MOLECULAR PHYLOGENY OF AMITOCHONDRIATE EXCAVATES

by

Martin Kolisko

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
October, 2011

DALHOUSIE UNIVERSITY

Department of Biology

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "MOLECULAR PHYLOGENY OF AMITOCHONDRIATE EXCAVATES" by Martin Kolisko in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated: October 12, 2011

External Examiner: _____

Research Co-Supervisor: _____

Research Co-Supervisor: _____

Examining Committee: _____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: October 12, 2011

AUTHOR: Martin Kolisko

TITLE: MOLECULAR PHYLOGENY OF AMITOCHONDRIATE EXCAVATES

DEPARTMENT OR SCHOOL: Biology

DEGREE: PhD CONVOCATION: May YEAR: 2012

_____

Signature of Author

Table of Contents

List of Tables

List of Figures

Abstract

Resolving the phylogenetic tree of eukaryotes is an ongoing challenge for evolutionary biologists. One of the most intriguing questions is the phylogenetic status of Excavata, a group that is well supported by morphological evidence, yet usually not recovered as a clade in molecular phylogenies. The most problematic group of excavates are diplomonads (e.g., *Giardia*), which tend to have very highly divergent gene sequences, making any phylogenetic analyses that include these protists very susceptible to long branch attraction artifact.

This thesis first explores which organisms are most closely related to diplomonads. Phylogenies of three marker genes demonstrate that enteromonads, formerly considered a possible sister group to diplomonads, are a polyphyletic group within diplomonads, suggesting complex evolution of cell morphology in this lineage. However, a large diversity of *Carpediemonas*-like organisms (CLOs) was discovered from marine/saline samples. Most of the major clades of CLOs had not been detected by previous environmental PCR studies. SSU rRNA gene phylogenies show that CLOs form a series of relatively short branches at the base of diplomonads. Phylogenomic analysis of eukaryotes (161 genes), incorporating EST data from 5 excavates, including 3 CLOs, shows that the non-monophyly of Excavata in phylogenomic studies is likely caused by long branch attraction artifact, since most of the methods used to suppress long branch attraction significantly weaken support for this topology. Furthermore, the shorter-branching CLOs represent valuable replacements for the long branching diplomonads; we recovered a robustly supported monophyletic Excavata, when long branches, including diplomonads (and parabasalids), were removed from the analysis. Subsequently, comparative analysis of the putative proteomes of three CLO isolates, the retortamonad *Chilomastix*, diplomonads and parabasalids was performed. Several putative evolutionary steps leading to the extremely reduced mitochondrial organelle of diplomonads were derived through the comparative analysis of predicted organellar proteomes.

This thesis shows the importance of taxon sampling for inferring deep eukaryotic evolution. The more robust understanding of the phylogeny of Excavata, especially diplomonads and parabasalids, and the new availability of a number of deep branching relatives of diplomonads, provide a framework for comparative analyses exploring the evolution of anaerobic organelles or parasitism.

List of Abbreviations Used

ACS	acetyl-CoA synthetase
ASCT	acetate:succinate CoA transferase
AU	approximately unbiased
BLAST	basic local alignment search tool
BP	bootstrap support
CL1-6	clade 1-6
CLO	*Carpediemonas*-like organism
CTAB	cetyl trimethylammonium bromide
EF1-α	elongation factor 1 alpha
EF2	elongation factor 2
EST	express sequence tags
Fdx	ferredoxin
Fe-Hyd	[FeFe]-hydrogenase
GCS-H	glycine cleavage system H
GSC-L	glycine cleavage system L
GCS-P	glycine cleavage system P
GCS-T	glycine cleavage system T
GTR	general time reversible
Hmp	hydrogensomal membrane protein
Hyd-E	hydrogenase maturase E
Hyd-F	hydrogenase maturase F

ICZN	international code of zoological nomenclature
LBA	long branch attraction
LGT	lateral gene transfer
MCF	mitochondrial carrier family
ML	maximum likelihood
MRO	mitochondrion-related organelle
PFO	pyruvate:ferredoxin oxidoreductase
PP	posterior probability
SSU rDNA	small subunit ribosomal DNA
SSU rRNA	small subunit ribosomal RNA
LSU rRNA	large subunit ribosomal RNA
RELL	re-sampling of estimated site log-likelihoods
SCS	succinyl-CoA synthetase
SHMT	serin hydroxymethyltransferase
TBR	tree bisection reconnection
TEM	transmission electron microscopy
TIM	inner membrane translocase
TOM	outer membrane translocase

Acknowledgements

First, I want to thank Jessica for adopting me as her brother and subsequently being the best sister ever. Thanks to Adrian, Jakub and Laura simply for being my friends with a capital F. Many thanks to Eleni for being there as solid as a rock during the past year and letting me yell at her whenever I needed to vent. Dan is a great friend, colleague, roommate, and co-brewer and I am glad to know him. Over the time I have known Manolis, he has become much more than the crazy Greek guy I met; he has become a reliable friend who has helped me whenever I needed it without asking for anything back. To Tasos, I want to thank him for his friendship and all the great times, whether it was hiking, drinking, or traveling in Brazil. As a former inhabitant of 1100 Lucknow, I have to thank everyone who stayed there and participated in all the legendary events.

I also want to acknowledge all the members of the Dr. Roger and Dr. Simpson labs. Special thanks goes to Matt for his excitement and talks about phylogenomics. To Courtney, Tasos and Michele I want to thank you for answering all my questions about the biochemistry of diverse organelles and your comments on my work. Aaron and Jong Soo, my friends and awesome colleagues, would keep my cultures alive, whenever needed. I also want to thank to Jacquelin de Mestral for help with lab work and dealing with bureaucrats. Gordon, Aaron, Kim and Eleni helped me a lot in the final stage of writing by proof reading my thesis.

Many thanks also go to the members of my committee. I learned an immense amount about phylogenetics and obtained bioinformatics skills during the special topic class with Joseph Bielawski. This work would not be possible without that knoweldge. To Sina Adl, I want to thank him for all his comments and help, but especially for being one of my closest friends, who never says “no” to a pint of beer.

I also wish to thank all my collaborators, especially Jeffrey Silberman, Yuji Inagaki, Kiyotaka Takishita, Akinori Yabuki, Naoji Yubuki, and Jan Andersson. Very special thanks go to Ivan Cepicka and Vladimir Hampl, who are not only my collaborators, but also great friends and mentors.

I also want to acknowledge the Nova Scotia Health Research Foundation for providing me with a scholarship.

Most importantly, I want to thank Alastair Simpson and Andrew Roger. There is simply not enough space to describe how amazing they are. Both Andrew and Alastair were supportive friends, enthusiastic mentors, and tolerant bosses with amazing senses of humour. Their passion for science and excitement for the research were contagious and highly motivating. They always showed a lot of appreciation for my work whether the experiments worked or not. Thanks to both of you for everything and especially for making the last six years so great that I wish I did not have to move on.

And last, I want to thank the ones who are closest to me. Many thanks to my mom and dad who gave me all of their support. This would not have been possible without them. They are truly the best parents one can ever wish for. To Rachel, I thank her for all her patience, gentleness and emotional support. Rachel, thank you for being who you are. You are the most amazing person.

Chapter 1

Introduction

This thesis explores some of the most intriguing and difficult questions in the evolution of eukaryotes – the unresolved phylogenetic status of the controversial 'supergroup' Excavata and specifically the phylogenetic affinities and evolution of diplomonads (e.g., the parasite *Giardia intestinalis*). Recently the most successful ways of improving the resolution of molecular phylogenies have been collecting data from large numbers of genes and improving taxon sampling. The presented work attempts to combine both of these approaches by isolating and molecularly characterizing organisms related to diplomonads, and to answer several questions regarding the phylogeny of Excavata and the evolutionary history of amitochondriate excavates.

1.1 Molecular Phylogenetics

In the mid-twentieth century, scientists started considering the use of molecular data to infer evolutionary relationships among organisms (Zuckerkandl and Pauling, 1965). By the early 1980s a number of molecular phylogenetic studies of eukaryotes had been published, but these were based on markers that had little deep historical signal (Kumazaki *et al.*, 1983). By the late 1980s, however, relatively information-rich markers such as SSU rRNA and some slow-evolving proteins began to be widely used (Vossbrinck *et al.*, 1987, Sogin *et al.*, 1989). Nonetheless, these pioneering phylogenetic studies of eukaryotes were based on single genes and constructed

using maximum parsimony or simple distance methods. This led to stochastic and systematic errors in the reconstructed phylogenies.

Stochastic errors are simply caused by the lack of data bearing useful information for phylogenetic reconstruction in the first place and/or by the majority of the positions in the alignment being saturated (Philippe and Laurent, 1998, Roger *et al.*, 1999, Philippe *et al.*, 2000). Saturated positions are those that have accumulated multiple changes over time within lineages, resulting in the loss by 'overwriting' of ancient phylogenetic signal. In datasets with large numbers of saturated positions, estimated phylogenetic affinities result from random rather than true historical signals (for example see Gribaldo and Philippe, 2000). Stochastic errors can be addressed by adding more data to the analyses – for example by the analysis of multiple genes rather than single markers.

Systematic errors are caused by misspecification of the true evolutionary processes operating at the sequence level. The most famous systematic error is long-branch attraction (LBA), which was mathematically demonstrated by Felsenstein (1978). He showed that sequences from distantly related taxa with high evolutionary rates (fast-evolving, or 'long branching' taxa) will end up grouping together artificially in a phylogeny with high statistical support when maximum parsimony methods are used. This is caused by the fact that two sequences that accumulate substitutions at a very high rate will end up similar to each other at some positions just by chance, as a result of the accumulation of multiple changes at that position. In a simplified manner, maximum parsimony cannot distinguish between two sequences that share the same character at a site as a result of

common ancestry and two sequences that have converged on the same state after multiple substitutions, resulting in topologies where the long branching taxa tend to branch together. In contrast to stochastic errors, long branch attraction artifacts, and other systematic errors, cannot be overcome by adding more data of the same kind. On the contrary, adding more data with the same properties will give even higher support to the erroneous topology (Felsenstein, 1978; Felsenstein, 2004).

Felsenstein (Felsenstein, 1981) also developed the use of an important model-based method for tree reconstruction, maximum likelihood. Maximum likelihood uses models of sequence evolution to obtain probabilities for change from one state to the other along the phylogenetic tree. It will allow for statistical accommodation of several aspects of sequence properties, for example, different relative rates of different substitutions or different rates of evolution at sites along the sequence alignment. Maximum likelihood will be robust to systematic errors as long as the statistical model fits closely the actual process of sequence evolution. However, the more that true evolutionary processes differ from the model assumptions the more prone the analysis will be to systematic error, such as LBA (Felsenstein, 2004). Similar problems exist with model-based distance methods (Susko *et al.*, 2004)

1.2 Early Phylogenetic Analyses of Eukaryotes: Stochastic and Systematic Errors

What are now seen as the pioneering molecular phylogenies of microbial eukaryotes were published in the 1980s and were largely based on small subunit ribosomal RNA (SSU rRNA) gene sequences (McCarroll *et al.*, 1983; Vossbrinck *et al.*, 1987;

Sogin *et al.*, 1989). One of the most influential phylogenetic studies of eukaryotes was published by (Sogin *et al.*, 1989), and this serves as a good general illustration of the view of eukaryote phylogeny and evolution that emerged at this time. In this study the eukaryotic tree (rooted using bacterial homologues) consisted of a crowning radiation of "short branching eukaryotes", with several much "longer branching" taxa attaching below this radiation to form an extended base of apparently early diverging lineages. This iconic portrayal of eukaryote relationships was recovered in many subsequent studies using rRNA genes and other markers such as the elongation factor 1 alpha (EF1-α) and elongation factor 2 (EF2) proteins (Van Keulen *et al.*, 1993; Hashimoto *et al.*, 1995; Cavalier-Smith and Chao, 1996; Kamaishi *et al.*, 1996). This tree seemed to fit well to the biological diversity of eukaryotic cells as it was understood at the time: In particular, all of the most 'basal' eukaryotic branches were populated by supposedly amitochondriate (mitochondrion-lacking) organisms, which agreed well with the 'Archezoa hypothesis' that had proposed that several contemporary mitochondrion-lacking eukaryotes had diverged before the acquisition of mitochondria (Cavalier-Smith, 1983). The most famous of these supposedly mitochondrion-lacking lineages were diplomonads, trichomonads and microsporidia. Interestingly all three groups are predominantly parasitic (microsporidia are entirely obligate parasites) and were represented by highly specialized parasites in these early studies, for example, the well-known human parasites *Giardia* (a diplomonad) and *Trichomonas* (a parabasalid). These three taxa usually formed the three most basal and longest branches on the SSU rRNA-based tree of eukaryotes.

Suspiciously, however, the shape of this eukaryotic tree was also consistent with a long-branch attraction artifact, where the longest branching eukaryotic sequences were branching close to the 'long branch' of the prokaryotic outgroup (Brinkmann and Philippe, 1999; Morin, 2000; Philippe *et al.*, 2000). Indeed overwhelming evidence was eventually assembled that indicated that this topology was a result of long-branch attraction artifacts, rather than a true reflection of historical relationships.

The most obvious demonstration of this problem came from further phylogenetic studies of microsporidia. Today we know that microsporidia are not an independent deep-branching eukaryote lineage, but instead extremely derived fungi. The relationship between microsporidia and fungi was first recovered by analyses of β-tubulin (Edlind *et al.*, 1996) and confirmed by analyses of other genes (α-tubulin-Keeling and Doolittle, 1996b; HSP70 –Germot *et al.*, 1997; Hirt *et al.*, 1997; RPB1 – Hirt *et al.*, 1999). When models incorporating among-site rate variation were used for tree reconstruction (Yang, 1996), microsporidia grouped with fungi in EF1-α and LSU rRNA analyses (Van de Peer *et al.*, 2000), as opposed to their previous deep-branching position close to the eukaryotic root. The fungal affinities of microsporidia were further corroborated by the fact that microsporidia, fungi and animals share a large insertion in the EF1-α protein (Hirt *et al.*, 1999).

The true phylogenetic position of parabasalids and diplomonads remains controversial. It is widely accepted that their basal position in phylogenetic studies is caused by long-branch attraction (Morin, 2000; Philippe *et al.*, 2000; Brinkmann *et al.*, 2005). However, some authors have still argued that diplomonads may be truly

deep branching and primitive eukaryotes despite the diplomonad position being affected by long branch attraction (Best *et al.*, 2004; Ciccarelli *et al.*, 2006; Morrison *et al.*, 2007).

Our understanding of other parts of the eukaryotic tree was less drastically compromised by phylogenetic systematic error. Unfortunately, the deep-level resolution of the trees based on single genes has turned out to be very low, probably because of limited data in these sequences combined with possible rapid radiation of the major eukaryote lineages and/or saturation erasing the phylogenetic signal for the deep phylogenetic splits (Philippe and Laurent, 1998; Roger *et al.*, 1999; Philippe *et al.*, 2000).

1.3 The Emergence of the 'Supergroups' View of Eukaryote Diversity

As sequencing became cheaper and more convenient, and genome-scale data from diverse organisms became available, more genes were included in phylogenetic studies of eukaryotes, and it became common to estimate trees from multi-gene datasets (Burger *et al.*, 1999; Baldauf *et al.*, 2000; Moreira *et al.*, 2000; Arisue *et al.*, 2002; Bapteste *et al.*, 2002). This approach, combined with greatly improved taxon sampling and better methods for analysis of rRNA sequences (Van Der Peer and De Wachter 1997), morphological data (Simpson and Patterson, 1999) and the discovery of rare genomic features shared by a few major groups (such as insertions in protein sequences, Archibald *et al.*, 2003, Baldauf and Palmer, 1993), resulted in a new synthesis of the eukaryotic tree of life (Cavalier-Smith, 2002; Simpson and Roger, 2002) that was formalized in a new classification of the eukaryotes by (Adl *et*

al., 2005). This synthesis hypothesized that almost all of eukaryotic diversity can be split into six "supergroups": Opisthokonta, Amoebozoa, Archaeplastida, Rhizaria, Chromalveolata and Excavata.

Opisthokonta consists of fungi, metazoa and several groups of single-celled eukaryotes, notably choanoflagellates. Molecular phylogenetic studies strongly and consistently support monophyly of opisthokonts (Baldauf and Palmer, 1993; Wainright *et al.*, 1993; Baldauf *et al.*, 2000; Zettler *et al.*, 2001; Hertel *et al.*, 2002; Ruiz-Trillo *et al.*, 2006). Opisthokonts also share a ~12 amino acid insertion in the sequence of EF1-α, in species where this gene is present (Baldauf and Palmer, 1993; Steenkamp *et al.*, 2006). The common morphological characters are flat mitochondrial cristae and, where present, a single flagellum that inserts at the posterior end of the cell in flagellated life stages (Cavalier-Smith and Chao, 2003b; Steenkamp *et al.*, 2006).

Amoebozoa primarily groups many organisms that possess amoeboid morphology, including slime molds, lobose amoebae, pelobionts and entamoebae. In the early period of molecular phylogenetics these organisms were not considered to form a monophyletic group, since they tended to branch as numerous independent lineages in early rRNA trees (Sogin, 1989; Hinkle *et al.*, 1994; Pawlowski *et al.*, 1996; Stiller and D., 1999; Milyutina *et al.*, 2001). However, relatively recent molecular phylogenetic studies with improved taxon/gene sampling have recovered a monophyletic Amoebozoa grouping (Bapteste *et al.*, 2002; Fahrni *et al.*, 2003; Nikolaev *et al.*, 2006; Tekle *et al.*, 2008; Minge *et al.*, 2009)

Archaeplastida groups together organisms with primary chloroplasts – Rhodophyta (red algae), Viridiplantae (green algae and land plants) and Glaucophyta. It is assumed that these groups arose from an initially non-photosynthetic eukaryote that enslaved a cyanobacterial endosymbiont, which ultimately became the chloroplast (Palmer, 2003). However, the monophyly of these groups was never supported using single gene phylogenies, with the exception of those of elongation factor 2, and of multigene datasets that included this protein (Moreira *et al.*, 2000; Hackett *et al.*, 2007; Inagaki *et al.*, 2009).

Rhizaria is an assemblage of predominantly amoeboid organisms including the Cercozoa (itself a diverse group including for example chlorarachniophytes, cercomonads, gromiids and euglyphids) as well as Foraminifera and Radiolaria. There is no morphological synapomorphy unifying this group and historically these groups were not considered to be specifically related. Cercozoa is a relatively recently established group first proposed by Cavalier-Smith based on SSU rRNA phylogenies (Cavalier-Smith, 1998; Cavalier-Smith and Chao, 2003a). A close relationship of Cercozoa with Foraminifera and Radiolaria was later supported by multiple molecular phylogenetic markers (Keeling, 2001; Longet *et al.*, 2003; Nikolaev *et al.*, 2004). There is also a shared ~1-2 amino acid insertion in the polyubiquitin sequence of these organisms (Archibald and Keeling, 2004; Burki *et al.*, 2010). It is important to note that some studies placed Foraminifera inside Cercozoa (Keeling, 2001; Archibald and Keeling, 2004).

Chromalveolata is a proposed group unifying the alveolates (including Apicomplexa, ciliates and dinoflagellates) with stramenopiles, cryptomonads and haptophytes. It was suggested that the common ancestor of these organisms obtained a plastid derived from a red alga via secondary endosymbiosis. This was based mainly on plastid morphology (i.e., numbers of bounding membranes) and common presence of chlorophyll c_2 in the plastids (Cavalier-Smith, 1999). Cavalier-Smith (1999) argued that secondary endosymbiosis should be a relatively rare event as plastid acquisition also requires development of complex protein import machineries. The 'chromalveolate hypothesis' gained support from the plastid and plastid-targeted gene phylogenies (Fast *et al.*, 2001; Yoon *et al.*, 2002; Patron *et al.*, 2004). However, phylogenies based on widely-used nuclear-encoded markers almost always failed to support the monophyly of Chromalveolata (Parfrey *et al.*, 2006).

Excavata is a diverse group consisting mostly of heterotrophic flagellates. It includes Parabasala, Fornicata (including diplomonads), Preaxostyla, Heterolobosea, Euglenozoa, Jakobida and *Malawimonas*. The proposed morphological synapomorphies for the group are the presence of a feeding groove used for suspension feeding and some underlying ultrastructural characters (Simpson, 2003). This group was proposed based on a combination of morphological and molecular data (Simpson and Patterson, 1999; Simpson *et al.*, 2000; Simpson and Patterson, 2001; Cavalier-Smith, 2002; Simpson, 2003). However, Excavata rarely formed a monophyletic group in phylogenetic analysis of data sets with broad

sampling of its proposed lineages (Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Hampl *et al.*, 2005; Simpson et al., 2006).

1.4 The Introduction of Phylogenomics

In recent years, analyses of larger multigene datasets have become the most important tools for estimating eukaryote phylogeny, especially 'phylogenomic' studies that incorporate data from tens and hundreds of genes (Bapteste *et al.*, 2002; Arisue *et al.*, 2005; Hampl *et al.*, 2005; Rodriguez-Ezpeleta *et al.*, 2005; Burki *et al.*, 2007; Hackett *et al.*, 2007; Nozaki *et al.*, 2007; Rodriguez-Ezpeleta *et al.*, 2007a; Burki *et al.*, 2008; Burki *et al.*, 2009; Hampl *et al.*, 2009; Minge *et al.*, 2009). These studies brought significantly more resolution to the eukaryotic tree (Figure 1.1). Some of the original 'supergroups' remain highly supported in these studies, for example Opisthokonta and Amoebozoa (e.g., Burki *et al.*, 2009; Hampl *et al.*, 2009; Minge *et al.*, 2009). Other groupings were altered – in particular, the chromalveolate lineages together with Rhizaria have been reshuffled into two new supergroups – the so-called "S.A.R. clade" (which contains stramenopiles and alveolates together with Rhizaria and is strongly supported) and Hacrobia (which includes haptophytes and cryptophytes, along with some obscure heterotrophic lineages, and not well supported) (Figure 1.1; Burki *et al.*, 2007; Hackett *et al.*, 2007; Okamoto *et al.*, 2009). Excavates remained non-monophyletic in most of these studies (Burki *et al.*, 2009; Hampl *et al.*, 2009), but their monophyly cannot be rejected by statistical tests comparing different topologies.

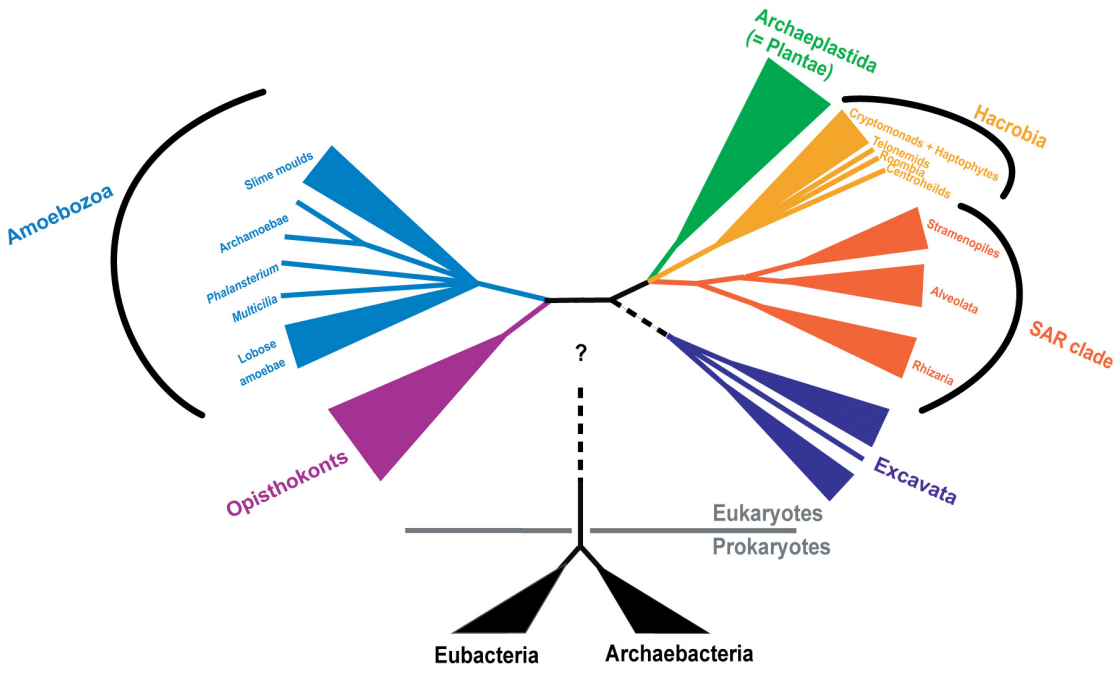


Figure 1.1 A current view of the evolutionary tree of major groups of Eukaryotes. The dashed line and '?' mark one of the possible position of the eukaryotic root (so called unikont/bikont rooting). Modified from Simpson and Roger (2009).

1.5 The Ongoing Impact of Organism Discovery

Phylogenomics represents one of two important elements in recent improvements in the resolution of and knowledge about eukaryotic phylogeny. At the same time as datasets used for phylogenetic analyses were getting bigger, a number of evolutionarily important major lineages of microbial eukaryotes were discovered and/or added to phylogenetic analyses for the first time. Through this discovery of new major lineages, we have gained a more thorough understanding of the evolutionary history of eukaryotes. A good example is *Capsaspora owcarzaki*, which is a single-celled protist living as a symbiont of the snail *Biomphalaria glabrata*. *Capsaspora* branches in the crucial spot on the eukaryotic tree between the

choanoflagellates-metazoa clade and the rest of the opisthokonts (Hertel *et al.*, 2002; Ruiz-Trillo *et al.*, 2004). This makes *Capsaspora* a central organism for understanding the origin of multicellularity in the opisthokont clade (Hertel *et al.*, 2002; Ruiz-Trillo *et al.*, 2004). Similarly, the protist *Chromera velia*, which possesses a functional chloroplast, is a close relative of apicomplexans (e.g., the malaria parasite *Plasmodium*). This makes *Chromera* very important for understanding the evolution of obligatory parasitic and non-photosynthetic apicomplexa and their apicoplast. (Moore *et al.*, 2008). More examples are *Tsukubamonas globosa* (Yabuki *et al.*, 2011) and *Palpitomonas bilix* (Yabuki *et al.*, 2010). *Tsukubamonas* is especially interesting from the point of evolution of the mitochondrial genome as it is a deep branching relative of Jakobids, which possess the largest mitochondrial genome known (Lang *et al.*, 1997). *Palpitomonas* seems to be a non-photosynthetic relative of Archaeplastida or cryptophytes and either way is potentially important for understanding plastid evolution. In addition, several other previously known important taxa recently became available for phylogenomics: for example the breviates (represented by *Breviata anthema*), the telonemids (*Telonema subtilis*), the centrohelids (*Raphidiophrys contractilis*), *Fonticula*, and nucleariids such as *Nuclearia simplex* (Brown *et al.*, 2009; Burki *et al.*, 2009; Liu *et al.*, 2009; Minge *et al.*, 2009).

1.6 Evolutionary History of Excavata and Fornicata

As mentioned above, the taxon Excavata is more-or-less the most contentious of the currently recognized supergroups. The monophyly of excavates, and hence the

phylogenetic validity of the taxon, is generally neither recovered nor highly supported, but at the same time cannot be rejected by molecular phylogenetics (Hampl *et al.*, 2009). Fornicata, along with their relatives the parabasalids, are the most problematic excavates, as they tend to form extremely long branches on the tree and therefore make the analyses very susceptible to long branch attraction artifacts (Brinkmann *et al.*, 2005; Hampl *et al.*, 2009).

Fornicata is a group consisting of anaerobic or microaerophilic heterotrophic flagellates. In addition to the mostly parasitic or commensalic diplomonads, mentioned above, they include the retortamonads, which are also mostly parasites or commensals (Kulda and Nohynkova, 1978) as well as the free-living *Carpediemonas membranifera* and *Dysnectes brevis* (Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Simpson, 2003; Yubuki *et al.*, 2007). Diplomonads – especially the infamous human parasite *Giardia intestinalis* – are the only (relatively) well-studied representatives of Fornicata at present. As mentioned above, the evolutionary history of diplomonads is poorly understood, yet extremely intriguing, as they were once considered the deepest branching eukaryotes and their position on the eukaryotic tree remains unclear. Furthermore they do not possess "classic" mitochondria - it is now known that they are not completely lacking mitochondrial organelles, but instead have extremely reduced versions of mitochondria, called mitosomes (Tovar *et al.*, 2003).

1.7 Aims of This Thesis.

The central aim of the research reported in this thesis is to resolve the phylogeny and evolution of Excavata in general, and 'long-branching' excavates in particular. by combining two approaches: 1) enhancing the sampling of understudied and completely new lineages related to diplomonads, then 2) conducting phylogenomic analyses using large multigene datasets that include this enhanced diversity, and that explore different methods of reducing systematic error, especially long branch attraction. A second aim is to explore the possible evolutionary steps leading to the extremely reduced mitochondrial organelles of diplomonads by using the better-resolved phylogeny in combination with *in-silico* predictions of the functions of the mitochondrial organelles of their newly discovered relatives.

Chapter Two explores the phylogenetic positions of enteromonads, which were previously proposed to be the closest relatives of diplomonads and therefore were possible candidates for helping resolve diplomonads' position on the tree. Three genes were used – SSU rRNA, HSP90 and α-tubulin – to estimate the phylogeny of diplomonads and enteromonads, incorporating 10 new enteromonad isolates. Enteromonads form a polyphyletic group within diplomonads, and thus do not represent a distinct major lineage or lineages within Fornicata.

Chapter Three analyzes phylogenetic relationships of 18 new isolates of *Carpediemonas*-like organisms (CLOs) isolated by me and my collaborators. Based on SSU rRNA gene phylogenies, CLOs form some six distinct lineages (CL1 – CL6) at the base of Fornicata. This wide, previously undetected diversity was captured through regular culturing techniques and had been largely missed by prior

environmental PCR approaches. These new organisms are excellent candidates for inclusion into large phylogenetic datasets: They appear to be shorter branching relatives of long branching diplomonads, and therefore, may be useful to suppress long branch attraction artifacts affecting estimates of the phylogenetic position of diplomonads.

In Chapter Four, the production of EST datasets for five new excavate isolates representing unstudied major lineages is reported, including three isolates of *Carpediemonas*-like organisms obtained during the research reported in Chapter Three, as well as the retortamonad *Chilomastix caulleryi* and the recently described *Tsukubamonas globosa* (Yabuki *et al.*, 2011). This data was incorporated into a new phylogenomic dataset of 161 genes and 85 taxa. The dataset was used for extensive phylogenetic study aimed at the understanding of the status of Excavata and exploration of long-branch attraction artifacts affecting the phylogeny of excavates. I find that the usually recovered topology in which Excavata is 'paraphyletic' rather than monophyletic is a result of long-branch attraction artifact as all long branch attraction suppression methods decrease the support for the paraphyletic topology. When *Carpediemonas*-like organisms are used as replacements for their long branching diplomonad and parabasalids, (and other long branches are excluded form the analysis), a strongly supported Excavata clade is recovered, while still keeping good representation of the group.

Chapter Five is an *in silico* analysis of proteins of possible mitochondrial origin detected in the EST datasets for the three *Carpediemonas*-like organisms and *Chilomastix caulleryi*. The putative partial proteomes of these organelles were

compared to proteomes of the reduced, anaerobic mitochondrial organelles of their relatives: the diplomonad *Giardia intestinalis*, and the parabasalid *Trichomonas vaginalis*. Using this data, and the new understanding of the phylogenetic relationships within Fornicata from Chapters Three and Four, I discuss the possible evolutionary steps leading from the ancestral organelle of the Fornicata-parabasalid group to the more reduced hydrogenosomes of *Trichomonas* and extremely reduced mitosomes of *Giardia*.

The conclusion chapter discusses, in the light of my obtained results, the most important future challenges for understanding the phylogeny and evolution of amitochondriate excavates in particular and microbial eukaryotes in general.

Chapter 2

Molecular Phylogeny of Enteromonads

This Chapter was published (Kolisko et al. 2008)

2.1 Introduction

Diplomonads and their close relatives - enteromonads, retortamonads and *Carpediemonas membranifera* - are small flagellates that tend to be found in low-oxygen habitats. Recently they were classified within Fornicata (Metamonada, Excavata) (Simpson, 2003). Very recently an additional member of Fornicata, *Dysnectes brevis,* was described (Yubuki *et al.*, 2007). Most diplomonads, all described enteromonads, and all described retortamonads except one are endobionts or parasites of animals, with several causing serious and highly prevalent diseases in fish, domestic animals and man (Kulda and Nohynkova, 1978).

Diplomonads and their relatives have been interesting for students of the evolution of the eukaryotic cell for several reasons. Firstly, they lack classical mitochondria. Secondly, diplomonads and their relatives branch at the base of the eukaryotic tree in the majority of phylogenies in which eukaryotes are rooted using prokaryotic outgroups (Sogin *et al.*, 1989; Bapteste *et al.*, 2002; Ciccarelli *et al.*, 2006). Thirdly, diplomonads, but not their relatives, possess a double karyomastigont, in other words, they have two similar or identical nuclei and two flagellar apparatuses per cell (Brugerolle, 1975; Siddall *et al.*, 1992). In the late 1980s and early 1990s it was widely supposed that the last common ancestor of

most or all living eukaryotes was a 'fornicate-like' amitochondriate organism (Cavalier-Smith, 1983; Sogin *et al.*, 1989), and some models even proposed that almost all living eukaryotes were descended from ancestors with a double karyomastigont (Cavalier-Smith, 1995). The best-studied diplomonad – *Giardia intestinalis* (= *G. lamblia*) was looked to as a model for understanding early eukaryotic cells.

Recent studies, however, have shown that Fornicata are secondarily amitochondriate (Tovar *et al.*, 2003), and that many, perhaps all, retain mitochondrion-related organelles. Using electron microscopy, a hydrogenosome-like double membrane-bounded organelle was observed superficially in *Carpediemonas membranifera* (Simpson and Patterson, 1999). Also their basal position in molecular phyogenies is now thought to be a long-branch attraction artifact (Felsenstein, 2004; Brinkmann *et al.*, 2005; Philippe *et al.*, 2005).

Diplomonads and enteromonads deserve special attention within Fornicata. These two groups were considered as closely related on the basis of ultrastructural studies (Brugerolle, 1975; Kulda and Nohynkova, 1978; Siddall *et al.*, 1992) and their close affinity was confirmed recently by molecular phylogenetic methods (Kolisko *et al.*, 2005). The morphology of diplomonads is extremely similar to the morphology of enteromonads – the main character distinguishing these two groups is the doubled karyomastigont of diplomonads (Brugerolle, 1975; Kulda and Nohynkova, 1978). In a very simplified way, the cell of diplomonads could be described as two enteromonad cells joined together (and conversely the enteromonad cell could be described as half of a diplomonad cell). The most

straightforward scenario explaining the evolution of the doubled karyomastigont is that diplomonads arose from enteromonads in a single evolutionary event. Siddall *et al.* (1992) proposed a mechanism of double karyomastigont formation from a single karyomastigont ancestor by secondary karyokinesis (mitosis) and mastigont duplication after delay or arrest of cytokinesis (cell division), resulting in a cell with four karyomastigonts. This cell could then have divided into two cells, each with a doubled karyomastigont. However, some other authors consider as plausible the opposite scenario – secondary simplification from the double karyomastigont morphology of diplomonads to the single karyomastigont morphology of enteromonads (Brugerolle and Taylor, 1977; Simpson, 2003).

Our understanding of the internal phylogeny of Fornicata is based to a large extent on molecular phylogenetic studies, especially with the relatively recent addition of several important taxa to the small subunit ribosomal RNA gene database, namely retortamonads (Silberman *et al.*, 2002), *Carpediemonas membranifera* (Simpson *et al.*, 2002c), one clade of enteromonads (Kolisko *et al.*, 2005), the diplomonad *Octomitus* sp. (Keeling and Brugerolle, 2006) and several new species from the diplomonad genera *Spironucleus* and *Hexamita* (Jørgensen and Sterud, 2007). The monophyly of Fornicata is very strongly supported by molecular phylogenetic studies and there is strong support for a position of *Carpediemonas membranifera* at the base of the Fornicata clade (Simpson *et al.*, 2002c; Simpson, 2003; Simpson et al., 2006). Meanwhile molecular phylogenies almost invariably divide diplomonads into two major clades, Hexamitinae and Giardiinae, that were already recognized on morphological grounds by Kulda and Nohýnková (Kulda and

Nohynkova, 1978; Silberman *et al.*, 2002; Kolisko *et al.*, 2005; Jørgensen and Sterud, 2007), with the former also identified by the synapomorphy of a non-canonical genetic code (Keeling and Doolittle, 1996a). Recently the diplomonad *Octomitus* sp. was confirmed as a sister branch of *Giardia* within Giardiinae (Keeling and Brugerolle, 2006), and at least one enteromonad group was surprisingly shown to fall within Hexamitinae (Kolisko *et al.*, 2005). Nonetheless, our understanding of the relationships amongst Fornicata is incomplete. For example, different analyses of SSU rRNA gene data place retortamonads (represented by isolates identified as *Retortamonas*) either as a sister group of the diplomonad-enteromonad clade (Keeling and Brugerolle, 2006), as predicted by morphology (Kulda and Nohynkova, 1978; Simpson, 2003), or as a sister branch of the *Giardia-Octomitus* clade, thereby making diplomonads appear paraphyletic (Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Kolisko *et al.*, 2005). The main problem in resolving the phylogeny of Fornicata is the limited amount of data, both in terms of taxon sampling and the amount of sequence information per taxon. For example, to date only one enteromonad genus has been studied by molecular means, using only a single gene (Kolisko *et al.*, 2005), whereas there are three genera of enteromonads already described, which are quite different in morphology, and enteromonads are generally not recovered as a clade in phylogenetic analyses of morphological data (Siddall *et al.*, 1992; Simpson, 2003).

In this study we aim to clarify relationships within fornicates, especially among enteromonads and Hexamitinae diplomonads, and to better understand the evolutionary history of single and double karyomastigonts. We introduce several

important taxa into our molecular analyses, including ten new isolates of enteromonads that represent at least three genera. We also introduce two protein-coding genes into the analyses, α-tubulin and HSP90.

2.2 Results

2.2.1 SSU rDNA Phylogeny

Our analyses of SSU rRNA genes include the broadest taxonomic sampling of Fornicates examined so far. In addition to previously available data I have included five new isolates of the enteromonad genus *Trimitus* (the previously published strain KRPO3 was shown to be very closely related to other representatives of the genus *Trimitus*, and therefore we consider KRPO3 as a member of *Trimitus*), three new isolates of the genus *Enteromonas*, two isolates of a new but undescribed enteromonad genus, an isolate of *Trepomonas steini* (Appendix C.1), a new isolate corresponding to the morphospecies *Trepomonas agilis* (PPS-6), a novel *Spironucleus* isolate (GEPA2H) and one uncultured eukaryote (CHESI2). We also extended the previously incomplete SSU rDNA sequence for *Spironucleus muris*. *Trimitus* sp. - IT1 and *Trimitus* sp. - KOMPKOJ represent the first observations and isolations of free-living enteromonads reported so far.

