-/59
Euglena tripteris
Phacus pyrum
100/100
Lepocinclis ovata
62/100
Strombomonas acuminata
100/100
Strombomonas triquetra
75/-
Strombomonas costata
100/100
Trachelomonas volvocina
-/98
Trachelomonas hisida
100/100
Trachelomonas echinata
Euglena gracilis
70/-
100/100
Euglena terricola
Euglena stellata
100/100
Euglena laciniata
Euglenids
100/100
Euglenids and Diplonemids
are sister groups
89/94
100/100
Eutreptiella gymnastica
100/100
Eutreptiella sp
Eutreptiella marina
Euglenozoa are
monophyletic
100/100
76/81
Petalomonas cantuscygni
Diplonema papillatum
100/100
Rhynchopus sp
Diplonemids
Cryptobia helicis
Trypanoplasma borreli
Naegleria (Percolozoa)
is sister to Euglenozoa
100/100
Bodo saliens
Rhynchobodo sp
Bodo uncinatus
100/100
Leishmania major
Crithidia fasciculata
89/100
Trypanosoma brucei
99/100
Trypanosoma cruzi
97/100
Trypanosoma grosi
88/-
Trypanosoma sp Pteromys
Trypanosoma otospermophili
Kinetoplastids
100/100
Naegleria gruberi
Naegleria fowleri
Percolozoa
-/84
Perkinsus andrewsi
Euplotes aediculatus
Tetrahymena pyriformis
Outgroups

Figure **2.5.a.** Bayesian phylogeny of 39 Euglenozoa and outgroup taxa using combined SSU and LSU rDNA sequences. Three data sets were defined as core conserved regions in both SSU and LSU rDNA, expansion segments I-III of LSU rDNA and expansion segment IV of LSU rDNA and SSU rDNA in MrBayes block. Each run contained 4 chains starting from random trees, and 500,000 generations with trees sampled at every one hundred generations. The majority rule consensus tree was calculated after removal of first 200 tree which were saved during burn-in period. The number at each node is posterior probabilities Bayesian produced. The text corresponds to a certain node as indicated by straight lines. Filled right arrows indicate the bootstrap supports for the three subgroups of Euglenozoa as Euglenids, Diplonemids and Kinetoplastids to be monoplyletic clades. Filled arrowhead points to the bootstrap support for that *Petalomonas cantuscygni* and Diplonemids are the closest relatives.

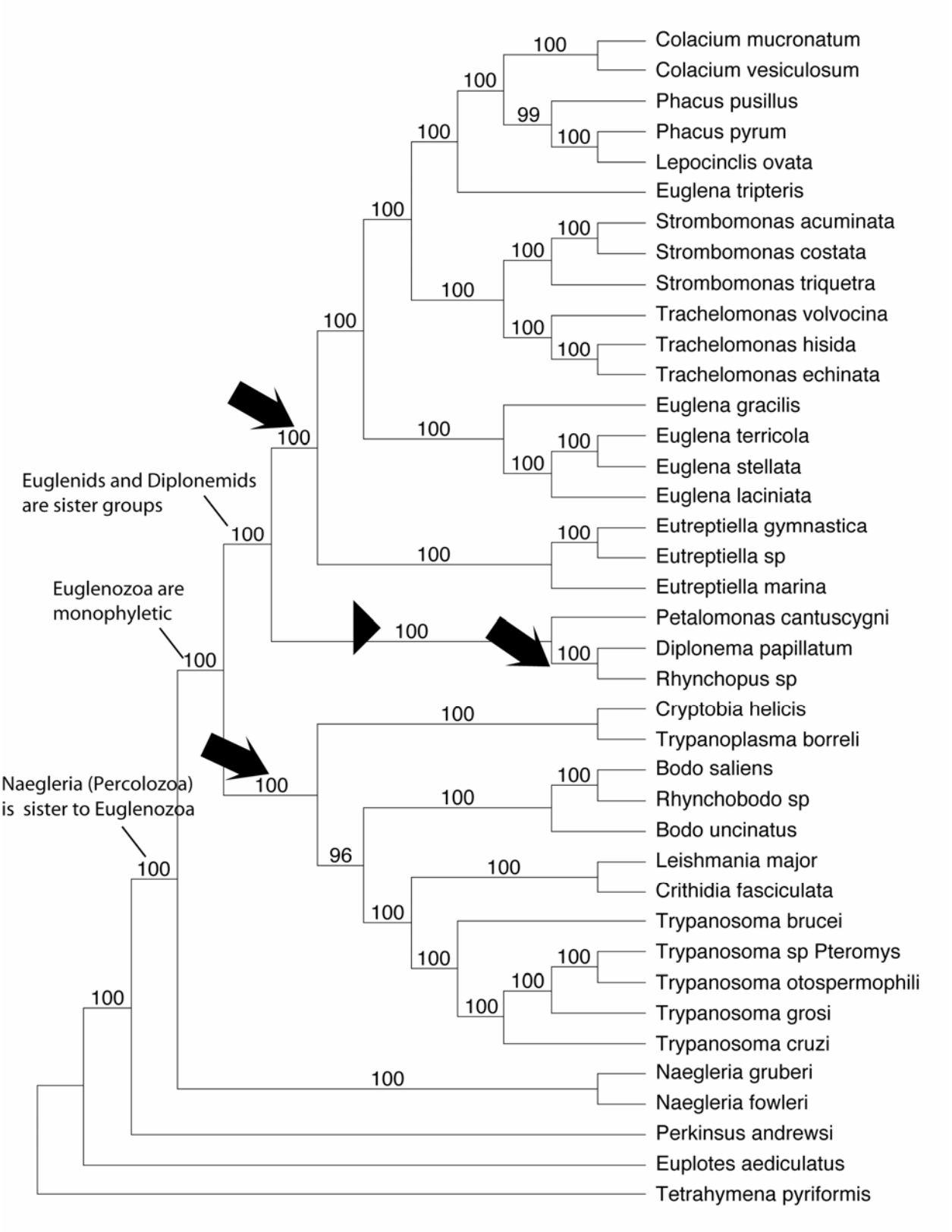
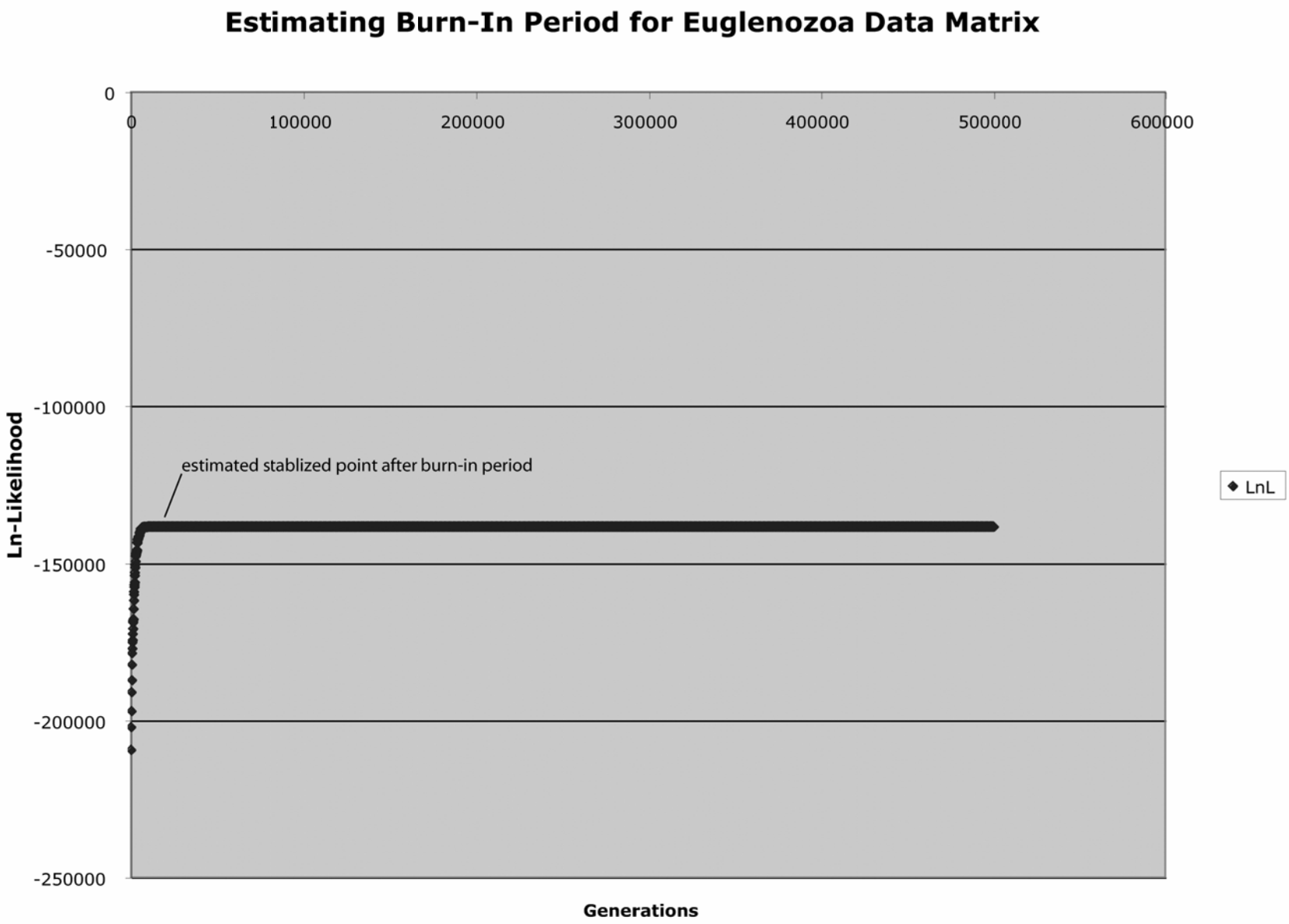