The SSU rDNA analyses were done using three different alignments, 'large' including a broad eukaryotic outgroup, 'main' including all Fornicata and a restricted outgroup, and 'small' including only Hexamitinae diplomonads and enteromonads. Figure 2.1 shows the ML tree based on SSU rRNA genes, with a

restricted outgroup (i.e., the 'main' data set). The overall topology of the SSU rDNA tree is as follows: Fornicata forms a clade with high statistical support (bootstrap support – BP 100/99/88 and Bayesian posterior probability – PP 1), enteromonads branch robustly within the Hexamitinae subtree (BP 100/99/100 and PP 1), and *Giardia* and *Octomitus* form the sister clade of the Hexamitinae-enteromonad subtree, thus rendering diplomonads plus enteromonads monophyletic to the exclusion of retortamonads, but usually with low or no statistical support (BP 43/40/85 and 0.43 PP). Very similar results were obtained when a broader outgroup sampling was employed ('large' dataset; statistical support for Fornicata monophyly: BP 98/97/82 and PP 1; for enteromonad-Hexamitinae monophyly: BP 99/98/99 and PP 1; for diplomonad/enteromonad monophyly: BP */*/64, and not recovered in MrBayes analyses).

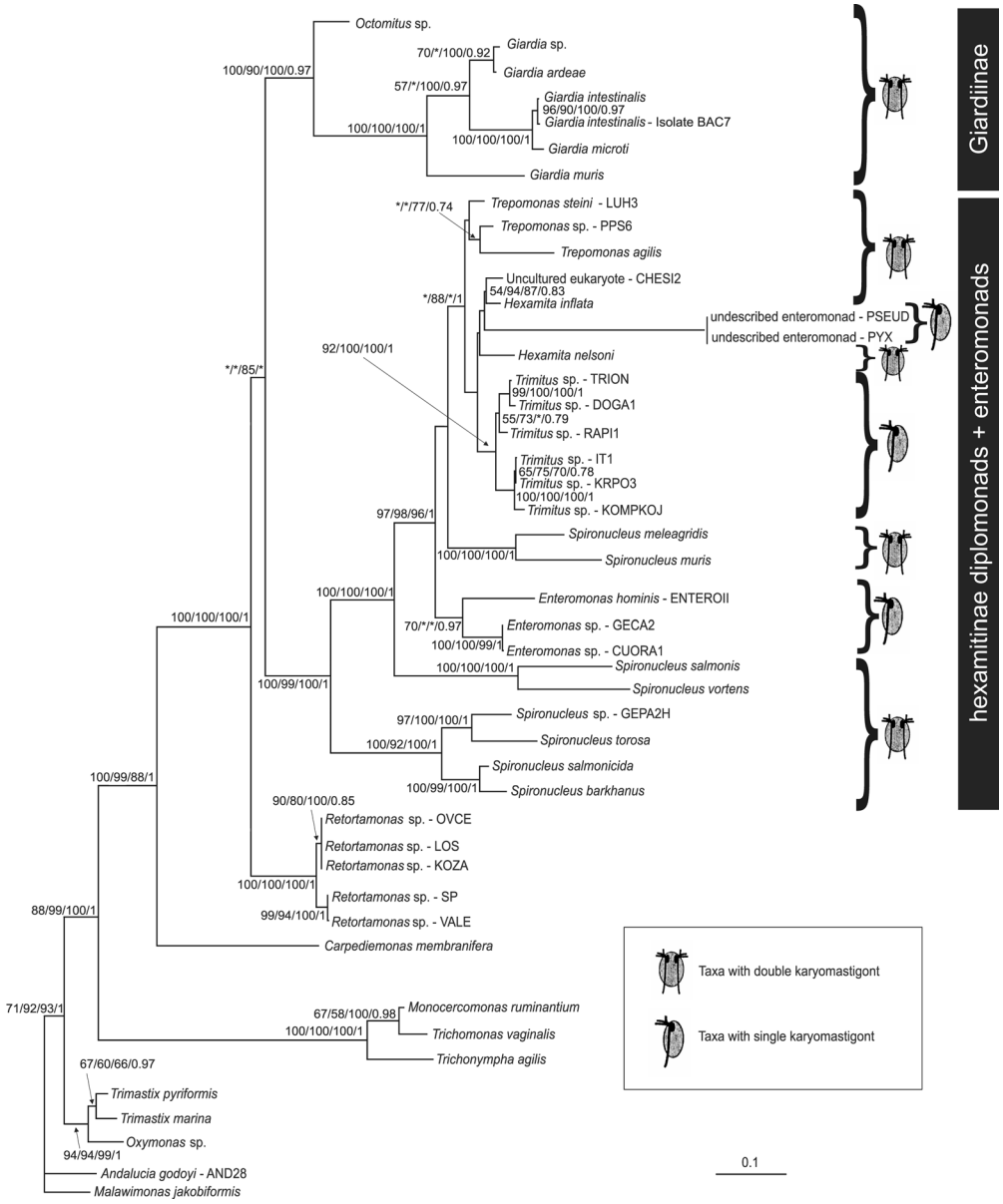


Figure 2.1. Maximum likelihood tree of Fornicata based on SSU rRNA genes (GTR + Γ + I model). Statistical support – ML bootstraps / RELL bootstraps / ML distance bootstraps / Bayesian posterior probability. Isolate PYX was identical in sequence with isolate PSEUD. Isolate PYX was therefore not included in the analysis but added to the tree by hand. Bootstrap support values <50% and posterior probabilities <0.7 are depicted by asterisks, or not shown.

Given the strong support for a clade consisting of all Hexamitinae and all enteromonads, we performed additional analyses with narrower taxon sampling to examine the internal relationships of this grouping (Figure 2.2). Within the Hexamitinae-enteromonad clade, members of the genus *Spironucleus* form three separated basal clades in all rooted analyses: i) *Spironucleus vortens* forms a clade with *S. salmonis*; ii) *S. barkhanus* branches with *S. torosa* and *S. salmonicida*; and iii) *S. muris* and *S. meleagridis* form a clade. Our isolate of *Spironucleus*, GEPA2H, branches as a sister of *S. torosa*, suggesting that it could be an isolate of this species, although the genetic distance between GEPA2H and *S. torosa* is greater than that between *S. barkhanus* and *S. salmonicida.* Representatives of the genus *Enteromonas* constitute a weakly supported or unsupported clade in analyses of the main and small datasets, but are polyphyletic in the ML tree estimated for the large dataset, because the isolate *Enteromonas hominis* branches as a sister to *Spironucleus muris* and *S. meleagridis.* The members of the genus *Hexamita* and uncultured eukaryote CHESI2 form a clade with the enteromonads of the genus *Trimitus*, as well as with the new enteromonad genus. All *Trimitus* isolates constitute a highly supported monophyletic group. *Hexamita*, by contrast, does not constitute a monophyletic group within that clade. Genus *Trepomonas* constitutes an unsupported clade (17% ML bootstrap support) in analyses of the main and small datasets and forms a paraphyletic group in analyses of the large data set.

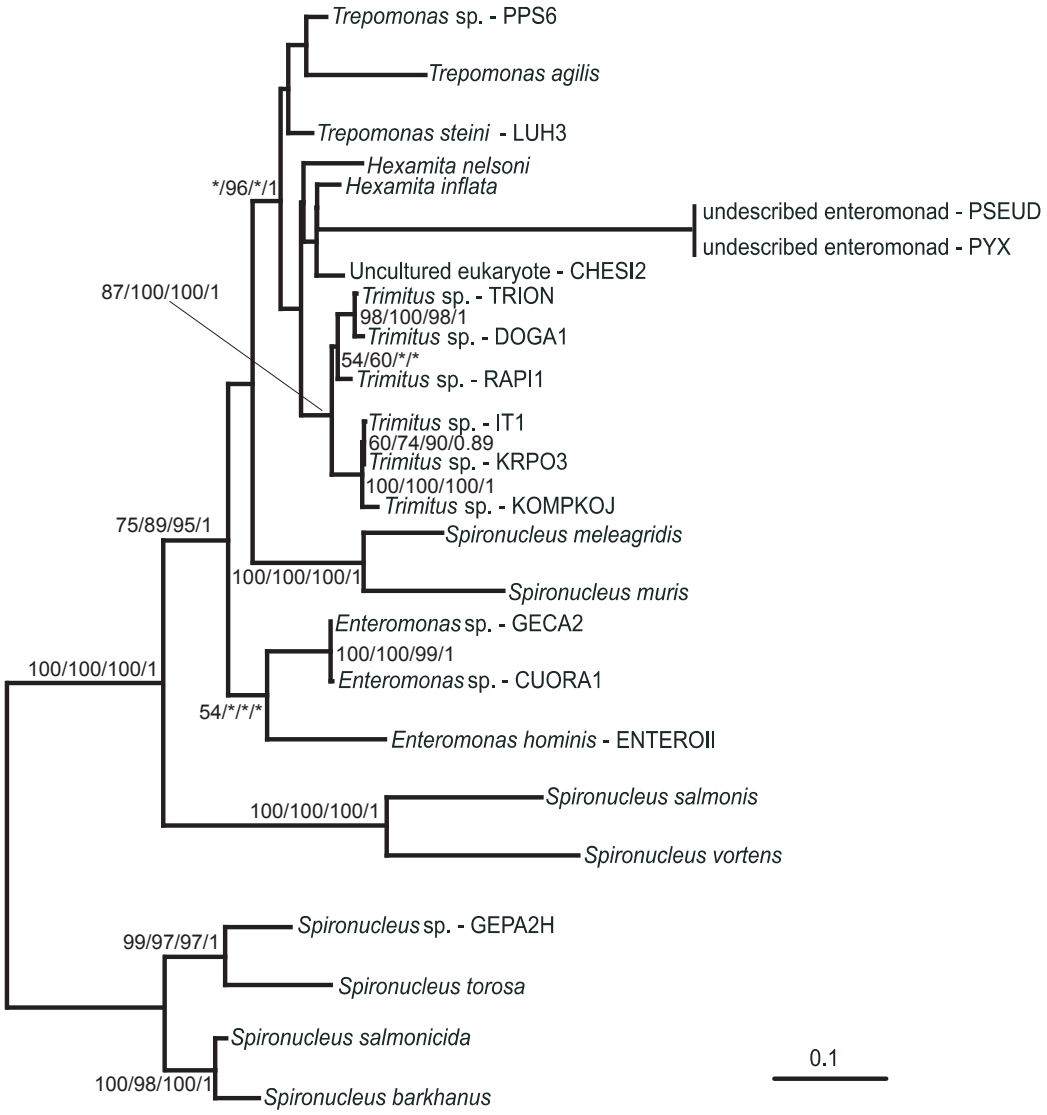


Figure 2.2. Maximum likelihood tree of the Hexamitinae-enteromonad clade based on SSU rRNA genes (GTR + Γ + I). Statistical support – ML bootstraps / RELL bootstraps / ML distance bootstraps/ Bayesian posterior probability. Isolate PYX was identical in sequence with isolate PSEUD. Isolate PYX was therefore not included in the analysis but added to the tree by hand. Bootstrap support values <50% and posterior probabilities <0.7 are depicted by asterisks, or not shown.

2.2.2. Concatenation of SSU rDNA, and Protein Gene Data

Sequences of HSP90 and/or α-tubulin coding genes for all outstanding sampled genera of diplomonads, except *Trepomonas* were obtained. For three taxa only one of the protein coding genes were obtained (Table 2.1). Interestingly, the alternative genetic code usage described earlier in all Hexamitinae (TAA and TAG encode glutamine instead of stop codons; 26), was also identified in the protein coding gene sequences from two out of three enteromonad genera (*Enteromonas* and the undescribed enteromonad genus). Analyses of HSP90 and α-tubulin genes support the monophyly of the genus *Enteromonas,* as *Enteromonas* GECA2 and *Enteromonas hominis* ENTEROII constitute a highly supported clade in both cases (Appendix C.2 and C.3).

The phylogenetic tree estimated for the concatenated SSU rDNA, HSP90 and α-tubulin data is shown in Figure 2.3. As in the SSU rDNA trees, *Spironucleus vortens* and *Spironucleus barkhanus* branch paraphyletically at the base of the Hexamitinae-enteromonad clade with high statistical support. *Spironucleus muris*, the genera *Enteromonas*, *Hexamita* and *Trimitus*, and the new enteromonad genus (represented by isolate PSEUD) constitute a clade with weakly resolved internal relationships. By contrast, *Giardia intestinalis* branches as a sister group of the Hexamitinae-enteromonad clade with high statistical support (BP 82), leaving retortamonads in a sister position to the whole 'diplomonads plus enteromonads' clade

Table 2.1. Sequences used in our analyses of concatenated genes. The sequences in bold font were sequenced for this study.

Isolates included in the analyses	Sequenced genes
Carpediemonas membranifera	SSU rRNA, HSP90, α-tubulin
Giardia intestinalis	SSU rRNA, HSP90, α-tubulin
Spironucleus vortens	SSU rRNA, **HSP90**, α-tubulin
Spironucleus barkhanus	SSU rRNA, **HSP90**, α-tubulin
Hexamita inflata	SSU rRNA, HSP90, α-tubulin
Spironucleus muris	SSU rRNA, α-tubulin
Retortamonas sp. – Vale	SSU rRNA, **HSP90**
Enteromonas hominis	**SSU rRNA, HSP90, α-tubulin**
Enteromonas sp. – GECA2	**SSU rRNA, HSP90, α-tubulin**
Trimitus sp. – TRION	**SSU rRNA, HSP90, α-tubulin**
undescribed enteromonad– PSEUD	**SSU rRNA, α-tubulin**

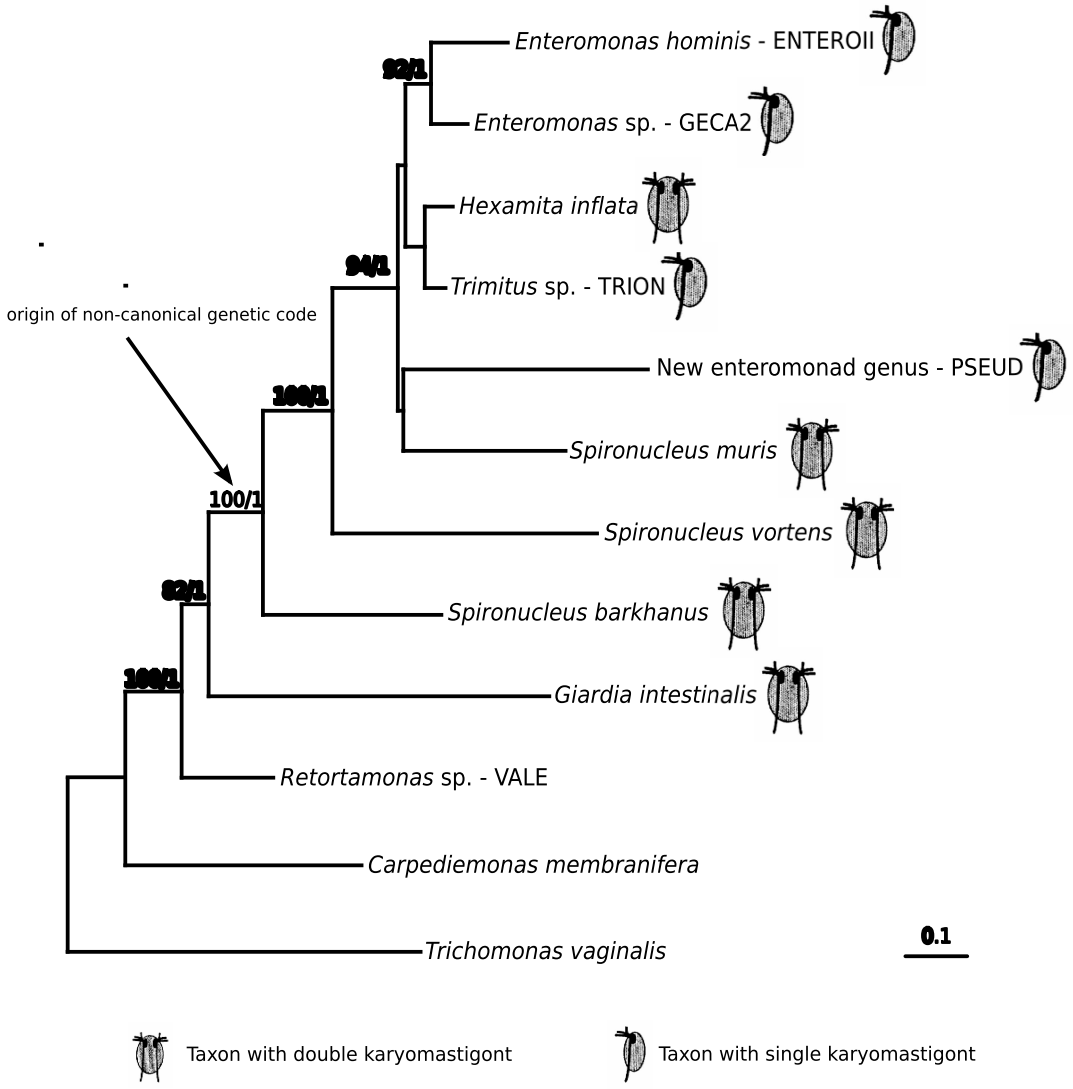


Figure 2.3. Bayesian tree of concatenated SSU rRNA, α-tubulin and HSP90 genes. Branch lengths shown are those estimated from the HSP90 partition of the concatenated data. Statistical support – Bayesian bootstraps / Bayesian posterior probability.

2.2.3 AU Tests

In order to explore further the evolutionary positions of enteromonads we performed AU tests of alternative topologies. In the case of the SSU rRNA gene data, we compared a set of 1000 reasonable trees, including both the ML tree and the topology of highest likelihood in which enteromonads were monophyletic. For the concatenated dataset, α-tubulin, and HSP90, all reasonable trees were examined (945, 945, and 15 trees respectively). The monophyly of enteromonads was not rejected by analyses of any one single gene. In the analysis of the concatenated dataset, however, the monophyly of enteromonads was rejected at the level of 5% ($p = 0.048$).

2.3. Discussion

2.3.1 Molecular Phylogeny of Fornicata

The analyses of SSU rRNA genes include the broadest taxonomic sampling of Fornicata published so far (as of 2008). These results are generally consistent with those of other recent studies (Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Kolisko *et al.*, 2005; Keeling and Brugerolle, 2006). The genus *Spironucleus* constitutes three separate branches close to the base of the Hexamitinae-enteromonad subtree. This topology is in good agreement with previous studies (Cavalier-Smith and Chao, 1996; Silberman *et al.*, 2002; Jørgensen and Sterud, 2007). However, all three clades of *Spironucleus* constitute long branches and their position at the base of the Hexamitinae-enteromonad subtree could be a long-branch attraction artifact

(Felsenstein, 2004). The presence of *Spironucleus* isolate GEPA2H within the clade of *Spironucleus barkhanus*, *S. torosa*, and *S. salmonicida* is noteworthy, since GEPA2H was isolated from the terrestrial tortoise *Geochelone pardalis,* while the latter three species infect marine teleosts. This argues against Jørgensen and Sterud's (2007) proposal that there are more-or-less distinct marine, freshwater, and terrestrial clades within *Spironucleus*. The internal relationships within the remainder of the Hexamitinae-enteromonad subtree were weakly or not supported and vary widely with the method of tree reconstruction and alignment. Further data will be required for a complete picture of the relationships within Hexamitinae.

Analyses of concatenated genes quite strongly support the monophyly of diplomonads plus enteromonads to the exclusion of retortamonads and *Carpediemonas*, albeit within the context of a smaller taxon sampling than is presently available for SSU rDNA alone. This supports previous morphological studies and analyses (Brugerolle, 1975; Siddall *et al.*, 1992; Simpson, 2003) but contrasts with some previous studies of SSU rDNA data, in which retortamonads branch weakly within diplomonads, as the sister group to Giardiinae (Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Kolisko *et al.*, 2005). The tendency for retortamonads to branch within diplomonads in SSU rDNA analyses is most probably an analysis artifact, but the cause of this artifact is not clear. Our preliminary analyses do not support the notion that either base composition heterogeneity (*Giardia* SSU rDNA is notable for its high GC content, however neither LogDet distance correction nor RY recoding results in a different well resolved topology) or simple long-branch attraction is to blame (data not shown). It is also

possible that SSU rDNA simply does not contain enough information to resolve the phylogenetic relationships among the ingroup taxa.

2.3.2 The Positions of Enteromonads

Enteromonads and Hexamitinae diplomonads form a monophyletic group to the exclusion of other Fornicata. This is consistent with previous molecular phylogenies including a much smaller sampling of enteromonads (Kolisko *et al.*, 2005) and one morphological study (Simpson, 2003). The clade has very strong statistical support in our analyses, and is further supported by a molecular synapomorphy: the non-canonical genetic code common to all studied Hexamitinae (Keeling and Brugerolle, 2006) appears also to be present in at least two of the three enteromonads for which we obtained protein-coding gene sequence data.

The most striking result of our study is the non-monophyly of enteromonads. The possibility of enteromonads being polyphyletic was suggested speculatively by Simpson (Simpson, 2003) on the basis of morphological data. The internal relationships of the Hexamitinae-enteromonad clade are weakly supported at present, and need further investigation. More extensive taxon sampling for studied protein-coding genes is necessary, as well as more protein-coding genes. Nonetheless, our results support the non-monophyletic status of enteromonads, as the monophyly of enteromonads was rejected by AU test with the concatenated data set. Interestingly, none of the single gene phylogenies themselves rejected enteromonad monophyly. The differing results from analysis of the concatenated dataset and the single-gene datasets could be caused either by an insufficient

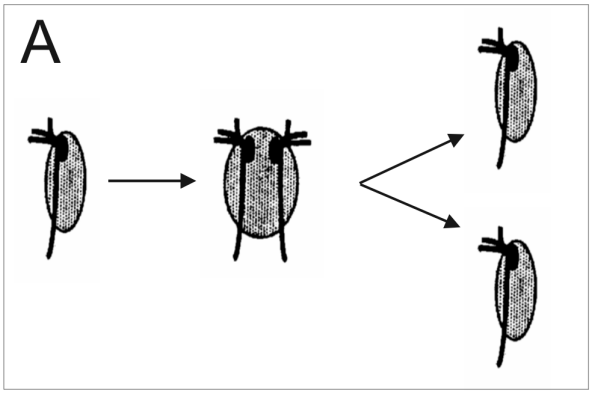
amount of data in single gene analyses or by conflicting signal between single gene phylogenies. However, none of the single gene ML trees was rejected in AU tests of the concatenated dataset, suggesting that there is no strong conflict between single gene phylogenies. It is reasonable to assume therefore that the rejection of enteromonad monophyly on the basis of concatenated data, as opposed to single gene analyses, is due to insufficient signal rather than conflicting single-gene data.

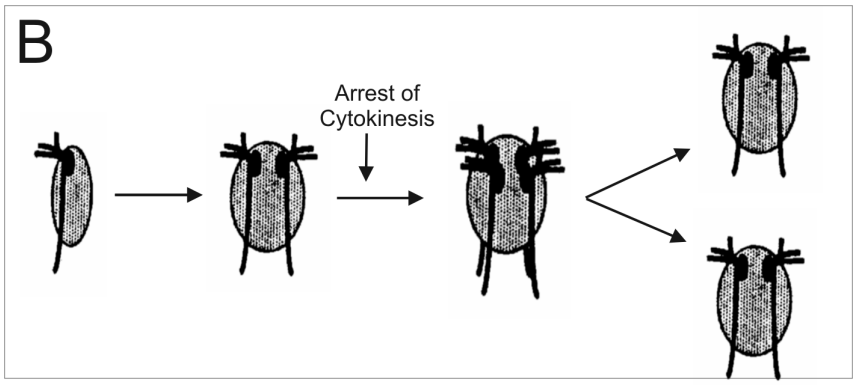
Many authors have treated diplomonads and enteromonads as taxa of equivalent rank (Brugerolle, 1975; Kulda and Nohynkova, 1978; Cavalier-Smith, 2003), implicitly or explicitly reflecting a widespread assumption that diplomonads and enteromonads are sister clades, or that enteromonads represent a paraphyletic group from which a monophyletic diplomonads group evolved (Siddall *et al.*, 1992; Cavalier-Smith, 2003). My analysis demonstrates, however, that many, and probably all enteromonads fall within diplomonads and specifically within Hexamitinae. Moreover, enteromonads represent a non-monophyletic group within Hexamitinae. There is no possible interpretation of these results that would allow recognition of diplomonads and enteromonads as separate taxa, without at least one of them being polyphyletic. Therefore, it is suggested that the taxon Diplomonadida and any of its effective synonyms be considered to include enteromonads, and that the taxa Enteromonadida and Enteromonadinae no longer be used. The term 'enteromonads' should be taken to have a purely descriptive, and not taxonomic, meaning.

2.3.3 Evolution of Single and Double Karyomastigont Cell Organization

In a previous study (Kolisko *et al.*, 2005) it was shown that at least one enteromonad clade branches within diplomonads, and branches with or within the diplomonad subgroup Hexamitinae, suggesting either that enteromonads are secondarily simplified, possibly in one evolutionary event, or that the double karyomastigont morphology characteristic of diplomonads arose more than once during their evolution. Our present study indicates that enteromonads do not constitute a monophyletic group within Hexamitinae, and thus switches between single- and double-karyomastigont morphology must have occurred several times during evolution, irrespective of the direction in which these switches occurred. If we assume only one direction of change within the diplomonad-enteromonad clade and assuming, for argument's sake, that our topologies are correct, there are two basic possibilities: either the double-karyomastigont morphology arose many times in closely related groups (at least seven times according to the SSU rRNA gene tree, Figures 2.1 and 2.2, or at least five times, according to the multigene analyses, Figure 2.3) or enteromonads arose by secondary reduction from double-karyomastigont ancestors at least three times independently.

There are three potential scenarios describing the switch between single and double karyomastigont morphologies (Figure 2.4). For comparison, the standard cell cycle of an enteromonad is depicted in Figure 2.4A.





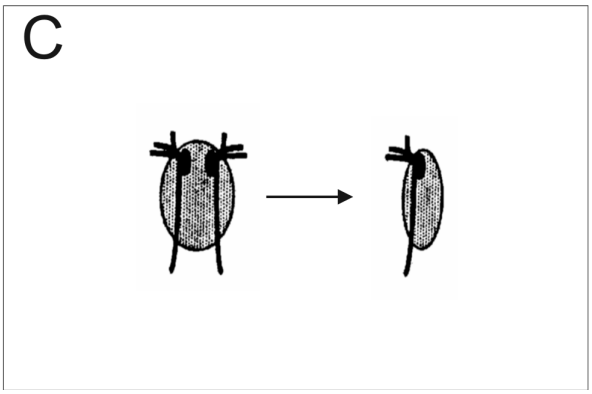


Figure 2.4. A. "Standard" cell cycle of an enteromonad cell; cell divides after karyokinesis. **B.** Model of evolutionary change from single karyomastigont morphology to double karyomastigont morphology by arrest of cytokinesis. The cell does not divide after the first karyokinesis and secondary karyokinesis results in a cell with four karyomastigonts. This cell then divides into two cells, each with a double karyomastigont. **C.** Model of evolutionary change from double karyomastigont morphology to single karyomastigont morphology, either by cytokinesis without karyokinesis, or by fusion of nuclei. (modified from Siddall, Hong and Desser 1992).

1. The switch from single karyomastigont morphology to double karyomastigont morphology was envisaged by Siddall, Hong and Desser (1992) as a change in the relative timing between karyokinesis and cytokinesis (Figure 2.4B). Under this model a cell with a single karyomastigont goes through nuclear and mastigont division and is prepared for cell division. Cell division is arrested, however, and the cell goes through another nuclear and mastigont division, resulting in a cell with four karyomastigonts. The cell with four karyomastigonts then divides into two cells, each with a double karyomastigont.
2. The opposite switch (figure 2.4C), from a double to single karyomastigont, could also be explained as a change in the timing of karyokinesis and cytokinesis. Cells with a double karyomastigont morphology could just go through cell division without nuclear and mastigont division. Another similar event would involve a cell with four karyomastigonts (i.e., a cell normally with a double karyomastigont prepared for cell division) dividing into four daughter cells instead of two.
3. The last scenario suggests fusion of the nuclei in a cell with the double-karyomastigont morphology. The resulting cell could then lose the second mastigont, resulting in a cell with a single-karyomastigont morphology.

These scenarios have different strengths and weaknesses in terms of plausibility. Scenario one seems implausible from a phylogenetic perspective, as it requires a very large number of parallel evolutions of the distinctive doubled cell morphology within the diplomonad-enteromonad clade. Scenario two invokes many

fewer evolutionary events, but requires that the two nuclei of the parental diplomonad cell be nearly identical (or, at a minimum, that at least one of the nuclei retains all essential genes), otherwise the single karyomastigont progeny would not be viable. Scenario three requires the same number of evolutionary transitions as scenario two, and is compatible with a diplomonad parent that had non-identical and essential nuclei, but is otherwise a more complex mechanism. Therefore, the key to understanding the evolution of nucleus number in diplomonads lies in knowing whether the nuclei of diplomonads are identical. Naïvely, it might be expected that the nuclei are non-identical, since one copy of each nucleus is transmitted to each daughter cell during division (Tumova *et al.*, 2006), and, in principle at least, the two nuclei would represent separate lineages once the diplomonad state has been fixed. This would allow essential genes to be lost from one or the other nucleus, such that both nuclei would soon be required for the cell lineage to persist and any uninucleate progeny would be unviable. However, this could be avoided if there were mechanisms that frequently generated cells in which both nuclei were copies of a single parental nucleus (most likely a sexual process), or genetic exchange between the two nuclei.

Empirical data on the nature of diplomonad nuclei are limited and conflicting. The human parasite *Giardia intestinalis* is the only diplomonad whose molecular and cellular biology has been studied in any particular detail. Another study (Yu *et al.*, 2002) used probes against selected genes to indicate that each nucleus of *G. intestinalis* contains a complete set of genetic information. Bernarder et al. (2001) have also deduced from results of FACS analysis that the two nuclei of *G. intestinalis*

are diploid. The genome sequence of *Giardia intestinalis (lamblia)* strain WB clone C6 was published by Morrison *et al*. (Morrison *et al.*, 2007), and minimal heterozygosity was detected. Recently, a process of physical transfer of DNA between nuclei was reported in *Giardia* cysts (Poxleitner *et al.*, 2008). However, Tumova *et al.* (2006) reported that the two nuclei of *G. intestinalis* possess different numbers of chromosomes. The focus on *Giardia* is unfortunate in some respects, as *Giardia* is a highly specialized parasite with an organization of cytoskeletal components (at least) that is substantially different from other diplomonads. It is possible that results obtained for *Giardia* may not be applicable to other diplomonads. *Giardia* also represents the diplomonad group that is most distantly related to the various enteromonad taxa. It is of great interest to compare the nuclei in one or more Hexamitinae diplomonads (e.g., *Spironucleus*, *Hexamita* or *Trepomonas*).

There are some observations that would support scenario 1, albeit indirectly. Firstly, the populations of enteromonads often contain some double individuals resembling *Hexamita* in their morphology, with two fully developed karyomastigonts and no apparent signs of cytokinesis (Brugerolle, 1975; Brugerolle, 1986, M. K. personal observation). This may suggest that there is some general tendency for delayed or arrested cytokinesis. Secondly, a recent study of the flagellar cycle of *Giardia intestinalis* has shown that the two karyomastigonts are not independent, as basal bodies migrate between the two karyomastigonts (Nohynkova *et al.*, 2006). This means that the flagellar maturation cycle would be corrupted in cells that switched back to a single karyomastigont.

Our molecular phylogenies contradict any scenario invoking just one unique evolutionary transition between single and double karyomastigont morphologies within diplomonads. According to the inferred topology, the most parsimonious scenario would have one transition from single to double karyomastigont morphology at the base of diplomonads and several independent reversals to the single karyomastigont morphology. However, taking into account possible inaccuracies in our estimated tree, principally involving *Spironucleus*, and the limited data on diplomonad cell biology, several independent transitions from single to double morphology (and no reversals) cannot be excluded. More data on the molecular and cellular biology of diplomonads in addition to *Giardia* will be necessary for understanding the enigmatic evolution of the double karyomastigont of diplomonads.

2.3.4 Conclusions

Our analyses of SSU rRNA, HSP90 and α-tubulin genes strongly positioned all enteromonads within Hexamitinae diplomonads and showed that enteromonads do not constitute a monophyletic group. These results suggest that transformations between single- and double-karyomastigont morphologies have occurred several times during the evolution of diplomonads, however, it is not possible to confidently determine the direction of these switches without more information about the cellular and molecular biology of diplomonads and enteromonads. We suggest that the high level taxa Enteromonadida, Enteromonadidae and Enteromonadinae should be abandoned and the genera *Enteromonas* and *Trimitus* should be considered as

members of Hexamitinae diplomonads. The term 'enteromonad' should have a purely utilitarian meaning - Diplomonadida with a single karyomastigont.

2.4 Materials and Methods

2.4.1 Cultures

All Enteromonas isolates (except KK and IT1) and isolate of *Trepomonas steini* were culture by I. Cepicka (Charles University, Prague); isolates KK and IT1 were obtained in direct collaboration between the author and I. Cepicka. Isolates used in this study are summarized in Table 2.2 All enteromonad isolates were obtained from animal guts or feces, except isolates KOMPKOJ and IT1, which were free living. *Trepomonas steini* and *Trepomonas* sp.-PPS6 were isolated from anoxic fresh water sediments. A culture of *Spironucleus vortens* was obtained from the American Type Culture Collection (ATCC #50386). DNA from *Spironucleus muris* was isolated from purified cysts obtained from the intestine of a SCID laboratory mouse (provided by J. Kulda). Xenic cultures of enteromonads were grown in Dobell-Leidlaw biphasic medium (Dobell and Leidlaw, 1926) and in TYSGM medium (Clark and Diamond, 2002) without tween and mucin at 21 °C, 27 °C and 37 °C. *Spironucleus vortens* was grown axenically in TYI-S33 medium as modified for *Giardia* at 27 °C (Keister, 1983). *Trepomonas* sp.-PPS6 and *Trepomonas steini* were grown in cerophyll medium (ATCC #802) at 21 °C (Table 2.2). DNA from *Spironucleus* sp. GEPA2H and uncultured eukaryote CHESI2 was isolated from crude cultures (provided by I. Cepicka and M. Uzlikova).

Table 2.2. Diplomonad and enteromonad isolates used in our study for sequencing. DL – Dobell and Leidlaw medium, 802 – cerophyll

Isolate	Strain	Source	Medium	T (°C)	Other Eukaryotes
Enteromonas sp.	GECA2	*Geochelone carbonaria*	DL	27	none
Enteromonas sp.	CUORA1	*Cuora amboinensis*	DL	27	none
Enteromonas hominis	ENTEROII	*Homo sapiens*	DL	37	none
Enteromonad	PSEUD	*Trachemis scripta elegans*	DL	27	*Retortamonas* sp.
Enteromonad	PYX	*Pyxidea mouhoti*	DL	27	Parabasalids
Trimitus sp.	KOMPKOJ	Compost, Kojčice, Czech Republic	TYSGM	21	none
Trimitus sp.	IT1	Pond in Italy	TYSGM	21	none
Trimitus sp.	DOGA1	*Doagania* sp.	DL	27	none
Spironucleus vortens	ATCC# 50386	ATCC	TYI	27	none
Spironucleus sp.	GEPA2H	*Geochelone pardalis*	DL	n/a	n/a
Trepomonas steini	LUH3	Flood, Vltava River, South Bohemia, Czech Republic	802	rt	Sawyeria sp.
Trepomonas sp.	PPS6	Point Pleasant Park pond, Halifax, NS, Canada	802	21	none
Uncultured eukaryote	CHESI2	*Chelodina* sp.	n/a	n/a	n/a

[1]The SSU rDNA sequence of isolate PYX was identical with isolate PSEUD. Isolate PYX was therefore not included in the phylogenetic analyses.

2.4.2 Gene Amplification and Sequencing

SSU rRNA sequences for isolates CHESI2, IT1, KK, LUH3, PPS6 and *S.muris* were obtained by the author. The author also completed partial HSP90 sequences of *S. vortens* and *S. barkhanus.* All other sequence data were obtained from previous work (M. Kolisko, Masters Thesis, Charles University, Prague) and by other collaborators. Genomic DNA was isolated using a High Pure PCR template kit (Roche Applied Science, UK) or using CTAB and organic extractions (Clark, 1992). SSU rDNA sequences were amplified by PCR using primers 'EntUnvF' and 'EntUnvR' (Kolisko *et al.*, 2005) or universal eukaryotic primers (Medlin *et al.*, 1988). In the case of the 'new enteromonad genus' isolate 'PSEUD', the culture also contained a retortamonad species. The SSU rDNA segments from both eukaryotes were amplified and partially sequenced. Specific primers for the 'new enteromonad genus', DimA (5'-AGTCAAAGATTAAAACATGCATAT-3') and DimB (5'-TCCTCTAAGCCTTCTAGTTCGTGCAAA-3') were then designed and used for amplification of the SSU rDNA from isolate 'PYX', which is an enteromonad closely related to isolate 'PSEUD'. A specific forward primer (SSUSmur20F 5'-AACTGCGGACGGCTCATT-3') was designed for *S. muris* and used with the universal eukaryotic reverse primer.

Genes for α-tubulin were amplified using primers AtubA and AtubB (Edgcomb *et al.*, 2001) and then by nested PCR with primers α-tubF1 and α-tubR1 (Moriya *et al.*, 2001). HSP90 genes were amplified using primers H90100X (Simpson *et al.*, 2002a) and H90910XR (Simpson et al., 2006). The annealing temperatures used were 45-53°C, 45-50°C and 48-53°C for SSU rDNA, α-tubulin and HSP90,

respectively. SSU rDNA amplicons were sequenced directly where possible. Otherwise, major PCR fragments of the expected sizes were subcloned (TOPO TA cloning kit for sequencing, pCR4-TOPO vector, Invitrogen, USA; or pGEM-T Easy vector cloning kit, Promega, USA) and several clones (2-6) were partially sequenced. Obtained sequences were then subjected to BLAST searches (Altschul *et al.*, 1990) to confirm their identity. At least one of the positive clones was fully sequenced bidirectionally by primer walking. All sequences obtained during this study are deposited in GenBank [GenBank: EF551168 - EF551186, EU043230, AY921407 and AY921408].

2.4.3 Phylogenetic Analyses

2.4.3.1 Alignments

All alignments used in this study were constructed using the program ClustalX 1.83 (Thompson *et al.*, 1997) followed by manual editing in the program BioEdit 7.0.5.3 (Hall, 1999) and are available upon request (see Appendix C.4 for details about the sequences used).

2.4.3.2 SSU rRNA Gene

Two data sets were constructed including all near-full-length Fornicata sequences, except some redundant close relatives within *Giardia* and retortamonads, plus an outgroup consisting of either i) a broad diversity of eukaryotes (large dataset), or ii) a few supposed close relatives of Fornicata – the excavate groups Parabasalia, *Trimastix*, Oxymonadida, *Malawimonas* and *Andalucia* (main dataset). These

datasets included 887 and 1041 well-aligned sites, respectively. An additional dataset was generated that included only Hexamitinae and enteromonads, and also included 1041 sites (small dataset).

Each dataset was analyzed using several likelihood-based phylogenetic methods. The model of sequence evolution was selected by the Akaike information criterion, as implemented in the program Modeltest 3.7 (Posada and Crandall, 1998). The general time reversible model of nucleotide substitution was used, with among-site rate variation modeled by a gamma distribution and a proportion of invariable sites (GTR + Γ + I model), with the gamma distribution approximated by 4 equiprobable discrete categories. Maximum likelihood (ML) analyses were performed using the program PAUP*4B10 (Swofford, 2002), with 10 random taxon additions followed by tree bisection and reconnection branch rearrangements, while ML bootstrap support (200 replicates) was estimated using PAUP*4B10 (10 random taxa additions followed by TBR; for the large dataset only the ML bootstrap analysis was instead performed using the program IQPNNI 3.0.1. (Vinh and von Haeseler, 2004), and LRSH-RELL bootstrapping (1000 replicates) was performed using the program Treefinder (version: February 2007) (Jobb *et al.*, 2004). The model of sequence evolution used was the same for all ML analyses. Least squares distance trees were estimated from ML distances using PAUP*4B10 and bootstrapped with 1000 replicates (each searched using 10 replicates of random taxon addition with TBR branch swapping). The Bayesian analysis was performed using the program MrBayes 3.1.2 (Huelsenbeck, 2000), using the GTR + Γ + I model with two runs, each

with four independent chains running for $3x10^6$ generations (a burn-in of $5x10^5$ generations was used), with default heating parameter and sampling frequency.

2.4.3.3 Analyses of Protein Coding Genes

The HSP90 and α-tubulin amino acid datasets included all available Fornicata sequences and an extensive eukaryotic outgroup consisting of representatives from major eukaryotic groups. The trimmed alignments included 370 sites for α-tubulin and 493 sites for HSP90. Both datasets were analyzed using the WAG + Γ + I model (Whelan and Goldman, 2001). The WAG matrix was selected over other substitution matrices by the Akaike information criterion, as implemented in the program ProtTest 1.4 (Abascal *et al.*, 2005). For each, the ML tree was estimated and bootstrap support (500 replicates) was estimated using IQPNNI 3.0.1, while LRSH-RELL bootstrap support (1000 replicates) was determined using Treefinder (version: February 2007). In addition, a Bayesian analysis (WAG + Γ + I model) was performed using MrBayes 3.1.2, with four independent chains running for 2 $x10^6$ generations, and with a conservative burn-in of $5x10^5$ generations.

2.4.3.4 Analyses of Concatenated SSU rDNA and Protein Sequences

The concatenated alignment of SSU rDNA, α-tubulin and HSP90 genes was analyzed using the program MrBayes 3.1.2 (two runs each with four independent chains running for 5 $x10^6$ generations with a burn-in of 1.5 $x10^6$ generations), with among-site rate variation for each gene modeled by a discrete approximation of a gamma distribution, proportion of invariable sites and a covarion model. The GTR

substitution model was used for the SSU rDNA partition and the WAG substitution matrix (Whelan and Goldman, 2001) was used for the protein coding genes. The branch lengths, α parameter, proportion of invariable sites and parameter for switching rates in the covarion model were estimated separately for each gene (both runs converged to the same level). The branch lengths in the depicted tree are those estimated for the HSP90 partition of the data (The alternative of displaying the average of the estimated branch lengths over all three genes was not followed on the grounds that this average does not reflect any actual parameter examined under the model of evolution we used). In addition to examining posterior probabilities we performed a full bootstrap analysis with 100 replicate samples. Each gene was re-sampled independently (using the program seqboot from the Phylip package (Felsenstein, 2005) and then each bootstrap sample was created by concatenating one replicate from each gene. Each bootstrap replicate was analyzed under the same conditions as the starting dataset but using only 2.5 $x10^5$ generations (a burn-in of 50000 generations was used). The consensus tree was made for each bootstrap replicate in MrBayes 3.1.2. The bootstrap consensus tree was then estimated from the 100 resulting trees using the program CONSENSE from the package Phylip 3.67 (Felsenstein, 2005).

2.4.3.5 Testing of Topologies

The topologies were compared using 'Approximately Unbiased' (AU) tests implemented in the program CONSEL 1.19 (Shimodaira and Hasegawa, 2001). We performed separate AU tests on four datasets – 1. SSU rRNA genes, 2. α-tubulin, 3.

HSP90, and 4. the concatenated dataset (with the missing genes treated as missing data). For the SSU rRNA test, the small dataset, which includes only Hexamitinae and enteromonads was used. For the α-tubulin, HSP90 and combined datasets, the alignments used for estimating the ML tree were used, except that taxa other than Hexamitinae and enteromonads were excluded. For AU tests using the SSU rRNA gene data we generated a set of reasonable trees (999 trees). This set was generated by saving the 999 trees with the highest likelihood found during ML analyses in PAUP*4B10 (10 random sequence additions plus TBR). The tree representing monophyletic enteromonads was generated using a constraint search in PAUP*4B10 (10 random addition replicates plus TBR). In the case of α-tubulin, HSP90, and concatenated analyses, we included all possible trees that were consistent with a constraint where nodes corresponding to those that had received 100% bootstrap support in the concatenated genes analysis were fixed. Site likelihoods were calculated using PAUP*4B10 for the SSU rRNA gene data, and using the PAML package (Yang, 1997) for protein data. For the concatenated gene analysis site likelihoods were generated separately for all three genes and then concatenated prior to analysis in CONSEL 1.19.

Chapter 3

New Diversity of Deep Branching Relatives of Diplomonads

The chapter was published (Kolisko et al. 2010)

3.1 Introduction

Over the last decade, conventional culturing approaches have led to the discovery of a selection of novel eukaryotic organisms of major evolutionary importance. For example *Breviata anathema*, a small amoeboid flagellate, was shown to be a deep branch attached to the supergroup Amoebozoa, and important for understanding the unikont/bikont hypothesis and consequently evaluating hypotheses for the location of the root of the eukaryote tree (Walker *et al.*, 2006; Minge *et al.*, 2009; Roger and Simpson, 2009). *Capsaspora owczarzaki* is a single-celled organism that is most closely related to choanoflagellates and/or ichthyosporeans and hence is one of the key taxa for understanding the evolution of multicellularity in animals and fungi (Hertel *et al.*, 2002; Ruiz-Trillo *et al.*, 2004). *Chromera velia* is a photosynthetic relative of the often-plastid-bearing, but non-photosynthetic apicomplexan parasites (Moore *et al.*, 2008). In addition, some organisms, such as centrohelids, *Telonema* and *Fonticula* were known for some time, but have only recently been studied using molecular techniques and have proven to be of particular phylogenetic importance (Burki *et al.*, 2009). Over a similar time-period environmental PCR approaches have revealed a number of additional and genuinely novel significant lineages (Massana and Pedrós-Alió, 2008). The most important perhaps include the several 'MAST

lineages' of uncultured, probably heterotrophic marine stramenopiles (Massana *et al.*, 2004; Massana *et al.*, 2006) and the mysterious picobiliphytes/biliphytes (Not *et al.*, 2007; Cuvelier *et al.*, 2008). On the other hand, the last decade has also seen the widespread incorporation of many morphologically distinct eukaryote lineages into existing major groups (e.g., Cercozoa and Bicosoecida – O'Kelly and Nerad, 1998; Cavalier-Smith and Chao, 2003a; Bass and Cavalier-Smith, 2004; Cavalier-Smith and Chao, 2006), as well as the refutation of several early claims of substantial novelty of major lineages from environmental PCR studies (Berney *et al.*, 2004; Cavalier-Smith, 2004). These latter trends tend to suggest that much of the major-lineage-level diversity of eukaryotes is already known. The extent to which this is accurate has important consequences for understanding eukaryote diversity and cell evolution.

Diplomonads, such as the human parasite *Giardia intestinalis*, are amongst the most interesting and problematic groups of microbial eukaryotes from an evolutionary perspective. Diplomonads are anaerobic or microaerophilic heterotrophic flagellates that live either in anoxic sediments or water bodies, or as parasites or commensals (Kulda and Nohynkova, 1978). They do not possess classical mitochondria and, for a long time, were considered to be ancestrally amitochondriate (Cavalier-Smith, 1983). This, in combination with their tendency to branch at the base of the eukaryotic trees estimated from small subunit ribosomal RNA (SSU rRNA) and translation elongation factor genes (Sogin *et al.*, 1989; Kamaishi *et al.*, 1996), led to a widespread view that diplomonads were "primitive eukaryotes". However, later studies have shown the presence of genes of mitochondrial origin in diplomonad genomes (Roger *et al.*, 1998; Tachezy *et al.*,

2001) and tiny mitochondrion-related organelles called mitosomes were subsequently discovered in *Giardia intestinalis* (Tovar *et al.*, 2003). Moreover, the position of diplomonads at the base of the eukaryotic tree is now widely considered to be the result of a long branch attraction artifact stemming from rapid gene sequence evolution in this group (Brinkmann *et al.*, 2005; Philippe *et al.*, 2005). Thus the true phylogenetic position and evolutionary history of diplomonads remains incompletely understood and there is considerable interest in using comparative genomics and cell biological approaches to better understand diplomonad evolution (Hampl *et al.*, 2009).

Until recently, the closest known relatives of diplomonads included retortamonads, which are poorly studied, mostly commensal organisms (Kulda and Nohynkova, 1978), and the more distantly related genus *Carpediemonas*. *Carpediemonas* is a small bacterivorous flagellate found in anoxic marine sediments that was described and characterized relatively recently (Ekebom *et al.*, 1996; Simpson and Patterson, 1999; Simpson *et al.*, 2002c). *Carpediemonas* tends to constitute a shorter branch than diplomonads in molecular phylogenies, and possesses double-membrane-bounded mitochondrion-like organelles that are considerably larger than the mitosomes of *Giardia* (Simpson and Patterson, 1999; Simpson *et al.*, 2002c; Simpson et al., 2006). This makes *Carpediemonas* potentially very important for resolving the phylogenetic position of diplomonads and understanding the reductive evolution of mitochondria-related organelles.

For a long time *Carpediemonas* appeared to be a phylogenetically isolated lineage, although, very recently two '*Carpediemonas*-like' organisms have been

described, *Dysnectes brevis* (Yubuki *et al.*, 2007) and *Hicanonectes teleskopos* (Park *et al.*, 2009). This study reports the isolation and culturing of eighteen new isolates of '*Carpediemonas*-like organisms' (CLOs) from oxygen-poor saline and marine habitats around the world. These new isolates are sufficiently distinct in morphology and/or in molecular comparisons to represent several new genus-level groups. We now must envisage CLOs as a phylogenetic cloud of at least six major lineages at the base of the diplomonad-retortamonad-*Carpediemonas* clade (e.g., Fornicata). The existence of such a wide diversity of CLOs was unanticipated, based on both historical microscopy/culturing efforts and recent environmental PCR surveys. This example suggests that a considerable number of evolutionarily important lineages of microbial eukaryotes may still be undiscovered and that culturing approaches remain a valuable avenue for understanding the scope of microbial eukaryotic diversity.

3.2 Results

3.2.1 New Isolates

Eighteen new isolates of '*Carpediemonas*-like organisms' (CLOs) were cultured from marine/saline locations around the world (Table 3.1). Light microscopy observations of the new isolates show that they usually have a typical excavate morphology, e.g., a visible feeding groove associated with the posterior flagellum (Figure 3.1). Most, but not all, broadly resemble *Carpediemonas membranifera* and *Dysnectes brevis* in that they are small bean- or crescent-shaped cells that swim

relatively slowly with slow rotation or no rotation. One isolate, BICM, is very similar in appearance to the original culture of *Carpediemonas membranifera* (isolate QB) (Figure 3.1 A and Table 3.2). Eleven of the new isolates (isolates GR1, PPP15C, LARNAKA2, NY0173, NY0166, ALLEPEYI, KR3, KR4, KR7, KR8 and GSML) are indistinguishable from the previously described morphospecies *Carpediemonas bialata*, which has not been cultured before, and is little-studied (Figure 3.1 C-D, Table 3.2). The other isolates all appear to belong to undescribed species as they neither correspond morphologically to previously described species, nor are they very similar at the molecular level (Table 3.3). Isolate SIVOTA, which has a short posterior flagellum, resembles Dy*snectes brevis,* except the cell shape tends to be more elongated (Figure 3.1 E-F, Table 3.2). Isolates CL and NC are bean-shaped cells with a visible groove and a free-trailing posterior flagellum that is approximately twice the length of the cell (Figure 3.1 H-J, Table 3.2). Isolates NY0171 and PCE differ substantially from the isolates discussed above – both are oval-shaped cells with a slightly curved feeding groove and they rotate when swimming (Figure 3.1 K-N, Table 3.2). Isolate PCS is rod-shaped with a flattened area at the anterior end of the cell (possibly the remnant of the excavate groove), where beats the single visible flagellum. Thus PCS differs substantially from previously described species and from all other new isolates (Figure 3.1 O-P, Table 3.2).

Table 3.1. Sampling sites and culture media for all cultured *Carpediemonas*-like organisms. Previously described isolates are marked with *. Media formulations are given in the Appendix D.1.

Isolate	Clade	Sampling site	Sampling environment	Media	Taxonomic classification
SIVOTA	CL1	Sivota, Greece (39°23'N, 20°14'E)	Littoral anoxic sediments	802SW	*Dysnectes* sp.
NY0165*	CL1	Kagoshima, Japan (31˚20'N, 130˚63'E)	Littoral anoxic sediments	NM	*Dysnectes brevis*
PCE	CL2	Prince cove, USA (41° 38'N, 70° 24'W)	Littoral anoxic sediments	SW1773	New genus A
NY0171	CL2	Ishigaki island, Japan, (24˚48'N, 124˚23'E)	Littoral anoxic sediments	NM	New genus A
PCS	CL3	Prince Cove, USA (41° 38'N, 70° 24'W)	Littoral anoxic sediments	SW1773	New genus B
SB*	CL3	BC, Canada (48°46'N, 123°28'W)	Littoral anoxic sediments	T/S	*Hicanonectes teleskopos*
BICM	CL4	BC, Canada (48°46'N, 123°28'W)	Littoral anoxic sediments	3%LB in 50%SW	*Carpediemonas membranifera*
QB*	CL4	Quibray bay, Australia (34°02'S, 151°10'E)	Littoral anoxic sediments	3%LB in 50%SW	*Carpediemonas membranifera*
NC	CL5	Nebraska, USA (40° 95'N, 96° 72'W)	Inland salt marsh sediments	802SW/ Horse serum	New genus C
CL	CL5	Mahone Bay, Canada (44°26'N, 64°21'W)	Littoral anoxic sediments	T/S	New genus C
GR1	CL6	Marmara, Greece (38°08'N, 22°21'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
PPP15C	CL6	Halifax, Canada (44°37', 63°33')	Littoral anoxic sediments	NM	*Kipferlia bialata*
LARNAKA 2	CL6	Larnaka, Cyprus (34°54'N, 33°38'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
NY0173	CL6	Sagami Bay, Japan, 35°0.09'N, 139°13.51'E	Deep sea anoxic sediments (~1.1km)	NM	*Kipferlia bialata*

Table 3.1. Continued

NY0166	CL6	Kagoshima, Japan (31˚20'N, 130˚63'E)	Littoral anoxic sediments	NM	*Kipferlia bialata*
ALLEPEYI	CL6	Allapuzha, India (28°07'N, 76°19'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
KR3	CL6	Adelianos Kampos, Greece (35°22'N, 24°32'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
KR4	CL6	Adelianos Kampos, Greece (35°22'N, 24°32'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
KR7	CL6	Adelianos Kampos, Greece (35°22'N, 24°32'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
KR8	CL6	Adelianos Kampos, Greece (35°22'N, 24°32'E)	Littoral anoxic sediments	802SW	*Kipferlia bialata*
GSML	CL6	Apalachee bay, USA (~30° 04'N, 84° 10'W)	Detritus in shipment of sea urchins	802SW	*Kipferlia bialata*

Figure 3.1. Light microscopic photographs of *Carpediemonas* and *Carpediemonas*-like organisms. A. *Carpediemonas membranifera* QB (Simpson, A.G.B., unpublished), B. *Carpediemonas membranifera* BICM, C. *Kipferlia bialata* n. gen, n. comb (source Micro*scope, original micrograph by Won Je Lee), D. *Kipferlia bialata* n. gen, n. comb., isolate KR8, E. *Dysnectes brevis* NY0165, F. *Dysnectes* sp., SIVOTA G. *Hicanonectes teleskopos* SB, H. *Carpediemonas*-like organism CL, I.-J. *Carpediemonas*-like organism NC, K.-L. *Carpediemonas*-like organism PCE, M.-N. *Carpediemonas*-like organism NY0171, O.-P. *Carpediemonas*-like organism PCS. Photographs of previously described organisms, *Carpediemonas membranifera* QB, *Kipferlia bialata* (source Micro*scope, original micrograph by Won Je Lee), *Dysnectes brevis* and *Hicanonectes teleskopos*, are included for comparative purposes. Scale bar is 5 μm for all figures.

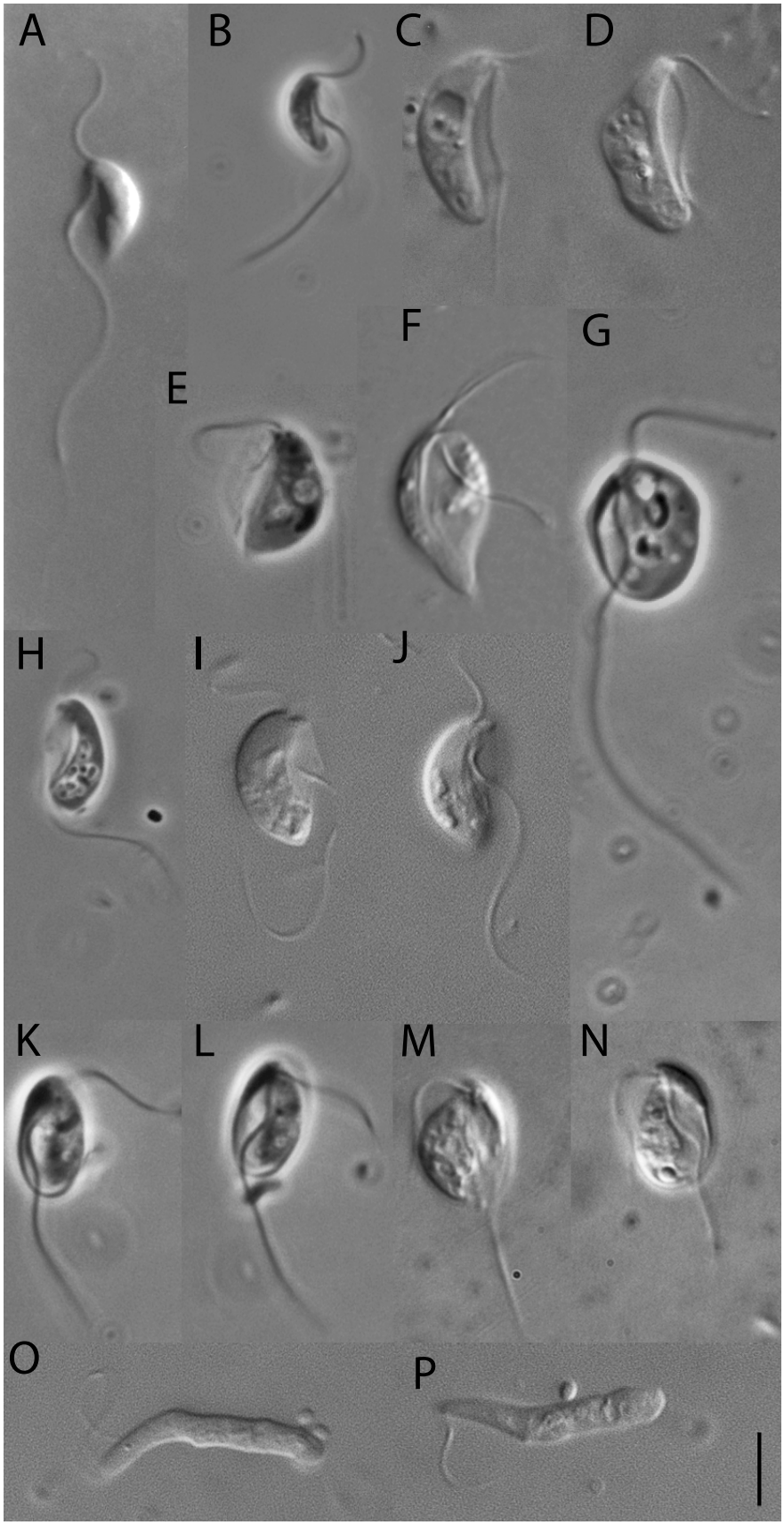

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P

Table 3.2. Morphological characteristics of *Carpediemonas*-like organisms.

Organism	Clade	Shape	# of Flagella	Posterior flagellum	Moving pattern
Ca. membranifera QB	CL4	bean shaped	2	~3.5x cell length	slow, with a slow wobbling
Ca. membranifera BICM	CL4	bean shaped	2	~3.5x cell length	slow, with a slow wobbling
Dysnectes brevis	CL1	bean shaped	2	~1x cell length	very slow, often adheres to surfaces
Hicanonectes teleskopos	CL3	oval shaped	2	~3x cell length	relatively fast with rapid rotation
CL	CL5	bean shaped	2	~2x cell length	slow with slow rotation/slow wobbling
NC	CL5	bean shaped	2	~2x cell length	slow with slow rotation/slow wobbling
PCS	CL3	spindle shaped	1	~1x cell length	slow, jerky
NY0171	CL2	oval shaped	2	~2x cell length	Relatively fast with slow, jerky rotation
PCE	CL2	oval shaped	2	~2x cell length	Relatively fast, with rapid rotation
SIVOTA	CL1	bean shaped	2	~1x cell length	very slow
Kipferlia bialata n. gen n. comb	CL6	bean shaped	2	~1.5x cell length	Very slow, often adheres to surfaces, rapid beating of the anterior flagellum

Table 3.3. Uncorrected genetic distance between and within each CLO clade (SSU rRNA gene). For context, genetic distances between all the CLO clades and the diplomonad *Octomitus* are also included.

	CL2	CL3	CL4	CL5	CL6	Octomitus	Internal
CL1	0.28	0.3	0.27	0.25	0.27	0.27	0.02
CL2		0.3	0.33	0.3	0.34	0.32	0.01
CL3			0.3	0.29	0.31	0.32	0.22
CL4				0.28	0.31	0.31	0.04
CL5					0.29	0.28	0.1
CL6						0.32	0.01

3.2.2 SSU rRNA Gene Phylogeny of New Isolates

In the phylogenetic analysis of SSU rRNA gene sequences I included the 18 new isolates of CLOs, plus *Carpediemonas membranifera* QB, *Dysnectes brevis*, and *Hicanonectes teleskopos*, 15 environmental SSU rRNA gene sequences similar to those from CLOs as identified by BLAST (Edgcomb *et al.*, 2002; Stoeck *et al.*, 2007; Takishita *et al.*, 2007b), 28 sequences representing diplomonads and retortamonads, and 31 outgroup taxa representing most other major eukaryotic groups. All the CLOs, diplomonads and retortamonads collectively constitute a monophyletic group that we equate with the taxon Fornicata, with high statistical support (97% bootstrap proportion (BP) and a posterior probability (PP) of 1). Diplomonads and the genus Retortamonas form a highly supported clade (100% BP and 1 PP), while the retortamonad *Chilomastix* branches as a sister group to the clade of diplomonads plus Retortamonas, but with a very low bootstrap support.

All CLOs branch basally to diplomonads and retortamonads as a non-monophyletic assemblage. The CLOs form six highly distinct and strongly supported clades, here called CL1 – CL6 (Figure 3.2, Table 3.1). Clade CL1 contains *Dysnectes brevis* and

isolate SIVOTA. Clade CL2 is formed by the very similar new isolates NY0171 and PCE. *Hicanonectes teleskopos* (isolate SB), the new isolate PCS and a single environmental sequence, D4P08A09, branch together as clade CL3, although PCS plus D4P08A09 are a group distinct from *H. teleskopos* within this clade. *Carpediemonas membranifera* and new isolate BICM constitute clade CL4, which represents the genus *Carpediemonas* itself. Clade CL5 contains only the new isolates CL and NC. Clade CL6 is a tight cluster containing the rest of the new isolates (GR1, PPP15C, LARNAKA2, NY0173, NY0166, ALLEPEYI, KR3, KR4, KR7, KR8 and GSML) and all environmental sequences except D4P08A09. While clades CL1 – CL6 form the basal branches within Fornicata, their interrelationships are essentially unresolved. In the maximum likelihood tree CL1 and CL2 form an unsupported monophyletic group (19 BP; 0.55 PP), and collectively constitute the closest relative of the diplomonads-retortamonads clade, with no statistical support. Clades CL3, CL4 and CL5 constitute a separate monophyletic group, but again with no statistical support (11 BP and 0.51 PP). Clade CL6, representing the *Carpediemonas bialata* morphospecies (here renamed *Kipferlia bialata* n gen. n. comb., see below), branches independently as the most basal group of Fornicata, with very weak support (46 BP and 0.91 PP). There is no evidence of a specific relationship between CL6 and CL4 (i.e., *Carpediemonas* proper).

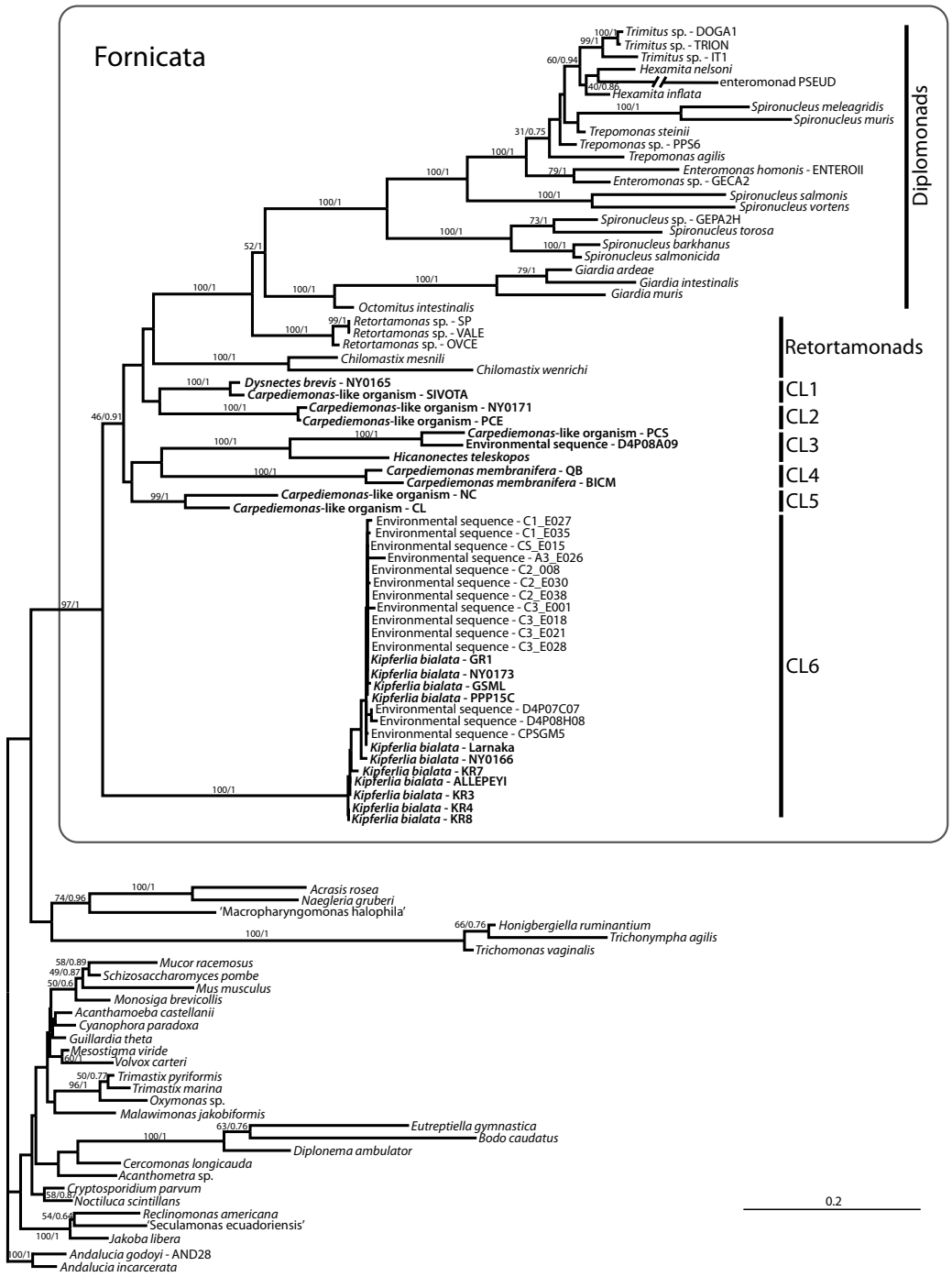


Figure 3.2 Maximum likelihood tree based on SSU rRNA genes from Fornicata including new isolates of *Carpediemonas*-like organisms (CLOs). The tree is rooted using a 31 taxon eukaryotic outgroup. The GTR + I + Γ model of sequence evolution was used. *Carpediemonas* and *Carpediemonas*-like organisms are depicted in bold. Statistical support is as follows: bootstrap proportion, based on 10000 replicates/MrBayes posterior probabilities. Statistical support is not shown for nodes with support lower than 50bp and 0.7pp.

3.2.3 Electron Microscopy of NY0173

Preliminary transmission electron microscopy of isolate NY0173 from clade CL6 was performed by Naoji Yubuki (University of British Columbia), and shows that the right margin of the groove is extended substantially into a thin membrane (Figure 3.3). The posterior flagellum bears a broad ventral vane, but a dorsal vane is absent or very small (Figure 3.3).

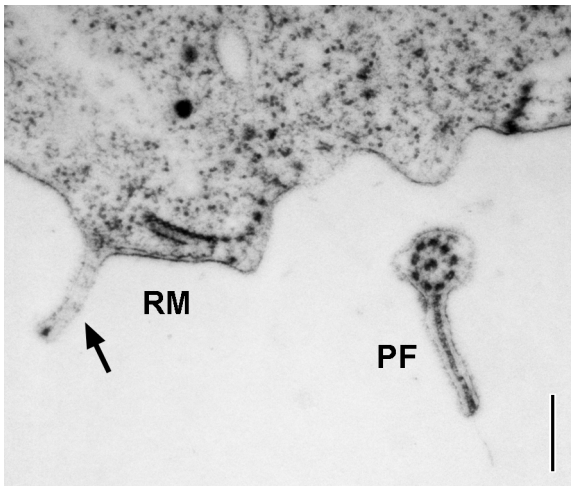


Figure 3.3. Transmission electron micrograph of isolate NY0173 of *Kipferlia bialata* n. gen. n. comb (clade CL6), showing the right portion of the ventral groove in transverse section. The image is shown looking from anterior to posterior, with ventral side of the cell facing downwards, thus the right side of the cell is towards the left side of the micrograph. The right margin of the groove (RM) is extended by a substantial membrane (arrow). The posterior flagellum (PF) bears a single broad vane on its ventral side. Scale bar represents 500 nm. This electron microscopy work was performed by Naoji Yubuki (University of British Columbia).

3.3 Discussion

Most of our new isolates apparently represent novel lineages at least at the 'genus level', based mainly on the dissimilarity of their SSU rRNA genes (Table 3.3) and pattern of phylogenetic relationships (Figure 3.2). The only new isolate that is assignable to a previously cultured species is BICM, which is indistinguishable by light microscopy from *Carpediemonas membranifera* isolate QB (Simpson and Patterson, 1999) and is very closely related in the SSU rRNA gene phylogeny (uncorrected genetic distance is 0.04). The new isolates in clade CL6 are indistinguishable in light microscopic appearance to the previously described, but never cultured or sequenced, *Carpediemonas bialata* (Lee and Patterson, 2000). The CL6 isolates are also nearly identical to each other in their SSU rRNA gene sequence (the average uncorrected genetic distance within the clade is 0.01), Therefore we consider that they all represent this one species. However, genetic distance between clade CL6 and *Carpediemonas membranifera*, the type species for the genus *Carpediemonas* (CL4), is considerable (0.31 - the same distance as between *C. membranifera* and the diplomonad genus *Octomitus*, Table 3.3), and the two groups do not constitute a clade in our SSU rRNA gene tree. There are also substantive ultrastructural differences: The membrane-like extension of the right margin of the groove in CL6 is not seen in *Carpediemonas*, nor in other CLOs examined to date. CL6 lacks the broad dorsal flagellar vane that is so-far characteristic of *Carpediemomas* amongst CLOs. The organism currently called *Carpediemonas bialata* should therefore be considered a member of a separate genus. We propose the new genus *Kipferlia* n. gen., and transfer *Carpediemonas*

bialata to this new genus as its type species, *Kipferlia bialata* (Ruinen, 1938) n. comb. (see taxonomic summary below).

Clades CL2 (isolates NY0171 and PCE) and CL5 (isolates CL and NC) are markedly dissimilar in SSU rRNA gene sequence from all formally described genera, and neither shows a reliable sistergroup relationship with members of a described genus. Isolate PCS does form a robust phylogenetic relation with *Hicanonectes* (both are within clade CL3), but the genetic dissimilarity between the two is still substantial (0.22), and they are unalike morphologically. It is very likely that each of these three groups will also be recognized as a new genus in the future. Isolate SIVOTA is most closely related to *Dysnectes brevis* but is molecularly distinguishable (Table 3.3), and differs slightly in appearance (Figure 3.1 E-F, Table 3.2), and probably is best considered as a separate species. The further characterization of these other new isolates and determination of their possible assignment into new genera will be the subject of the future work. Recently isolate 'CL' from clade CL5 has been formally described as *Ergobibamus cyprinoides* (Park et al., 2010; Appendix H), and is referred to by that name in Chapters 4 and 5 of this thesis.

The availability of a wide diversity of basal lineages within Fornicata will be valuable for understanding the evolution of diplomonads and their mitochondrial organelles. For example, it will now be possible to perform comparative analyses of inferred mitochondrial proteins in several different *Carpediemonas*-like lineages, together with diplomonads/retortamonads. An important prerequisite for such comparative analyses is a robust understanding of the actual phylogenetic relationships amongst the *Carpediemonas*-like lineages, and their relationship to

diplomonads and retortamonads. Unfortunately, these deep relationships are poorly and/or inconsistently supported in SSU rRNA gene phylogenies. For example, in our analysis, clade CL6 is recovered as the most basal clade of Fornicata, consistent with previous analyses in which CL6 is represented by environmental sequences (Takishita *et al.*, 2007b; Park *et al.*, 2009), but support decreases as taxon sampling increases. Meanwhile, like Park et al. (2009) and (Cepicka *et al.*, 2008) we recover a clade that includes *Dysnectes* as the closest relatives of the diplomonad-retortamonad clade, but this conflicts with the analysis of Yubuki et al. (2007). It seems that considerably more sequence data (i.e., additional genes for analysis) will be necessary to reliably resolve the relationships among the major clades of CLOs. In very recent work, phylogenetic analyses based on 4-7 genes confirm that CLOs are a paraphyletic assemblage and that among CLOs *Dysnectes* is especially closely related to diplomonads (Appendix I). The paraphyly of CLOs is also discussed in chapter 4 of this thesis.

Until very recently (2007), *Carpediemonas membranifera* was the only species other than diplomonads and retortamonads within the clade Fornicata, and appeared to be a phylogenetically isolated organism. It is now clear that *Carpediemonas membranifera* is merely one representative of large assemblage of free-living *Carpediemonas*-like organisms. CLOs were mostly undetected by both historical microscopy-based studies and more recent environmental PCR approaches. The very limited detection of CLO sequences in clone libraries generated from environmental PCR is particularly striking, as the habitats in which these organisms have been isolated – low oxygen marine/saline sediments – have

been frequently sampled (Dawson and Pace, 2002; Edgcomb *et al.*, 2002; Stoeck *et al.*, 2003; Takishita *et al.*, 2005; Behnke *et al.*, 2006; Takishita *et al.*, 2007a; Takishita *et al.*, 2007b; Epstein and López-García, 2008). Only three of these studies recovered CLO sequences and only two reported them in their results (Edgcomb *et al.*, 2002; Stoeck *et al.*, 2007; Takishita *et al.*, 2007b). Environmental PCR-based studies of low-oxygen marine/saline water column sites have also not recovered CLO sequences (Stoeck et al., 2003; Behnke et al., 2006).

Some very recent studies use 454 sequencing of SSU rDNA environmental PCR samples to examine protist diversity (Stoeck *et al.*, 2009; Stoeck *et al.*, 2010), potentially providing a much deeper coverage of diversity than sequencing of clone libraries. We performed a detailed search for CLOs in two 454 environmental surveys of anaerobic environments (Stoeck *et al.*, 2009; Stoeck *et al.*, 2010) and still identified representatives of just two CLO clades – CL1 and CL6 – all from a single sampling site (Appendix D.2).

It is likely that the true diversity of major lineages of CLOs is still greater than we have described in this study. All but one of our six major clades are represented by only two isolates, leaving the strong possibility that additional readily cultivable lineages have been missed through pure chance. Other lineages could occur in environments other than oxygen-poor saline sediment, or may simply require different culturing approaches. Still others may be difficult to culture and may be detected most effectively through environmental PCR with taxon-specific primers.

By far the most commonly encountered clade is CL6 (*Kipferlia bialata* n. gen., n. comb.), which includes over half (11/18) of our new isolates and all but one of the

previously reported environmental sequences. This may suggest that clade CL6 is much more abundant in the environment than the other clades. Alternatively, clade CL6 may be over-represented due to the conditions for culturing and/or environmental PCR. Culturing bias cannot be ruled out as the majority of isolates from clade CL6 were isolated using 802SW media, while other isolates were usually obtained using various other types of media (Table 3.1). It is possible that 802SW media selects for CL6 over the other clades. By contrast we found little evidence to suspect a PCR bias towards CL6. SSU rRNA gene sequences from this clade do not constitute better targets for the particular PCR primers used by the environmental studies that yielded CL6 sequences (Edgcomb *et al.*, 2002; Takishita *et al.*, 2007b). Also we have performed preliminary experiments on mixtures of DNA from different CL clades, and did not find a strong PCR bias towards representatives of clade CL6 (data not shown).

3.3.1 Concluding Remarks

The current understanding of the diversity of single-celled eukaryotes is based on microscopy and culturing going back more than 150 years, and more recently on environmental PCR surveys (Bass and Cavalier-Smith, 2004; Groisillier *et al.*, 2006; Not *et al.*, 2007). We have explored an important 'region' of the eukaryotic tree that was, until recently, seemingly populated by a single isolated lineage. Our application of straightforward culturing techniques revealed a large number of very distinct lineages in this region of the tree. Moreover, these were isolated from marine anoxic sediments (except isolate GSML), a relatively easily accessible and often-sampled

habitat type. The bulk of these lineages had been completely missed by both the historical microscopy/culturing efforts, and by environmental PCR endeavors targeting similar habitats.

The reasons for this limited prior detection of the diversity of CLOs are not clear, but might involve a low abundance of most of these organisms in the environment. Indeed, some of the environmental sequences (CPSGM5) in clade CL6 were detected only after crude enrichment (Takishita *et al.*, 2007b). In addition, we cannot exclude the possibility that a role is played by an experimental bias in environmental PCR studies other than primer-target mismatch.

I see no good reason to assume that the overlooking of major-lineage-level diversity we report is unique to the base of Fornicata. I suggest it is more likely that undersampling at the level of major lineages could still be widespread for microbial eukaryotes. Our understanding of eukaryotic evolution would be greatly advanced if this were overcome. It would be particularly interesting to see whether other phylogenetically isolated but evolutionarily important lineages such as Chromera (Moore et al., 2008) are in fact the tips of large 'icebergs' of high-level lineage diversity. A combination of raw culturing effort and much deeper environmental PCR sampling, perhaps coupled with the use of enrichments (i.e., 'semi-environmental' samples), and/or taxon-specific primers may help to capture a larger portion of the diversity (Takishita *et al.*, 2007b; Lara *et al.*, 2009).

3.4 Taxonomic Summary.

The new genus *Kipferlia* is described here in accordance with the International Code of Zoological Nomenclature (ICZN, 1999).