Colacium mucronatum
Colacium vesiculosum
Phacus pusillus
Phacus pyrum
Lepocinclis ovata
Euglena tripteris
Strombomonas acuminata
Strombomonas costata
Strombomonas triquetra
Trachelomonas volvocina
Trachelomonas hisida
Trachelomonas echinata
Euglena gracilis
Euglena terricola
Euglena stellata
Euglena laciniata
Eutreptiella gymnastica
Eutreptiella sp
Eutreptiella marina
Petalomonas cantuscygni
Diplonema papillatum
Rhynchopus sp
Cryptobia helicis
Trypanoplasma borreli
Bodo saliens
Rhynchobodo sp
Bodo uncinatus
Leishmania major
Crithidia fasciculata
Trypanosoma brucei
Trypanosoma sp Pteromys
Trypanosoma otospermophili
Trypanosoma grosi
Trypanosoma cruzi
Naegleria gruberi
Naegleria fowleri
Perkinsus andrewsi
Euplotes aediculatus
Tetrahymena pyriformis
Euglenids and Diplonemids are sister groups
Euglenozoa are monophyletic
Naegleria (Percolozoa) is sister to Euglenozoa

Figure **2.5.b.** Estimating of burn-in period of combined SSU and LSU rDNA data matrix. Number of generation (x-axis) and Ln-likelihood number (y-axis) are plotted against each other, and the point indicated by straight line is the estimated stabilized point (around 20,000 generations). The generations after that point are what we use to generate the majority consensus tree.


Estimating Burn-In Period for Euglenozoa Data Matrix
0
0
100000
200000
300000
400000
500000
600000
-50000
-100000
-150000
-200000
-250000
Ln-Likelihood
estimated stablized point after burn-in period
Generations
LnL

CHAPTER 3

Surface morphological structures of *Petalomonas cantuscygni*[1]

[1] Bing Ma, Mark A. Farmer. To be submitted to *Journal of Phycology*.

ABSTRACT

Petalomonas cantuscygni is a small, colorless flagellates found in marine environments. Many of its morphological features suggest that it may resemble the ancestral form of the Euglenozoa. Previous studies on *Petalomonas cantuscygni* focused on its morphological structures and have used SSU rDNA sequence data to infer its phylogenetic position. However, phylogenies based on morphological data conflicted with the molecular phylogenetic analyses, which makes the position of *Petalomonas cantuscygni* even more complicated. In this study, we focused on *Petalomonas cantuscygni* from its basic shape to its pellicle strip morphology and ontogeny. Here we present morphological data of pellicle strip development in *Petalomonas cantuscygni* and provide ideas for future study to infer its evolutionary position in Euglenozoa.

INTRODUCTION

The Euglenozoa is a natural, monophyletic assemblage of early branching Eukaryotes. It is composed of around 1,600 species; most of which are unicellular but some are colonial. Three distinct groups of flagellates: Kinetoplastids, Euglenids and Diplonemids share a number of morphological features and form the Euglenozoa (Cavalier-Smith 1981; Simpson 1997). Euglenids (Butschli 1884) are distinguished by the presence of a pellicle that is organized by supporting microtubules underneath the cell membrane. An additional glycoprotein layer is appressed to the inner surface of the plasma membrane forming the ribbon-shaped pellicle strips, which extend the length of the cell and in species such as *Euglena gracilis* facilitate a variety of unique cell shapes ranging from rigid to flexible. Most Euglenids are larger than 30µm in length although a few species may exceed 500µm. Euglenids are one of the best-known groups of

photosynthetic flagellates, but many Euglenids species are primitively or secondarily colorless. They are commonly found in freshwater, while there are some marine species such as *Eutreptiella*.

Kinetoplastids (Honigberg 1963) are distinguished from other protists by the presence of a large accumulation of mitochondrial DNA called the *kinetoplast*, which is found in close association with flagellar bases. Like Euglenids, the cytoskeletal microtubules of Kinetoplastids underlie the plasma membrane and form a continuous or discontinuous supporting corset. Unlike the Euglenids, Kinetoplastids do not have an additional glycoprotein layer. Most Kinetoplastids are smaller than Euglenids and many species are less than 30μm in length. Kinetoplastids include a number of parasites responsible for diseases in humans (i.e. sleeping sickness, Chagas' disease), animals and plants, while others are free-living organisms (e.g., Bodonids).

In addition to the Kinetoplastids and Euglenids, there are several other taxa that belong to the Euglenozoa. One of these is the Diplonemids (Cavalier-Smith 1993). They also have a corset of supporting cytoskeletal microtubules under the plasmamembrane, but unlike the Euglenids these microtubules are not organized into pellicle strips. Diplonemids are also smaller than Euglenids with length of less than 30μm. Most Diplonemids are thought to be free-living organisms, but recently have been proposed to be a group of pathogenic organisms that may target a number of hosts (von der Heyden, Chao et al. 2004). Outside these three subgroups are a number of protists that are within the Euglenozoa (i.e. *Postgaardi* and *Petalomonas cantuscygni*), but their phylogenetic affinities to one of the three well-defined major groups are unclear.

Petalomonas cantuscygni is a small, colorless flagellates found in marine environments, and its lots of morphological features that suggest that it may resemble the ancestral form of Euglenozoa. It has a single emergent flagellum, relatively few longitudinally

arranged pellicular strips, mitochondrial inclusions that resemble a kinetoplast and it is osmotrophic and/or phagotrophic which is considered to be the state of the ancestral form of Euglenozoa (Cann and Pennick 1986). It does not have plastids, an eyespot or a paraflagellar swelling, which are believed to be derived features associated with the photosynthetic Euglenids. *Petalomonas cantuscygni* has a Type I feeding apparatus supported by a few microtubules (Triemer and Farmer 1991). There are a total of four types of Euglenid feeding apparatus: Type I (MTR/pocket), Type II (plicate type), type III (short extensive type) and type IV (siphon type) (Triemer and Farmer 1991). The type I feeding apparatus consists of a cytoplasmic pocket that arises adjacent to the flagellar opening and extends towards the posterior of the cell. It is believed to be the simplest and most ancestral feeding type among the four (Triemer and Fritz 1986; Farmer and Triemer 1988; Triemer and Ott 1990; Triemer and Farmer 1991). Besides the Type I feeding apparatus, *Petalomonas cantuscygni* also has two basal bodies with three asymmetrically distributed microtubular rootlets and a mitotic spindle, which forms within a closed nuclear envelope (Triemer and Farmer 1991). In this way *Petalomonas cantuscygni* is similar in morphology to other genera within the major subgroups of Euglenozoa.

Previous studies on *Petalomonas cantuscygni* focused on its morphological structures and have used SSU rDNA sequence data to infer its phylogenetic position. However, the morphological data conflict with the molecular phylogenetic analyses, which makes the position of *Petalomonas cantuscygni* even more complicated. Homologous structures in flagellar apparatus, cytoskeleton, feeding apparatus and mitotic apparatus show that the Euglenids, even the photosynthetic ones, and Kinetoplastids share a common ancestor and are more closely related to each other than either is to other eukaryotes (Triemer and Farmer, 1991). Interestingly, *Petalomonas cantuscygni* has a euglenid-like pellicle composed of longitudinal strips, and a

number of mitochondrial inclusions near the flagellar apparatus that resemble a kinetoplast. In many ways *Petalomonas cantuscygni* bridges the gap between the Euglenids and Kinetoplastids and can be thought of as a "missing-link" between the two groups (Triemer and Farmer 1991; Leander, Triemer et al. 2001). However, the molecular systematic studies suggest that Euglenids and Diplonemids are more closely related groups than either is to the Kinetoplastids (Busse and Preisfeld 2003; von der Heyden, Chao et al. 2004). This contradicts with the idea that *Petalomonas cantuscygni* is a "missing-link" between Euglenids and Kinetoplastids.