***Kipferlia* n. gen.**

Diagnosis. Free-living, biflagellated, and colourless cells with a conspicuous ventral groove. The right margin of the groove is markedly extended by a fine membrane, visible by electron microscopy. The posterior flagellum beats within the groove, and bears a single broad vane, located ventrally. Inhabits low oxygen marine environments. Similar to *Carpediemonas* and *Dysnectes* in typical habitat and general appearance when viewed by light microscopy, but distinct from both in SSU rRNA gene phylogenies (see Figure 3.2).

Type species. *Kipferlia bialata* (Ruinen, 1938) n. comb.

Other species. None described

Etymology. Kipferln (sing. Kipferl; German) are small crescent-shaped cookies from southern Germany and Austria. The name refers to the shape of the type species. The name *Kipferlia* is considered to be of feminine gender, in agreement with the species epithet for the type species.

Taxonomic Assignment. Eukaryota; Excavata; Fornicata

***Kipferlia bialata* (Ruinen, 1938) n. comb.**

Basionym. *Cryptobia bialata* Ruinen, 1938

Other synonyms. *Carpediemonas bialata* (Ruinen, 1938) Lee and Patterson 2000.

Comments. Originally described in 1938 as *Cryptobia bialata* (Ruinen, 1938), this organism was next identified as a distinct morphospecies by Lee and Patterson (2000), who renamed the organism *Carpediemonas bialata* (Ruinen 1938).

3.5 Materials and Methods

3.5.1 Culture Isolation and Light Microscopy

Isolates SB, BICM, CL and PPP15C were isolated and documented personally by the author; other new isolates were obtained by other collaborators. All isolates except isolate GSML were cultured from anoxic sediments; the locations of the sampling sites, as well as media used for cultivation, are listed in Table 3.1 (Media formulations are given in the Appendix D.1). Isolate GSML was cultured from detritus accompanying a shipment of sea urchins collected in an estuarine bay (Gulf Specimen Marine Lab cat# E-1610) and received at the University of Arkansas. Monoeukaryotic cultures were usually established via transferring the cultures at the point where CLOs were the most common eukaryotes, which slowly diluted out other eukaryotes. Isolates CL and BICM were purified away from ciliates by filtering the culture through 3 μm filters. A single cell of each four isolates: NY0165 (CL1), NY0166 (CL6), NY0171 (CL2) and NY0173 (CL6) was isolated by micropipeting

from the enrichment culture and inoculated into the medium, which was prepared as the low oxygen environment beforehand. All cultures were xenic and grown in nutrient-rich media (see Appendix D.1). The low-oxygen environment was maintained by high bacterial growth and by the large volume of media relative to the size of the culturing tubes (i.e., a small headspace). The actual oxygen levels were not monitored. Light microscopy observations utilized a Zeiss Axiovert 200M microscope equipped with an Axiocam HR digital camera, a Leica DMR light microscope (Leica, Germany) equipped with a Keyence VB6010 digital chilled CCD camera (Keyence, Osaka, Japan), a Zeiss Axioskop 2 equipped with a JVC KY-F75U color digital camera using Automontage (Syncroscoy, Frederick, MD) and an Olympus Microscope BX51 and camera Olympus DP70 (Olympus America inc.).

Transmission electron microscopy of strain NY0173 was performed by Naoji Yubuki (University of British Columbia). For transmission electron microscopy (Figure 3.3), cells were high-pressure frozen using a Leica HPM100. The procedure for the high pressure freezing fixation, dehydration and embedding was same as that described by (Yubuki *et al.*, 2010). Ultra-thin sections were cut on a Leica EM UC6 ultra-microtome and double stained with 2% (w/v) uranyl acetate and lead citrate and observed using a Hitachi H7600 electron microscope.

3.5.2 DNA Isolation and Sequencing

Molecular sequences were obtained from isolates SB, BICM, CL, PPP15C personally by the author and sequences from other new isolates were obtained by other collaborators. The DNA was isolated from the cultures using the CTAB

(cetyltrimethylammonium bromide) protocol of Clark (Clark, 1992), a modified CTAB protocol (Ishida *et al.*, 1999), a simple phenol/chloroform protocol (Garriga *et al.*, 1984), a High Pure PCR template kit (Roche Applied Science, UK), or a Gentra PureGene DNA isolation kit (Qiagen, USA). Universal eukaryotic primers 5' primer A, 3' primer B (Medlin *et al.*, 1988), or 18S Fw (5' aacctggttgatcttgccag 3') and 18S Re (5' cygcaggttcacctacggaa 3') were used to amplify the SSU rRNA gene of all but one isolate (PCS). The SSU rRNA gene of isolate PCS was amplified using 5' primer A and PCS_1600R (5' ccatgtccaaacaacttgcc 3'). Fragments of expected size were purified from agarose gels using the Qiagen Gel extraction kit (Qiagen, USA) or GeneElute Gel extraction kit (Sigma-Aldrich, USA) and either directly sequenced or cloned using the TOPO-TA cloning kit (Invitrogen, USA) or Promega T-easy Vector system (Promega, USA). In the latter cases, several clones were partially sequenced and their identity was checked using BLAST (Altschul *et al.*, 1990) before one to ten pooled clones were fully bidirectionally sequenced by an oligonucleotide primer-walking approach. All 18 new sequences are deposited in GenBank database under accession numbers GU827588 - GU827605.

3.5.3 Phylogenetic Analyses

A eukaryotic secondary structure-based alignment was downloaded from the European SSU rRNA gene database (http://bioinformatics.psb.ugent.be/webtools/-rRNA/). Missing and new taxa were realigned to the downloaded alignment with the program ClustalX (Thompson *et al.*, 1997). The final dataset contained 63 Fornicata sequences and 31 sequences from other eukaryotes. The resulting alignment was

edited by eye and ambiguously aligned regions were discarded, leaving 914 nt positions. The relatively low number of truly unambiguously aligned positions was due to the divergent nature of diplomonad SSU rRNA genes.

The phylogenetic trees were constructed using Maximum Likelihood (ML) and Bayesian methods. The GTR + I + Γ model of sequence evolution was selected by the Akaike information criterion implemented in the program Modeltest 7.0 (Posada and Crandall, 1998). The maximum likelihood tree was constructed using PAUP*4b10 (Swofford, 2002) with 20 random sequence stepwise addition replicates and tree-bisection-reconnection rearrangements. Bootstrap support was estimated from 10000 bootstrap replicates using RAxML 7.0 (GTR + I + Γ) (Stamakis, 2006). The Bayesian analyses was performed with MrBayes 3.1 (Huelsenbeck, 2000) using a GTR + I + Γ model and was run for 20 million generations (stationarity was reached after 500 000 and burnin was set to 500 000 generations, while other parameters were left at their default values).

Chapter 4

Phylogenomic Analysis of Excavata

4.1. Introduction

In the early 2000s the standard working model for describing the deep-level diversity of eukaryotes was a system of six supergroups – Opisthokonta, Amoebozoa, Archaeplastida, Chromalveolata, Rhizaria and Excavata (Simpson and Roger, 2002; Adl *et al.*, 2005). The support and confidence in the phylogenetic validity of these groups varied strongly. For example, Opisthokonta – the lineage that includes Animals and Fungi - received high support from molecular phylogenetic analyses of a number of genes, and this was corroborated by both morphological characters, and a shared insertion in EF1-α gene (Baldauf and Palmer, 1993; Steenkamp *et al.*, 2006). Some other groups by contrast, especially Chromalveolata and Excavata, were usually not recovered as monophyletic in molecular phylogenies of one or a few genes (Parfrey *et al.*, 2006; Simpson et al., 2006).

In the last half-decade, datasets consisting of 70-200 concatenated genes have become the major resource for inferring the deep evolutionary relationships within eukaryotes (Rodriguez-Ezpeleta *et al.*, 2005; Patron *et al.*, 2007; Rodriguez-Ezpeleta *et al.*, 2007a; Burki *et al.*, 2008; Burki *et al.*, 2009; Hampl *et al.*, 2009; Minge *et al.*, 2009). Such 'phylogenomic' analyses brought more clarity to the tree of eukaryotes. For example, Amoebozoa and Opisthokonta have generally received

very high support, and also formed a clan referred to as 'Unikonta'. By contrast, the supergroup Chromalveolata turned out not to be strictly monophyletic. Currently the chromalveolate groups are divided amongst two new supergroups: the 'S.A.R.' clade (also known as Harosa) and Hacrobia. S.A.R. consists of Stramenopiles and Alveolates (both former chromalveolates) along with Rhizaria, which was previously considered as supergroup on its own (Burki *et al.*, 2007; Hackett *et al.*, 2007). Hacrobia includes haptophytes, cryptophytes and several more obscure lineages (Burki *et al.*, 2009). The S.A.R. clade generally receives high support on phylogenomic analyses (Burki *et al.*, 2008; Burki *et al.*, 2009) while Hacrobia receives mediocre-to-no support, and is in clear need of further study.

The one supergroup that is almost never recovered as monophyletic in global phylogenomic analyses, yet cannot be firmly rejected at the same time, is Excavata (Hampl *et al.*, 2009). Excavata is a diverse assemblage, mostly consisting of free-living and parasitic heterotrophic flagellates, but also including the euglenophyte algae and the heterolobosean amoebae. The common synapomorphic features are the presence of a suspension feeding groove and several underlying ultrastructural characters (Simpson, 2003). 'Typical excavates' – Jakobida, *Malawimonas*, *Carpediemonas* and *Carpediemonas*-like organisms, Retortamonadida and *Trimastix* - possess all of these morphological characters (Bernard *et al.*, 1997; Simpson and Patterson, 1999; Simpson *et al.*, 2000; Simpson and Patterson, 2001; Simpson *et al.*, 2002b; Yubuki *et al.*, 2007; Park *et al.*, 2009; Park *et al.*, 2010; Appendix G and H). Other members of Excavata - Diplomonadida, Parabasalida, Discicristata, *Tsukubamonas* and Oxymonadida - possess none or only some of these characters,

but are united with one or other of the 'typical excavates' by molecular phylogenetic evidence (Dacks *et al.*, 2001; Edgcomb *et al.*, 2001; Silberman *et al.*, 2002; Simpson *et al.*, 2002c; Simpson et al., 2006; Yabuki *et al.*, 2011). The excavates diplomonads and parabasalids are both notorious for forming extremely long branches in molecular phylogenetic trees (Sogin *et al.*, 1989; Vankeulen *et al.*, 1993; Kamaishi *et al.*, 1996).

In recently published phylogenomic analyses, Excavata have tended to form two groups that branch sequentially ('paraphyletically') at the base of Unikonta (Burki *et al.*, 2008; Burki *et al.*, 2009; Hampl *et al.*, 2009). One is formed by the very short-branching taxon *Malawimonas,* while the second group includes all other excavates, including the notoriously long-branching taxa. In one study that focused specifically on the monophyly of Excavata, it was argued that this topology may be a long branch attraction artifact between long-branching excavates and long-branching representatives of the S.A.R. clade (Hampl *et al.*, 2009). Indeed, after the removal of 14 longest branches from the analyses Hampl et al. (2009) saw the bootstrap support for excavate monophyly rise to >90% (a similar result had been also reported by Rodriguez-Ezpeleta *et al.* (2007a). At this point, however, all representatives of diplomonads, parabasalids, oxymonads and *Trimastix* had been removed from the analyses. This exclusion of a major proportion of excavate diversity (the entire group Metamonada) significantly weakens this analysis as an argument for the monophyly of Excavata.

As mentioned above, it was suggested that the cause of non-monophyly of Excavata in phylogenomic analyses is a long branch attraction (LBA) artifact. This is a type of systematic error where taxa that have accumulated a large number of

substitutions tend to branch together, irrespective of their true evolutionary relationships (Felsenstein, 1978). Broadly speaking, the phenomenon is a result of model violation, and because it is a systematic error, statistical support for the 'LBA topology' will increase as more similar data is added to the analyses. Therefore, it is of particular concern in data-rich phylogenomic analyses.

Several methods have been proposed to suppress suspected LBA artifacts. The ones most commonly used include i) exclusion of long-branching taxa, ii) exclusion of fast-evolving sites or genes and iii) recoding of the data into a smaller number of possible states such that that the data only include the most infrequent types of substitutions (Rodriguez-Ezpeleta *et al.*, 2007b). It has also been argued that including deep, short-branching relatives of long-branching taxa will 'break' that long branch making the branch leading to these two taxa shorter and less prone to LBA artifact (Hendy and Penny, 1989; Poe, 2003; Brinkmann *et al.*, 2005; Geuten *et al.*, 2007). Another possibility is to use short-branching relatives to represent the group as a whole, instead of the long-branching taxa.

Two excavate lineages - diplomonads and parabasalids – almost invariably form extremely long branches and are probably a major cause of LBA artifact in any phylogenomic study that includes these taxa (Brinkmann *et al.*, 2005; Simpson et al., 2006; Hampl *et al.*, 2009). Recently, however, a large diversity of previously unknown shorter branching relatives of diplomonads has been characterized. These organisms are collectively called *Carpediemonas*-like organisms, or CLOs, and include the new taxa *Dysnectes, Hicanonectes, Kipferlia and Ergobibamus,* as well as some yet undescribed lineages (Yubuki *et al.*, 2007; Park *et al.*, 2009; Kolisko *et al.*,

2010; Park *et al.*, 2010; Chapter 3; Appendix G and H). It was suggested that inclusion of CLOs into phylogenomic analyses may help to suppress LBA artifact (Kolisko *et al.*, 2010; Chapter 3): It is possible that these organisms will help to break the long branches leading to diplomonads, or that they can be used as substitutes for diplomonads (and perhaps parabasalids, which are related, more distantly, to diplomonads)

In this study we have obtained large amounts of sequence data from several recently isolated CLOs, namely *Carpediemonas membranifera* strain BICM, *Ergobibamus cyprinoides*, and *Kipferlia bialata* (Kolisko *et al.*, 2010; Park *et al.*, 2010; see Chapter 3 and Appendix H), as well as the retortamonad *Chilomastix caulleryi* (Cepicka *et al.*, 2008) and the deep-branching discoban excavate *Tsukubamonas globosa* (Yabuki *et al.*, 2011). These data were added to a dataset consisting of >150 genes with a broad sampling of excavates and other eukaryotes. Extensive analyses were conducted to bring better understanding to the phylogenetic status of Excavata, and probable LBA artifact affecting the placement of major subgroups in this taxon.

4.2. Results

4.2.1. Assembled Dataset

The 'Main' dataset consisted of 43516 sites (161 genes) and 85 eukaryotic taxa. A second dataset was assembled that included only 32995 sites (144 genes) where at least one of the CLO species (i.e., *Carpediemonas membranifera*, *Ergobibamus*

cyprinoides or *Kipferlia bialata*) was represented by non-gap characters ('CLO_rep' dataset). All analyses were performed on both datasets, except the long gene removal and the Bayesian analysis (performed on 'Main' only - see below).

4.2.2. Tree Topology

Identical phylogenies were estimated for the 'Main' and 'CLO_rep' datasets (Figure 4.1, Appendix E.1). Opisthokonta and Amoebozoa were each recovered as monophyletic with high bootstrap support (BP), with the apusomonad *Thecamonas* forming sister branch to opisthokonts with high support. These three groups formed a monophyletic 'unikont' grouping with 94% or 91% bootstrap support. Alveolates, stramenopiles and rhizarians formed a highly supported 'S.A.R.' clade ('Main' – BP=94% and 'CLO_rep' – BP=91%). Cryptophytes, haptophytes, *Telonema*, red algae, green algae and glaucophytes (i.e., archaeplastidans and 'hacrobians') formed a second monophyletic group with high support (BP>95%). Most basal relationships within this heterogeneous assemblage received little support, although the cryptophyte *Guillardia theta* was a sister group to red algae with moderate support ('Main' – BP=75% and 'CLO_rep'=73%), while haptophytes and *Telonema* also formed a moderately supported clade ('Main' – BP=84% and 'CLO_rep' – BP=78%).

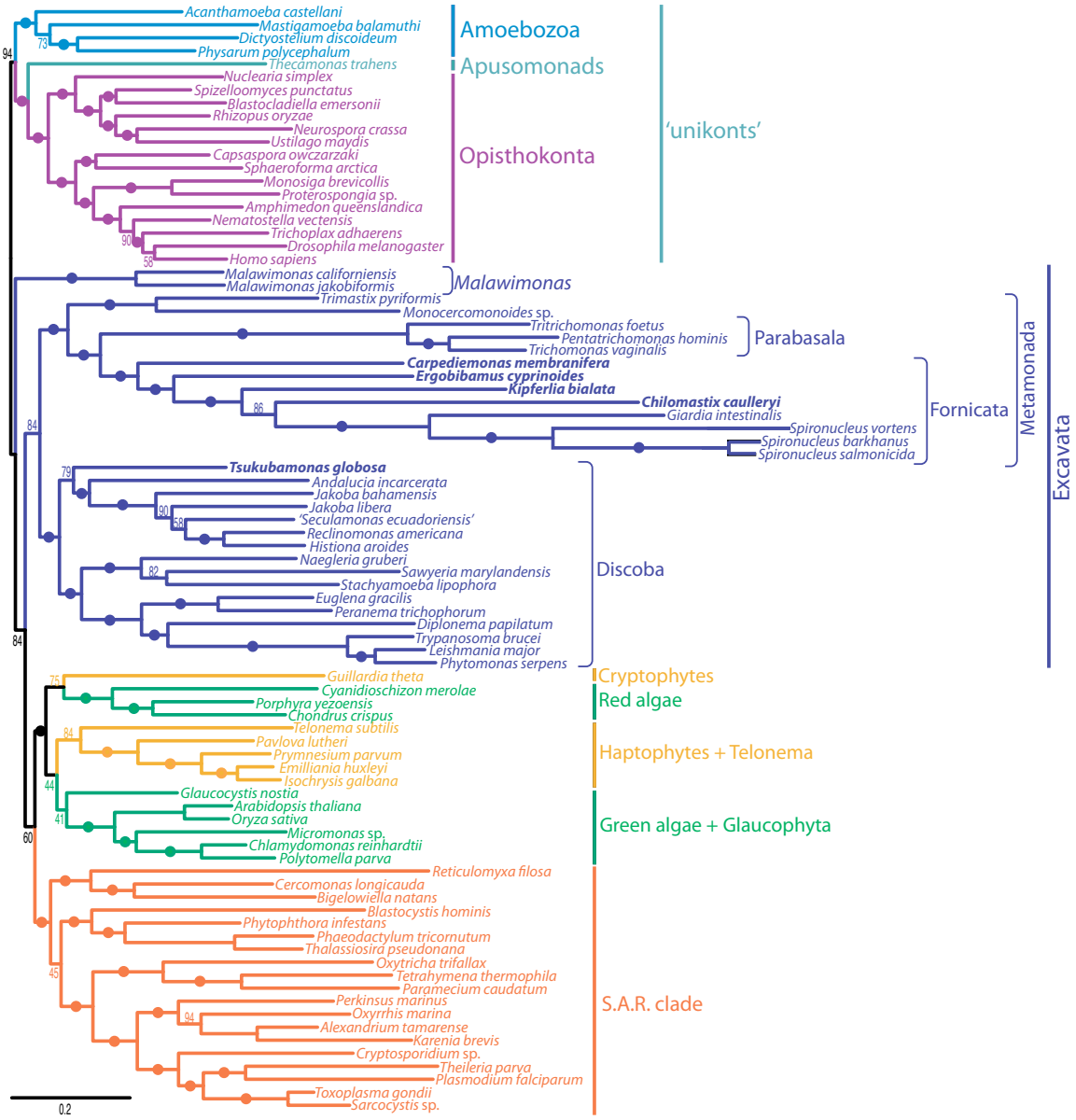


Figure 4.1 Maximum likelihood tree based on the 'Main' dataset (161 genes; 43516 sites), estimated using the program RAxML (model settings PROTGAMMALGF). The scale bar represents 0.2 expected substitutions per position. Statistical support was estimated using 500 bootstrap replicates. Dots on internal branches represent bootstrap support >95%. Newly sequenced taxa are depicted in bold font.

Excavates were not recovered as a monophyletic group in the initial analyses. All excavates except *Malawimonas* formed a single clade ('Main' – BP=84%, 'CLO_rep' – BP=66%). This was split into two maximally supported subgroups: Discoba (Jakobida, Heterolobosea, Euglenozoa and *Tsukubamonas*) and Metamonada (*Trimastix,* Oxymonadida, Parabasalida, Diplomonadida, *Chilomastix* and the CLOs). The newly sequenced *Tsukubamonas globosa* branches at the base of Jakobida with moderate support ('Main' – BP=79% and 'CLO_rep' – BP=76%). The CLOs and *Chilomastix caulleryi*, meanwhile, branch as a perfect comb at the base of diplomonads, to form a maximally supported clade that is equivalent to the taxon Fornicata. *Chilomastix* formed the closest relative of diplomonads, followed by *Kipferlia*, then *Ergobibamus*, and finally *Carpediemonas,* thus placing *Carpediemonas* as the deepest branching lineage within Fornicata. All of these relationships were strongly supported (BP=86% for the *Chilomastix*-diplomonad clade; all other nodes received 100% BP). Parabasalids branch with maximum support as the closest relatives of Fornicata.

Malawimonas was placed adjacent to the other excavates within the tree, but not in a unique clade/clan with them – for convenience this relationship is described as 'paraphyly of excavates'. *Malawimonas* instead branches between unikonts and the rest of eukaryotes (i.e., the tree contains a *Malawimonas*-unikont clan), with relatively high support in the 'Main' dataset (BP=84%), but substantially lower support in the 'CLO_rep' dataset (BP=63%).

The topologies recovered in the PhyloBayes analyses were consistent with the results of the ML analyses. However, only three of the four chains converged.

Both results (one based on the three converged chains, the other on the single rogue chain) are depicted in supplementary materials (Appendix E.2 and E.3). The only difference between them is the position of *Reticulomyxa filosa*, which branched with other Rhizaria in the three converged chains but as sister to the stramenopile *Blastocystis hominis* in the rogue chain.

4.2.3. LBA Suppression

4.2.3.1. Long-Branching Taxa Removal

The inclusion of shorter-branching CLOs in our analysis allowed me to perform analyses where I sequentially removed either long-branching taxa or long-branching genes (see below), while still keeping meaningful taxonomic representation from the Metamonada clade (and the Fornicata clade within metamonads). The sequential removal of long-branching taxa causes a general decline of support for paraphyly of excavates (i.e., a unikont-*Malawimonas* clan). Support fell from BP=73% in the 'Main' analysis, and BP=48% for 'CLO_rep' to BP=0% after removal of 18 and 11 taxa respectively (Figure 4.2, Appendix E.4; bootstrap support estimated by rapid bootstrapping). Over the same series support for monophyly of Excavata rose to BP>90% in the 'Main' dataset (Figure 4.3) and BP>95% in CLO_rep dataset. Support for a clade of *Malawimonas* specifically with Metamonada rose to BP=77% when 15 taxa were removed and then declined to BP=55% when 20 taxa were removed. Support values for Opisthokonta and Metamonada were tracked as a control for overall phylogenetic signal, and remained very high throughout the deletion series.

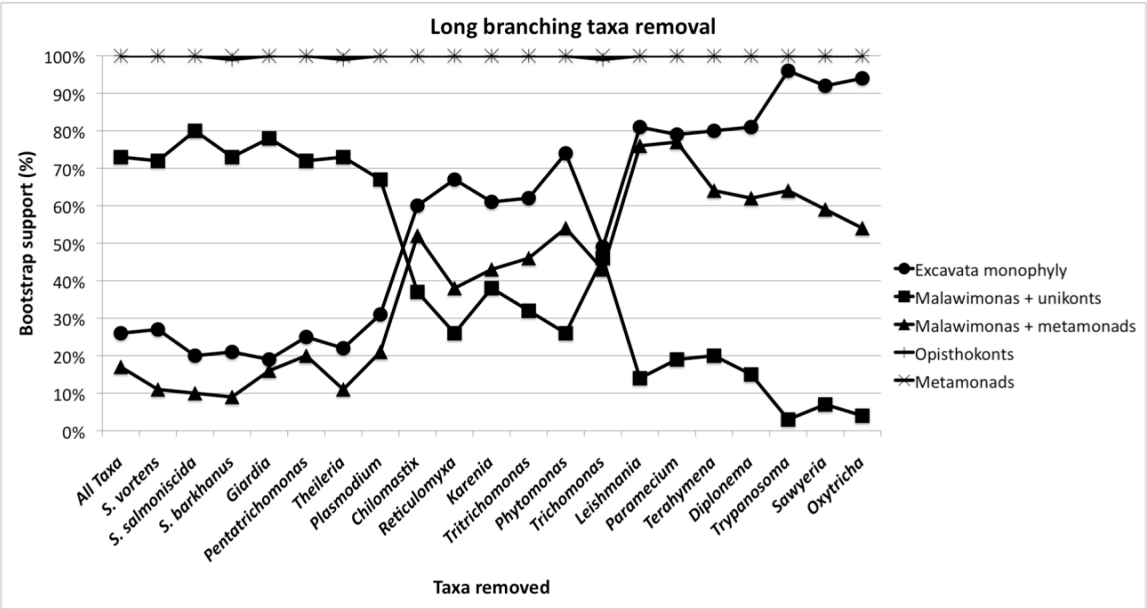


Figure 4.2. Graph depicting the change of bootstrap support for different topologies as fast-evolving taxa are removed sequentially from the 'Main' dataset. After removal of 18 taxa the support for paraphyly of Excavata (i.e., a *Malawimonas*-unikonts grouping) is close to zero, and Excavata monophyly is strongly supported.

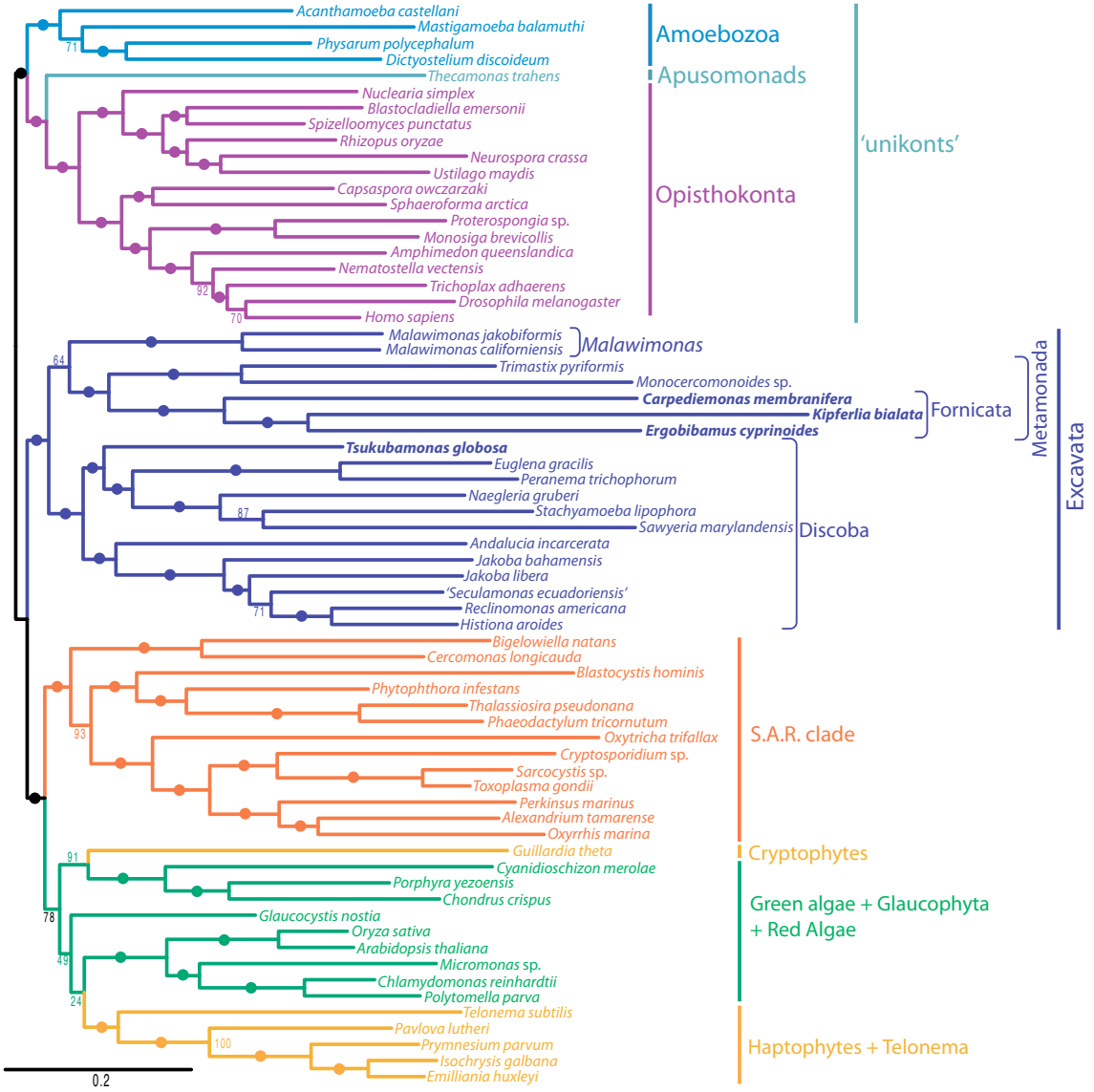


Figure 4.3 Maximum likelihood tree based on the 'Main' dataset after removal of 18 longest branching taxa, showing a strongly supported Excavata clade. The tree was estimated using the program RAxML (model settings PROTGAMMALGF). The scale bar represents 0.2 expected substitutions per position. Statistical support was estimated using 100 rapid bootstrap replicates. Dots on internal branches represent bootstrap support >95%. Newly sequenced taxa are depicted in bold font.

4.2.3.2. Long-Branching Gene Removal

The sequential removal of long-branching genes had a similar, but smaller, effect to the removal of long-branching taxa. Support for the 'excavate paraphyly' topology eventually declined to BP=17% and support for the monophyly of Excavata increased to BP = 75% (Figure 4.4). After 6500 single gene sequences were removed from the dataset the support for all tracked groups declined, indicating that there is no longer enough signal left in the dataset to recover robust phylogenetic relationships.

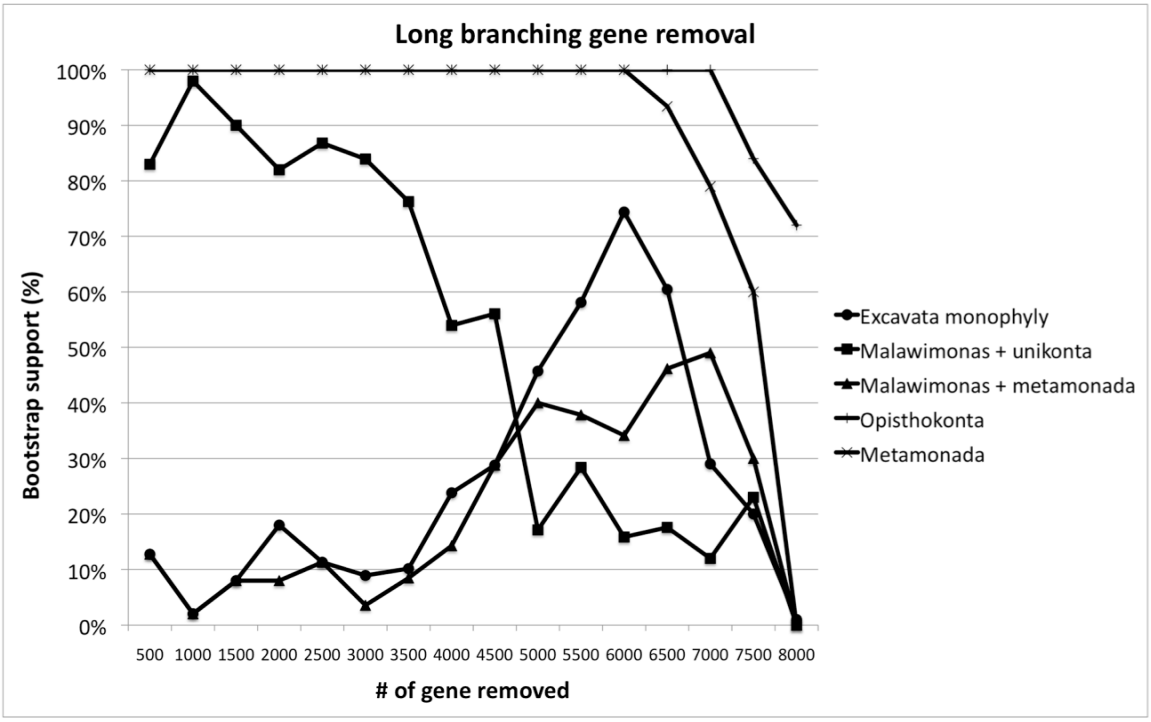


Figure 4.4. Graph depicting the change of bootstrap support for different topologies as fast-evolving genes are removed from the dataset. A declining trend is seen in the support for excavate paraphyly (*Malawimonas*-unikont clan), and a modest increase in support for Excavate monophyly. Note that after removal of 6500 single gene sequences the support declines for all topologies, suggesting that there is no longer enough signal left for tree reconstruction.

4.2.3.3. Fast-Evolving Sites Removal

I examined the effect of removing different numbers of fast-evolving sites in concert with removing different numbers of fast-evolving taxa. The results of these analyses are depicted in heat maps in Figure 4.5 and Appendix E.5. There are two different major effects observed: (i) As sites are removed the originally well-supported grouping of Discoba and Metamonada breaks up (ii) Removing fast-evolving taxa diminishes support for the 'excavate paraphyly' topology and leads to the monophyly of *Malawimonas*-Metamonada. This results in four different topologies, with each topology being most highly supported in a different quadrant of the heatmap (summarized in Figure 4.6). The maximum likelihood topology of the full dataset (i.e., with a *Malawimonas*-unikonts grouping) is supported when most or all sites and taxa are included in the analyses (Figure 4.6, tree 1). Excavate monophyly becomes strongly supported when only fast-evolving taxa are removed, but most sites retained (Figure 4.6, tree 2; see also 4.2.3.1). Excavata forms three groups - Discoba, Metamonada and *Malawimonas* - branching sequentially at the base of unikonts when fast-evolving sites are removed in quantity, but most taxa are retained (Figure 4.6, tree 3). When both fast-evolving sites and taxa are removed Excavata forms two groups arranged paraphyletically at the base of unikonts, but now the two groups are Discoba and *Malawimonas*-Metamonada (Figure 4.6, tree 4).

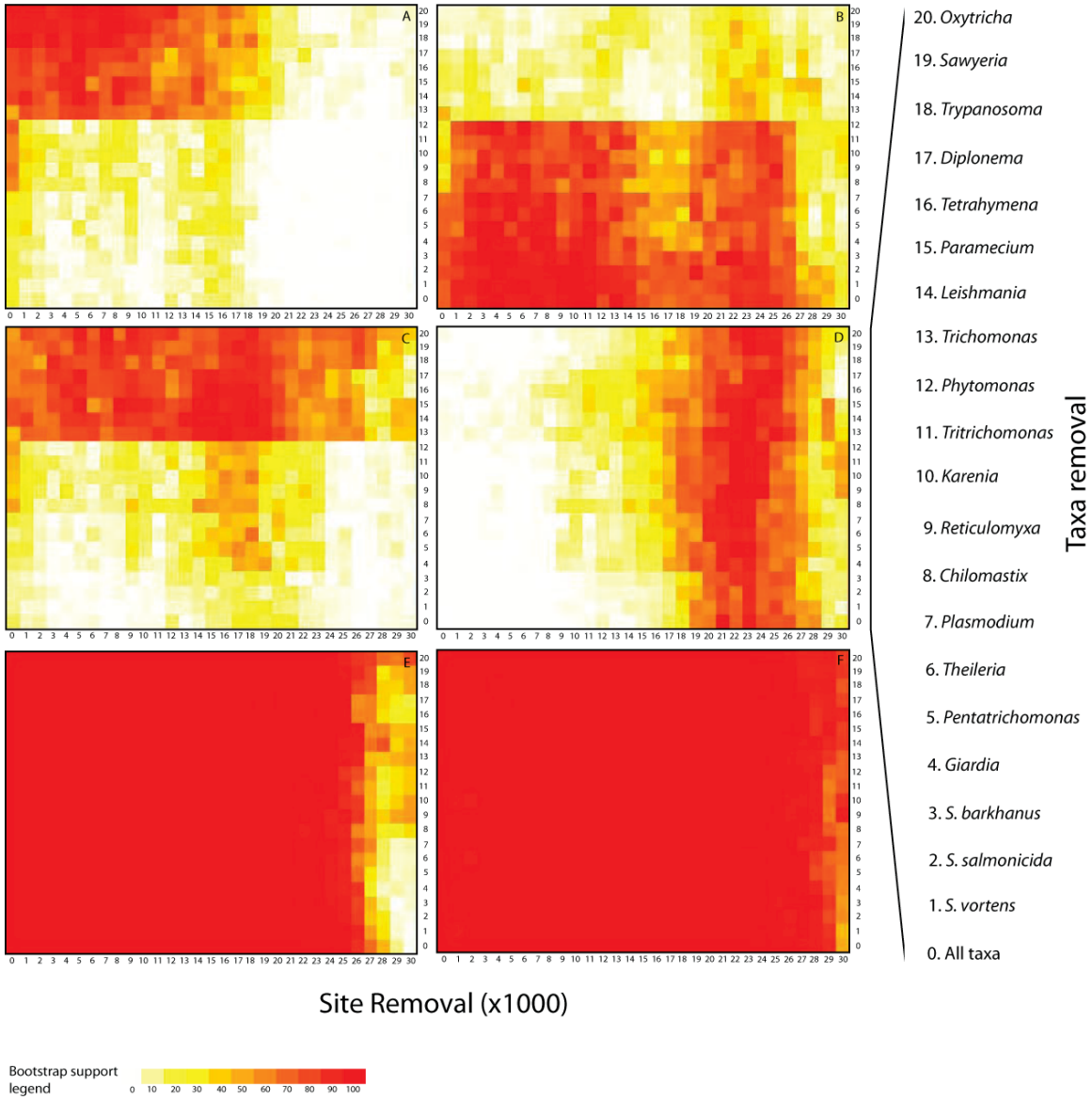


Figure 4.5 Heat map graphs depicting the bootstrap support for different topologies as fast-evolving sites are removed together with fast-evolving taxa from the 'Main' dataset. Numbers along the x axis represent the number of removed sites (x1000) and numbers on the y axis represent the number of removed taxa. The actual taxa removed are listed on the far right. A. Bootstrap support for monophyletic Excavata. B. Bootstrap support for paraphyletic Excavata (i.e., the ML topology in Figure 4.1). C. Bootstrap support for *Malawimonas*-Metamonada monophyly. D. Bootstrap support for *Malawimonas*-Metamonada-unikonts monophyly. E. Bootstrap support for Metamonada. F. Bootstrap support for unikont clade.

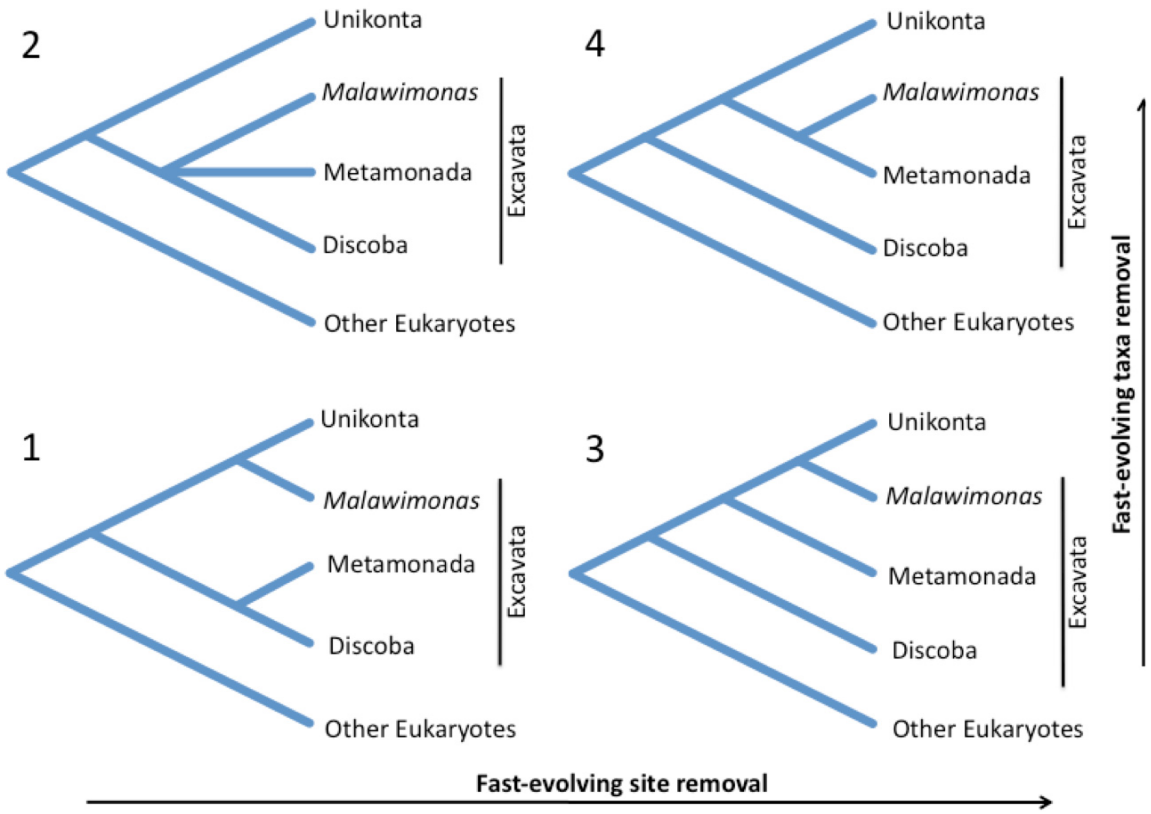


Figure 4.6. Summary showing the four different topologies recovered in with removal of fast-evolving sites and taxa. The positioning of the trees matches the four zones of high support for each topology in the heatmaps in Figure 4.5.

4.2.3.4. 'Ratio of Rates' Analysis

In this analysis sites with highest relative difference between rates in slow-evolving and fast-evolving taxa were removed from fast-evolving taxa only. As with removal of fast-evolving sites (see section 4.2.3.3) the removal was done in concert with the sequential exclusion of fast-evolving taxa. Results of this analysis are depicted as heatmaps in Figure 4.7 and Appendix E.6. Here the support is split somewhat diagonally between the *Malawimonas*-Unikonta topology (paraphyly of excavates) and Excavata monophyly. The support for *Malawimonas*-unikonts topology

generally decreased as the sites with highest ratio were removed, and this effect was more pronounced as long-branching taxa were excluded. In other words, the more fast-evolving taxa are removed, the greater the effect of the ratio-of-rates-based removal analyses.

4.2.3.5. Amino Acid Recoding

The rapid bootstrap analyses of a dataset where the amino acid alphabet was reduced to 5 different categories showed a decline of support for excavate paraphyly (BP=55%, compared to BP=73% in the full dataset). Accordingly, the support for monophyly of Excavata rose, but remained low in absolute terms (BP=43%, compared to BP=26% in the full dataset).

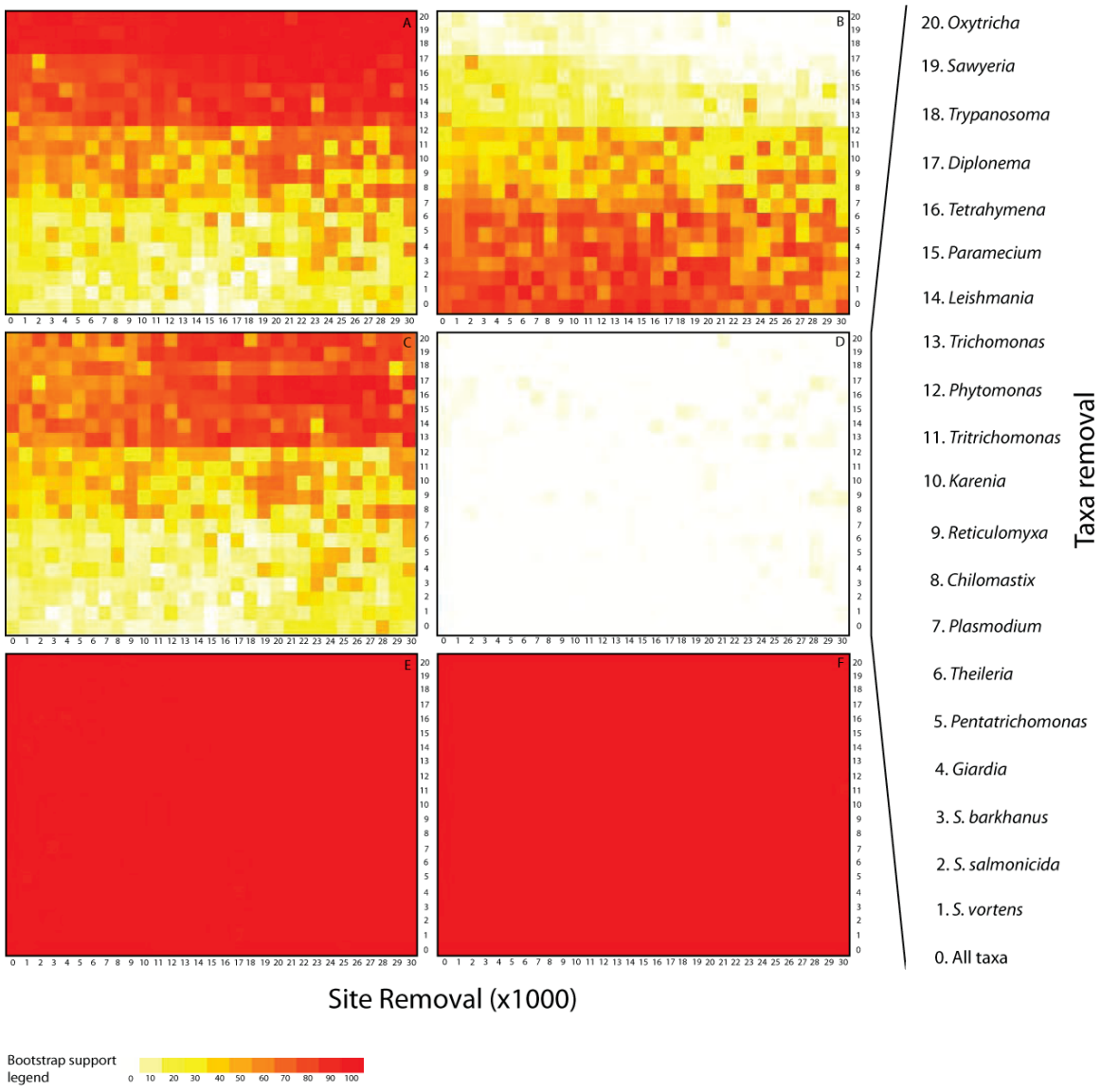


Figure 4.7. Heat map graphs depicting the bootstrap support for different topologies as sites are removed from fast-evolving taxa according to 'ratio of rates', together with removal of fast-evolving taxa from the 'Main' dataset. Numbers along the x axis represent the number of sites replaced with '-' character in fast-evolving taxa (x1000) and numbers on the y axis represent the number of removed taxa. The actual taxa removed are listed on the far right. A. Bootstrap support for monophyletic Excavata. B. Bootstrap support for paraphyletic Excavata (i.e., the ML topology). C. Bootstrap support for *Malawimonas*-Metamonada monophyly. D. Bootstrap support for *Malawimonas*-Metamonada-unikonts monophyly. E. Bootstrap support for Metamonada. F. Bootstrap support for unikont clade.

4.3. Discussion

4.3.1. Topology of the Tree of Eukaryotes

The overall tree topologies estimated are similar to those from previously published phylogenomic analyses in the recovery of strongly supported 'unikont' and 'S.A.R.' clades, and the grouping of the S.A.R. clade with Archaeplastida and the 'Hacrobia' lineages (Burki *et al.*, 2007; Burki *et al.*, 2008; Burki *et al.*, 2009; Hampl *et al.*, 2009). The mixing together of lineages typically assigned to Archaeplastida and Hacrobia mirrors the results of some other analyses (Burki *et al.*, 2007; Burki *et al.*, 2009; Hampl *et al.*, 2009). This could be potentially caused by our inability to distinguish endosymbiotically transferred genes in cryptophytes and/or haptophytes (Martin and Herrmann, 1998; Lane and Archibald, 2008), but could also be true historical signal that indicates the non-monophyly of 'Hacrobia', and possibly Archaeplastida.

4.3.2. The Phylogeny of Fornicata and Metamonada

Fornicata (diplomonads, retortamonads and CLOs) received maximal support, as in previously published studies (Hampl *et al.*, 2005; Simpson et al., 2006; Hampl *et al.*, 2009; Kolisko *et al.*, 2010; Parfrey *et al.*, 2010; Chapter 3), and our analyses also confirmed the close relationship and the paraphyly of the CLOs (Kolisko *et al.*, 2010; Chapter 3 and Appendix I). Metamonada – the group consisting of fornicates, *Trimastix*, oxymonadids and parabasalids – also received the highest possible support throughout the analyses, as did the grouping of Fornicata plus parabasalids within metamonads. In evaluating the evidence for Metamonada monophyly, it is

important to consider the possibility of LBA, since diplomonads and parabasalids constitute the longest branches on the tree (Simpson et al., 2006; Hampl *et al.*, 2009). In this study, however, the very high support for the monophyly of Metamonada was not reduced by any of the several analyses designed to suppress LBA (see Figures 4.2, 4.4, 4.5, 4.6 and 4.7). This strongly suggests that the robust Metamonada grouping seen with this dataset is due to true historical signal.

4.3.3. Monophyly of Excavata and Position of *Malawimonas*

4.3.3.1. Long Branch Breaking/Replacement

One of the main aims of this study was to employ the CLOs, which are relatively short-branching compared to their diplomonad and parabasalid relatives, to improve estimates of the evolutionary position of these phylogeneticaly problematic groups. The added CLOs could 'break' the extremely long branches leading to diplomonads and parabasalids (Poe, 2003; Brinkmann *et al.*, 2005; Geuten *et al.*, 2007), or could be used to completely replace these long-branching taxa within the analysis.

The initial analyses of the main dataset recovered Excavata as a 'paraphyletic' group, with *Malawimonas* branching as sister to unikonts with high support (BP=84%). This suggested that long branch breaking was not effective in this case, as this particular topology is suspected to represent a long branch attraction (LBA) artifact (Hampl *et al.*, 2009). Interestingly, the paraphyly of excavates received lower support in the 'CLO_rep' dataset (BP=63%). Based on this result I initially speculated

that a potential for long branch breaking indeed existed, but had been strongly affected by the missing data in the CLOs (i.e., in the 'breaking' taxa). This would have been caused by the fact that maximum likelihood is evaluated independently for each position in the alignment, which means that the long branch breaking cannot take effect at positions where none of the short-branching taxa were sampled. However, analyses where CLOs were excluded from the 'CLO_rep' dataset, and jackknifing analyses where sites were excluded at random to make datasets the same size as 'CLO_rep' (Appendix E.7) often showed similarly low support for Excavata paraphyly. Therefore the reduced length of the dataset rather then the long-branch breaking effect could explain the lower bootstrap support for LBA topology in the 'CLO_rep' dataset. This is reasonable, as LBA gets stronger with increasing amounts of similar data and may thus be reduced as data is excluded (Felsenstein, 1978). However the same would apply to true historical signal as well. In other words the long-branch-breaking strategy on its own does not seem to have been effective at overcoming suspected LBA in this particular case.

On the other hand, it seems that complete replacement of long branches (diplomonads and parabasalids) by shorter branches (CLOs) was an effective strategy to suppress LBA. The removal of long-branching taxa led to progressively lower support for the 'excavate paraphyly' topology, and correspondingly increased support for excavate monophyly. Excavate paraphyly received almost no bootstrap support, and monophyly of excavates received very high support (BP>90%), after removal of the 18 longest branching taxa, at which point all of the diplomonads and parabasalids, as well as *Chilomastix* were excluded, but all the CLOs remained

included. At this time CLOs were effectively a replacement for their long-branching diplomonad and parabasalid relatives.

Bearing these results in mind, perhaps it should become standard practice to exclude diplomonads (and potentially parabasalids) from global phylogenomic analyses of eukaryotes now that shorter-branching CLOs are available to represent the larger group. This is similar to the case of microsporidia, another important group of eukaryotes that consistently form extremely long branches (Vossbrinck *et al.*, 1987; Brinkmann *et al.*, 2005). Microsporidia are usually not included in global phylogenomic analyses of eukaryotes today because it is well-accepted that they belong to fungi (Edlind *et al.*, 1996; Van de Peer *et al.*, 2000) and their inclusion carries a strong risk of introducing systematic error into an analysis, without adding meaningfully to its phylogenetic informativeness (Brinkmann *et al.*, 2005).

4.3.3.2. Complex LBA Artifact Suppression

To understand better LBA artifact suppression we also employed other methods that are less drastic than complete removal of taxa from the analyses (Rodriguez-Ezpeleta *et al.*, 2007b). Most of these approaches substantially weakened the support for Excavata paraphyly (the inferred LBA topology). This is consistent with the hypothesis that the *Malawimonas*-unikonts clan indeed results from long branch attraction artifact, as suggested by previous studies (Hampl *et al.*, 2009).

The ratio-of-rates analyses seemed to perform as a relatively effective method for LBA artifact suppression, as it lowered support for assumed LBA topology even with full taxon sampling. The performance of this LBA suppression

method was slightly uneven along the series of data removal (Figure 4.7). This could be explained either as a result of introducing a large amount of missing data (which is inherent to the method) or of suboptimal splitting of the dataset into "long" and "short" taxa for the rate computations. The properties of this method and the most appropriate level of data removal and partitioning of taxa should be further explored in future studies.

The only method of LBA suppression that did not materially weaken the support for the supposed LBA topology is the removal of fast-evolving sites. However, these analyses revealed the existence of another potential phylogenetic signal for another alternative topology – the non-monophyly of Metamonada plus Discoba. It is therefore possible that the relatively close relationship between Metamonada and Discoba is a result of LBA attraction as well. However, it is important to note that the relationship between Metamonada and Discoba was not even slightly weakened by other LBA suppression analyses. Further, it is a common result to recover a strongly supported relationship between *Malawimonas* and Discoba when no metamonad species are included in the analysis (Rodriguez-Ezpeleta *et al.*, 2007a; Hampl *et al.*, 2009). This strongly contradicts the idea that the signal supporting Excavate monophyly is due to LBA between *Malawimonas*-Metamonada and Discoba. These results could also be explained by there not being enough data left for resolving the branch uniting the Metamonada and Discoba.

4.3.4. Concluding Remarks

In this study I explored the evolutionary history of Excavata through inclusion of *Carpediemonas*-like organisms, which are shorter-branching relatives of extremely long-branching diplomonads and parabasalids. It was inferred that the replacement of the long-branching diplomonads and parabasalids by CLOs may be an effective strategy to suppress long branch attraction artifact, while still retaining a broad sampling of the diversity of major excavate lineages.

Four different topologies were recovered over the course of our analyses, suggesting complex phylogenetic signals (Figure 4.6). Possible LBA suppression methods provide strong evidence that the Excavata paraphyly (*Malawimonas*-unikonts) topology recovered in the ML tree of the 'Main' dataset is the result of LBA artifact. Under certain circumstances the monophyly of Excavata receives very high support (removing long-branching taxa and ratio of rates analysis) and seems to be the most plausible estimate of the true topology (Figure 4.6, tree 2). However, it is also important to note that some analyses intended to suppress LBA support other topologies in which the Discoba are not specifically related to the rest of Excavata (i.e., Metamonads and *Malawimonas* - Figure 4.6, tree 4). Sampling of additional deep-branching excavates, especially deep-branching relatives of Metamonada may be useful to resolve the phylogenetic status of Excavata robustly.

This study highlights the general importance of taxon sampling for the better resolution of deep eukaryotic phylogeny. Importantly, the newly available CLOs are an attractive alternative to diplomonads and parabasalids for phylogenomic studies aimed at accurately estimating the tree of eukaryotes while maintaining a good

sampling of major taxonomic groups. In the near term, one of the main goals for eukaryotic phylogenomics will be to sample and obtain data from more species representing currently undersampled major lineages throughout the tree. In addition to groups within Excavata, sampling more taxa from 'Hacrobia' and 'Apusozoa', and from other mystery lineages, such as Collodictyonids, will be crucial for determining the correct topology of so-far unresolved parts of the eukaryotic tree.

4.4. Materials and Methods

4.4.1. EST Sequencing and Assembly

EST data were obtained from five previously unsampled isolates: *Carpediemonas membranifera* isolate BICM*, Ergobibamus cyprinoides* isolate CL*, Kipferlia bialata* Sagami isolate*, Chilomastix caulleryi* and *Tsukubamonas globosa* isolate TKB055*.* Mass culturing was performed through inoculating a large number of 50ml Falcon tubes or 500 ml flat culture flasks (Corning, USA) containing 40ml or 400ml of the particular culture media for each organism (Kolisko *et al.*, 2010; Yabuki *et al.*, 2011; Chapter 3). Approximately 10^9 cells per strain were processed. Cells were harvested via centrifugation and pelleted cells were lysed in Trizol (Invitrogen, Carlsbad, USA). For *E. cyprinoides* total RNA (±3.5mg) was isolated using the Trizol protocol, then subjected to two rounds of polyA selection (Poly(A)Purist™ MAG Kit, Ambion, Austin, USA) to isolate mRNA. The mRNA was then sent to Express Genomics (Baltimore, USA) for plasmid cDNA library construction. From this 5000 Sanger

sequences were obtained (Agencourt, Beverly, USA) and a single run of 454 sequencing (Roche, Indianapolis, USA) was performed that yielded 158696 reads (454 sequencing performed by Genome Quebec, Montreal, Canada). From the remaining four taxa total RNA was isolated using Trizol (Invitrogen, USA), then poly-A selection (2 rounds) and cDNA synthesis was performed by Vertis (Freising, Germany). The resulting cDNA was then sequenced by Genome Quebec using 454 technology. The number of reads obtained is as follows: 575603 from *Ca. membranifera*, 281500 from *K. bialata*, 283187 from *Ch. caulleryi* and 269568 from *T. globosa*.

The obtained data were assembled using Newbler 2.0 (Roche, Indianapolis, USA) and Mira 2.9 (Chevreux *et al.*, 2004) and the best contig for each targeted gene was selected from the combination of both assemblies. 10103 Mira and 5175 Newbler contigs were obtained from *Ca. membranifera*, 6020 Mira and 3618 Newbler contigs from *E. cyprinoides,* 5555 Mira and 1097 Newbler contigs from *K. bialata*, 8163 Mira and 5058 Newbler contigs from *Ch. caulleryi* and 7408 Mira and 4674 Newbler contigs from *T. globosa*.

4.4.2. Core Dataset Assembly

The dataset was assembled through the combination of two previously published phylogenomic datasets, compiled by Hampl et al. (2009) and by Burki et al. (2009), resulting in 161 genes. Several additional newly available taxa were added to the dataset. Reciprocal blast against the Swissprot (Bairoch and Apweiler, 1997) and OrthoMLC (Li *et al.*, 2003) databases was used to semi-automatically identify

paralogs that were subsequently removed from the analyses. Alignments for each gene were built using the FSA alignment program (Bradley *et al.*, 2009) and ambiguously aligned positions were masked by eye in the program Bioedit (Hall, 1999). The function '--add' was used in the program MAFFT (setting 'linsi', Katoh *et al.*, 2002; Katoh *et al.*, 2005) to preserve the existing mask when adding new species. From the main dataset we constructed a second dataset that contained only the sites that were represented by an amino acid residue in at least one of the CLO isolates (*Ca. membranifera*, *E. cyprinoides* or *K. bialata*). All the taxa included possess at least 15% of the alignment positions. Two taxa, *Breviata anathema* and *Raphidiophrys contractilis*, which were missing 78% and 71% of aligned positions respectively, were removed from the dataset because preliminary analysis identified them as 'rogue taxa' according to an algorithm for uncovering hidden phylogenetic consensus (Pattengale et al. 2010). The final two datasets, referred to as 'Main' and 'CLO_rep', consisted of 85 taxa, and 43516 sites or 32995 sites respectively, representing data from 161 and 144 genes respectively. For gene sampling per taxon and for additional data sources see Hampl *et al.* (2009) and Appendix E.8 and E.9.

4.4.3. Phylogenetic Analyses

4.4.3.1. Tree Construction

The matrix was analyzed as one un-partitioned supergene in the program RAxML 7.2.6. (Stamakis, 2006) The LG model with estimated amino acid frequencies was

used along with a gamma distribution for modeling variable rates across sites (setting PROTGAMMALGF, with 4 categories). The search for the maximum likelihood tree was performed with 20 taxon addition replicates (setting –N 20). The statistical support was estimated using non-parametric bootstrapping, with 500 replicates.

The 'Main' dataset was also analyzed with a Bayesian method using PhyloBayes3.2 (Lartillot and Philippe, 2004; Lartillot *et al.*, 2009), Four independent chains were run with model set for CAT+Γ for 40000 generations, with burnin set to 4000. Convergence was assessed using the program bpcomp. However, only three chains converged to the same result. Both topologies with associated posterior probabilities are depicted in supplementary materials (Appendices G.1 and G.2)

4.4.3.2. Fast-Evolving Taxa Removal

To identify the longest-branching taxa we used the maximum likelihood trees estimated for the two original datasets ('Main' and 'CLO_rep'). The degree to which each terminal taxon represented a long branch was measured as an average of the ten largest distances between it and the other taxa in the dataset. This allowed an evaluation of the overall branch lengths of each taxon regardless of the rooting of the tree. The taxa were then sorted by this metric, and the 20 longest-branching taxa were removed sequentially from the dataset. The 21 resulting datasets (including zero taxa removed) were analysed by rapid bootstrapping with 100 replicates in the program RAxML (model settings: PROTCATLG).

4.4.3.3. Fast-Evolving Sites Removal

Rates of evolution for every site in each of the 21 datasets generated by removing fast-evolving taxa (see 4.4.3.2., above) were estimated using the program 'distest' (Susko *et al.*, 2003). Sites were then sorted from slowest to fastest and the fastest sites were sequentially removed in blocks of 1000, until 30000 were removed (31 steps, including zero sites removed). This resulted in 651 datasets with a range of taxa and site inclusions. These datasets were analyzed by rapid bootstrapping with 100 replicates in the program RAxML (model settings: PROTCATLG). The support for topologies of interest was plotted on heatmap graphs using the R package.

4.4.3.4. 'Ratio of Rates' and Site Removal

This analysis also started with the 21 datasets generated by sequential fast-evolving taxa removal. In each of these 21 datasets the branch length of each taxon was estimated the same way as in the fast-evolving taxa removal analyses (see 4.4.3.2, above). Two sub-datasets were then constructed, each containing one third of all taxa; sub-dataset 1 contained the fastest-evolving taxa; sub-dataset 2 contained the slowest-evolving taxa. For each of these two datasets rates of evolution at each site were estimated using the program distest (Susko *et al.*, 2003). For each site the ratio between the rates in sub-datasets 1 and 2 was computed (the higher this ratio, the faster the long-branching taxa are evolving relative to short-branching taxa). All sites were sorted according to this ratio, and sites from fast-evolving taxa were then sequentially replaced by missing characters in blocks of 1000 sites (e.g., step1 – 1000 sites with the highest ratio replaced by '-' in the fast-evolving taxa, step2 –

2000 sites with the highest ratio replaced by '-' in the fast-evolving taxa, etc.). For each step the full datasets were then reassembled, resulting in datasets with a higher proportion of missing data in long-branching taxa. This was done up till 30000 sites were replaced by missing data in long-branching taxa, resulting in 651 datasets total (as in the previous analysis – see 4.4.3.3). These datasets were again analysed by rapid bootstrapping analyses in RAxML (model setting PROTCATLG). The support values for topologies of interest were plotted on heatmap graphs using R.

4.4.3.5. Fast Gene Removal

Maximum likelihood trees were constructed for each single gene dataset in the program RAxML (settings: PROTGAMMALGF). The branch length for each gene from each taxon was estimated by the average of the 10 longest pairwise distances, as described above (see 4.4.3.2). All gene-taxon pairs were sorted according to this branch length metric (resulting in a list: "gene2_taxon5, gene5_taxon67, gene125_taxon1…", where gene2_taxon5 has an overall longest branch lengths). The genes were then sequentially removed from the dataset 500 at a time up to 8000 genes removed. Each dataset was then rapidly bootstrapped as described above.

4.4.3.6. Amino Acid Recoding

Amino acids were recoded into five categories ('A', 'T', 'C', 'G' and '-') as in Hrdý *et al.* (2004) and the dataset was analyzed using rapid bootstrapping as described above.

4.4.3.7. Additional Examination of the 'CLO_rep' Dataset

To test the potential for 'long-branch breaking' in the 'CLO_rep' dataset two supplementary analyses were performed. (i) The analysis of the 'CLO_rep' dataset was repeated as described above (see 4.4.3.1), but with the three CLOs excluded. (ii) Ten datasets were constructed by removing 10526 sites at random from the 'Main' dataset (this is the same number of sites as was excluded from the 'CLO_rep' dataset). These ten datasets were then analyzed by bootstrapping as for the 'CLO_rep' dataset (See 4.4.3.1), but with 300 replicates rather than 500.

Chapter 5

Reductive Evolution of Mitochondria-Related Organelles in Metamonada

5.1 Introduction

A key event in the evolution of the eukaryotic cell was the enslavement of an α-proteobacterium to form the organelle that became the modern mitochondrion (Margulis, 1970; Martin and Müller, 1998; Cavalier-Smith, 2002). The presence of mitochondria is considered as one of the signatures of eukaryotic organisms and in some theories the acquisition of mitochondria is even considered to be the "founding step" of eukaryotic cells (Sagan, 1967; Margulis, 1970; Martin and Müller, 1998). Mitochondria are crucial organelles for eukaryotic cells since they function in respiration (which can supply most of a cell's energy), iron-sulfur cluster synthesis, the urea cycle and the metabolism of amino acids and fatty acids.

However, several anaerobic groups of protists do not possess classical mitochondria. Several of these were, at one point, classified as members of the taxon Archezoa (Cavalier-Smith, 1983), a group that was erected to house eukaryotes that were supposed to have arisen before the acquisition of mitochondria. Later studies have shown, however, that all of the putative archezoans possess genes of mitochondrial origin (i.e., genes that were transferred from mitochondrial to the nuclear genome of the host) and/or actual remnant mitochondrial organelles (Roger *et al.*, 1998; Tovar *et al.*, 2002; Tovar *et al.*, 2003). Today, there is no major eukaryotic group known that does not possess a strong candidate mitochondrion-

related organelle (MRO)(Tovar *et al.*, 2002; Tovar *et al.*, 2003, Barbera *et al.*, 2007). In most studied cases, though not all, these anaerobic MROs now lack their own genome (Barbera *et al.*, 2007).

The most famous putative archezoans were diplomonads and parabasalids, as they also emerged as very deep eukaryote branches in early molecular phylogenetic studies (Sogin *et al.*, 1989; Van Keulen *et al.*, 1993), which seemed to corroborate their primitive status. This 'basal' position is today known to be a result of phylogenetic analysis artifacts (Gribaldo and Philippe, 2000; Philippe *et al.*, 2000) and there is relatively strong evidence for parabasalids and diplomonads being somewhat closely related (Hampl *et al.*, 2005; Simpson et al., 2006). Parabasalids possess hydrogenosomes while the diplomonad *Giardia* possesses mitosomes – these are the two main kinds of anaerobic mitochondria-related organelles.

Parabasalids, such as the human parasite *Trichomonas*, possess relatively large (~500 nm) organelles bounded by two membranes but lack cristae and genomes. These organelles are called hydrogenosomes as they produce molecular hydrogen (Lindmark and Müller, 1973). Biochemically similar organelles are also known from anaerobic ciliates and chytrid fungi, but each of these MROs have evolved independently from aerobic mitochondria (Yarlett *et al.*, 1981; Yarlett *et al.*, 1984; Yarlett *et al.*, 1986; Müller, 1993).

The hydrogenosomes of *Trichomonas* have several known functions. They generate ATP anaerobically through the conversion of pyruvate to acetate. This metabolic pathway uses pyruvate:ferredoxin oxidoreductase to convert pyruvate to acetyl-CoA and carbon dioxide, while reducing ferredoxin, and then

acetate:succinate CoA transferase (ASCT) to convert acetyl-CoA to acetate. Other enzymes involved in this pathway are [FeFe]-hydrogenase, that oxidizes ferredoxin and reduces protons to form molecular hydrogen, and succinyl-CoA synthetase that converts succiny-CoA to succinate (succinate is necessary for function of ASCT), while generating ATP. A decarboxylating malate dehydrogenase (also known as the malic enzyme) is also present in hydrogenosomes. This converts malate to pyruvate and CO_2 while reducing NAD+ to NADH. In turn, the 51 and 24 kDa subunits of NADH dehydrogenase (Complex I of the mitochondrial electron transport chain) have been shown to function in hydrogenosomes to regenerate NAD^+, reducing ferredoxin in the process, ultimately feeding into hydrogen production via [FeFe]-hydrogenase (Lahti *et al.*, 1992; Müller, 1993; Lahti *et al.*, 1994; Brugerolle *et al.*, 2000; Doležal *et al.*, 2004; Hrdý *et al.*, 2004; Pütz *et al.*, 2006; Van Grinsven *et al.*, 2008). Like mitochondria, hydrogenosomes also function in Fe-S cluster synthesis using IscU, IscS, ferredoxin and several other proteins (Lill *et al.*, 1999; Sutak *et al.*, 2004; Lill and Mühlenhoff, 2005; Tachezy and Doležal, 2007). They also house two of the four glycine cleavage system proteins (the H- and L- subunits), as well as the serine hydroxymethyl transferase enzyme (Mukherjee *et al.*, 2006a; Mukherjee *et al.*, 2006b) that, in other organisms transform glycine into serine. Finally, hydrogenosomes also possess an oxygen scavenging system that acts in oxygen defence (Lindmark and Müller, 1974; Ellis *et al.*, 1994; Pütz *et al.*, 2005) a mitochondrial-like protein import and folding machinery as well as transporters that move metabolites and ATP across the bounding membranes (Dyall *et al.*, 2000; Dyall *et al.*, 2003; Doležal *et al.*, 2005).

In contrast to the large and relatively complex hydrogenosomes of parabasalids, the diplomonad *Giardia intestinalis*, has mitosomes, that are amongst most reduced MROs observed to date. These organelles are so small that localization of iron sulfur (Fe-S) cluster assembly proteins to the organelle was required for the mitosomes of *Giardia* to even be detected in the first place (Tovar *et al.*, 2003). Based on these initial experiments it was proposed that Fe-S cluster assembly might be the sole primary function of *Giardia's* mitosome (Lill *et al.*, 1999; Tovar *et al.*, 2003). This was very recently supported and extended by proteomic studies, which did not identify a suite of mitosomal proteins that would be related to primary functions other than Fe-S cluster assembly. The one possible exception is oxidoreductase GiOR-1 that localized into the mitosome. Its function is unknown, but based on the measured activity of the expressed protein, it was speculated that it may be involved in electron transport (Jedelský *et al.*, 2011). There are also known some basic mechanisms for protein and metabolite transport to the *Giradia* mitosome. However neither the inner membrane pore complex nor ATP/ADP transporters have, so far, been identified in *Giardia* mitosomes, nor have candidate homologs been identified in its nuclear genome.

Recent research on previously poorly studied species highlights the functional and organizational diversity of anaerobic mitochondrial organelles in eukaryotes (Gill *et al.*, 2007; Hampl *et al.*, 2008; Stechmann *et al.*, 2008; Barbera *et al.*, 2010). For example, the organelles of both *Blastocystis* sp. and *Nyctotherus ovalis* still possess a genome (Akhmanova *et al.*, 1998; Wawrzyniak *et al.*, 2008), and house a very diverse proteome suggesting that they host a wide variety of metabolic

pathways. In *Blastocystis* sp. proteins typical for both regular mitochondria and hydrogenosomes were discovered, for example both Pyruvate dehydrogenase and Pyruvate:ferredoxin oxidoreductase (Stechmann *et al.*, 2008; Denoeud *et al.*, 2011). Other anaerobic protists with mitochondria-related organelles of more or less unknown functions include *Sawyeria* and *Trimastix* (Hampl *et al.*, 2008; Barbera *et al.*, 2010). They possess organelles that generally resemble hydrogenosomes, but appear have a more complex metabolism than the *Trichomonas* hydrogenosome described above. For example they both possess a full glycine cleavage system and some proteins of mitochondrial fatty acid metabolism. Based on these studies it seems that there is a functional "continuum" of organelles ranging from typical mitochondria to mitosomes (Hjort *et al.*, 2010).

As shown in Chapter 4 as well as several previous studies (Hampl *et al.*, 2005; Simpson et al., 2006), diplomonads and parabasalids are related to one another, but quite distantly. However, it remains unclear whether the mitosomes of diplomonads evolved from organelles of greater, lesser or equal complexity to the hydrogenosomes of *Trichomonas*. Also both *Trichomonas* and *Giardia* are highly specialized parasites and some of the reduction in organelle function observed in these species may be the direct result of independent adaptation to their parasitic life style.

The diversity and phylogenetic positions of *Carpediemonas*-like organisms revealed in the previous two chapters (and Appendix G, H and I) make it conceivable to investigate the possible properties of the ancestral MRO of diplomonads and to propose a sequence of evolutionary steps that led to the mitosomes of *Giardia*. The

goal of this study was to explore the predicted proteomes of four basal lineages related to diplomonads (e.g., *Giardia intestinalis*). Three are groups of free-living *Carpediemonas*-like organisms, CLOs, represented by *Carpediemonas membranifera*, *Ergobibamus cyprinoides* and *Kipferlia bialata*. The fourth is *Chilomastix caulleryi*, which is a parasitic/commensalic retortamonad, and, relative to the three CLOs included in the study, it represents the most closely related lineage to diplomonads (see chapters 3 and 4, and appendix I). Both *Carpediemonas* and *Ergobibamus* possess relatively large (~300 nm) double bound membrane bound organelles that superficially resemble hydrogenosomes when viewed by transmission electron microscopy (TEM) (Simpson and Patterson, 1999; Park *et al.,* 2010; Simpson unpublished; appendix H). There are no data available for *C. caulleryi,* but the presumably related *Chilomastix cuspidata* also possesses relatively large double-membrane bound organelles somewhat reminiscent of parabasalid hydrogensomes when viewed using TEM, although these seem to be rare in the cell (Hampl and Simpson, 2007). By *in-silico* analyses of the EST datasets (see Chapter 4) of the four lineages, the partial organellar proteomes were predicted. These predictions were then compared to the organelle proteomes of *Trichomonas* and *Giardia.*

5.2 Results

I analyzed 13006 clusters of ESTs from *Carpediemonas*, 6421 from *Ergobibamus*, 7330 from *Kipferlia* and 9056 from *Chilomastix*. Out of these, 29, 35, 23 and 21 clusters respectively were identified as putative mitochondrial/organellar function

using CBOrg and four mitochondrial prediction algorithms: Mitoprot (Claros and Vincens, 1996), TargetP (Nielsen *et al.*, 1997; Emanuelsson *et al.*, 2000), Mitopred (Guda *et al.*, 2004a; Guda *et al.*, 2004b) and Predotar (Small *et al.*, 2004). For details of localization predictions see Table 5.1 and Appendix F.1.

Table 5.1. Table showing the putative MRO proteomes of *Trichomonas*, *Giardia*, *Carpediemonas*, *Ergobibamus*, *Kipferlia* and *Chilomastix*. In *Trichomonas,* 'H' indicates that the protein is localized in hydrogenosomes. In *Giardia,* 'C' marks localization in cytosol and 'M' in mitosome. In the four studied organisms, any number, including a zero, indicates the predicted presence of a protein. The actual number represents how many of the four localization software tools predicted that the protein was localized in an MRO (see appendix F.1); the tp means that a putative targeting peptide was identified (see also Figure 5.1). The '-' means that a homolog of this protein was not identified in this particular organism. The '*' means that no homolog so far was identified among the ESTs

Protein name	***Trichomonas***	***Giardia***	***Carpediemonas***	***Ergobibamus***	***Kipferlia***	***Chilomastix***
Energy metabolism						
Pyruvate:ferredoxin oxidoreductase	H	C	*	0	*	0
[FeFe]-hydrogenase	H	C	0	0	*	*
Malic enzyme	H	C	0	0	1	0
NADH dehydrogenase 51 kDa	H	-	*	0	1	*
NADH dehydrogenase 24 kDa	H	-	1	0	*	*
Adenylate kinase	H	C	0	0	0	0
Acetate:succinate CoA transferase	H	-	*	*	0	*
Succinate CoA synthetase – alpha subunit	H	-	1	2	3(tp)	*
Succinate CoA synthetase – beta subunit	H	-	1(tp)	1	0	*
Acetyl-CoA synthetase	-	C	*	*	*	0
Hydrogenase maturase protein F	H	-	*	1	0	*
Hydrogenase maturase protein E	H	-	0	*	*	*
Hydrogenase maturase protein G	H	-	*	*	*	*

Table 5.1. Continued

Amino acid metabolism						
Serine hydroxymethyl transferase	H	-	2(tp)	0	*	*
T-protein	-	-	0	0	0	*
P-protein	-	-	*	2	0	*
L-protein	H	-	0	0	0	*
H-protein	H	-	0	3(tp)	*	*
L-protein (Spirochaete version)	-	-	0	0	*	*
Alanine transaminase	H	C	*	0	0	1
Branched chain amino acid aminotransferase	C	C	*	2	0	0
Phosphatidylserine decarboxylase	C	C	*	2	*	*
Aspartate aminotransferase mitochondrial	C	C	*	0	0	1
Glutamate dehydrogenase	C	C	*	0	1	1
Ornithine aminotransferase	C	C	*	0	*	*
Ornithine carbamoyltransferase	C	C	*	0	0	1
Iron-Sulfur cluster synthesis						
Scaffold protein IscU	H	M	3(tp)	4(tp)	4(tp)	4(tp)
Cystein desulfurase IscS	H	M	0	3	*	0
Ferredoxin	H	M	1(tp)	*	2(tp)	0
Isa1/Isa2 (only Isa2 known from *Trichomonas* and *Giardia*)	H	M	*	*	*	*
ISD11	H	-	*	*	*	*
Frataxin	H	-	2	0	*	*
Glutaredoxin	-	C	0	0	0	0
DNAJ Jac1	H	M	*	*	*	*
Iron-sulfur cluster co-chaperone HscB	-	M	*	2	*	*
Oxygen scavenging system						
Hydroperoxide reductase ruberythrin	H	-	*	*	*	*
Thioredoxin	H	C	1(tp)	0	*	*
Thioredoxin peroxidase	H	C	0	3	0	0
Thioredoxin reductase	H	C	*	0	*	2
Superoxide dismutase	H	C	0	0	0	0
Import machinery						
Inner membrane translocase 23 (Tim23)	H	-	2	*	*	*
Outer membrane translocase 40 (Tom40)	H	M	*	*	*	*
Inner membrane translocase Pam18	H	M	4(tp)	*	*	2(tp)
Mdj11	H	-	*	*	*	*
Mge	H	-	*	*	*	*
Sam50	H	-	*	*	*	*

Table 5.1. Continued

Hydrogenosomal integral membrane protein 35	H	-	*	*	*	*
Hydrogenosomal integral membrane protein 31 putative ATP/ADP translocase	H	-	1	2	*	2
Mitochondrial carrier protein	H	-	2	1	1	*
Mitochondrial folate transporter	-	-	2	*	*	*
Mitochondrial processing peptidase beta-subunit	H	M	2	0	*	*
Heat shock protein 70	H	M	2	1	0	*
Heat shock protein CPN10	H	M	2	*	*	*
Heat shock protein CPN60	H	M	*	2	*	*
Fatty acid metabolism						
Long chain fatty acid CoA ligase	-	-	0	2	0	0

5.2.1 Identification of Targeting Peptides

Thirteen of the clusters encoding proteins of possible organellar function possess putative targeting peptides that could be identified by Mitoprot and TargetP prediction software and by comparison to homologous prokaryotic sequences (Figure 5.1). As an example, Figure 5.2 depicts a comparison of the N-terminus of the protein IscU in the four studied organisms, *Giardia, Trichomonas*, and the α-proteobacterium *Rickettsia*. As expected, the putative targeting peptides represent N-terminal extensions relative to the bacterial sequence, and possess a high proportion of basic and hydroxylated amino acids.