The question remains as to whether *Petalomonas cantuscygni* is representative of the ancestral form of the Euglenozoa? Obviously, there is no absolute answer to these questions, and more advanced ultrastructural and molecular data were needed to address this question. This study was carried out to uncover the special characters of *Petalomonas cantuscygni* in an effort to identify its evolutionry position within this group and help us to better understand the origins of the Euglenozoa. In this study, we focused on *Petalomonas cantuscygni* from its basic shapes to pellicle strip morphology and ontogeny. Here we present morphological data of pellicle strip development in *Petalomonas cantuscygni* and provide more references for future study to infer its evolutionary position in Euglenozoa.

MATERIALS AND METHODS

Strain and culture condition

The strain of *Petalomonas cantuscygni* used in this study is CCAP1259/1 (CCAP: Culture Collection of Algae and Protozoa, Scottish Association for Marine Science, Oban, Scotland). Cells were grown in ESNW medium (recipe below). Cultures were maintained at 20˚C incubator on a 12:12 light dark cycle.

ESNW medium (Enriched Soil extract Natural Seawater medium) is a variation of natural seawater medium to which soil extract (1:250) and ES vitamins (1:250) (Harrison, Waters et al. 1980) are added. The recipe for soil extract is as follows: to 1L distilled water add: 50g Garden soil, 0.2g $NH_4MgPO_4 \cdot 6H_2O$, 0.2g $CaCO_3$, 0.2g crushed barley, and 10 pieces of dry split peas. Heat to 70 ˚C and maintain for 5 hours, remove from heart and cover with cheesecloth. Let it stand for 2 days at room temperature, decant the supernatant, filter through a 0.2 μm filter, autoclave for 30 min, and cool down for use. The recipe for ES vitamin solution is: Thiamine 0.1 $g \cdot L^{-1}$, Cyanocobalamin 2 $mg \cdot L^{-1}$, Biotin 1 $mg \cdot L^{-1}$.

Scanning Electron Microscope (SEM) Sample Preparation

Fixation: Around 10ml of liquid culture medium containing log-phase cells was collected and transferred into a small Petri dish that contained a piece of filter paper which saturated with 4% OsO4 (0.1M Cacodylate buffer), mounted on the inner surface of the lid. The lid was placed over the chamber and the cells were fixed by OsO4 vapors for 30 min at room temperature. Four to five drops of 4% OsO4 (0.1M Cacodylate buffer) were added directly into the liquid medium and the cells were fixed for another 30 min.

Dehydration: The cells were transferred onto 8 μm polycarbonate membrane filters (Corning Incorporated Separations Division, Acton, MA). Cells were dehydrated with a graded series of ethyl alcohol at the concentration of 30%, 50%, 70%, 85%, 95% and 100% for 15 min each.

CPD: After cells were dehydrated using 100% ethyl alcohol for another two times, samples were ready for Critical Point Drying (CPD). Filters containing cells were transferred to the CPD (Samdri 780A Critical Point Dryer, Tousimis, Rockville, MD) and dried according to manufacturer's instructions using CO_2.

Mounting and Coating: Filters were mounted on aluminum stubs with carbon adhesive tabs. The SPI sputter coater (SPI Module Sputter Coater, Structure Probe Incorporated, West Chester, PA) for 60 seconds (~153Å) was used for gold coating according to manufacturer's instructions.

Observations: The cells were viewed using a LEO 982 Scanning Electron Microscope (LEO Electron Microscopy Incorporated, Thornwood, NY). Samples were kept in a desiccator.

In situ SEM sample preparation

Log-phase cells in culture flasks were selected and fixed *in situ.* Excess medium was removed and only a thin layer medium was left to cover the surface of the culture flask to prevent cells from prematurely drying out. A piece of cotton with 1%OsO4 (in Cacodylate buffer) was placed in the neck of culture flask, and vapor fixation progressed for 1 hour. Following vapor fixation a drop of 1% OsO4 (in 0.1M Cacodylate buffer and ESNW culture medium) was added for each 1 ml of medium and allowed to fix for another 30 min. Fixed cells were rinsed in ESNW and then dehydrated with graded ethanol. The cultured flasks were broken open and individual pieces with attached cells ere dried following the procedures previously described. After drying the samples containing attached cells were mounted, coated, and observed under SEM.

RESULTS AND DISCUSSIONS

Morphological development of Petalomonas cantuscygni

Petalomonas cantuscygni is a fusiform or flattened cell, with a length of 9 to 15 μm and width at 8 to 10 μm. Observation of 176 SEM images of *Petalomonas cantuscygni,* showed that more than 50% the cells were pear-shaped (Fig 1A and Fig 1B) and this represents the most

common form of *Petalomonas cantuscygni* in culture. Living cells were observed to move forward along the surface of the culture flask with their emergent flagellum positioned in front of the cell, and the cell body angled up from the surface. Approximately 30% of the observed cells were flattened dorsoventrally and some of the cells could be very flat or leaf-shaped (Fig 1C), and a few others were fusiform or less flattened as in Fig 1D.

From the range of different shapes and sizes of *Petalomonas cantuscygni*, we assume that cell morphology and age are correlated with each other. It appears that fusiform cells are the result of a recent cell division, pear-shaped cells are typically interphase cells and cells become more broadly flattened as they near mitosis. Mature cells divide longitudinally to form two fusiform cells each of which then starts a new cell cycle. The evidence in support of this progression is as follows: first, there are primarily three cell shapes in culture (pear-shaped, flattened and fusiform) and these occur in a certain ratio as mentioned above. Second, from the size of each cell shape, the flattened cells are roughly twice as large as the fusiform cells, and the size of pear-shaped sized cells are in between. Third, the SEM images of cells division show that the very flattened cells are longitudinally divided evenly from the anterior end (shown in Fig 3). Finally, the likelihood of fusiform cells being found side by side was more than 60%, and the chance of finding side by side fusiform cells increased to 80% in the preparation of cells fixed *in situ*. In addition, the appearance and size of these two fusiform cells are very similar, while other cells all have very distinguishable look even when they are in similar shape.

Surface morphological structures of Petalomonas cantuscygni

SEM images of the surface features of *Petalomonas cantuscygni* can be seen in figure 2. The most prominent cell surface feature is the pellicle strips, each of which was assigned a number in a clockwise order for reference as marked in Fig 2A. Fig 2B shows *Petalomonas*

cantuscygni with a single emergent leading flagellum. Fig 2C shows the feeding pocket which is discontinuously formed at the anterior end of cell. We also demonstrate the maximum number of pellicle strips (P value) in fig 2D.

The euglenid pellicle makes up the cortex of the cell and is composed of four components: the plasma membrane, proteinaceous strips arranged in parallel, subtending microtubules and tubular cisternae of endoplasmic reticulum (Leander, Triemer et al. 2001; Leander, Witek et al. 2001). The most obvious components of the pellicle are the pellicle strips, which are long, ribbon-shaped structures typically articulating with adjacent strips along their lateral margins to form either a groove or a ridge in different directions (Leander, Triemer et al. 2001). Pellicle strips may extend either helically or longitidinally from the flagellar opening to the cell's posterior end (Triemer and Farmer 1991; Leander and Farmer 2000). The pellicle of different species may vary from rigid to flexible depending on the way the pellicle strips are arranged. In the case of rigid shape, cells can not change their shapes; but in the case of being flexible, the pellicle strips allow the cell to perform a contorting motion called *metaboly*. The P value for *Petalomonas cantuscygni* is eight from our forty-nine SEM images of the posterior view. We assigned an identifying the number to pellicle strips starting at the point of the "disruption" of anterior feeding pocket, and count the number of pellicle strips in a clockwise order. As Leander and co-workers described (Leander, Witek et al. 2001), P value is the value of maximum number of pellicle strips surrouding the periphery of the cell. It reflects an evolutionary mode within variation: as a member of the colorless euglenids with a longitidinal pellicles, *P. cantuscygni* possesses between eight and twelve strips, a number which is believed to be closer to that found in the ancestral euglenid form. There appears to be a tendency to increase strip number during evolutionary process. The appearance of a helical pellicle may be

associated with the strip-doubling event that lead to the value of pellicle strips changing to above eighteen; and the second strip-doubling events that lead to the value of pellicle strips up to about fifty. This permanent strip doubling event is inferred to be a consequence of an ancestral cell that duplicated its cytoskeleton without undergoing cytokinesis.

There is posterior strip reduction found in *Petalomonas cantuscygni*. The pellicle strips designated as number seven terminated prior to converging with other remaining pellicle strips at the posterior end of flattened cells (Fig 2a). Pellicle strip number seven is considered to be the lagging strip. However, even the lagging strip was found to posteriorly terminate with the others in flat and pear-shape cells (above 95%); it is not found in any very fusiform newly-divided cells, indicating that this lagging event may be related to the growth or reproduction process. More advanced work needs to be accomplished to address this question.