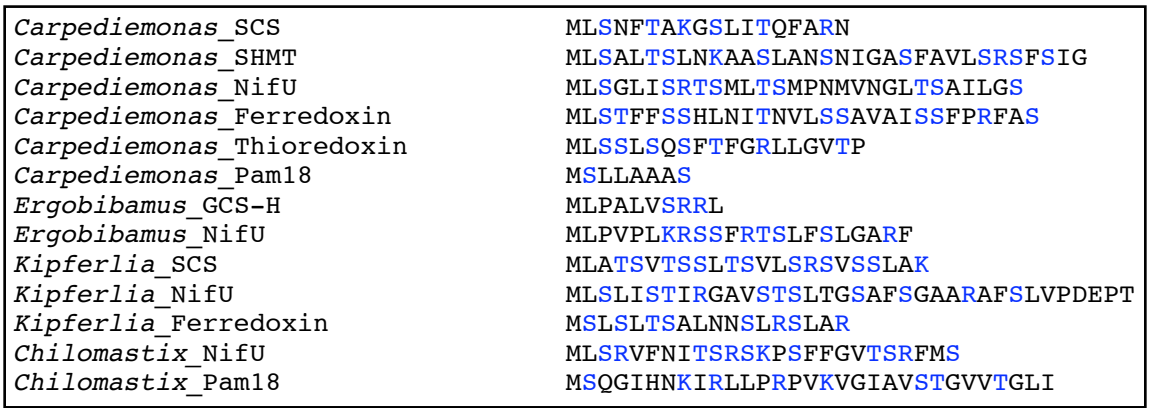


Figure 5.1. Predicted targeting peptides from the four EST projects. Basic and hydroxylated amino acids are highlighted in blue (see appendix F.1 for actual prediction results).

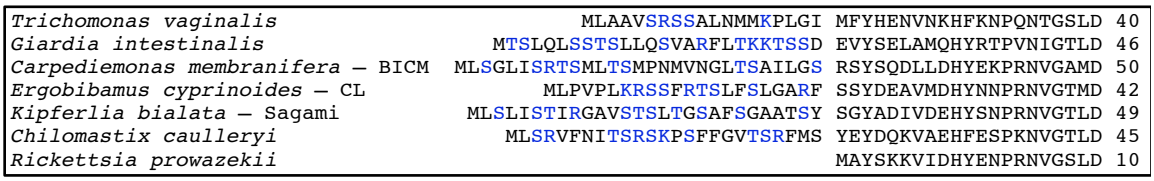


Figure 5.2. Predicted targeting peptides of IscU proteins of *Carpediemonas*-like organisms and *Chilomastix* and their comparison to targeting peptides of *Trichomonas and Giardia*. *Rickettsia* represents bacterial sequences, with no targeting peptides. Basic and hydroxylated amino acids are highlighted in blue.

5.2.2 Energy Metabolism

I identified numerous transcripts encoding enzymes for anaerobic pyruvate metabolism, as predicted for the hydrogenosome of *Trichomonas*. Transcripts encoding the three central proteins involved in hydrogenosome-type anaerobic energy metabolism, pyruvate:ferredoxin oxidoreductase (PFO), [FeFe]-hydrogenase(Fe-hyd) and ferredoxin (Fdx), were all identified in *Ergobibamus.* PFO

transcripts were not detected in *Carpediemonas,* and Fe-hyd transcripts were not found in the data from *Chilomastix*. Neither PFO nor Fe-hyd were found in *Kipferlia*. Two different proteins involved in maturation of [FeFe]-hydrogenase, HydF and HydE were identified in the data from the three CLOs (however only one of those in each, table 5.1). Malic enzyme and adenylate kinase were also identified in all four organisms (Adenylate kinase is also found in hydrogenosome of *Trichomonas* and participates in energy metabolism). Both subunits of Complex I (24 kDa and 51 kDa) were identified in *Ergobibamus*, the 24 kDa subunit was also found in *Carpediemonas* and the 51kDa subunit in *Kipferlia*. Acetate:succinate CoA transferase (ASCT) was identified only in *Kipferlia*, however the enzyme that acts in concert with ASCT, succinyl-CoA synthetase (SCS) was identified in all three CLOs. Both the alpha and beta subunits of SCS were detected in each species. The SCS subunits beta of *Carpediemonas* and alpha of *Kipferlia* appear to have targeting peptides (Figure 5.1). A different enzyme, ADP-forming Acetyl-CoA synthetase (ACS), was identified in *Chilomastix*. Both ASCT and ACS convert acetyl-CoA to actetate, though the latter directly generates ATP in the process, while in the former case ATP is generated only upon regeneration of succinate via SCS (Steinbüchel and Müller, 1986; Tielens and Van Hellemond, 2007).

5.2.3 Iron Sulfur Cluster Assembly

Major parts of the iron-sulfur cluster assembly pathway were identified in the three CLOs and/or *Chilomastix*. The scaffold protein IscU, cysteine desulfurase (IscS) and ferredoxin were found in all four organisms (except IscS in *Kipferlia* and ferredoxin

in *Ergobibamus*). There is good evidence that IscU is targeted to the organelle in all four organisms since all four mitochondrial localization prediction tools predicted the translated cluster as mitochondrial, and the encoded polypeptide possesses a putative targeting peptide in all four species (Figure 5.2). Other proteins involved in Fe-S cluster assembly are frataxin and the co-chaperone HscB; these were found in some of the four studied organisms (Appendix F.1).

5.2.4 Amino Acid Metabolism

Transcript clusters encoding several enzymes of amino acid metabolism were found in the analyzed data. Perhaps the most important are the four parts of the glycine cleavage complex (glycine cleavage system T protein, GCS-T; glycine cleavage system P protein; GCS-P; glycine cleavage system L protein, GCS-L; glycine cleavage system H protein, GCS-H) and serine hydroxymethyl transferase (SHMT). Transcripts encoding all five proteins were identified in *Ergobibamus*. Four and three homologs of subunits of this system were also found in *Carpediemonas* and *Kipferlia* respectively, but interestingly, none of them were detected in *Chilomastix*. Two different GCS-L transcripts were identified in *Ergobibamus* and *Carpediemonas*, one that seemed to be the regular eukaryotic version, based on similarity, and one that is most similar to GCS-L from the spirochaete bacterium *Leptospira*, that likely represent a lateral gene transfer (Table 5.1, Appendix F.1). There is some evidence that the GCS-P and GCS-H homologs of *Ergobibamus* are localized to the organelle, as two and four of the localization prediction tools suggested these to be mitochondrial

(Table 5.1). Furthermore, the GCS-H protein from *Ergobibamus* possesses a putative targeting peptide (Figure 5.1).

Other enzymes involved in amino acid metabolism of possible mitochondrial origin, identified in the analyzed data, are branched-chain-amino-acid aminotransferase, phosphatidylserine decarboxylase, aspartate aminotransferase, glutamate dehydrogenase, ornithine aminotransferase and ornithine carbamoyltransferase.

5.2.5 Import Machinery

Homologs of a few parts of the mitochondrial-type protein import machinery were identified in our EST data. Notably, transcripts encoding the inner membrane translocase Pam18 and TIM23 proteins were found in *Carpediemonas*, while a Pam18 homolog was also found in *Chilomastix*. A mitochondrial processing peptidase beta subunit homolog was identified in *Carpediemonas* and *Ergobibamus*. Several molecular chaperones involved in transport machinery were identified. Mitochondrial-type HSP70 was found in all three CLOs, chaperonin-10 (CPN10) in *Carpediemonas* and chaperonin-60 (CPN60) in *Ergobibamus*. Homologs of a few other transporters and carriers were also found, predominantly in *Carpediemonas* and *Ergobibamus* (Mitochondrial carrier protein, hydrogenosomal integral membrane protein 31, Table 5.1).

5.2.6 An Oxygen Scavenging System

Proteins putatively involved in oxygen stress protection were found encoded by transcripts in our data: thioredoxin, thioredoxin reductase and peroxidase and superoxide dimutase. These are present in both *Trichomonas* and *Giardia* and while they are putatively localized to the hydrogenosome of *Trichomonas,* only one protein possibly involved in oxygen stress defence was found in *Giardia* mitosomes (Jedelský *et al.*, 2011). We have identified parts of this pathway in all four organisms. Only thioredoxin of *Carpediemonas* and thioredoxin peroxidase of *Ergobibamus* show evidence for an organellar localization.

5.2.7 Other

As mentioned above, "typical" mitochondria also function in fatty acid synthesis. One protein involved in fatty acid metabolism, long chain fatty acid CoA ligase, was identified in all four organisms.

5.3 Discussion

The main aim of this study was to compare partial putative organellar proteomes of representatives of four deep-branching relatives of diplomonads to the proteomes of diplomonad *Giardia* and the parabasalid *Trichomonas* and thus shed light on the evolution of these anaerobic organelles since their divergence from a common ancestor.

5.3.1 Anaerobic Energy Metabolism

Two enzymes, PFO and [FeFe]-hydrogenase, form an important element of energy metabolism in anaerobic protists (Tielens and Van Hellemond, 2007). In *Trichomonas* both of these proteins function within hydrogenosomes (Williams *et al.*, 1987; Bui and Johnson, 1996), while in *Giardia* they are localized in the cytosol (Tielens and Van Hellemond, 2007). We failed to recover the sequence encoding the N-terminus for either of these two proteins in any of our study organisms, which makes *in-silico* prediction of their subcellular localization difficult. However, these two enzymes are metabolically connected with the activities of several other enzymes including: (i) acetate:succinate CoA transferase (ASCT – converts Acetyl-CoA to acetate and transfers CoA on to succinate), (ii) succinyl-CoA synthetase (SCS – converts succinyl-CoA to succinate and produces ATP (or GTP) in the process), (iii) Malic enzyme (converts malate and NAD+ to pyruvate, $C0_2$ and NADH), and (iv) the 54 and 21 kDa subunits of complex I in *Trichomonas* that deoxidize NADH to NAD+ (Brugerolle *et al.*, 2000; Doležal *et al.*, 2004; Hrdý *et al.*, 2004; Van Grinsven *et al.*, 2008). At least one of these was predicted to be localized in the organelle in all three CLOs and some possesses a putative targeting peptide. Moreover, ASCT and SCS are exclusively organelle-localized in eukaryotes that have been studied to date (Van Grinsven *et al.*, 2008). This suggests that a hydrogenosomal-type energy-generating metabolic pathway is localized in the organelles in the three studied CLOs. Interestingly the ASCT of *Trichomonas vaginalis* is different from other known mitochondrial ASCT types (e.g., in trypanosomatids and *Ascaris*), as it is most similar to the Acetyl-CoA hydrolase enzyme family (Van Grinsven *et al.*, 2008). The putative

ASCT of *Kipferlia*, I have identified was also most similar to the same Acetyl-CoA hydrolase family as the ASCT of *Trichomonas* belong to (identified by blast search, ASCT from *Trichomonas* was not the best hit). By contrast none of these energy-generating functions are found in the organelle of *Giardia* (Jedelský *et al.*, 2011). In *Giardia* pyruvate is converted to acetyl-CoA in the cytosol, again using PFO and Fe-Hyd, and acetyl-CoA is then converted to acetate using Acetyl-CoA synthetase generating ATP in the process (alternatively pyruvate is converted to ethanol using alcohol dehydrogenase E) (Tielens and Van Hellemond, 2007). *Giardia* also lacks both the ASCT and SCS enzymes (Morrison *et al.*, 2007; Jedelský *et al.*, 2011). Interestingly we see a similar pattern in *Chilomastix*: ASC was identified in the *Chilomastix* ESTs, but neither ASCT nor SCS were found.

5.3.2 Iron-Sulfur Cluster Assembly

Iron-sulfur cluster assembly is the most broadly conserved function of mitochondria and related organelles (Lill *et al.*, 1999; Tovar *et al.*, 2003; Lill and Mühlenhoff, 2005). It has been suggested that iron-sulfur cluster synthesis is a universally present function of mitochondria and mitochondria-like organelles and is, in some cases, the only reason for the maintenance of these organelles in extremely specialized parasites like *Giardia* (Lill *et al.*, 1999; Tovar *et al.*, 2003). The only possible exception is represented by *Mastigamoeba balamuthi* and *Entamoeba histolytica*, which possesses versions of these proteins derived by lateral gene transfer from epsilon-proteobacteria, that may be partially localized in the organelle (Van der Giezen *et al.*, 2004; Gill *et al.*, 2007; Aguilera *et al.*, 2008; Maralikova *et al.*,

2010). In *Giardia,* iron-sulfur cluster synthesis is the only known cellular function performed by the mitosome (Tovar *et al.*, 2003; Jedelský *et al.*, 2011). The data presented in this chapter support the near-universality of iron-sulfur cluster assembly function of mitochondria and MROs, as these proteins are present in all four studied organisms. Also IscU at least possess targeting peptides in all four studied organisms (Figure 5.2), again suggesting strongly that it functions inside the organelle.

5.3.3 Import Machinery

Import machinery is crucial for any MRO as most organellar proteins are nucleus-encoded, and must be synthesized in the cytosol and then imported into the organelle (all proteins must be imported if the organelle has no genome). In both *Trichomonas* and *Giardia* there is a known homolog of TOM40, a protein complex that transports proteins across the outer membrane of the organelle. Pam18, Pam16 and mtHSP70 drive translocation across the inner membrane and are also present in both species (Bozner, 1997; Tovar *et al.*, 2003; Doležal *et al.*, 2005; Šmíd *et al.*, 2008; Jedelský *et al.*, 2011). The inner membrane pore protein complex, formed by TIMs 17, 22 and 23, is apparently present in *Trichomonas* (Doležal *et al.*, 2006), but interestingly, is not found in the *Giardia* mitosomal proteome, and genes encoding homologs of these TIM proteins were not detected in the *Giardia* genome (Jedelský *et al.*, 2011). Our data suggest that the loss of the TIM17/22/23 complex happened sometime after the divergence of *Carpediemonas* (*Carpediemonas* is the most basal sampled clan of Fornicata – see appendix I), as TIM23 was detected in the

Carpediemonas data. It is possible that the TIM pore complex was lost early in the evolution of Fornicata and the majority of Fornicata are missing this protein. Another alternative is that the pore complex was lost after divergence of diplomonads or even after divergence of just *Giardia*. Our data are most consistent with early loss events, as homologs of TIM23 were only found in *Carpediemonas*., although it is also possible that these proteins are present in other CLOs and/or *Chilomastix*, but were simply missed in our EST samples. The lack of any transcripts encoding the TOM40 complex in our ESTs probably just reflects incomplete data as well, rather than a true absence of the TOM40 complex in CLOs and *Chilomastix*.

In *Trichomonas vaginalis*, ATP/ADP translocation is putatively processed by Hmp31 (hydrogenosomal membrane protein 31) (Dyall *et al.*, 2000; Tjaden *et al.*, 2004). In *Giardia* no ATP/ADP transporter is known (Jedelský *et al.*, 2011), although one must exist because of the ATP-dependent proteins within its mitosomes. We have identified homologues of Hmp31 in all organisms except *Kipferlia*, suggesting the apparent loss of an Hmp31-like protein occurred after the divergence of *Chilomastix*.

There were only few targeting peptides identified in *G. intestinalis* so far and they tend to be relatively short, with only one positively charged amino acid relatively close to the cleavage site (Tachezy and Doležal, 2011). In eukaryotes with classical mitochondria the positively charged residues located further from the cleavage site may increase affinity of the peptide to the membrane for the organelle (Jedelský *et al.*, 2011). It is speculated that the lack of these residues in *Giardia* may be caused by the low membrane potential of the mitosomes (Jedelský *et al.*, 2011;

Tachezy and Doležal, 2011). Interestingly, some of the targeting peptides identified in our ESTs databases possess positively charged amino acids further from the putative cleavage site, suggesting the possibility of existence of some proton gradient in the organelle of CLOs and *Chilomastix*.

5.3.4 Glycine Cleavage System

The identification of all four components of glycine cleavage system (GCS) and serine hydroxymethyl transferase (SHMT) in *Ergobibamus* strongly suggests the presence of this pathway in the last common ancestor of Parabasalids and Fornicata. *Giardia* does not possess any of the GCS enzymes (Jedelský *et al.*, 2011) and we were unable to find them in our *Chilomastix* data. *Trichomonas* possesses only SHMT and three proteins of GCS complex – GCS-L and two copies of GCS-H (Mukherjee *et al.*, 2006a; Mukherjee *et al.*, 2006b). Interestingly the GCS-L protein of *Trichomonas* was apparently acquired via LGT from a firmicute bacterium (Mukherjee *et al.*, 2006a), consistent with loss of the original mitochondrial version and secondary gain of this foreign gene along the lineage leading to *Trichomonas*. The evolution of the GCS-L protein seems also to be rather complicated, as both *Carpediemonas* and *Ergobibamus* possess two copies of GCS-L protein, one possibly acquired trough LGT from a spirochete bacterium (Table 5.1, Appendix F.1). As GCS-L also functions within three other complexes – pyruvate dehydrogenase, oxoglutarate dehydrogenase and branched-chain-alpha-keto acid dehydrogenase – it is possible that the two different copies act in two different complexes. However, none of the other subunits of these three complexes were found in any of the four EST projects.

It is possible the one of these other complexes is present either in cytosol or in the organelle, and just was not recovered in our ESTs, and that it is using the spirochaete version of the L-protein. It is also possible that the spirochaete version of the protein acts in a yet unknown function of this protein, or that it acts in the glycine cleavage system along with the eukaryotic protein (in other words, that they are interchangeable).

5.3.5 Oxygen Scavenging System

An oxygen scavenging system (superoxide dismutase, thioredoxin, thioredoxin reductase, thioredoxin oxidase, hydroperoxide reductase ruberythrin, Table 5.1) is putatively present in *Trichomonas* hydrogenosome and acts in oxygen defense inside hydrogensomes, which contain oxygen-sensitive enzymes such as [FeFe]-hydrogenase (Lindmark and Müller, 1974; Ellis *et al.*, 1994; Pütz *et al.*, 2005). In *Giardia* Parts of this oxygen defense system were found in the mitosomal fraction using mass spectrometry analysis, however localization experiments did not clearly confirm localization of these into the mitosome (Jedelský *et al.*, 2011). Regardless of their localization, our data suggest the presence of these enzymes in the last common ancestor of parabasalids and Fornicata, as we found components in all four organisms.

5.3.6 Synthesis

Based on comparisons discussed above it is possible to propose a tentative partial history of organellar reduction and modification within the Parabasala-Fornicata

grouping (Figure 5.3). It is probable that the last common ancestor of parabasalids and Fornicata, and the common ancestor of all Fornicata, possessed a rather hydrogenosome-like organelle. This organelle was probably involved in energy generation through a similar pathway to that found in the hydrogenosomes of *Trichomonas*. It possessed protein import and Fe-S cluster synthesis machineries and the associated proteins. Unlike the *Trichomonas* hydrogenosome, it also possessed a complete glycine cleavage system, and possibly other proteins involved in amino acid metabolism, which is further corroborated by identification of these enzymes in *Trimastix* and putative targeting to its MROs (Hampl *et al.*, 2008). *Trimastix* belongs to the taxon Preaxostyla, which is the next closest known relative of the Parabasala-Fornicata clade (Simpson et al., 2006; Chapter 4) I also speculate that this organelle would also have an oxygen scavenging system and possibly other mitochondrial biochemical pathways.

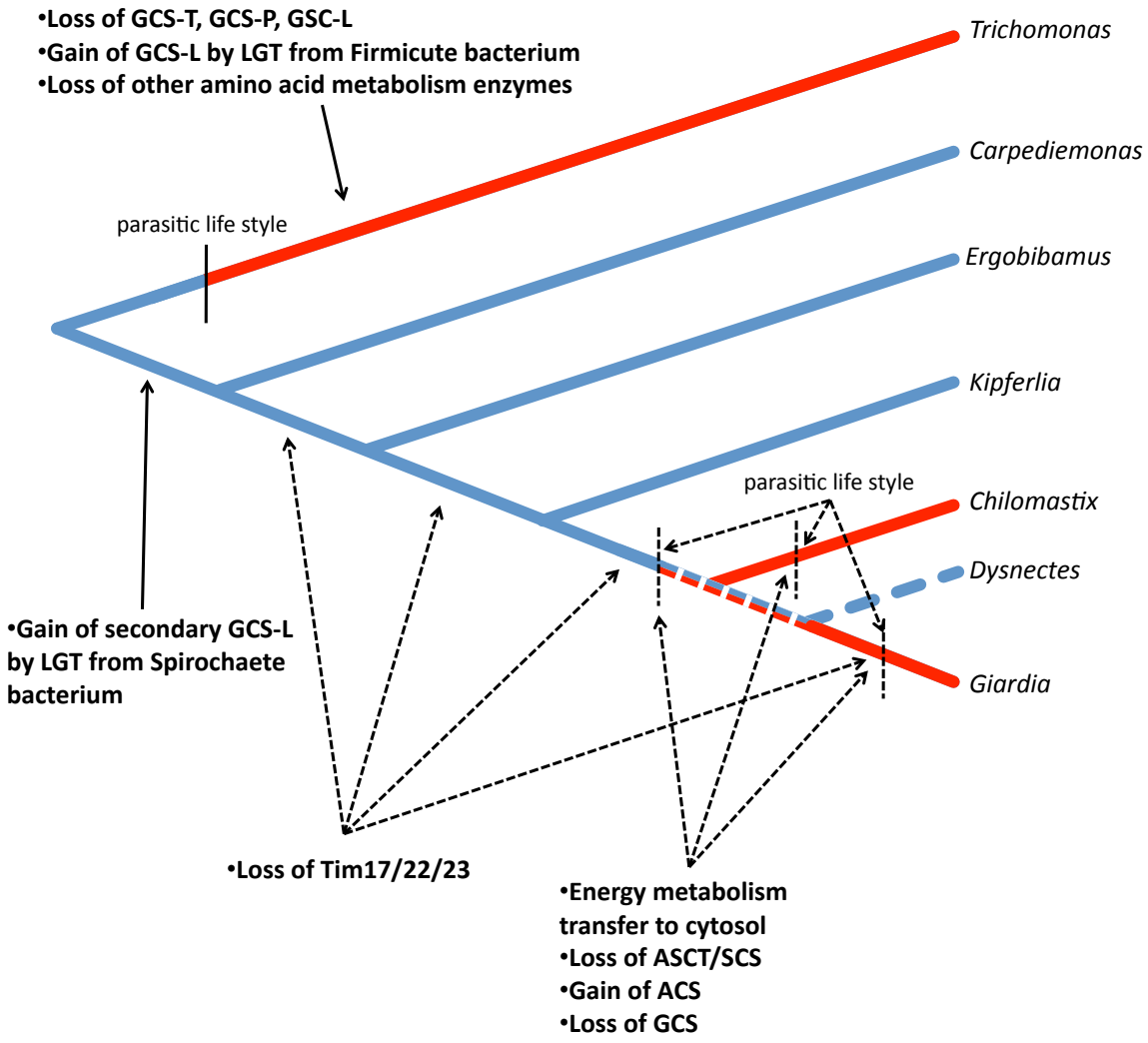


Figure 5.3 Proposed evolutionary changes in the Fornicata/Parabasala organellar proteome. Dotted branches on the tree mark uncertain phylogenetic positions. Red coloured branches highlight parasitic and commensalic species, while blue ones highlight free-living species. Dotted arrows mark alternative scenarios.

After the divergence of parabasalids, their hydrogenosomes lost a significant number of proteins involved in amino acid metabolism, including some parts of glycine cleavage system. The GCS-L protein was replaced by bacterial copy via LGT from a firmicute bacterium. Independently, a deep-branching fornicate ancestor also obtained a novel version of GCS-L via LGT from a spirochaete bacterium, but retained the ancestral form as well (or at least some extant representatives did). Both *Chilomastix* and *Giardia* completely lost GCS, suggesting that the reduction/loss of glycine cleavage may be connected to a change in life style from free living to parasitic or commensalic (Figure 5.3). This is a reasonable speculation as *Giardia* and *Chilomastix* are intestinal endoparasites, and may be scavenging their amino acids from the host. As diplomonads (*Giardia*) and retortamonads (*Chilomastix*) diverged, the energy metabolism was transferred from organelle to cytosol as suggested by the loss of the ASCT/SCS system and gain of ACS instead.

Among the organisms I have focused on, *Chilomastix* is the closest relative of *Giardia,* It is possible that *Chilomastix* and *Giardia* arose from a common parasitic ancestor, in which case it can be reasonably suggested that the loss of GCS and transfer of energy metabolism to the cytosol happened during their common ancestry. However, our previous work (Appendix I) has suggested that another free-living CLO, *Dysnectes brevis*, is the closest relative of diplomonads and that *Chilomastix* diverges prior to the clade of *Dysnectes*-diplomonads. This would suggest the possibility of independent reductive evolution in *Chilomastix* and *Giardia* or that the organelle of the free-living *Dysnectes* is also very reduced, biochemically speaking. It is also important to note the problems associated with incomplete

coverage of the transcriptome based on the ESTs we have obtained; a more complex organellar proteome may exist in *Chilomastix* than our initial survey revealed.

The above-mentioned comparisons and proposed evolutionary transitions are based solely on the results of *in-silico* analyses of EST data and should be considered as a working hypothesis to show fruitful directions for future research. The majority of the proposed transitions in Figure 5.3 are relatively easily testable through localization experiments of key enzymes in the metabolic pathways I have discussed. Also obtaining more transcriptomic (and possibly genomic) data would help to deepen the sequence coverage, to detect more of the transciptome, and obtain more full-length sequences. Given the similarities between *Chilomastix* and *Giardia* it will be very important to obtain data for the organelles of *Dysnectes*, the possible closest relative of diplomonads. Gaining better insight into the functions of MROs in other diplomonads will be very important as well. *Giardia* is an extremely specialized parasite and conclusions drawn from *Giardia* may not hold for other diplomonads, especially the (possibly secondarily) free-living species of diplomonads such as *Hexamita* and *Trepomonas* (Kolisko *et al.*, 2008; Chapter 2). Obtaining more detailed genetic and cell-biological data for these species will also shed more light on the evolution of mitochondria-related organelles in Fornicata.

5.4. Materials and Methods

5.4.1 Data Assembly

We examined EST data from isolates of four species, *Carpediemonas membranifera*, *Ergobibamus cyprinoides*, *Kipferlia bialata* and *Chilomastix caulleryi*. The data generation is described in the previous chapter (see Chapter 4, Materials and methods). Raw EST reads were newly assembled using a recent version of the program Mira3 (Chevreux *et al.*, 2004). In the case of *Ergobibamus* data, a true hybrid assembly of Sanger and 454 reads was produced (i.e., both Sanger reads and 454 reads were fully utilized in one assembly).

5.4.2 Selection of Proteins of Putative Mitochondrial Origin

The primary search for proteins of possible mitochondrial origin was done using the program CBOrg (Gaston *et al.*, 2009), which uses a BLAST search against the mitochondrial and non-mitochondrial proteomes of man, *Tetrahymena*, yeast and *Trichomonas* to predict possible organellar proteins. The CBOrg analysis flagged 860 clusters from *Carpediemonas*, 541 clusters from *Ergobibamus*, 369 clusters from *Kipferlia* and 574 clusters from *Chilomastix*. These were then BLASTed against the genbank 'nr' database and all hits with expectation values lower than 10^{-5} were discarded. All sequences with only bacteria amongst the top 20 blast hits were also discarded as potential contamination from prokaryotic prey. The rest of the BLAST records were manually searched and annotated (again using nr BLAST), leaving 29 proteins from *Carpediemonas*, 35 proteins from *Ergobibamus*, 23 proteins from

Kipferlia and 21 proteins from *Chilomastix* as proteins of possible mitochondrial origin.

5.4.3 Targeting and Localization Predictions

For each sequence we attempted to determine the probability of its targeting to mitochondria. We used four algorithms for localization and targeting predictions: Mitoprot (Claros and Vincens, 1996), TargetP (Nielsen *et al.*, 1997; Emanuelsson *et al.*, 2000), Mitopred (Guda *et al.*, 2004a; Guda *et al.*, 2004b) and Predotar ((Small *et al.*, 2004). Results of these four methods are summarized for each organism in Appendix F.1. Predicted putative targeting peptides were confirmed to be extensions by BLASTing against bacterial sequences, which would lack targeting peptides. In the case of IscU, to serve as an example, the putative targeting peptides detected in our study organisms were also compared to those in *Giardia* and *Trichomonas* (see Figure 5.2). The other targeting peptides can be seen in Figure 5.1.

Chapter 6

Conclusions

In this thesis, I explored several aspects of the evolution of anaerobic excavates, especially the group Fornicata, namely diversity, internal relationships, and position on the eukaryotic tree. The work combined two fundamental aspects that improve phylogenetic inference: enriching taxon sampling and collecting more sequence data. The main aims were to resolve some of the hardest questions of eukaryotic phylogeny: the phylogenetic position of diplomonads and subsequently the phylogenetic status of Excavata.

I analyzed the phylogenetic position of enteromonads, which were previously thought to be a sister group to diplomonads. However, the results showed that enteromonads branch within diplomonads, which raised new questions regarding the evolution of the doubled karyomastigont morphology of diplomonads, especially whether it evolved several times or if enteromonads are secondarily reduced. Providing answers to these questions will require a much greater understanding of the cell biology of these organisms; nevertheless, it is an exciting research area to explore.

I isolated several new lineages of deep branching relatives of diplomonads that prove to be very important for resolving the phylogeny of eukaryotes. These organisms - CLOs – in most cases had not been detected by previous environmental PCR studies, which has two important consequences. Firstly, it suggests that our understanding of eukaryotic diversity is still far from being complete and

discoveries of new important lineages are to be expected (Yabuki *et al.*, 2010; Kim *et al.*, 2011; Yabuki *et al.*, 2011). Secondly, it suggests that classical culturing approaches remain a valuable tool for uncovering novel eukaryote diversity and should be employed along with culture-independent methods. An important approach in this search is the recent modification of environmental PCR techniques such as using group-specific primers and next generation sequencing (Lara *et al.*, 2009; Stoeck *et al.*, 2009). Another future alternative is provided by the current development of single cell genomics, which promises large amounts of data from microorganisms that cannot be cultured (Ishoey *et al.*, 2008).

Phylogenomics has improved the resolution of deep eukaryotic relationships because the large datasets are robust to stochastic errors; however, they are not robust to systematic errors. Such large datasets are expected to be very sensitive to systematic errors. For robust results, it is crucial to include data that are reasonably fit in their properties to the rest of the dataset. For example, the presented work shows improvement of the results when shorter-branching CLOs substituted the long-branching parabasalids and diplomonads.

The future development of phylogenomics includes modification and optimization of some programs for parallel computing, which will allow usage of more complex models of evolution robust to systematic errors (Quang *et al.*, 2008; Wang *et al.*, 2008; Wang *et al.*, 2009). Currently, the computational demands of these models and programs are too high to be of any practical use for phylogenomic analyses. Another problem is the potential impact of missing data on the multi-gene analyses. I believe this is currently given less attention than it deserves. Simulation

studies conclude, sometimes vaguely, that missing data is not a problem as long as there is enough data to position any particular taxon on the tree (Wiens, 2006). The impact of missing data needs to be investigated more using real-life phylogenomic datasets with long branches and suboptimal models to better understand the impact of missing data in practice.

Another "phylogenomic" approach for researching deep level relationships between eukaryotes is genome-wide analyses of rare evolutionary events like lateral gene transfer or deletions/insertions within sequences (Archibald *et al.*, 2003; Rice and Palmer, 2006). For example, with the currently available data and computer power it may be relatively easy to perform a supergroup-wide search for lateral gene transfer candidates and look for shared patterns supporting the monophyly of particular groups. The major issue with such analyses would be correctly interpreting the results, as there is currently no statistical framework for assessing the significance of such an event as support for the monophyly of any particular group.

I also explored and suggested possible evolutionary steps in the functional reduction of mitochondria in the diplomonad lineage. However, the evolutionary importance of Fornicata does not end here. Fornicata also represent an excellent system for studying the evolution of parasitism. There are several basal free-living lineages represented by CLOs, followed by transition to parasitic lifestyle. Within diplomonads, there are highly specialized parasites (*Giardia*, *Spironucleus*), commensals (Enteromonads, some *Hexamita* species) and probably secondary free-living representatives (some *Hexamita* species and *Trepomonas*). Comparative

analyses of these organisms may reveal the genomic changes associated with a transition from free-living to parasitism and *vice versa*. Future plans includes sequencing the genome of *Carpediemonas membranifera* strain BICM (isolated during this project) and such analyses would be one of the main aims of the project.

The study of microbial eukaryotes has just entered a new era. We are now beginning to benefit from the power of large-scale phylogenomic and comparative analyses. Using the current, relatively cheap, sequencing technologies (e.g., 454, illumina, SOLiD) and bioinformatics tools, it has become relatively easy and economical to obtain large amounts of data from almost any cultured organisms. Combining these with culturing of 'new' lineages and the possible future usage of single cell genomics/transcriptomics promises steady progress towards a robust understanding of deep eukaryotic evolution.

References

Abascal, F., Zardoya, R. & Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics,* **21:** 2104-2105.

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., McCourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A. V., Spiegel, F. W. & Taylor, M. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology,* **52:** 399-451.

Aguilera, P., Barry, T. & Tovar, J. (2008) *Entamoeba histolytica* mitosomes: Organelles in search for a function. *Experimental Parasitology,* **118:** 10-16.

Akhmanova, A., Voncken, F., Van Alen, T., Van Hoek, A., Boxma, B., Vogels, G., Veenhuis, M. & Hackstein, H. P. (1998) A hydrogenosome with a genome. *Nature,* **396:** 527-528.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology,* **215:** 403-410.

Archibald, J. M. & Keeling, P. J. (2004) Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. *Journal of Eukaryotic Microbiology,* **51:** 113-118.

Archibald, J. M., Longet, D., Pawlowski, J. & Keeling, P. J. (2003) A novel polyubiquitin structure in Cercozoa and Foraminifera: Evidence for a new eukaryotic supergroup. *Molecular Biology and Evolution,* **20:** 62-66.

Arisue, N., Hashimoto, T., Yoshikawa, H., Nakamura, Y., Nakamura, G., Nakamura, F., Yano, T. & Hasegawa, M. (2002) Phylogenetic position of *Blastocystis hominis* and of stramenopiles inferred from multiple molecular sequence data. *Journal of Eukaryotic Microbiology,* **49:** 42-53.

Arisue, N., Hasegawa, M. & Hashimoto, T. (2005) Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Molecular Biology and Evolution,* **22:** 1157-1157.

Bairoch, A. & Apweiler, R. (1997) The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Research,* **25:** 31-36.

Baldauf, S. L. & Palmer, J. D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proceedings of the National Academy of Sciences of the United States of America,* **90:** 11558-11562.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science,* **290:** 972-977.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Durufle, L., Gaasterland, T., López, P., Müller, M. & Philippe, H. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proceedings of the National Academy of Sciences of the United States of America,* **99:** 1414-1419.

Barbera, M. J., Ruiz-Trillo, I., Leigh, J., Hug, L. & Roger, A. J. (2007) The Diversity of mitochondrion-related organelles amongst eukaryotic microbes. IN Martin, W. F. & Müller, M. (Eds.) *Origin of Mitochondria and Hydrogenosomes.* Berlin, Springer-Verlag.

Barbera, M. J., Ruiz-Trillo, I., Tufts, J. Y. A., Amandine, B., Silberman, J. D. & Roger, A. J. (2010) *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties. *Eukaryotic Cell,* **9:** 1913-1924.

Bass, D. & Cavalier-Smith, T. (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *International Journal of Systematic and Evolutionary Microbiology,* **54:** 2393-2404.

Behnke, A., Bunge, J., Barger, K., Breiner, H. W., Alla, V. & Stoeck, T. (2006) Microeukaryote community patterns along an O_2/H_2S gradient in a supersulfidic anoxic Fjord (Framvaren, Norway). *Applied and Environmental Microbiology,* **72:** 3626-3636.

Bernander, R., Palm, J. E. & Svärd, S. G. (2001) Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cellular Microbiology,* **3:** 55-62.

Bernard, C., Simpson, A. G. B. & Patterson, D. J. (1997) An ultrastructural study of a free-living retortamonad, *Chilomastix cuspidata* (Larsen & Patterson, 1990) n. comb. (Retortamonadida, Protista). *European Journal of Protistology,* **33:** 254-265.

Berney, C., Fahrni, J. & Pawlowski, J. (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biology,* **2:** 13.

Best, A. A., Morrison, H. G., McArthur, A. G., Sogin, M. L. & Olsen, G. J. (2004) Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. *Genome research,* **14:** 1537-1547.

Bozner, P. (1997) Immunological detection and subcellular localization of HSP70 and HSP60 homologs in *Trichomonas vaginalis*. *Journal of Parasitology,* **83:** 224-229.

Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I. & Pachter, L. (2009) Fast statistical alignment. *PLoS Computational Biology,* **5:** 15.

Brinkmann, H. & Philippe, H. (1999) Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution,* **16:** 817-825.

Brinkmann, H., Van der Giezen, M., Zhou, Y., De Raucourt, G. P. & Philippe, H. (2005) An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Systematic Biology,* **54:** 743-757.

Brown, M. W., Spiegel, F. W. & Silberman, J. D. (2009) Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution,* **26:** 2699-2709.

Brugerolle, G. (1975) Contribution a l'etude cytologique et phyletique des diplozoaires (Zoomastigophorea, Diplozoa, Dangeard 1910). VI. Caracteres generaux des diplozoaires. *Protistologica,* **11:** 111 - 118.

Brugerolle, G. (1986) Separation des genres *Trimitus* (Diplomonadida) and *Tricercomitus* (Trichomonadida) d'apres leur ultrastructure. *Protistologica,* **22:** 31-37.

Brugerolle, G. & Taylor, F. J. R. (1977) Taxonomy, cytology and evolution of the Mastigophora. *Proceedings of the Fifth International Congress of Protozoology New York***:** 14 - 28.

Brugerolle, G., Bricheux, G. & Coffe, G. (2000) Immunolocalization of two hydrogenosomal enzymes of *Trichomonas vaginalis*. *Parasitology Research,* **86:** 30-35.

Bui, E. T. & Johnson, P. J. (1996) Identification and characterization of [Fe]-hydrogenase in the hydrogensome of *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology,* **76:** 305-310.

Burger, G., Saint-Louis, D., Gray, M. W. & Lang, B. F. (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. *The Plant Cell,* **11:** 1675-1694.

Burki, F., Shalchian-Tabrizi, K., Minge, M. A., Skjaevland, A., Nikolaev, S. I., Jakobsen, K. S. & Pawlowski, J. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE,* **2:** e790.

Burki, F., Shalchian-Tabrizi, K. & Pawlowski, J. (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biology Letters,* **4:** 366-369.

Burki, F., Inagaki, Y., Bråte, J., Archibald, J. M., Keeling, P. J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., Klaveness, D., Jakobsen, K. S., Pawlowski, J. & Shalchian-Tabrizi, K. (2009) Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biology and Evolution,* **1:** 231-238.

Burki, F., Kudryavtsev, A., Matz, M. V., Aglyamova, G. V., Bulman, S., Fiers, M., Keeling, P. J. & Pawlowski, J. (2010) Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evolutionary Biology,* **10:** 377.

Cavalier-Smith, T. (1983) A 6-kingdom classification and a unified phylogeny. IN Schwemmler, W. & Schenk, H. E. A. (Eds.) *Endocytobiology II.* Berlin, De Gruyter.

Cavalier-Smith, T. (1995) Cell cycles, diplokaryosis and the archezoan origin of sex. *Archiv Fur Protistenkunde,* **145:** 189-207.

Cavalier-Smith, T. (1998) A revised six-kingdom system of life. *Biological Reviews of the Cambridge Philosophical Society,* **73:** 203-266.

Cavalier-Smith, T. (1999) Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *Journal of Eukaryotic Microbiology,* **46:** 347-366.

Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology,* **52:** 297-354.

Cavalier-Smith, T. (2003) The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher

taxa. *International Journal of Systematic and Evolutionary Microbiology,* **53:** 1741-1758.

Cavalier-Smith, T. (2004) Only six kingdoms of life. *Proceedings of the National Academy of Sciences of the United States of America,* **40:** 21-48.

Cavalier-Smith, T. & Chao, E. E. (1996) Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. *Journal of Molecular Evolution,* **43:** 551-562.

Cavalier-Smith, T. & Chao, E. E. (2003a) Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist,* **154:** 341-358.

Cavalier-Smith, T. & Chao, E. E. (2003b) Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution,* **56:** 540-563.

Cavalier-Smith, T. & Chao, E. E. (2006) Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *Journal of Molecular Evolution,* **62:** 388-420.

Cepicka, I., Kostka, M., Uzlikova, M., Kulda, J. & Flegr, J. (2008) Non-monophyly of Retortamonadida and high genetic diversity of the genus *Chilomastix* suggested by analysis of SSU rDNA. *Molecular Phylogenetics and Evolution,* **48:** 770-775.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W., E., Wetter, T. & Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome research,* **14:** 1147-1159.

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science,* **311:** 1283-1287.

Clark, C. G. (1992) DNA purification from polysaccharide-rich cells. *Protocols in protozoology,* **1:** D-3.1-D-3.2.

Clark, C. G. & Diamond, L. S. (2002) Methods for cultivation of luminal parasitic protists of clinical importance. *Clinical Microbiology Reviews,* **15:** 329-341.

Claros, M. G. & Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry,* **241:** 779-786.

Cuvelier, M. L., Ortiz, A., Kim, E., Moehlig, H., Richardson, D. E., Heidelberg, J. F., Archibald, J. M. & Worden, A. Z. (2008) Widespread distribution of a unique marine protistan lineage. *Environmental Microbiology,* **10:** 1621-1634.

Dacks, J. B., Silberman, J. D., Simpson, A. G. B., Moriya, S., Kudo, T., Ohkuma, M. & Redfield, R. J. (2001) Oxymonads are closely related to the excavate taxon *Trimastix*. *Molecular Biology and Evolution,* **18:** 1034-1044.

Dawson, S. C. & Pace, N. R. (2002) Novel kingdom-level eukaryotic diveristy in anoxic environments. *Proceedings of the National Academy of Sciences of the United States of America,* **99:** 8324-8329.

Denoeud, F., Roussel, M., Noel, B., Wawrzyniak, I., Da Silva, C., Diogon, M., Viscogliosi, E., Brochier-Armanet, C., Couloux, A., Poulain, J., Segurens, B., Anthouard, V., Texier, C., Blot, N., Poirer, P., Ng, G. C., Tan, K. S. W., Artiguenave, O. J., Aury, J. M., Delbac, F., Wincker, P., Vivares, C. P. & Alaoui, H. E. (2011) Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parsite. *Genome Biology,* **12:** R29.

Dobell, C. & Leidlaw, P. P. (1926) On the cultivation of *Entamoeba histolytica* and some other parasitic amoebae. *Parasitology,* **18:** 283-318.

Doležal, P., Vanácová, S., Tachezy, J. & Hrdý, I. (2004) Malic enzymes of *Trichomonas vaginalis*: two enzyme families, two distinct origins. *Gene,* **329:** 81-92.

Doležal, P., Šmíd, O., Rada, P., Zubacova, Z., Bursac, D., Sutak, R., Nebesarova, J., Lithgow, T. & Tachezy, J. (2005) *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. *Proceedings of the National Academy of Sciences of the United States of America,* **102:** 10924-10929.

Doležal, P., Likic, V., Tachezy, J. & Lithgow, T. (2006) Evolution of the molecular machines for protein import into mitochondria. *Science,* **313:** 314-318.

Dyall, S. D., Koehler, C. M., Delgadillo-Correa, M. G., Bradley, P. J., Plumper, E., Leuenberger, D., Turck, C. W. & Johnson, P. J. (2000) Presence of a member of the mitochondrial carrier family in hydrogenosomes: Conservation of membrane-targeting pathways between hydrogenosomes and mitochondria *Molecular and Cellular Biology,* **20:** 2488-2497.

Dyall, S. D., Lester, D. C., Schneider, R. E., Delgadillo-Correa, M., Plumper, E., Martinez, A., Keohler, C. M. & Johnson, P. J. (2003) *Trichomonas vaginalis* Hmp35, a putative pore-forming hydrogenosomal membrane protein, can form a complex in yeast mitochondria. *The Journal of Biological Chemistry,* **278:** 30548-30561.

Edgcomb, V. P., Roger, A. J., Simpson, A. G. B., Kysela, D. T. & Sogin, M. L. (2001) Evolutionary relationships among 'jakobid' flagellates as indicated by alpha- and beta-tubulin phylogenies. *Molecular Biology and Evolution,* **18:** 514-522.

Edgcomb, V. P., Kysela, D. T., Teske, A., Gomez, A. D. & Sogin, M. L. (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences of the United States of America,* **99:** 7658-7662.

Edlind, T. D., Li, J. & Visvesvara, M. H. (1996) Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Molecular Phylogenetics and Evolution,* **5** 359-367.

Ekebom, J., Patterson, D. J. & Vørs, N. (1996) Heterotrophic flagellates from coral reef sediments (Great Barrier Reef Australia). *Archiv Für Protistenkunde,* **146:** 251-272.

Ellis, J. E., Yarlett, N., Cole, D., Humpreys, M. J. & Lloyd, D. (1994) Antioxidant defenses in the microaerophilic protozoan *Trichomonas vaginalis*: comparison of metronidazol-resistant and sensitive strains *Microbiology,* **140:** 2489-2494.

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology,* **300**.

Epstein, S. S. & López-García, P. (2008) 'Missing' protists: a molecular perspective. *Biodiversity and Conservation,* **17:** 261-276.

Fahrni, J. F., Bolivar, I., Berney, C., Nassonova, E., Smirnov, A. & Pawlowski, J. (2003) Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. *Molecular Biology and Evolution,* **20:** 1881-1886.

Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Molecular Biology and Evolution,* **18:** 418-426.

Felsenstein, J. (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology,* **27:** 401-410.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution,* **17:** 368-376.

Felsenstein, J. (2004) *Inferring phylogenies,* Sunderland, Massachusetts, Sinauer Associates, Inc.

Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Garriga, G. H., Bertrandt, H. & Lambowitz, A. (1984) RNA splicing in *Neurospora* mitochondria: nuclear mutants defective in both splicing and 3' end synthesis of large rRNA. *Cell,* **36:** 623-634.

Gaston, D., Tsaousis, A. D. & Roger, A. J. (2009) Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Methods in Enzymology,* **457:** 21-47.

Germot, A., Philippe, H. & Le Guyader, H. (1997) Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Molecular and Biochemical Parasitology,* **87:** 159-168.

Geuten, K., Massingham, T., Darius, P., Smets, E. & Goldman, N. (2007) Experimental design criteria in phylogenetics: Where to add taxa. *Systematic Biology,* **56:** 609-622.

Gill, E. E., Diaz-Trivino, S., Barbera, M. J., Silberman, J. D., Stechmann, A., Gaston, D., Tamas, I. & Roger, A. J. (2007) Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. *Molecular Microbiology,* **66:** 1306-1320.

Gribaldo, S. & Philippe, H. (2000) Pitfalls in tree reconstruction and the phylogeny of eukaryotes. IN Hirt, R. P. & Horner, D. S. (Eds.) *Organelles, genomes, and eukaryote phylogeny: an evolution synthesis in the age of genomics.* New York, NY, CRC press.

Groisillier, A., Massana, R., Valentin, K., Vaulot, D. & Guillou, L. (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquatic Microbial Ecology,* **42:** 277-291.

Guda, C., Fahy, E. & Subramaniam, S. (2004a) MITOPRED: A genome-scale method for prediction of nuclear-encoded mitochondrial proteins. *Bioinformatics,* **20**.

Guda, C., Guda, P., Fahy, E. & Subramaniam, S. (2004b) MITOPRED: A web server for the prediction of mitochondrial proteins. *Nucleic Acids Research,* **32:** W372-W374.

Hackett, J. D., Yoon, H. S., Li, S., Reyes-Prieto, A., Rummele, S. E. & Bhattacharya, D. (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Molecular biology and evolution,* **24:** 1702-13.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* **41:** 95-98.

Hampl, V. & Simpson, A. G. B. (2007) Possible mitochondria-related organelles in poorly-studied "Amitochondriate" eukaryotes. IN Tachezy, J. (Ed.) *Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes.* Berlin, Springer-Verlag.

Hampl, V., Horner, D. S., Dyal, P., Kulda, J., Flegr, J., Foster, P. G. & Embley, T. M. (2005) Inference of the phylogenetic position of oxymonads based on nine genes: Support for Metamonada and Excavata. *Molecular Biology and Evolution,* **22:** 2508-2518.

Hampl, V., Silberman, J. D., Stechmann, A., Diaz-Trivino, S., Johnson, P. J. & Roger, A. J. (2008) Genetic evidence for a mitochondriate ancestry in the 'amitochondriate' flagellate *Trimastix pyriformis*. *PLoS ONE,* **3:** e1383.

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G. B. & Roger, A. J. (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic 'supergroups'. *Proceedings of the National Academy of Sciences of the United States of America,* **106:** 3859-3864.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K. & Hasegawa, M. (1995) Phylogenetic place of mitochondrion-lacking protozoan *Giardia lamblia* inferred from amino acid sequences of elongation factor 2. *Molecular Biology and Evolution,* **15:** 782-793.

Hendy, M. D. & Penny, D. (1989) A framework for the quantitative study of evolutionary trees. *Systematic Zoology,* **38:** 297-309.

Hertel, L. A., Bayne, C. J. & Loker, E. S. (2002) The symbiont *Capsaspora owczarzaki*, nov gen. nov sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoea. *International Journal for Parasitology,* **32:** 1183-1191.

Hinkle, G., Leipe, D. D., Nerad, T. A. & Sogin, M. L. (1994) The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. *Nucleic Acids Research,* **22:** 465-469.

Hirt, R. P., Healy, B., Vossbrinck, C. R., Canning, E. U. & Embley, T. M. (1997) A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Current Biology,* **7:** 995-998.

Hirt, R. P., Logsdon, J. M., Healy, B., W., D. M., Doolittle, W. F. & Embley, T. M. (1999) Microsporidia are related to Fungi: evidence from the largest subunit of RNA

polymerase II and other proteins. *Proceedings of the National Academy of Sciences of the United States of America,* **96:** 580-585.

Hjort, K., Goldberg, A. V., Tsaousis, A. D., Hirt, R. P. & Embley, T. M. (2010) Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philosophical Transactions of the Royal Society Biological Sciences,* **365:** 713-727.

Hrdý, I., Hirt, R. P., Doležal, P., Bardonová, L., Foster, P. G., Tachezy, J. & Embley, T. M. (2004) *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature,* **432:** 618-622.

Huelsenbeck, J. P. (2000) MrBayes: Bayesian inference of phylogeny. Distributed by the author.

ICZN (1999) International code of zoological nomenclature.

Inagaki, Y., Nakajima, Y., Sato, M., Sakaguchi, M. & Hashimoto, T. (2009) Gene sampling can bias multi-gene phylogenetic inferences: The relationship between red algae and green plants as a case study. *Molecular Biology and Evolution,* **26:** 1171-1178.

Ishida, K., Green, B. R. & Cavalier-Smith, T. (1999) Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nulcear and nucleomorph small-subunit rRNA genes. *Molecular Biology and Evolution,* **16:** 321-331.

Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. & Lasken, R. S. (2008) Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology,* **11:** 198-204.

Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Šmíd, O., Hrdý, I., Sedinová, M., Marcinciková, M., Voleman, L., Perry, A. J., Beltran, N. C., Lithgow, T. & Tachezy, J. (2011) The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS ONE,* **6:** e17285.

Jobb, G., von Haeseler, A. & Strimmer, K. (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology,* **4:** 18.

Jørgensen, A. & Sterud, E. (2007) Phylogeny of *Spironucleus* [Eopharyngia: Diplomonadida: Hexamitinae]. *Protist,* **158:** 247-254.

Kamaishi, T., Hashimoto, T., Nakamura, Y., Nakamura, F., Murata, S., Okada, N., Okamoto, K., Shimizu, M. & Hasegawa, M. (1996) Protein phylogeny of translation elongation factor EF-1α suggests microsporidians are extremely ancient eukaryotes. *Journal of Molecular Evolution,* **42:** 257-263.

Katoh, K., Kei-ichi, K., Hiroyuki, T. & Takashi, M. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research,* **33:** 511-518.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research,* **30:** 3059-3066.

Keeling, P. J. (2001) Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Molecular Biology and Evolution,* **18:** 1551-1557.

Keeling, P. J. & Doolittle, W. F. (1996a) A non-canonical genetic code in an early diverging eukaryotic lineage. *The EMBO journal,* **15:** 2285-2290.

Keeling, P. J. & Doolittle, W. F. (1996b) Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Molecular Biology and Evolution,* **13:** 1297-1305.

Keeling, P. J. & Brugerolle, G. (2006) Evidence from SSU rRNA phylogeny that *Octomitus* is a sister lineage of *Giardia*. *Protist,* **157:** 205-212.

Keister, D. B. (1983) Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **77:** 487-488.

Kim, E., Harrison, J. W., Sudek, S., Jones, M. D. M., Wilcox, H. M., Richards, T. A., Worden, A. Z. & Archibald, J. M. (2011) Newly identified and diverse plastid-bearing branch on th eukaryotic tree of life. *Proceedings of the National Academy of Sciences of the United States of America,* **108:** 1496-1500.

Kolisko, M., Cepicka, I., Hampl, V., Kulda, J. & Flegr, J. (2005) The phylogenetic position of enteromonads: a challenge for the present models of diplomonad evolution. *International Journal of Systematic and Evolutionary Microbiology,* **55:** 1729-1733.

Kolisko, M., Cepicka, I., Hampl, V., Leigh, J., Roger, A. J., Kulda, J., Simpson, A. G. B. & Flegr, J. (2008) Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: Implications for the evolutionary history of the double karyomastigont of diplomonads. *BMC Evolutionary Biology,* **8:**205

Kolisko, M., Silberman, J. D., Cepicka, I., Yubuki, N., Takishita, K., Yabuki, A., Leander, B. S., Inouye, I., Inagaki, Y., Roger, A. J. & Simpson, A. G. B. (2010) A wide diversity of previously undetected free-living relatives of diplomonads

isolated from marine/saline habitats. *Environmental Microbiology,* **12:** 2700-2710.

Kulda, J. & Nohynkova, E. (1978) Flagellates of the human intestine and of intestines of other species. IN Kreier, J. P. (Ed.) *Parasitic protozoa: Intestinal flagellates, Histomonads, Trichomonads, Amoeba, Opalinids, and Ciliates.* New York, NY, Academic Press inc.

Kumazaki, T., Hori, H. & Osawa, S. (1983) Phylogeny of protozoa deduced from 5S rRNA sequences. *Journal of Molecular Evolution,* **19:** 411-419.

Lahti, C. J., d'Oliviera, C. E. & Johnson, P. J. (1992) Beta-Succinyl-coenzyme A synthetase from *Trichomonas vaginalis* is a soluble hydrogenosomal protein with an amino-terminal sequence that resembles mitochondrial presequences. *Journal of Bacteriology,* **174:** 6822-6830.

Lahti, C. J., Bradley, P. J. & Johnson, P. J. (1994) Molecular characterization of the alpha-subunit of *Trichomonas vaginalis* hydrogenosomal succinyl-CoA synthetase. *Molecular and Biochemical Parasitology,* **66:** 309-318.

Lane, C. E. & Archibald, J. M. (2008) The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends in Ecology & Evolution,* **23:** 268-275.

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M. & Grey, M. W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature,* **387:** 493-497.

Lara, E., Moreira, D., Vereshchaka, A. & López-García, P. (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. *Environmental Microbiology,* **11:** 47-55.

Lartillot, N. & Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution,* **21:** 1095-1109.

Lartillot, N., Lepage, T. & Blanquart, S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics,* **25:** 2286-2288.

Lee, W. J. & Patterson, D. J. (2000) Heterotrophic flagellates (Protista) from marine sediments of Botany Bay, Australia. *Journal of Natural History,* **34:** 483-562.

Li, L., Stoeckert, C. J. & Roos, D. S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research,* **13:** 2178-2189.

Lill, R. & Mühlenhoff, U. (2005) Iron-sulfur-protein biogenesis in eukaryotes. *Trends in Biochemical Sciences,* **30:** 133-141.

Lill, R., Diekert, K., Kaut, A., Lange, H., Pelzer, W., Prohl, C. & Kispal, G. (1999) The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. *Biological Chemistry,* **380:** 1157-1166.

Lindmark, D. G. & Müller, M. (1973) Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *The Journal of Biological Chemistry,* **248:** 7724-7728.

Lindmark, D. G. & Müller, M. (1974) Superoxide-dismutase in anaerobic flagellate *Tritrichomonas foetus,* and *Monocercomonas* sp. *Journal of Biological Chemistry,* **249:** 4634-4637.

Liu, Y., Steenkamp, E. T., Brinkmann, H., Forget, L., Philippe, H. & Lang, B. F. (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BMC Evolutionary Biology,* **9:** 272.

Longet, D., Archibald, J. M., Keeling, P. J. & Pawlowski, J. (2003) Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. *International Journal of Systematic and Evolutionary Microbiology,* **55:** 1735-1739.

Maralikova, B., Ali, V., Nakada-Tsukui, K., Nozaki, T., Van der Giezen, M., Henze, K. & Tovar, J. (2010) Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. *Cellular Microbiology,* **13:** 331-342.

Margulis, L. (1970) *Origin of eukaryotic cells; evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth,* New Haven, Yale University Press.

Martin, W. & Herrmann, R. G. (1998) Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiology,* **118:** 9-17.

Martin, W. & Müller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature,* **440:** 41-45.

Massana, R. & Pedrós-Alió, C. (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current Opinion in Microbiology,* **11:** 213-218.

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. & Pedrós-Alió, C. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and Environmental Microbiology,* **70:** 3528-3534.

Massana, R., Terrado, R., Forn, I., Lovejoy, C. & Pedrós-Alió, C. (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environmental Microbiology,* **8:** 1515-1522.

McCarroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R. & Sogin, M. L. (1983) Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. *Biochemistry,* **22:** 5858-5868.

Medlin, L., Elwood, H. J., Stickles, S. & Sogin, M. L. (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene,* **71:** 491-499.

Milyutina, I. A., Aleshin, V. V., Mikrjukov, K. A., Kedrova, O. S. & Petrov, N. B. (2001) The unusually long small subunit ribosomal RNA gene found in amitochondriate amoebo-flagellate *Pelomyxa palustris*: its rRNA predicted secondary structure and phylogenetic implication. *Gene,* **272:** 131-139.

Minge, M. A., Silberman, J. D., Orr, R. J. S., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., Skjaevland, A. & Jakobsen, K. S. (2009) Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proceedings of the Royal Society of London Series B-Biological Sciences,* **276:** 597-604.

Moore, R. B., Obornik, M., Janouskovec, J., Chrudimsky, T., Vancova, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M. & Carter, D. A. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature,* **451:** 959-963.

Moreira, D., Le Guyader, H. & Philippe, H. (2000) The origin of red algae and the evolution of chloroplasts. *Nature,* **405:** 69-72.

Morin, L. (2000) Long branch attraction effects and the status of 'basal eukaryotes': Phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad *Trepomonas agilis*. *Journal of Eukaryotic Microbiology,* **47:** 167-177.

Moriya, S., Tanaka, M., Ohkuma, M., Sugano, S. & Kudo, T. (2001) Diversification of the microtubule system in the early stage of eukaryotic evolution: elongation factor 1 alpha and α-tubulin protein phylogeny of termite symbiotic oxymonad and hypermastigote protists. *Journal of Molecular Evolution,* **52:** 6-16.

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C.,

Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E. J., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svärd, S. G. & Sogin, M. L. (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science,* **317:** 1921-1926.

Mukherjee, M., Brown, M. T., McArthur, A. G. & Johnson, P. J. (2006a) Proteins of the glycine decarboxylase complex in the hydrogenosome of *Trichomonas vaginalis*. *Eukaryotic Cell,* **5:** 2062-2071.

Mukherjee, M., Sievers, S. A., Brown, M. T. & Johnson, P. J. (2006b) Identification and biochemical characterization of serine hydroxymethyl transferase in the hydrogenosome of *Trichomonas vaginalis*. *Eukaryotic Cell,* **5:** 2072-2078.

Müller, M. (1993) The hydrogenosome. *Journal of General Microbiology,* **139:** 2879-2889.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering,* **10:**1-6

Nikolaev, S. I., Berney, C., Fahrni, J. F., Bolivar, I., Polet, S., Mylnikov, A. P., Aleshin, V. V., B., P. N. & Pawlowski, J. (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America,* **101:** 8066-8071.

Nikolaev, S. I., Berney, C., Petrov, N. B., Mylnikov, M. B., Fahrni, F. J. & Pawlowski, J. (2006) Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. *International Journal of Systematic and Evolutionary Microbiology*, **56:** 1449-1458

.

Nohynkova, E., Tumova, P. & Kulda, J. (2006) Cell division of *Giardia intestinalis*: Flagellar developmental cycle involves transformation and exchange of flagella between mastigonts of a diplomonad cell. *Eukaryotic Cell,* **5:** 753-761.

Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Töbe, K., Vaulot, D. & Medlin, L. (2007) Picobiliphytes: A marine picoplanktonic algal group with unknown affinities to other Eukaryotes. *Science,* **315:** 253-255.

Nozaki, H., Iseki, M., Hasegawa, M., Misawa, K., Nakada, T., Sasaki, N. & Watanabe, M. (2007) Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Molecular Biology and Evolution,* **24:** 1592-1595.

O'Kelly, C. J. & Nerad, T. A. (1998) Kinetid architecture and bicosoecid affinities of the marine heterotrophic nanoflagellate *Caecitellus parvulus* (Griessmann, 1913) Patterson et al., 1993. *European Journal of Protistology,* **34:** 369-375.

Okamoto, N., Chantangsi, C., Horak, A., Leander, B. S. & Keeling, P. J. (2009) Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the *Hacrobia* taxon nov. *PLoS ONE,* **4:** e7080.

Palmer, J. D. (2003) The symbiotic birth and spread of plastids: How many times and whodunit? *Journal of Phycology,* **39:** 4-12.

Parfrey, L. W., Barbero, E., Lasser, E., Dunthorn, M., Bhattacharya, D., Patterson, D. J. & Katz, L. A. (2006) Evaluating support for the current classification of eukaryotic diversity. *PLoS Genetics,* **2:** e220.

Parfrey, L. W., Grant, J., Tekle, Y. I., Lasek-Nesselquist, E., Morrison, H. G., Sogin, M. L., Patterson, D. J. & Katz, L. A. (2010) broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Systematic Biology,* **59:** 518-533.

Park, J. S., Kolisko, M., Heiss, A. A. & Simpson, A. G. B. (2009) Light microscopic observations, ultrastructure, and molecular phylogeny of *Hicanonectes teleskopos* n. gen., n. sp., a deep-branching relative of diplomonads. *Journal of Eukaryotic Microbiology,* **56:** 373-384.

Park, J. S., Kolisko, M. & Simpson, A. G. B. (2010) Cell morphology and formal description of *Ergobibamus cyprinoides* n. g., n. sp., another Carpediemonas-like relative of diplomonads. *Journal of Eukaryotic Microbiology,* **57:** 520-528.

Patron, N. J., Inagaki, Y. & Keeling, P. J. (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Current Biology,* **17:** 887-891.

Patron, N. J., Rogers, M. B. & Keeling, P. J. (2004) Gene replacement of Fructose-1,6-Bisphosphate Aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryotic Cell,* **3:** 1169-1175.

Pattengale, N. D., Swenson, K. M. & Moret, B. M. E. (2010) Uncovering hidden phylogenetic consensus. *Bioinformatics Research*, **6053:** 123-139.

Pawlowski, J., Bolivar, I., Fahrni, J. F., Cavalier-Smith, T. & Gouy, M. (1996) Early origin of foraminifera suggested by SSU rRNA gene sequences. *Molecular Biology and Evolution,* **13:** 445-450.

Philippe, H. & Laurent, J. (1998) How good are deep phylogenetic trees? *Current Opinion in Genetics & Development,* **8:** 616-623.

Philippe, H., López, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M. & Le Guyader, H. (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proceedings of the Royal Society of London Series B-Biological Sciences,* **267:** 1213-1221.

Philippe, H., Delsuc, F., Brinkmann, H. & Lartillot, N. (2005) Phylogenomics. *Annual Review of Ecology, Evolution and Systematics,* **36:** 541-562.

Poe, S. (2003) Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. *Systematic Biology,* **52:** 423-428.

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

Poxleitner, M. K., Carpenter, M. L., Mancuso, J. J., Wang, C. J., Dawson, S. C. & Cande, W. Z. (2008) Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite *Giardia intestinalis*. *Science,* **319:** 1530-1533.

Pütz, S., Gelius-Dietrich, G., Piotrowski, M. & Henze, K. (2005) Rubrerythrin and peroxiredoxin: two novel putative peroxidases in the hydrogenosomes of the microaerophilic protozoon *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology,* **142:** 212-223.

Pütz, S., Doležal, P., Gelius-Dietrich, G., Bohacova, L., Tachezy, J. & Henze, K. (2006) Fe-hydrogenase maturases in the hydrogenosomes of *Trichomonas vaginalis*. *Eukaryotic Cell,* **5:** 579-586.

Quang, L. S., Gascuel, O. & Lartillot, N. (2008) Empirical perofile micture models for phylogenetic reconstruction. *Bioinformatics,* **24:** 2317-2323.

Rice, D. W. & Palmer, J. D. (2006) An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *Current Biology,* **15:** 1325-1330.

Rodriguez-Ezpeleta, N., Brinkman, H., C., B. S., Roure, B., Brurger, G., Loffelhardt, W., Bohnert, H. J., Philippe, H. & Lang, B. F. (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Current Biology,* **15:** 1325-1330.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A. J., Gray, M. W., Philippe, H. & Lang, B. F. (2007a) Toward resolving the eukaryotic tree: The phylogenetic positions of jakobids and cercozoans. *Current Biology,* **17:** 1420-1425.

Rodriguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B. F. & Philippe, H. (2007b) Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic Biology,* **56:** 389-399.

Roger, A. J. & Simpson, A. G. B. (2009) Evolution: Revisiting the root of the eukaryote tree. *Current Biology,* **19:** R165-R167.

Roger, A. J., Sandblom, O., Doolittle, W. F. & Philippe, H. (1999) An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. *Molecular Biology and Evolution,* **16:** 218-233.

Roger, A. J., Svärd, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. & Sogin, M. L. (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proceedings of the National Academy of Sciences of the United States of America,* **95:** 229-234.

Ruinen, J. (1938) Notizen über Salzflagellaten. II. Über die Verbereitung der Salzflagellaten. *Archiv Für Protistenkunde,* **90:** 210-258.

Ruiz-Trillo, I., Inagaki, Y., Davis, L. A., Sperstad, S., Landfald, B. & Roger, A. J. (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Current Biology,* **14:** R946-R947.

Ruiz-Trillo, I., Lane, C. E., Archibald, J. M. & Roger, A. J. (2006) Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. *Journal of Eukaryotic Microbiology,* **53:** 379-384.

Sagan, L. (1967) On the origin of mitosing cells. *Journal of Theoretical Biology,* **14:** 255-274.

Shimodaira, H. & Hasegawa, M. (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics,* **17:** 1246-1247.

Siddall, M. E., Hong, H. & Desser, S. S. (1992) Phylogenetic analysis of the Diplomonadida (Wenyon, 1926) Brugerolle, 1975: evidence for heterochrony in protozoa and against *Giardia lamblia* as a "missing link". *Journal of Protozoology,* **39:** 361-367.

Silberman, J. D., Simpson, A. G. B., Kulda, J., Cepicka, I., Hampl, V., Johnson, P. J. & Roger, A. J. (2002) Retortamonad flagellates are closely related to diplomonads - Implications for the history of mitochondrial function in eukaryote evolution. *Molecular Biology and Evolution,* **19:** 777-786.

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. & Patterson, D. J. (1999) The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the 'Excavate hypothesis'. *European Journal of Protistology,* **35:** 353-370.

Simpson, A. G. B. & Patterson, D. J. (2001) On core jakobids and excavate taxa: The ultrastructure of *Jakoba incarcerata*. *Journal of Eukaryotic Microbiology,* **48:** 480-492.

Simpson, A. G. B. & Roger, A. J. (2002) Eukaryotic evolution: Getting to the root of the problem. *Current Biology,* **12:** R691-R693.

Simpson, A. G. B., Bernard, C. & Patterson, D. J. (2000) The ultrastructure of *Trimastix marina* Kent, 1880 (Eukaryota), an excavate flagellate. *European Journal of Protistology,* **36:** 229-251.

Simpson, A. G. B., Lukes, J. & Roger, A. J. (2002a) The evolutionary history of kinetoplastids and their kinetoplasts. *Molecular Biology and Evolution,* **19:** 2071-2083.

Simpson, A. G. B., Radek, R., Dacks, J. B. & O'Kelly, C. J. (2002b) How oxymonads lost their groove: An ultrastructural comparison of *Monocercomonoides* and excavate taxa. *Journal of Eukaryotic Microbiology,* **49:** 239-248.

Simpson, A. G. B., Roger, A. J., Silberman, J. D., Leipe, D. D., Edgcomb, V. P., Jermiin, L. S., Patterson, D. J. & Sogin, M. L. (2002c) Evolutionary history of 'early-diverging' eukaryotes: The excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Molecular Biology and Evolution,* **19:** 1782-1791.

Simpson, A. G. B., Inagaki, Y. & Roger, A. J. (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of 'primitive' eukaryotes. *Molecular Biology and Evolution,* **23:** 615-625.

Small, I., Peeters, N., Legeai, F. & Lurin, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics,* **4:** 1581-1590.

Šmíd, O., Matusková, A., Harris, S. R., Kucera, S. R., Novotný, M., Horvathová, L., Hrdý, I., Kutejová, E., Hirt, R. P., Embley, T. M., Janata, J. & Tachezy, J. (2008) Reductive evolution of the mitochondrial processing peptidase of the unicellular parasites *Trichomonas vaginalis* and *Giardia intestinalis*. *PLoS Pathogens,* **4:** e1000243.

Sogin, M. L. (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *American Zoologist,* **29:** 487-499.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science,* **243:** 75-77.

Stamakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* **22:** 2688-2690.

Stechmann, A., Hamblin, K., Perez-Brocal, V., Gaston, D., Richmond, G. S., Van der Giezen, M., Clark, C. G. & Roger, A. J. (2008) Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. *Current Biology,* **18:** 580-585.

Steenkamp, E. T., Wright, J. & Baldauf, S. L. (2006) The protistan origins of animals and fungi. *Molecular Biology and Evolution,* **23:** 93-106.

Steinbüchel, A. & Müller, M. (1986) Anaerobic pyruvate metabolism of *Tritrichomonas foetus* and *Trichomonas vaginalis* hydrogenosomes. *Molecular and Biochemical Parasitology,* **20:** 57-65.

Stiller, J. W. & D., H. B. (1999) Long-branch attraction and the rDNA model of early eukaryotic evolution. *Molecular Biology and Evolution,* **16:** 1270-1279.

Stoeck, T., Taylor, T. T. & Epstein, S. S. (2003) Novel eukaryotes from the premanently anoxic Cariaco Basin (Caribbean Sea). *Applied and Environmental Microbiology,* **69:** 5656-5663.

Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V. & Epstein, S. (2007) Protistan diversity in the arctic: a case of paleoclimate shaping modern biodiversity? *PLoS ONE,* **2:** e728.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Christoserdov, A., Orsi, W. & Edgcomb, V. P. (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biology,* **7:** 1-20.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W. & Richards, T. A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology,* **19:** 21-31.

Susko, E., Field, C., Blouin, C. & Roger, A. J. (2003) Estimation of rates-across-sites distributions in phylogenetic substitution models. *Systematic Biology,* **52:** 594-603.

Susko, E., Inagaki, Y. & Roger, A. J. (2004) On inconsistency of the Neighbor-Joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled. *Molecular Biology and Evolution,* **21:** 1629-1642.

Sutak, R., Doležal, P., Fiumera, H. L., Hrdý, I., Dancis, A., Delgadillo-Correa, M., Johnson, P. J., Müller, M. & Tachezy, J. (2004) Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proceedings of the National Academy of Sciences of the United States of America,* **101:** 10368-10373.

Swofford, D. L. (2002) PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4*B10. *Sinauer Associates, Sunderland, Massachusetts*.

Tachezy, J. & Doležal, P. (2007) Iron-Sulfur Proteins and Iron-Sulfur Cluster Assembly in Organisms with Hydrogenosomes and Mitosomes. IN Martin, W. & Müller, M. (Eds.) *Origin of Mitochondria and Hydrogenosomes.* Berlin, Springer-Verlag.

Tachezy, J. & Doležal, P. (2011) The *Giardia* mitosomes. IN Lujan, H. D. & Svärd, S. G. (Eds.) *Giardia: A Model Organism.* Berlin, Springer-Verlag.

Tachezy, J., Sánchez, L. B. & Müller, M. (2001) Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Molecular Biology and Evolution,* **18:** 1919-1928.

Takishita, K., Miyake, H., Kawato, M. & Maruyama, T. (2005) Genetic diversity of microbial eukaryotes in anoxic sediment around fumaroles on a submarine caldera floor based on the small-subunit rDNA phylogeny. *Extremophiles,* **9:** 185-196.

Takishita, K., Tsuchiya, M., Kawato, M., Oguri, K., Kitazato, H. & Maruyama, T. (2007a) Genetic diversity of microbial eukaryotes in anoxic sediment of the saline meromictic lake Namako-ike (Japan): On the detection of anaerobic or anoxic-tolerant lineages of eukaryotes. *Protist,* **158:** 51-64.

Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y. & Maruyama, T. (2007b) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles,* **11:** 563-576.

Tekle, Y. I., Grant, J. S., Anderson, O. R., Nerad, T. A., Cole, J. C., Patterson, D. J. & Katz, L. A. (2008) Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. *Molecular Phylogenetics and Evolution*, **47:** 339-352.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* **24:** 4876-4882.

Tielens, A. G. M. & Van Hellemond, J. J. (2007) Anaerobic mitochondria: properties and origins. IN Martin, W. & Müller, M. (Eds.) *Origin of Mitochondria and Hydrogenosomes.* Berlin, Springer-Verlag.

Tjaden, J., Haferkamp, I., Boxma, B., Tielens, A. G. M., Huynen, M. & Hackstein, H. P. (2004) A divergent ADP/ATP carrier in the hydrogenosomes of *Trichomonas gallinae* argues for an independent origin of these organelles. *Molecular Microbiology,* **51:** 1439-1446.

Tovar, J., Fischer, A. & Clark, C. G. (2002) The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Molecular Microbiology,* **32:** 1013-1021.

Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., Van der Giezen, M., Hernández, M., Müller, M. & Lucocq, J. M. (2003) Mitochondrial remnant organelles of *Giardia* function in iron-sulfur protein maturation. *Nature,* **426:** 172-176.

Tumova, P., Hofstetrova, K., Nohynkova, E., Hovorka, O. & Kral, J. (2006) Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of *Giardia*. *Chromosoma,* **116:** 65-78.

Van de Peer, Y. & De Wachter, R. (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *Journal of Molecular Evolution*, **45:** 619-630.

Van de Peer, Y., Ali, A. B. & Meyer, A. (2000) Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious Fungi. *Gene,* **246:** 1-8.

Van der Giezen, M., Cox, S. & Tovar, J. (2004) The iron-sulfur cluster assembly genes iscS ans iscU of *Entamoeba histolytica* were acquired by horizontal gene transfer. *BMC Evolutionary Biology,* **4:** 7.

Van Grinsven, K. W. A., Rosnowsky, S., Van Weelden, S. W. H., Pütz, S., Van der Giezen, M., Martin, W., Van Hellemond, J. J., Tielens, A. G. M. & Henze, K. (2008) Acetate:succinate CoA-transferase in the hydrogenosomes of *Trichomonas vaginalis*. *The Journal of Biological Chemistry,* **283:** 1411-1418.

Van Keulen, H., Gutell, R. R., Gates, M. A., Campbell, S. R., Erlandsen, S. L., Jarroll, E. L., Kulda, J. & Meyer, E. A. (1993) Unique phylogenetic position of Diplomonadida based on the complete small subunit ribosomal-RNA sequence of *Giardia ardae*, *G. muris*, *G. duodenalis* and *Hexamita* sp. *FASEB Journal,* **7:** 223-231.

Vinh, L. S. & von Haeseler, A. (2004) IQPNNI: moving fast through tree space and stopping in time. *Molecular Biology and Evolution,* **8:** 1565-1571.

Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vosbrinck, B. A. & Woese, C. R. (1987) Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature,* **326:** 411-414.

Wainright, P. O., Hinkle, G., Sogin, M. L. & Stickel, S. K. (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science,* **260:** 340-342.

Walker, G., Dacks, J. B. & Embley, T. M. (2006) Ultrastructural description of *Breviata anathema*, n. gen., n. sp., the organism previously studied as "*Mastigamoeba invertens*". *Journal of Eukaryotic Microbiology,* **53:** 65-78.

Wang, H. C., Li, K., Susko, E. & Roger, A. J. (2008) A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evolutionary Biology,* **8**:331.

Wang, H. C., Susko, E. & Roger, A. J. (2009) PROCOV: maximum likelihood estimation of protein phylogeny under covarion models and site-specific covarion pattern analysis. *BMC Evolutionary Biology,* **9**:225.

Wawrzyniak, I., Roussel, M., Diogon, M., Couloux, A., Texier, C., Tan, K. S. W., Vivares, C. P., Delbac, F., Wincker, P. & El Alaoui, H. (2008) Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. *International Journal for Parasitology,* **38:** 1377-82.

Whelan, S. & Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. *Molecular Biology and Evolution,* **18:** 691-699.