Pellicle development during Petalomonas cantuscygni ontogeny

Figure 3 includes a series of images of pellicle development during the *Petalomonas cantuscygni* cell cycle. Fig 3A shows a pear-shaped *Petalomonas cantuscygni*, which is a middle-aged cell according to our developmental model. The cells become progressively more flattened as they age (Fig 4B). Figure 4C shows a posterior shot of the pellicle strip duplication, in this figure, every single pellicle strip is doubled. In fig 4D, the pellicle strips have finished doubling and are already separated from each other. The cell starts to divide itself longitudinally from anterior end (Fig 4E) and the cleavage furrow becomes deeper and deeper as cytokinesis progresses (Fig 4E to 4H). As seen in Fig 4I near the end of cell division only the posterior tips hold the two daughter cells together. In the last figure 4J, the two cells are completely separated from each other and each is fusiform in shape.

Some assumptions can be made based these observations. The first is that cell division starts from the anterior end and cytokinesis begins after the pellicle strips have been duplicated and the P value is doubled. Out of hundreds of examined cells only 14 were found to be in an obvious state of cell division. Based on the relatively few cells that were fixed during active cytokinesis, we assume that the time spent in cell division time is relatively short, occupying between 5%-10% of the cell cycle. There are many questions that we cannot answer using available data for now (e.g. how the feeding and flagellar apparatus are duplicated) and it will require further studies to address these questions.

The taxonomic position of Petalomonas cantuscygni

Most of the evidence for determining the phylogenetic position of *Petalomonas cantuscygni* is based on morphological and molecular studies. Previous studies based on ultrastructural data suggested that *Petalomonas cantuscygni* resembles what is thought to be the ancestral form of the Euglenozoa and *Petalomonas cantuscygni* was proposed to be a missing-link between the Euglenids and Kinetoplastids. However, these studies did not account for the morphological features of diplonemids and did not perform comparative morphology on all of three Euglenozoan subgroups at the same time. Diplonemids do not have pellicle strips like euglenids, but they have a corset of supporting microtubules under the plasmamembrane. Whether the pellicle strips in *Petalomonas cantuscygni* were acquired independently during the evolutionary process and thus are homoplastic features; or whether strips were lost in diplonemids is not known. Further morphological data are needed to determine whether there is any vestige of pellicle strips in diplonemids. These data would be helpful in re-evaluating the position of *Petalomonas cantuscygni*.

The fact that Petalomonas cantuscygni falls clearly within the Euglenozoa is not a mystery, but it becomes debatable when our phylogenic studies (in accompanied chapter) indicate that it groups most closely with the diplonemids. Previous molecular data suggested that *Petalomonas cantuscygni* is basal to the Euglenids and it is mainly based on SSU rDNA phylogeny (Preisfeld, Berger et al. 2000; Leander, Witek et al. 2001; Mullner, Angeler et al. 2001; Preisfeld, Busse et al. 2001; Busse and Preisfeld 2003). However, our own data (see accompanied chapter) uses combination of SSU and LSU rDNA sequence data that contain over three times as many as phylogenetically informative characters than does the SSU rDNA only data set. The resulting phylogeny strongly supported that *Petalomonas cantuscygni* is grouped with Diplonemids. In addition, from the result of pairwise distance comparisons, *Petalomonas cantuscygni* shows the shorter distance to diplonemids than to other species in euglenids and kinetoplastids. But our taxon sampling does not include the phagotrophic euglenid *Notoselenus*, which has been suggested to be sister group to *Petalomonas cantuscygni* and they were group in the order Petalomonadida (von der Heyden, Chao et al. 2004). Future more basal Euglenids taxa such as *Notoselenus, Entosiphon* and other basal colorless euglenids probably could be more helpful to address the evolutionary position of *Petalomonas cantuscygni*.

There has been a tradition that *Petalomonas cantuscygni* was selected to be the outgroup taxon in the systematic studies of Euglenozoa (MontegutFelkner and Triemer 1997; Linton, Hittner et al. 1999; Preisfeld, Berger et al. 2000; Leander and Farmer 2001). However, based on our phylogeny that euglenids and diplonemids are closely related than to kinetoplastids, *Petalomonas cantuscygni* is a suitable outgroup taxa for diplonemids. Possibly, it could be an appropriate outgroup taxa to basal in-group euglenids or kinetoplastids depending upon the relative position of these taxa; but, it may not be suitable to root derived clades such as

phototrophic euglenids alone or trypanosomatids without other outgroup taxa. Additionally, it may not be appropriate to map the characters onto *Petalomonas cantuscygni* when it is the outgroup taxon to study character evolution.

CONCLUSIONS

Although our data cannot provide enough evidences to resolve all the problems, some tentative conclusions (or assumptions) can be drawn.

1. The assumption that the fusiform cells are recently divided, the pear-shaped cells are typical of interphase cells, and cells become more flattened as they approach cell division. Flattened or leaf-shaped cell divide longitudinally into two fusiform cells and begin a new cell cycle.
2. The anterior feeding pocket is formed discontinuously, with one single emergent flagellum leading out.
3. The P value for *Petalomonas cantuscygni* is eight. Posterior strips reductions occur, especially pellicle strips number 7 that does not converge with other pellicle strips but ends itself alone at pellicle strip number 8.
4. Cell division starts from the anterior end, possibly in the feeding pocket, and cytokinesis happens after all the pellicle strips get duplicated and P value is doubled.
5. The time for cell division is assumingly relatively short in the life cycle of *Petalomonas cantuscygni*, roughly 5%-10% of a complete life cycle.

ACKNOWLEDGEMENTS

Financial support was provided by the U.S. National Science Foundation PEET (Partnerships for Enhancing Expertise in Taxonomy) grant no. DEB 4-21348. We also thank Dr. John Shields for technical assistance.

REFERENCES

Andersson, J. O. and A. J. Roger (2002). "A cyanobacterial gene in nonphotosynthetic protists - An early chloroplast acquisition in eukaryotes?" Current Biology **12**(2): 115-119.

Atkinson, R. G., K. M. Bolitho, et al. (1998). "Apple ACC-oxidase and polygalacturonase: ripening-specific gene expression and promoter analysis in transgenic tomato." Plant Molecular Biology **38**(3): 449-460.

Baldauf, S. L. (2003). "The deep roots of eukaryotes." Science **300**(5626): 1703-6.

Baldauf, S. L., A. J. Roger, et al. (2000). "A kingdom-level phylogeny of eukaryotes based on combined protein data." Science **290**(5493): 972-7.

Brosnan, S., W. Shin, et al. (2003). "Phylogeny of the photosynthetic euglenophytes inferred from the nuclear SSU and partial LSU rDNA." Int J Syst Evol Microbiol **53**(Pt 4): 1175-86.

Brown, J. R. and W. F. Doolittle (1997). "Archaea and the prokaryote-to-eukaryote transition." Microbiol Mol Biol Rev **61**(4): 456-502.

Brown, P., Zakry's B., and Farmer, M. A. (2003). "Plastid morphology, ultrastructure, and development in Colacium and the loricate euglenophytes (Euglenophyceae)." J. Phycol. **39**: 115-121.

Busse, I., D. J. Patterson, et al. (2003). "Phylogeny of phagotrophic euglenids (Euglenozoa): A molecular approach based on culture material and environmental samples." Journal of Phycology **39**(4): 828-836.

Busse, I. and A. Preisfeld (2002). "Phylogenetic position of Rhynchopus sp. and Diplonema ambulator as indicated by analyses of euglenozoan small subunit ribosomal DNA." Gene **284**(1-2): 83-91.

Busse, I. and A. Preisfeld (2002). "Unusually expanded SSU ribosomal DNA of primary osmotrophic euglenids: molecular evolution and phylogenetic inference." J Mol Evol **55**(6): 757-67.

Busse, I. and A. Preisfeld (2003). "Application of spectral analysis to examine phylogenetic signal among euglenid SSU rDNA data sets (Euglenozoa)." Organisms Diversity & Evolution **3**(1): 1-12.

Busse, I. and A. Preisfeld (2003). "Discovery of a group I intron in the SSU rDNA of Ploeotia costata (Euglenozoa)." Protist **154**(1): 57-69.

Busse, I. and A. Preisfeld (2003). "Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa)." Int J Syst Evol Microbiol **53**(Pt 2): 617-24.

Busse, I. and A. Preisfeld (2003). "Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa)." International Journal of Systematic and Evolutionary Microbiology **53**: 617-624.
Butschli, O. (1884). "Protozoa in Bronn's Klassen und Ordnung des." Tierreiches. **2**: 617-1097.
Cann, J. P. and N. C. Pennick (1986). "Observations on Petalomonas-Cantuscygni, N-Sp, a New Halo-Tolerant Strain." Archiv Fur Protistenkunde **132**(1-2): 63-71.
Cavalier-Smith, T. (1981). "Eukaryote kingdoms: seven or nine?" Biosystems **14**(3-4): 461-81.
Cavalier-Smith, T. (1993). "Kingdom protozoa and its 18 phyla." Microbiol Rev **57**(4): 953-94.
Cavalier-Smith, T. (1998). "A revised six-kingdom system of life." Biological Reviews **73**(3): 203-266.
Cavalier-Smith, T. (2000). "Flagellate megaevolution: the basis for eukaryote diversification." 361-390.
Cavalier-Smith, T. (2002). "The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa." Int J Syst Evol Microbiol **52**(Pt 2): 297-354.
Clark, C. G., B. W. Tague, et al. (1984). "Xenopus-Laevis-28S Ribosomal-Rna - a Secondary Structure Model and Its Evolutionary and Functional Implications." Nucleic Acids Research **12**(15): 6197-6220.
Daubin, V., E. Lerat, et al. (2003). "The source of laterally transferred genes in bacterial genomes." Genome Biology **4**(9): -.
Doolittle, W. F. (2000). "Uprooting the tree of life." Scientific American **282**(2): 90-95.
Elbrachter, M., E. Schnepf, et al. (1996). "Hemistasia phaeocysticola (SCHERFFEL) comb nov, redescription of a free-living, marine, phagotrophic kinetoplastid flagellate." Archiv Fur Protistenkunde **147**(2): 125-136.
Farmer, M. A. and R. E. Triemer (1988). "Flagellar Systems in the Euglenoid Flagellates." Biosystems **21**(3-4): 283-291.
Fast, N. M., L. Xue, et al. (2002). "Re-examining alveolate evolution using multiple protein molecular phylogenies." J Eukaryot Microbiol **49**(1): 30-7.
Goldman, N., J. P. Anderson, et al. (2000). "Likelihood-based tests of topologies in phylogenetics." Syst Biol **49**(4): 652-70.
Hancock, J. M. and G. A. Dover (1988). "Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs." Mol Biol Evol **5**(4): 377-91.
Hannaert, V., E. Saavedra, et al. (2003). "Plant-like traits associated with metabolism of Trypanosoma parasites." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 1067-1071.
Harrison, P. J., R. E. Waters, et al. (1980). "A Broad-Spectrum Artificial Seawater Medium for Coastal and Open Ocean Phytoplankton." Journal of Phycology **16**(1): 28-35.
Hassouna, N., B. Michot, et al. (1984). "The Complete Nucleotide-Sequence of Mouse 28S Ribosomal-Rna Gene - Implications for the Process of Size Increase of the Large Subunit Ribosomal-Rna in Higher Eukaryotes." Nucleic Acids Research **12**(8): 3563-3583.
Hillis, D. M. and J. P. Huelsenbeck (1992). "Signal, noise, and reliability in molecular phylogenetic analyses." J Hered **83**(3): 189-95.
Honigberg, B. M., Balamuth, W. (1963). "Subphylum Sarcomastigophora nom. Nov. to embrace the flagellate and amoiboid assemblages of protozoans." Journal of Protozoology **10 (suppl.) 27**.
Huelsenbeck, J. P. and J. P. Bollback (2001). "Empirical and hierarchical Bayesian estimation of ancestral states." Systematic Biology **50**(3): 351-366.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.
Huelsenbeck, J. P., F. Ronquist, et al. (2001). "Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology." Science **294**(5550): 2310-2314.
Inagaki, Y., Y. HayashiIshimaru, et al. (1997). "Algae or protozoa: Phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences." Journal of Molecular Evolution **45**(3): 295-300.
Inderbitzin, P., S. R. Lim, et al. (2004). "The phylogenetic position of Spathulospora based on DNA sequences from dried herbarium material." Mycological Research **108**: 737-748.
Kapoor, M., C. A. Curle, et al. (1995). "The Hsp70 Gene Family of Neurospora-Crassa - Cloning, Sequence-Analysis, Expression, and Genetic-Mapping of the Major Stress-Inducible Member." Journal of Bacteriology **177**(1): 212-221.
Keller M. D., S., R. C., Claus, W., and Guillard, R. R. L. (1987). "Media for the culture of oceanic ultraphytoplankton." J. Phycol. **23**: 633-638.
Kerrigan, J., M. T. Smith, et al. (2003). "Ascobotryozyma cognata sp. nov., a new ascomycetous yeast associated with nematodes from wood-boring beetle galleries." Mycological Research **107**: 1110-1120.
Kivic, P. A. and P. L. Walne (1984). "An Evaluation of a Possible Phylogenetic Relationship between the Euglenophyta and Kinetoplastida." Origins of Life and Evolution of the Biosphere **13**(3-4): 269-288.
Korall, P. and P. Kenrick (2004). "The phylogenetic history of Selaginellaceae based on DNA sequences from the plastid and nucleus: extreme substitution rates and rate heterogeneity." Molecular Phylogenetics and Evolution **31**(3): 852-864.
Kuzoff, R. K., J. A. Sweere, et al. (1998). "The phylogenetic potential of entire 26S rDNA sequences in plants." Molecular Biology and Evolution **15**(3): 251-263.
Larsen, J. and D. J. Patterson (1990). "Some Flagellates (Protista) from Tropical Marine-Sediments." Journal of Natural History **24**(4): 801-937.
Leander, B. S. (2004). "Did trypanosomatid parasites have photosynthetic ancestors?" Trends in Microbiology **12**(6): 251-258.
Leander, B. S. and M. A. Farmer (2000). "Comparative morphology of the euglenid pellicle. I. Patterns of strips and pores." Journal of Eukaryotic Microbiology **47**(5): 469-479.
Leander, B. S. and M. A. Farmer (2001). "Evolution of Phacus (Euglenophyceae) as inferred from pellicle morphology and SSU rDNA." Journal of Phycology **37**(1): 143-159.
Leander, B. S., R. E. Triemer, et al. (2001). "Character evolution in heterotrophic euglenids." European Journal of Protistology **37**(3): 337-356.
Leander, B. S., R. P. Witek, et al. (2001). "Trends in the evolution of the euglenid pellicle." Evolution **55**(11): 2215-2235.
Lin, C. P. and B. N. Danforth (2004). "How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets." Molecular Phylogenetics and Evolution **30**(3): 686-702.
Linton, E. W., D. Hittner, et al. (1999). "A molecular study of euglenoid phylogeny using small subunit rDNA." Journal of Eukaryotic Microbiology **46**(2): 217-223.
Maddison, W. P. a. D. R. M. (1992). "MacClade. Version 3.0."
Marin, B., A. Palm, et al. (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**(1): 99-145.

Martin, W. and P. Borst (2003). "Secondary loss of chloroplasts in trypanosomes." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 765-767.
Maslov, D. A., S. Yasuhira, et al. (1999). "Phylogenetic affinities of Diplonema within the euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences." Protist **150**(1): 33-42.
Montegutfelkner, A. E. and R. E. Triemer (1994). "Phylogeny of Diplonema-Ambulator (Larsen and Patterson).1. Homologies of the Flagellar Apparatus." European Journal of Protistology **30**(2): 227-237.
MontegutFelkner, A. E. and R. E. Triemer (1997). "Phylogenetic relationships of selected euglenoid genera based on morphological and molecular data." Journal of Phycology **33**(3): 512-519.
Moreira, D., P. Lopez-Garcia, et al. (2001). "New insights into the phylogenetic position of diplonemids: G+C content bias, differences of evolutionary rate and a new environmental sequence." International Journal of Systematic and Evolutionary Microbiology **51**: 2211-2219.
Mullner, A. N., D. G. Angeler, et al. (2001). "Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence." International Journal of Systematic and Evolutionary Microbiology **51**: 783-791.
Hansen G. and Daugbjerg N. (2004). "Ultrastructure of Gyrodinium spirale, the type species of Gyrodinium (Dinophyceae), including a phylogeny of G. dominans, G. rubrum and G. spirale deduced from partial LSU rDNA sequences." Protist **155**: 271-294.
Needlema.Sb and C. D. Wunsch (1970). "A General Method Applicable to Search for Similarities in Amino Acid Sequence of 2 Proteins." Journal of Molecular Biology **48**(3): 443-&.
Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.
Passamaneck, Y. J., C. Schander, et al. (2004). "Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences." Molecular Phylogenetics and Evolution **32**(1): 25-38.
Porter, D. (1973). "Isonema-Papillatum Sp-N, a New Colorless Marine Flagellate - Light-Microscopic and Electron-Microscopic Study." Journal of Protozoology **20**(3): 351-356.
Posada, D. and T. R. Buckley (2004). "Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests." Systematic Biology **53**(5): 793-808.
Posada, D. and K. A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**(9): 817-818.
Preisfeld, A., S. Berger, et al. (2000). "Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data." Journal of Phycology **36**(1): 220-226.
Preisfeld, A., I. Busse, et al. (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**(Pt 3): 751-8.
Pupko, T., D. Huchon, et al. (2002). "Combining multiple data sets in a likelihood analysis: which models are the best?" Mol Biol Evol **19**(12): 2294-307.
Schnare, M. N., J. R. Cook, et al. (1990). "14 Internal Transcribed Spacers in the Circular Ribosomal DNA of Euglena-Gracilis." Journal of Molecular Biology **215**(1): 85-91.

Schnare, M. N. and M. W. Gray (1990). "Sixteen discrete RNA components in the cytoplasmic ribosome of Euglena gracilis." J Mol Biol **215**(1): 73-83.
Shimodaira, H. (2001). "Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection." Communications in Statistics-Theory and Methods **30**(8-9): 1751-1772.
Shimodaira, H. and M. Hasegawa (1999). "Multiple comparisons of log-likelihoods with applications to phylogenetic inference." Molecular Biology and Evolution **16**(8): 1114-1116.
Simpson, A. G., J. Lukes, et al. (2002). "The evolutionary history of kinetoplastids and their kinetoplasts." Mol Biol Evol **19**(12): 2071-83.
Simpson, A. G. B. (1997). "The identity and composition of the Euglenozoa." Archiv Fur Protistenkunde **148**(3): 318-328.
Simpson, A. G. B. (2003). "Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota)." International Journal of Systematic and Evolutionary Microbiology **53**: 1759-1777.
Simpson, A. G. B. and A. J. Roger (2004). "Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa." Molecular Phylogenetics and Evolution **30**(1): 201-212.
Snoeyenbos-West, O. L., J. Cole, et al. (2004). "Molecular phylogeny of phyllopharyngean ciliates and their group I introns." J Eukaryot Microbiol **51**(4): 441-50.
Sofia, H. J., G. Chen, et al. (2001). "Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods." Nucleic Acids Res **29**(5): 1097-106.
Sogin, M. L., H. J. Elwood, et al. (1986). "Evolutionary Diversity of Eukaryotic Small-Subunit Ribosomal-Rna Genes." Proceedings of the National Academy of Sciences of the United States of America **83**(5): 1383-1387.
Soltis, D. E. and P. S. Soltis (1997). "Phylogenetic relationships in Saxifragaceae sensu lato: A comparison of topologies based on 18S rDNA and rbcL sequences." American Journal of Botany **84**(4): 504-522.
Spencer, D. F., J. C. Collings, et al. (1987). "Multiple Spacer Sequences in the Nuclear Large Subunit Ribosomal-Rna Gene of Crithidia-Fasciculata." Embo Journal **6**(4): 1063-1071.
Stechmann, A. and T. Cavalier-Smith (2003). "Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90." J Mol Evol **57**(4): 408-19.
Stechmann, A. and T. Cavalier-Smith (2003). "Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90." Journal of Molecular Evolution **57**(4): 408-419.
Suh, S. O., H. Noda, et al. (2001). "Insect symbiosis: derivation of yeast-like endosymbionts within an entomopathogenic filamentous lineage." Mol Biol Evol **18**(6): 995-1000.
Sung, D. Y., E. Vierling, et al. (2001). "Comprehensive expression profile analysis of the Arabidopsis hsp70 gene family." Plant Physiology **126**(2): 789-800.
Swofford, C. D., J. W. Kasckow, et al. (1996). "Substance use: A powerful predictor of relapse in schizophrenia." Schizophrenia Research **20**(1-2): 145-151.
Swofford, D. L. (1993). Phylogenetic Analysis using Parsimony (PAUP) 3.1.1. Washington, D.C., Smithsonian Institute.
Swofford, D. L. (2002). PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, MA, Sinauer.

Talke, S. and A. Preisfeld (2002). "Molecular evolution of euglenozoan paraxonemal rod genes par1 and par2 coincides with phylogenetic reconstruction based on small subunit rDNA data." Journal of Phycology **38**(5): 995-1003.
Tamura, K. and M. Nei (1993). "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees." Molecular Biology and Evolution **10**(3): 512-526.
Tessier, L. H., H. vanderSpeck, et al. (1997). "The cox1 gene from Euglena gracilis: A protist mitochondrial gene without introns and genetic code modifications." Current Genetics **31**(3): 208-213.
Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.
Thompson, M. D., D. W. Copertino, et al. (1995). "Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus Euglena." Nucleic Acids Res **23**(23): 4745-52.
Triemer, R. E. (1986). "Light and Electron-Microscopic Description of a Colorless Euglenoid, Serpenomonas-Costata N-G, N-Sp." Journal of Protozoology **33**(3): 412-415.
Triemer, R. E. and M. A. Farmer (1991). "An Ultrastructural Comparison of the Mitotic Apparatus, Feeding Apparatus, Flagellar Apparatus and Cytoskeleton in Euglenoids and Kinetoplastids." Protoplasma **164**(1-3): 91-104.
Triemer, R. E. and L. Fritz (1986). "Investigations of the Feeding Apparatus in a Phagotrophic Euglenoid." Journal of Cell Biology **103**(5): A526-a526.
Triemer, R. E. and D. W. Ott (1990). "Ultrastructure of Diplonema-Ambulator Larsen and Patterson (Euglenozoa) and Its Relationship to Isonema." European Journal of Protistology **25**(4): 316-320.
Van de Peer, Y., J. M. Neefs, et al. (1993). "Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock." J Mol Evol **37**(2): 221-32.
Veldman, G. M., J. Klootwijk, et al. (1981). "The primary and secondary structure of yeast 26S rRNA." Nucleic Acids Res **9**(24): 6935-52.
von der Heyden, S., E. E. Chao, et al. (2004). "Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa." Journal of Eukaryotic Microbiology **51**(4): 402-416.
Ware, V. C., B. W. Tague, et al. (1983). "Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis." Nucleic Acids Res **11**(22): 7795-817.
White, T. C., G. Rudenko, et al. (1986). "Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs." Nucleic Acids Res **14**(23): 9471-89.
Willey, R. L., Walne, P.L. Kivic, P. (1988). "Phagotrophic and the origins of euglenoid flagellates." CRC Critical Reviews in Plant Sciences **7**: 303-340.
Winchell, C. J., A. P. Martin, et al. (2004). "Phylogeny of elasmobranchs based on LSU and SSU ribosomal RNA genes." Molecular Phylogenetics and Evolution **31**(1): 214-224.
Wuyts, J., P. De Rijk, et al. (2000). "Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA." Nucleic Acids Research **28**(23): 4698-4708.

Wuyts, J., Y. Van de Peer, et al. (2001). "Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA." Nucleic Acids Res **29**(24): 5017-28.
Yang, Z. H. (1996). "Among-site rate variation and its impact on phylogenetic analyses." Trends in Ecology & Evolution **11**(9): 367-372.
Yang, Z. H. (1998). "On the best evolutionary rate for phylogenetic analysis." Systematic Biology **47**(1): 125-133.
Yasuhira, S. and L. Simpson (1997). "Phylogenetic affinity of mitochondria of Euglena gracilis and kinetoplastids using cytochrome oxidase I and hsp60." Journal of Molecular Evolution **44**(3): 341-347.

Figure **3.1.** SEM of different shapes of *Petalomonas cantuscygni*. **1A.** A typical pear-shaped *Petalomonas cantuscygni* (Bar = 2 μm). **1B.** A typical cell in culture, noticed that the cell is observed to move forward along the surface of the culture flask with their emergent flagellum positioned in front of the cell and the cell body angled up from the surface (Bar = 2 μm). **1C.** A flattened dorsoventrally or leaf-shaped cell (Bar = 2 μm). **1D.** A fusiform or less flattened cell (Bar = 2 μm).

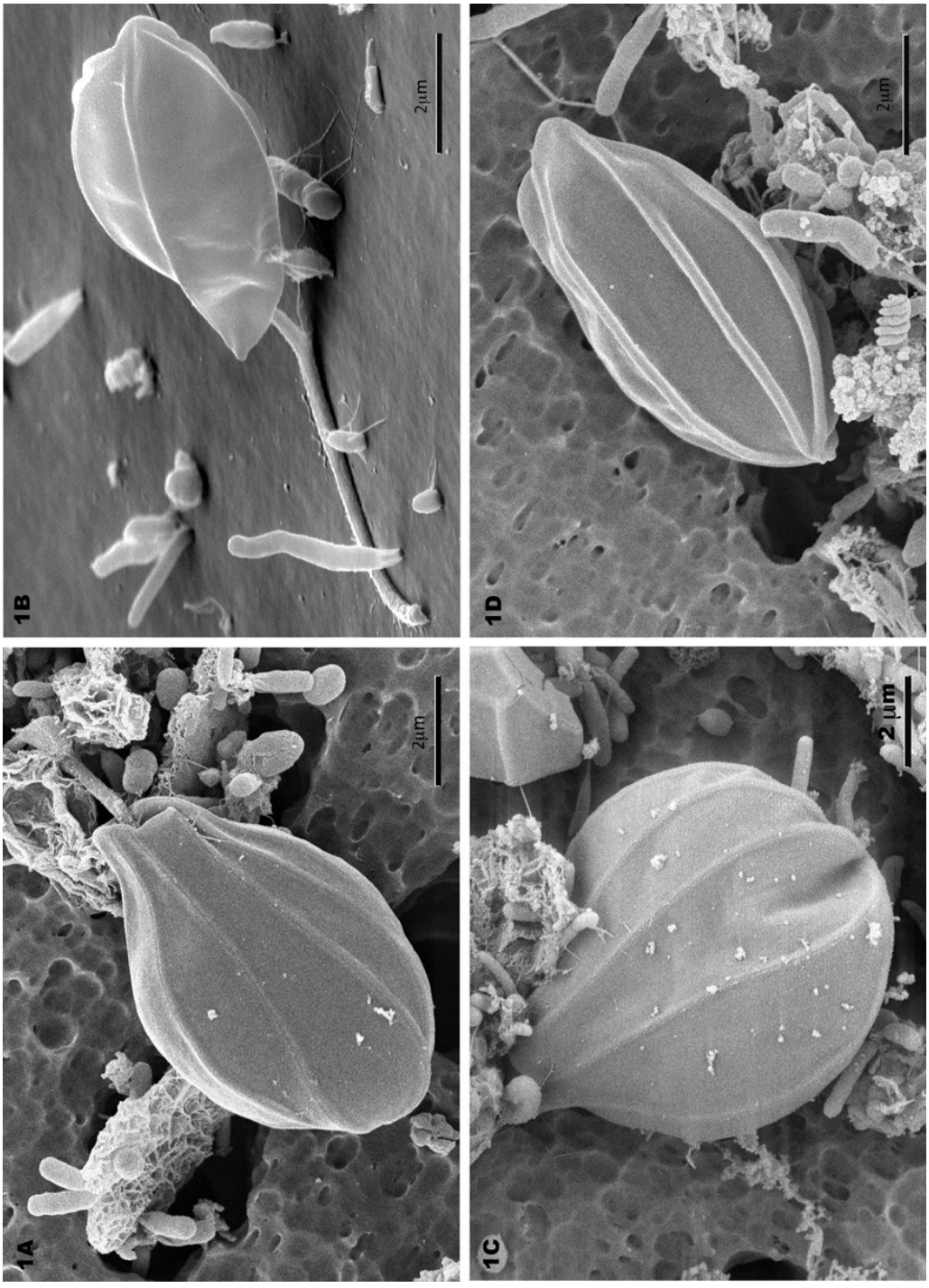


Figure **3.2.** Surface morphological structures of *Petalomonas cantuscygni*. **2A.** A anterior end of a cell with pellicle strips being assigned a number in a clockwise order for reference as marked (Bar = 2 μm). **2B.** *Petalomonas cantuscygni* with a single emergent leading flagellum and feeding pocket angled up from the surface in culture (Bar = 2 μm). **2C.** The feeding pocket which is discontinuously formed at the anterior end of cell (Bar = 2 μm). **2D.** Counting the maximum number of pellicle strips (P value) from the posterior end of the cell (Bar = 2 μm).

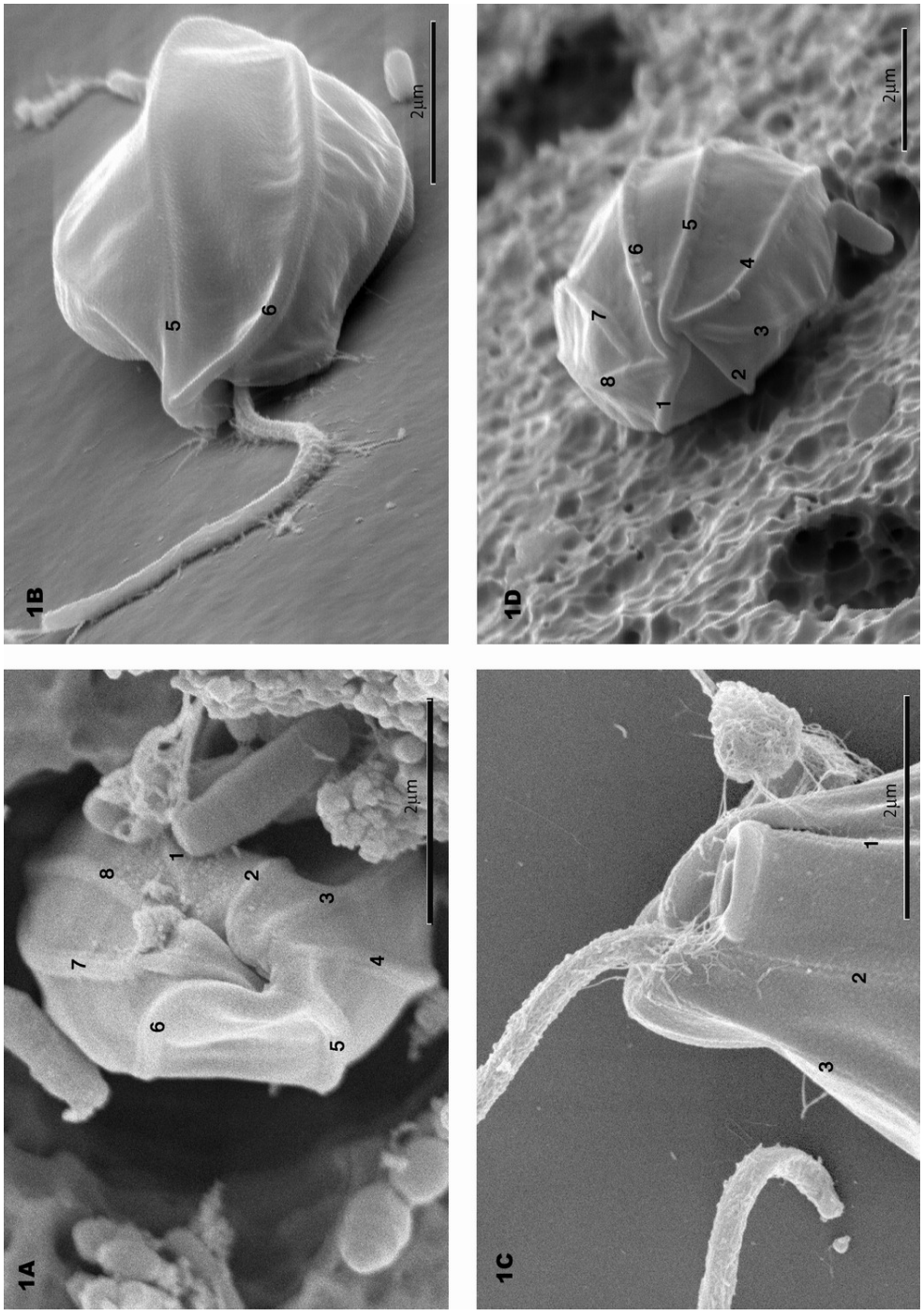


Figure **3.3.** SEM of Pellicle development during *Petalomonas cantuscygni* ontogeny. **3A.** A pear-shaped *Petalomonas cantuscygni*, which is a middle-aged cell according to our developmental model (Bar = 5 μm). **3B.** The cell become progressively more flattened as they age (Bar = 5 μm). **3C.** A posterior shot of the pellicle strip duplication (Bar = 2 μm). **3D.** The pellicle strips have finished doubling and are already separated from each other (Bar = 2 μm). **3E.** The cell starts to divide itself longitudinally from anterior end (Bar = 5 μm). **3F-3H.** The cleavage furrow becomes deeper and deeper as cytokinesis progresses (Bar = 2 μm). **2I.** Near the end of cell division, only the posterior tips hold the two daughter cells together (Bar = 2 μm). **2J.** The two cells are completely separated from each other and each is fusiform in shape (Bar = 2 μm).

3A
5µm
3B
5µm
3C
2µm
3D
2µm
3E
2µm
3F
2µm
3G
2µm
3H
2µm
3I
2µm
3J
2µm

CHAPTER 4

CONCLUSION

Euglenozoa and its three subgroups, euglenids, kinetoplastids and diplonemids, were all suggested to be monophyletic groups based on their morphological features and phylogenies using different molecular markers, but relationships among these subgroups remain controversial. Based on morphological similarities including the shape of the feeding and flagellar apparati, patterns of mitosis, and cytoskeletal composition, Euglenids and Kinetoplastids were proposed to be more closely related to each other than either is to Diplonemids (Triemer and Farmer 1991; Snoeyenbos-West, Cole et al. 2004). In contrast, recent neighbor-joining (Busse, Patterson et al. 2003) and distance-based (von der Heyden, Chao et al. 2004) phylogenetic analyses of SSU rDNA suggested that the Euglenids and Diplonemids were more closely related. Other recent studies using heat shock protein marker Hsp 70 and 90 (Simpson and Roger 2004) suggested that Kinetoplastids and Diplonemids are sister clades. In addition, it has recently been proposed that Kinetoplastids are perhaps descended from a photosynthetic Euglenids ancestor (Hannaert, Saavedra et al. 2003; Martin and Borst 2003; Leander 2004). Because of the evolutionary, ecological and medical significance of this group and conflicting results obtained to date, relationships among the Euglenozoa are an area of intensive ongoing research.

Selection of appropriate molecular markers is essential to reconstruct the phylogeny of Euglenozoa. A variety of markers have been sampled previously for systematic studies in this group. Studies based on paraxonemal rod 1 and 2 (Talke and Preisfeld 2002), heat shock protein 70 and 90 (Baldauf, Roger et al. 2000; Simpson, Lukes et al. 2002; Simpson and Roger 2004), *Cox*1 (Tessier, vanderSpeck et al. 1997); chloroplast-based *rbcL* (ribulose-1,5-bisphosphate

carboxylase large subunit; (Thompson, Copertino et al. 1995), and mitochondria-based COI (cytochrome oxidase subunit I; (Maslov, Yasuhira et al. 1999) have yielded conflicting results. Additionally, systematic studies of Euglenozoa using SSU rDNA have not provide adequate resolution or confident support for inferred clades, especially deep-level relationships. Although SSU rDNA has several advantages including being effectively a single-copied gene, highly conserved, and easy to amplify, these merits are partly offset by its other properties. SSU rDNA is relatively short and contains fewer phylogenetically informative characters, and these informative sites are prone to base substitution saturation (Busse and Preisfeld 2002; Busse, Patterson et al. 2003; Busse and Preisfeld 2003).

This study began with selection of an appropriate molecular marker. LSU rDNA in combination with SSU rDNA was selected based on several attractive properties. Since LSU rDNA comprises a mosaic of slowly-evolving conserved core (CC) regions and faster-evolving expansion segments (ES) that facilitates study of relationships at both higher and lower taxonomic levels. We evaluate its phylogenetic potential using partitioned data matrices and compare their relative sequence dissimilarity and phylogenetic informativeness. Using this marker, a robust and highly resolved phylogeny was constructed and bootstrap analysis was performed. In this phylogeny, the position of *Petalomonas cantuscygni* is different from previous studies. Therefore, we focus on *Petalomonas cantuscygni* from its basic shapes, natural motion pattern, feeding pocket, emergent flagellum, pellicle strips and ontogeny, in order to provide a basis for future comparative studies in Euglenozoa.

Using LSU rDNA combined with SSU rDNA as our molecular marker, we also explore different algorithms to infer the phylogenetic relationships more accurately. One advantages of using molecular data is to model-based approaches. We could utilize program to

find the most suitable model of sequence evolution, which is especially important for the anciently diverged sequences that have multiple substations at sites. Based on the selected model and correlated parameters, we infer the phylogeny using different algorithms including Bayesian analysis, uncorrected and maximum likelihood corrected neighbour joining; unweighted parsimony and parsimony with substitutions weighted according to the instantaneous rate matrix or characters weighted according to their RC values. The maximum likelihood tree has been selected as the best tree by SH test (Shimodaira and Hasegawa 1999) and Bayesian consensus tree has been recognized as the second best tree, while others are all significantly worse than the best tree. However, a lot of related studies utilize certain algorithms without topology test, and in many cases, the algorithm being selected is less computational intensive but not the best one.

This study optimized the molecular marker selection as well as computational algorithm selection, but it could be further explored and the phylogenetic relationships could be further tested in the Euglenozoa by the following ways: 1) the taxa sampling could be improved by adding more basal taxa to each of the subgroups or orphan taxa; 2) different molecular markers independent from ribosomal DNA could be further used to provide another lines of evidences; 3) more advanced anatomical data could be explored to test the resulting relationships. This study provides a frame for the future research of anatomical data and a promising foundation for future studies of character evolution.

LSU rDNA IS A USEFUL MOLECULAR MARKER

We conclude that Euglenozoa LSU rDNA sequences contain significant phylogenetic signal in both conserved core regions and all classes of expansion segments. LSU rDNA sequences evolve 1.16 to 2.67 times as fast as SSU rDNA, and yield 2.41 times as many

informative characters. Different partitions of LSU rDNA presumably are useful for phylogeny retrieval at different taxonomic levels especially when analyzed with suitable models of sequence evolution and appropriately estimated parameters for LSU rDNA. Collectively, expansion segments I-III have a similar evolutionary rate with SSU rDNA, therefore should be appropriate for phylogeny retrieval at taxonomic levels similar to those investigated with SSU rDNA. Depending on the subgroups of Euglenozoa studies, expansion segment IV evolves between 11.89 to 90.49 times as fast as the conserved core regions of LSU rDNA. Accordingly, they need to be appropriately weighted, or perhaps excluded, from phylogenetic analyses at much higher taxonomic levels, but they appear to be informative, especially for analyses that target lower-level relationship (e.g. at the intrageneric level) in Euglenozoa. In general, LSU rDNA has great phylogenetic potential for phylogenetic analyses of different taxonomic levels with appropriate partitions and weighting strategies.

KINETOPLASTIDS ARE AT THE BASE OF EUGLENOZOA

Several conclusions on deep-level relationships can be drawn based on our phylogeny. Euglenozoa is a monophyletic group and its three main subgroups, euglenids, diplonemids and Kinetoplastids, are all monophyletic. *Petalomonas cantuscygni* is part of a sister clade to Diplonemids. Together these form a clade that is closer to euglenids than to kinetoplastids. Trypanosomatids are monophyletic while core bodonids are paraphyletic. Within the Euglenids, *Strombomonas* and *Trachelomonas* are monophyletic and are closest relatives to each other. In contrast, *Phacus* is paraphyletic. *Eutreptiella* form a monophyletic group at the base of euglenids.

PLACE-HOLDER EUGLENOZOA IN THE TOL

Finally, based on the branch lengths and phylogenetic diversities in our phylogeny, *Eutreptiella marina, Diplonema papillatum, Trypanoplasma borreli* are selected as place-holder taxa to represent Euglenozoa and *Naegleria gruberi* was selected as outgroup taxon in broader studies of the TOL.

ORPHAN TAXON *PETALOMONAS CANTUSCYGNI*

The fusiform cells are newly born. The pear-shaped cells are roughly middle-aged, and cell becomes more flattened when they get older. A flattened or leaf-shaped cell divides longitudinally into two fusiform cells and starts a new life cycle. Cell division starts from the anterior end, possibly feeding pocket, and cytokinesis happens after all the pellicle strips get duplicated and the P value is doubled. The P value for *Petalomonas cantuscygni* is eight. There are posterior strips reductions, especially number 7 (based on our naming) that does not converge with other pellicle strips but terminates itself alone at pellicle strip number 8.

REFERENCES

Baldauf, S. L., A. J. Roger, et al. (2000). "A kingdom-level phylogeny of eukaryotes based on combined protein data." Science **290**(5493): 972-7.

Busse, I., D. J. Patterson, et al. (2003). "Phylogeny of phagotrophic euglenids (Euglenozoa): A molecular approach based on culture material and environmental samples." Journal of Phycology **39**(4): 828-836.

Busse, I. and A. Preisfeld (2002). "Phylogenetic position of Rhynchopus sp. and Diplonema ambulator as indicated by analyses of euglenozoan small subunit ribosomal DNA." Gene **284**(1-2): 83-91.

Busse, I. and A. Preisfeld (2003). "Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of Distigma and Astasia (Euglenozoa)." International Journal of Systematic and Evolutionary Microbiology **53**: 617-624.

Hannaert, V., E. Saavedra, et al. (2003). "Plant-like traits associated with metabolism of Trypanosoma parasites." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 1067-1071.

Leander, B. S. (2004). "Did trypanosomatid parasites have photosynthetic ancestors?" Trends in Microbiology **12**(6): 251-258.
Martin, W. and P. Borst (2003). "Secondary loss of chloroplasts in trypanosomes." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 765-767.
Maslov, D. A., S. Yasuhira, et al. (1999). "Phylogenetic affinities of Diplonema within the euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences." Protist **150**(1): 33-42.
Shimodaira, H. and M. Hasegawa (1999). "Multiple comparisons of log-likelihoods with applications to phylogenetic inference." Molecular Biology and Evolution **16**(8): 1114-1116.
Simpson, A. G., J. Lukes, et al. (2002). "The evolutionary history of kinetoplastids and their kinetoplasts." Mol Biol Evol **19**(12): 2071-83.
Simpson, A. G. B. and A. J. Roger (2004). "Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa." Molecular Phylogenetics and Evolution **30**(1): 201-212.
Snoeyenbos-West, O. L., J. Cole, et al. (2004). "Molecular phylogeny of phyllopharyngean ciliates and their group I introns." J Eukaryot Microbiol **51**(4): 441-50.
Talke, S. and A. Preisfeld (2002). "Molecular evolution of euglenozoan paraxonemal rod genes par1 and par2 coincides with phylogenetic reconstruction based on small subunit rDNA data." Journal of Phycology **38**(5): 995-1003.
Tessier, L. H., H. vanderSpeck, et al. (1997). "The cox1 gene from Euglena gracilis: A protist mitochondrial gene without introns and genetic code modifications." Current Genetics **31**(3): 208-213.
Thompson, M. D., D. W. Copertino, et al. (1995). "Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus Euglena." Nucleic Acids Res **23**(23): 4745-52.
Triemer, R. E. and M. A. Farmer (1991). "An Ultrastructural Comparison of the Mitotic Apparatus, Feeding Apparatus, Flagellar Apparatus and Cytoskeleton in Euglenoids and Kinetoplastids." Protoplasma **164**(1-3): 91-104.
von der Heyden, S., E. E. Chao, et al. (2004). "Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa." Journal of Eukaryotic Microbiology **51**(4): 402-416.