Wiens, J. J. (2006) Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics,* **39:** 34-42.

Williams, K., Lowe, P. N. & Leadlay, P. F. (1987) Purification and characterization of pyruvate: ferredoxin oxidoreductase from the anaerobic protozoon *Trichomonas vaginalis*. *Biochemical Journal,* **246:** 529-536.

Yabuki, A., Inagaki, Y. & Ishida, K. (2010) *Palpitomonas bilix* gen. et sp. nov.: A novel deep-branching heterotroph possibly related to Archaeplastida or Hacrobia. *Protist,* **161:** 523-538.

Yabuki, A., Nakayama, T., Yubuki, N., Hashimoto, T., Ishida, K. & Inagaki, Y. (2011) *Tsukubamonas globosa* n. gen., n. sp., a novel excavate flagellate possibly holding a key for the early evolution in "Discoba". *Journal of Eukaryotic Microbiology,* **58:** 319-331.

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

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences,* **13:** 555-556.

Yarlett, N., Hann, A. C., Lloyd, D. & Williams, A. (1981) Hydrogenosomes in the rumen protozoon *Dasytricha ruminantium* Schuberg. *Biochemical Journal,* **200:** 365-372.

Yarlett, N., Coleman, G. S., Williams, A. & Lloyd, D. (1984) Hydrogenosomes in known species of rumen entodiniomorphid protozoa. *FEMS Microbiology Letters,* **21:** 15-19.

Yarlett, N., Orpin, C. G., Munn, E. A. & Yarlett, C. (1986) Hydrogenosomes in the rumen fungus *Neocallimastix patriciarum*. *Biochemical Journal,* **236:** 729-739.

Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. (2002) The single, ancient origin of chromist plastids. *Proceedings of the National Academy of Sciences of the United States of America,* **99:** 15507-15512.

Yu, L. Z., Birky, C. W. & Adam, R. D. (2002) The two nuclei of *Giardia* each have complete copies of the genome and are partitioned equationally at cytokinesis. *Eukaryotic Cell,* **1:** 191-199.

Yubuki, N., Inagaki, Y., Nakayama, T. & Inouye, I. (2007) Ultrastructure and ribosomal RNA phylogeny of the free-living heterotrophic flagellate *Dysnectes brevis* n. gen., n. sp., a new member of the fornicata. *Journal of Eukaryotic Microbiology,* **54:** 191-200.

Yubuki, N., Leander, B. S. & Silberman, J. D. (2010) Ultrastructure and molecular phylogenetic position of a novel phagotrophic stramenopile from low oxygen environments: *Rictus lutensis* gen. et sp. nov. (Bicosoecida, incertae sedis) *Protist,* **161:** 264-278.

Zettler, L. A. A., Nerad, T. A., O'Kelly, C. J. & Sogin, M. L. (2001) The nucleariid amoebae: more protists at the animal-fungal boundary. *Journal of Eukaryotic Microbiology,* **48:** 293-297.

Zuckerkandl, E. & Pauling, L. (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology,* **8:** 357-366.

Appendix A

Copyright permission for Chapter 2

Appendix B

Copyright permission for Chapter

Kolisko, M., Silberman, J. D., Cepicka, I., Yubuki, N., Takishita, K., Yabuki, A., Leander, B. S., Inouye, I., Inagaki, Y., Roger, A. J. & Simpson, A. G. B. (2010) A Wide Diversity of Previously Undetected Free-Living Relatives of Diplomonads Isolated from Marine/Saline Habitats. *Environmental Microbiology,* **12**(10)**:** 2700-2710.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Sep 07, 2011

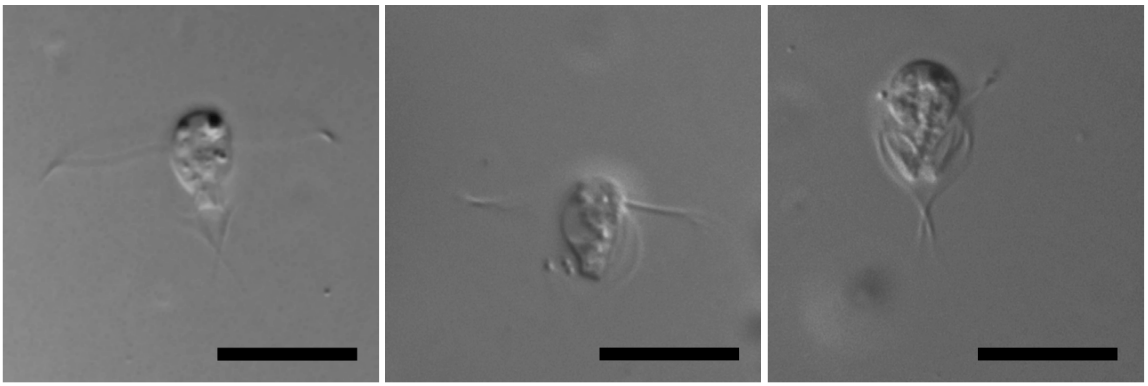
This is a License Agreement between Martin Kolisko ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

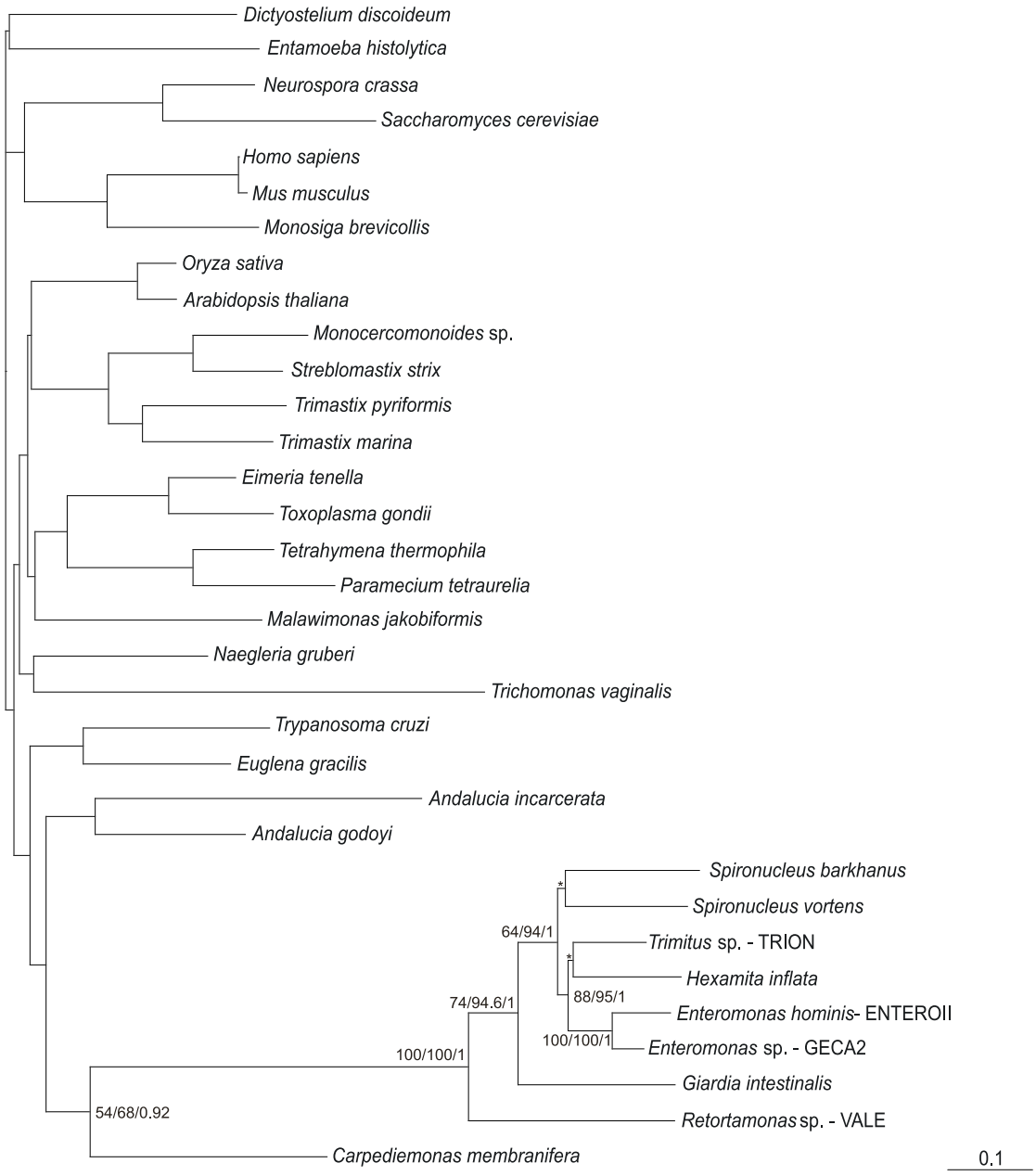
License Number	2743731132849
License date	Sep 07, 2011
Licensed content publisher	John Wiley and Sons
Licensed content publication	Environmental Microbiology
Licensed content title	A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats
Licensed content author	Martin Kolisko,Jeffrey D. Silberman,Ivan Cepicka,Naoji Yubuki,Kiyotaka Takishita,Akinori Yabuki,Brian S. Leander,Isao Inouye,Yuji Inagaki,Andrew J. Roger,Alastair G. B. Simpson
Licensed content date	Oct 1, 2010
Start page	2700
End page	2710
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Total	0.00 USD
Terms and Conditions	

Appendix C

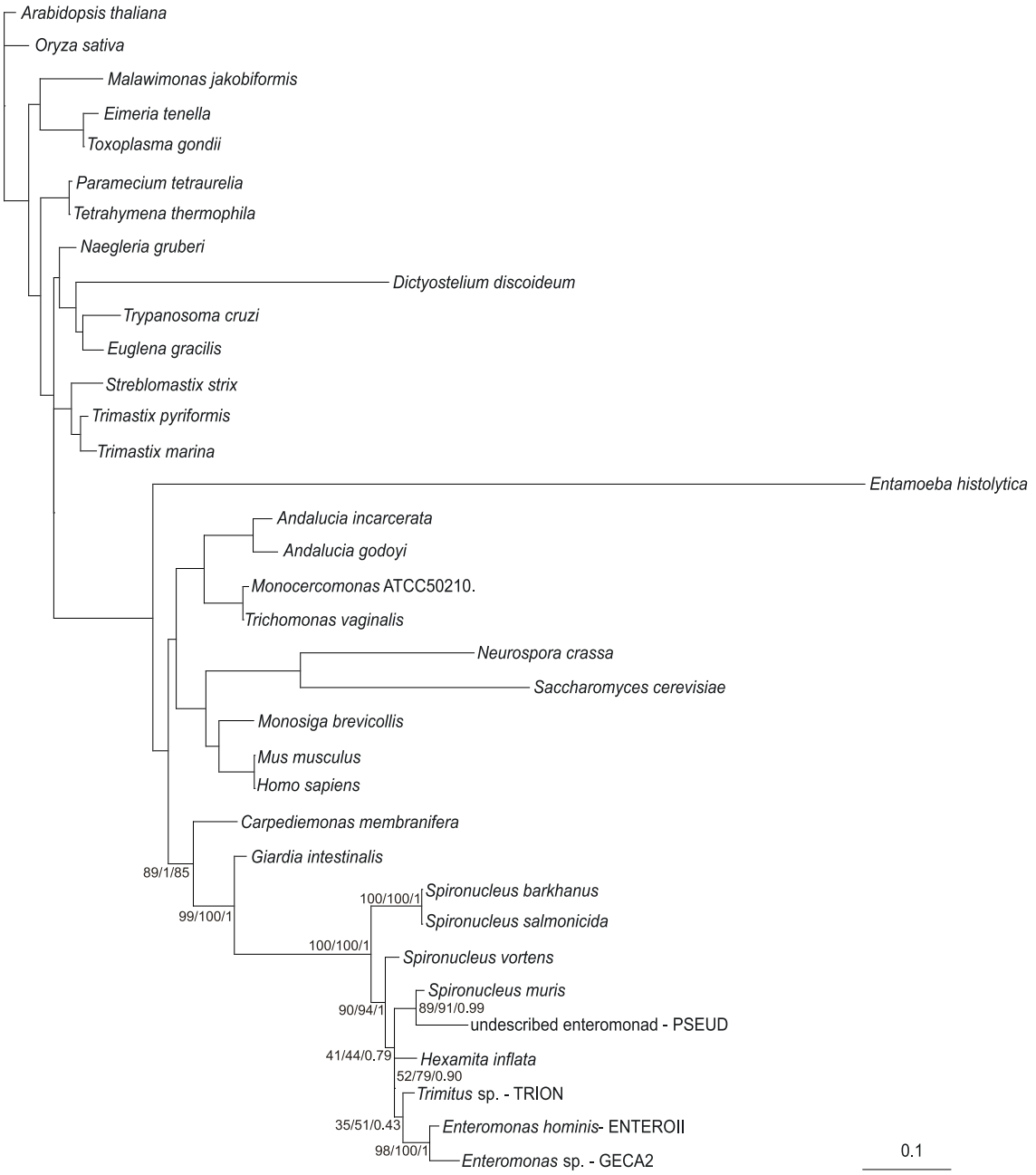
Supplementary Materials for Chapter 2



Appendix C.1. DIC light microscopy photographs of *Trepomonas steini*. Two quite long flagella are directed laterally while two posterior flagella extend beyond the end of the cell (four shorter flagella do not extend past the posterior end of the cell), which is typical for T. steini. As is typical for *T.steini*, the cell moves in one of two alternating modes i) slowly with a regular jerky rotation around the longitudinal axis, ii) a faster smooth movement. Scale bar is 10 μm.



Appendix C.2. Maximum likelihood (WAG + Γ + I) tree of Fornicata based on HSP90 protein sequences. Statistical support is as follows IQPNNI 500 bootstraps / 1000 RELL bootstraps / Bayesian posterior probability. ***** means that branch was not recovered in the majority rule consensus tree of the bootstrap analyses. Only statistical support within Fornicata shown.



Appendix C.3. Maximum likelihood (WAG + Γ + I) tree of Fornicata based on tubulin protein sequence. Statistical support is as follows IQPNNI 500 bootstraps / 1000 RELL bootstraps / Bayesian posterior probability. Only statistical support within Fornicata shown.

Appendix C.4. Table of all sequences used for the phylogenetic analyses described in chapter 2.

Taxon	Isolate	SSU rRNA	α-tubulin	hsp90
Acanthometra sp.		AF063240.1		
Acrasis rosea		AF011458.1		
Andalucia godoyi AND28		AY965870.1	EU334874	EU33480
Andalucia incarcerata			AAK27844.1	EU334886
Arabidopsis thaliana			NP_193232.1	AAN31859
Carpediemonas membranifera		AY117416.1	AAM77193.1	ABC54644
Cercomonas ATCC50317		U42449.1		
Clathrulina elegans		AY305009.1		
Cryptosporidium parvum		L16996.1		
Cyanophora paradoxa		X68483.1		
Dermamoeba algensis		AY294148.1		
Dictyostelium discoideum			XP_637058.1	P54651
Eimeria acervulina			CAA61255.1	
Eimeria tenella			CAA61255.1	O44001
Entamoeba histolytica			P31017	XP_653132
Enteromonad	PYX	AY921407		
Enteromonad	PSEUD	AY921408	EF551185	
Enteromonadidae sp.	KR-PO3	AY701872.1		
Enteromonas hominis	ENTEROII	EF551180	EF551184	EF551168
Enteromonas sp.	GECA2	EF551178	EF551186	EF551169
Enteromonas sp.	CUORA1	EF551179		
Euglena gracilis			P33625	AAQ24862.1
Giardia ardeae		Z17210.1		
Giardia intestinalis		M54878.1	XP_001705720	BAD83616
Giardia intestinalis isolate BAC7		AF199444.1		
Giardia microti		AF006676.1		
Giardia muris		X65063.1		
Giardia sp.		U20351.1		
Guillardia theta		X57162.1		
Hexamita inflata		L07836.1	AAC47085.1	AAR26696
Hexamita nelsoni		EF050053.1		
Homo sapiens			NP_006073	NP031381
Ichthyobodo sp.		AY229972.1		
Jakoba libera		AF411288.1		
Macropharyngomonas halophila		AF011465.1		

Malawimonas jakobiformis		AY117420.1	AAK27846.1	ABC54645
Mesostigma viride		AJ250108.1		
Monocercomonas ATCC50210			AAD05022.1	
Monocercomonas ruminantium	KOJ14	AY319280.1		
Monocercomonoides sp.				AAW221773
Monosiga brevicollis		AF100940.1	AAK27410.1	AAP51213
Mus musculus			NP_035783.1	P1149xxxx
Naegleria gruberi		M18732.1	P11237	AAM937561
Neurospora crassa			XP_963223.1	XP323482
Noctiluca scintillans		AF022200.1		
Octomitus intestinalis		DQ366277.1		
Oryza sativa			NP_001051132.1	P33126
Paramecium tetraurelia			XP_001454509	AAG00568
Podocoryne carnea		AF358092.1		
Reclinomonas americana		AY117417.1		
Retortamonas sp.	KOZA1	AF439344.1		
Retortamonas sp.	LOS	AF439345.1		
Retortamonas sp.	OVCE	AF439346.1		
Retortamonas sp.	ATCC 50375	AF439347.1		
Retortamonas sp.	VALE	AF439348.1		EF551172
Saccharomyces cerevisiae			NP_013625.1	NP_015084
Schizosaccharomyces pombe		X58056.1		
Spironucleus barkhanus		DQ186581.1	AAC47212	ABC54647
Spironucleus meleagridis		EF050054.1		
Spironucleus muris		EU043230	AAC47088.1	
Spironucleus salmonicida		DQ186595.1	ABB18140.1	
Spironucleus salmonis		DQ394703.1		
Spironucleus sp.	GEPA2H	EF551181		
Spironucleus torosa		EF050055.1		
Spironucleus vortens		U93086.1	AAB81021.1	EF551170
Streblomastix strix			ABC97356.1	AAO46123
Tetrahymena termophyla			XP_001022424	AAD937561
Toxoplasma gondii			P10873.1	AAP44977
Trepomonas agilis		AF015455.1		
Trepomonas sp.	PPS6	EF551174		
Trepomonas steinii	LUH3	EF551173		
Trichomonas vaginalis		U17510.1	AAK83156.1	XP_001317545
Trichonympha agilis		AB003920.1		

Trimastix marina		AF244905.1	ABC54661.1	ABC54648
Trimastix pyriformis		AF244903.1	PEP, *Trimastix pyriformis* Cluster TPL00000212	EU327684
Trimitus sp.	RAPI1	AY701873.1		
Trimitus sp.	TRION	AY701874.1	EF551183	EF551171
Trimitus sp.	IT1	EF551176		
Trimitus sp.	DOGA1	EF551177		
Trimitus sp.	KOMPKOJ	EF551182		
Trypanosoma cruzi			XP_802499.1	A26125
Uncultured eukaryote	CHESI2	EF551175		
Uncultured *Oxymonas* sp.		AB092931.1		
Volvox carteri		X53904.1		

11. Appendix D

Supplementary Materials for Chapter 3

Appendix D.1. Media formulations used to cultivate *Carpediemonas*-like organisms

802SW	Boil 5g of cerophyll in 1l of seawater for 5 min. Filter the medium and autoclave. Add 10 - 12 ml per 15 ml tube.
NM	In 15 ml tube combine: 1 sterile rice grain, 0.5ml of modified ATCC medium 1171 (prepared with heat inactivated horse serum) and 10 ml of sterile seawater
SW1773	In 15 ml tube mix: 9ml of 802SW and 3ml of sterile ATCC medium 1171
T/S	Mix: 485 ml of sterile seawater, 485 ml of sterile modified TYSGM-9 medium (prepared without serum – see below) and 30ml of heat inactivated horse serum. Add 10-12 ml per 15ml tube
3%LB	In 15ml tube mix: 300ul of LB media and 10 ml of sterile seawater
802SW/horse serum	Prepare horse serum slant: add 3 ml of horse serum into 15 ml tube. Incubate tubes on side at 80°C for 2 hours. The horse serum solidifies and forms slanted surface at the bottom of the tube. Repeat twice: incubate the horse serum slants overnight in the room temperature followed by incubation at 80°C for 2 hours. Add 3-4 ml of 802SW media over horse serum slant.
Modified TYSGM-9 medium	In 485ml of distilled water dissolve 1g of Tryptone, 0.5g of yeast extract, 1.4g of K_2HPO_4, 0.2g of KH_2PO_4 and 3.75g of NaCl. Autoclave. Add 15 ml of heat inactivated bovine or horse serum.
Pre-inoculation	Media for isolates PCE, PCS, NC and GSML were pre-inoculated with *Klebsiella* sp.

Appendix D.2

Analyses of 454 data

We searched for sequences from *Carpediemonas*-like organisms (CLOs) in two environmental PCR datasets from anoxic marine material (Stoeck *et al.*, 2009) with ~250,000 454 reads, and (Stoeck *et al.*, 2010) with ~660,000 454 reads). The first includes sequences derived from material from the Framvaren Fjord, and Cariaco Basin, the second from Framvaren Fjord only.

The sequences in these datasets are quite short (~150bp), and mostly encompass a variable region of the SSUrRNA gene. We were concerned, therefore, that simple BLAST analyses would not be a very sensitive method for identifying CLO sequences as they are quite divergent from each other. Therefore we analyzed all 454 reads one by one, using a combination of phylogenetic methods and similarity searches. The workflow was as follows:

1. Each 454 read was added to a reference alignment with similar taxon sampling to the one presented in the main paper, and aligned using the program mafft (Katoh et al., 2005) with the fastest possible set up ('mafft –intree 1 infile outfile')
2. Each alignment was then analyzed with program RAxML 7.0.4 (Stamakis, 2006) using the '–f p' option. The program used maximum parsimony to classify the new sequence to the existing reference tree (the new sequence is not present in the reference tree)
3. Program PHAT (part of the PhyloGenie package; (Frickey and Lupas, 2004) was then used to filter the trees where the new sequence was branching within or sister to Fornicata, e.g.,: sequences from possible CLO organisms or diplomonads. After this we were left with ~ 33,000 potential Fornicata sequences. However, as Fornicata sequences are long-branching, this set was presumed to include many divergent sequences from organisms unrelated to Fornicata, in addition to genuine Fornicata sequences.
4. The potential Fornicata sequences were then extracted into a fasta file and the program blastclust (from the NCBI blast suite) was used to cluster the sequences that were nearly identical (similarity set to 0.95) sequences. This step grouped the ~33,000 sequences into 1490 clusters.
5. These sequences were analyzed by BLAST and all sequences with obvious high similarity (similarity over 90%) to organisms other than CLOs were discarded. Around 90% of the sequences were excluded by this step.
6. The remaining sequences (150) were re-aligned to the dataset from the main paper with program mafft (einsi seting) and a tree was constructed using maximum likelihood program RAxML 7.0.4 (with GTRGAMMAI model).

Results: We have identified 22 reads closely related to clade CL6 from (only from Stoeck *et al.*, 2009, Framvaren Fjord), and 8 reads closely related to clade CL1 (only from Stoeck *et al.*, 2009, Framvaren Fjord). A further 11 sequences appear to be from diplomonads (8 from Framvaren fjord and 3 from Cariaco basin, from Stoeck *et al.*, 2009).

We have also identified one sequence that branches amongst CLOs in the ML phylogeny, but is not closely related to any of this known sequences from CL1-6. It is

possible that this sequence represents an additional CLO lineage, but more likely that it represents an unrelated sequence that is misplaced in this phylogeny. The 454 sequences are simply too short to make a definitive statement about position of this sequence.

Discussion: Analysis of 454 sequencing did allows us to recover sequences from two CLO clades from environments that did not yield any CLO sequences in previous studies that employed clone libraries and Sanger sequencing. It is possible that this was due to the much deeper sampling available with 454 sequencing. Interestingly we did not recover any CLO sequences in the larger 454 dataset (Stoeck et al. 2009). Overall this suggests that shallow sequence coverage is not the only reason for the limited recovery of CLO by previous environmental studies. It supports the idea that that CLOs are often extremely-rare-to-nonexistent in suboxic marine systems, or that there is some strong bias against their sequences in PCR studies.
A downside of 454 sequencing is the limited lengths of the reads, which makes it difficult to place some sequences on the tree, especially if the 454 sequence is not very similar to any available near-full-length sequence. It is quite possible that sequences from novel CLO lineages may have been missed by our analyses for this reason. Until longer sequences become available, we do not expect analysis of such datasets to be a particularly effective way of identifying additional major lineages within groups with divergent rRNA genes, such as Fornicata.

References

Frickey, T. & Lupas, N. L. (2004) Phylogenie: Automated Phylome Generation and Analysis. *Nucleic Acids Research,* **32**(17)**:** 5231-5238.

Katoh, K., Kei-ichi, K., Hiroyuki, T. & Takashi, M. (2005) Mafft Version 5: Improvement in Accuracy of Multiple Sequence Alignment. *Nucleic Acids Research,* **33**(2)**:** 511-518.

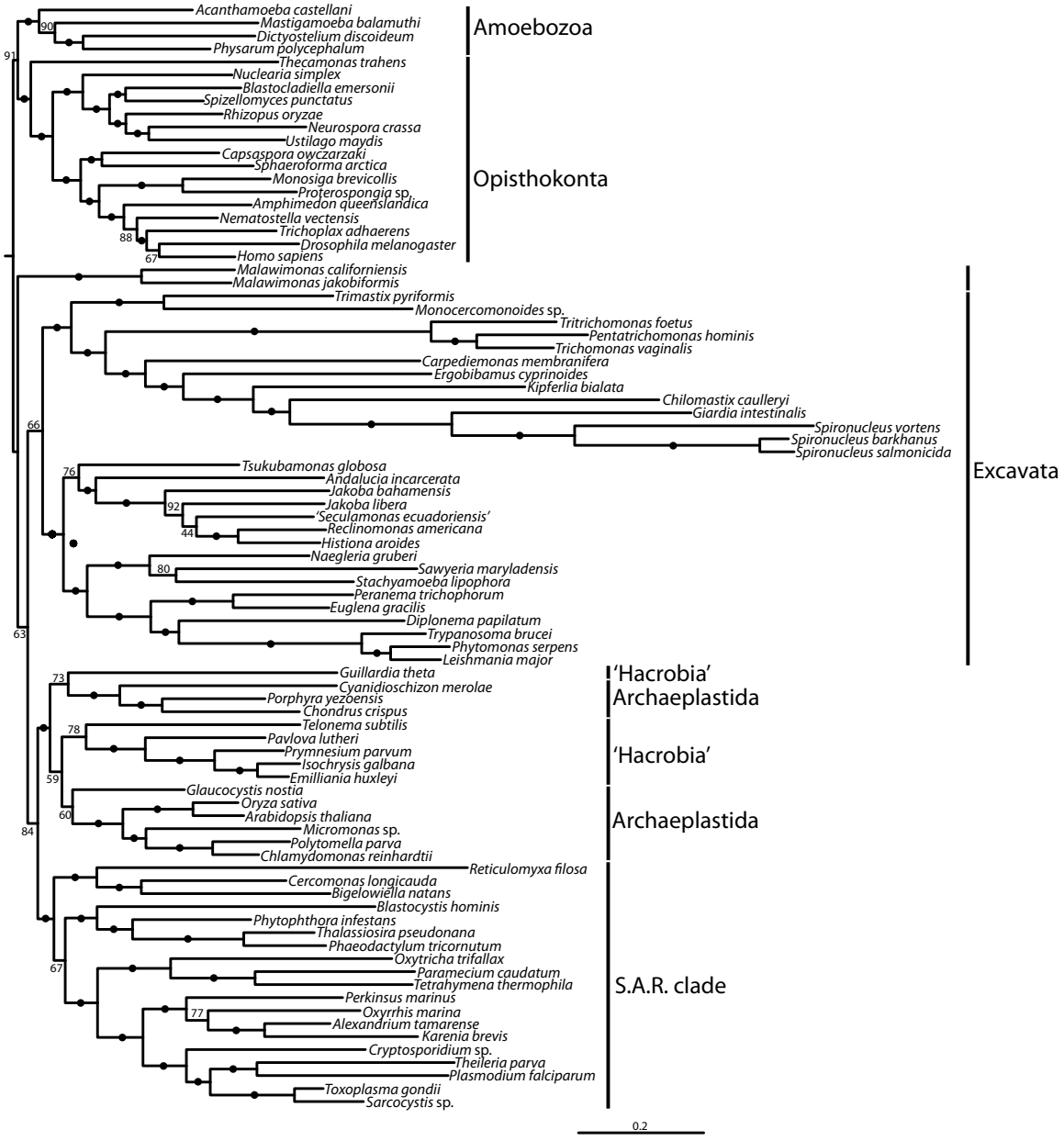
Stamakis, A. (2006) Raxml-Vi-Hpc: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics,* **22**(21)**:** 2688-2690.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W. & Richards, T. A. (2010) Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water. *Molecular Ecology,* **19:** 21-31.

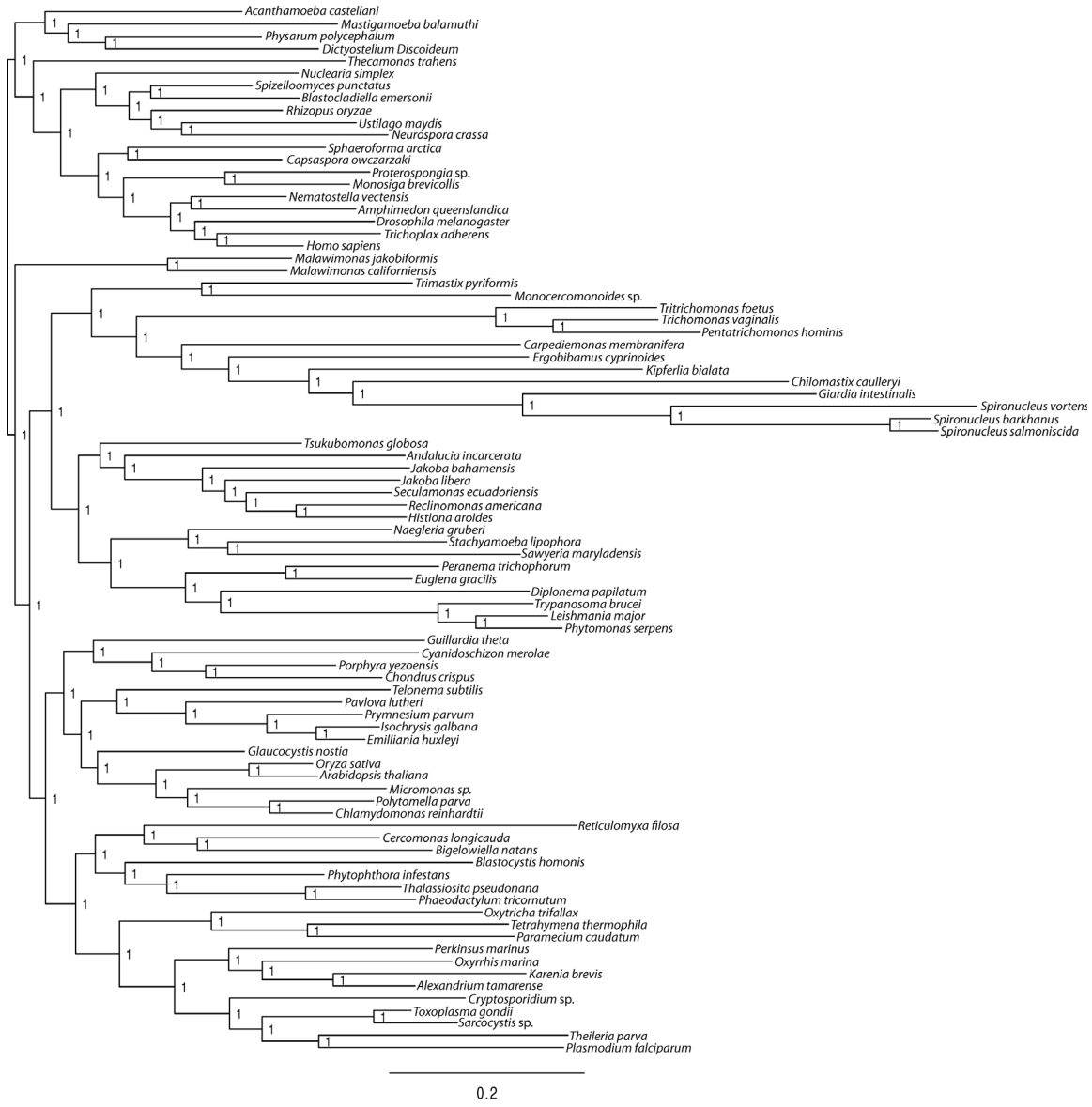
Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Christoserdov, A., Orsi, W. & Edgcomb, V. P. (2009) Massively Parallel Tag Sequencing Reveals the Complexityof Anaerobic Marine Protistan Communities. *BMC Biology,* **7**(72)**:** 1-20.

12. Appendix E

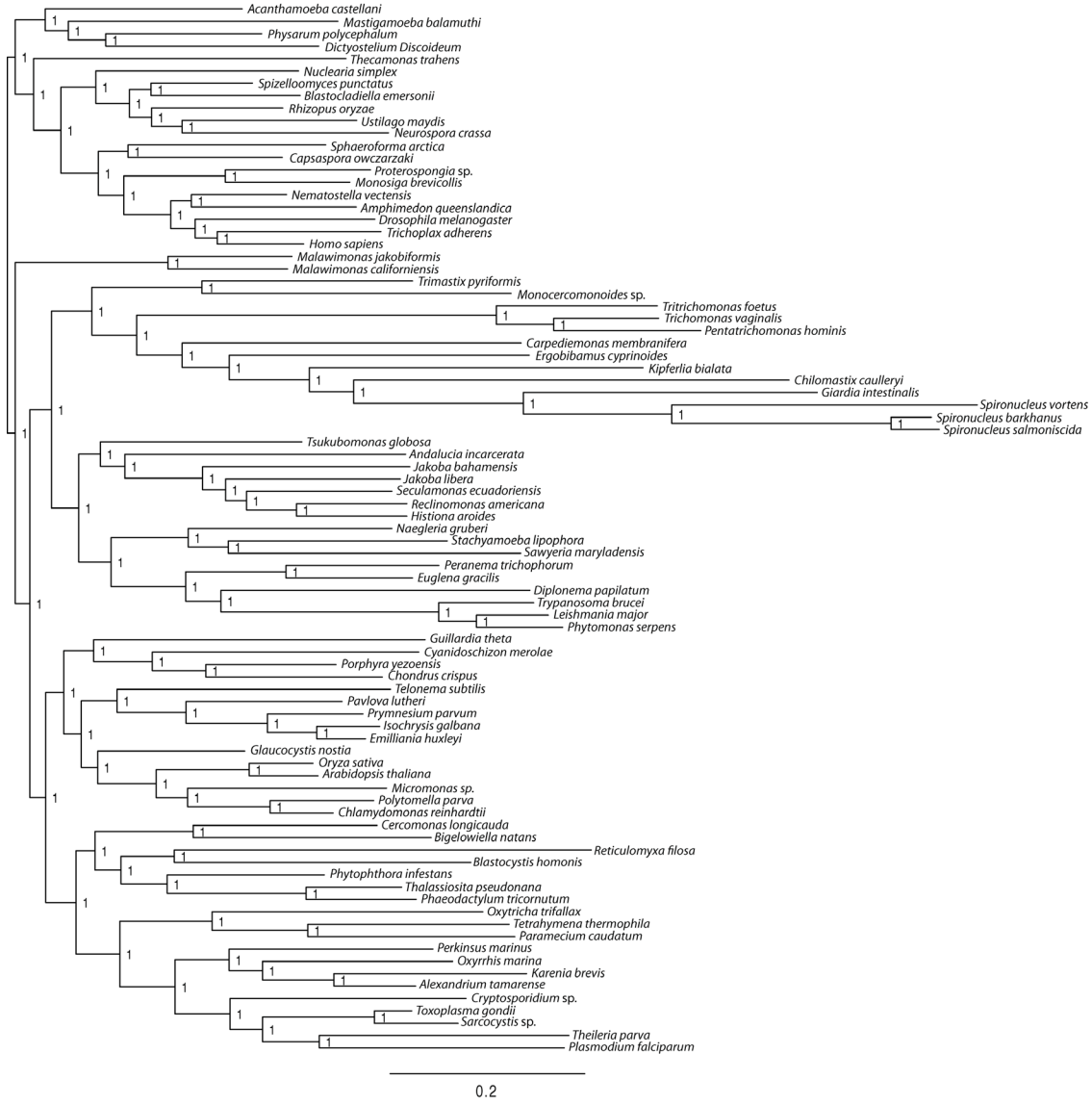
Supplementary Materials for Chapter 4



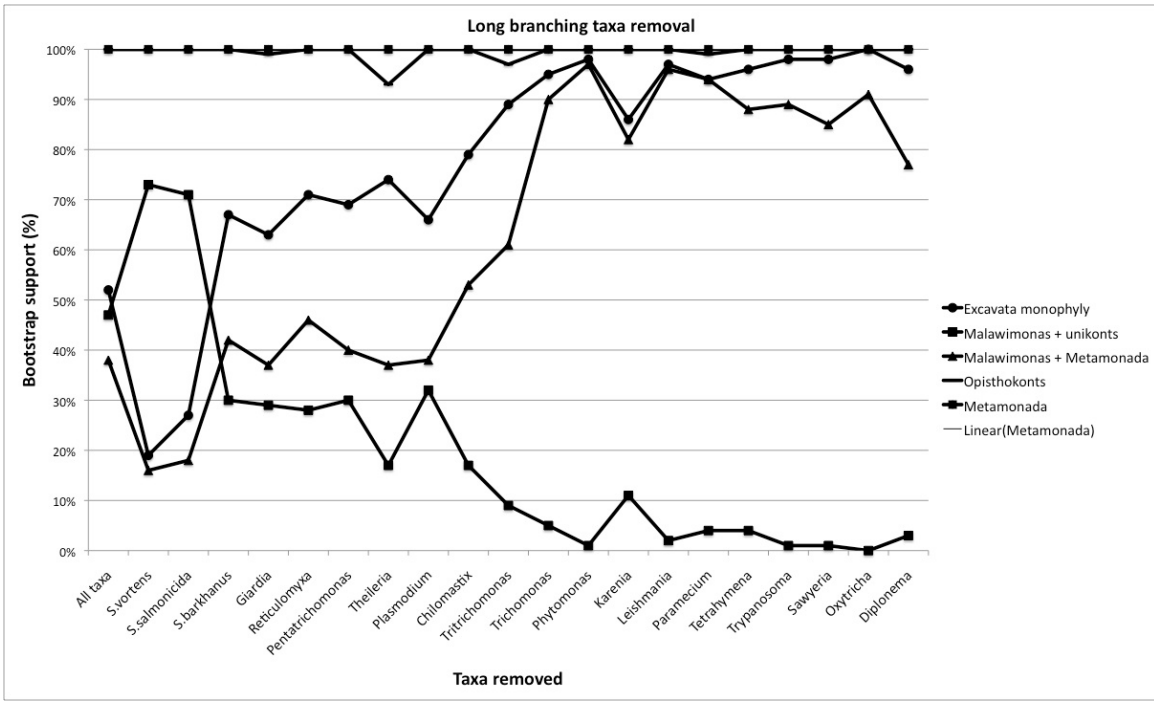
Appendix E.1 Maximum likelihood tree based on the 'CLO_rep' dataset (144 genes; 32995 sites), estimated using the program RAxML (model settings PROTGAMMALGF). The scale bar represents 0.2 expected substitutions per position. Statistical support was estimated using 500 bootstrap replicates. Dots on internal branches represent bootstrap support >95%. Newly sequenced taxa are depicted in bold font.



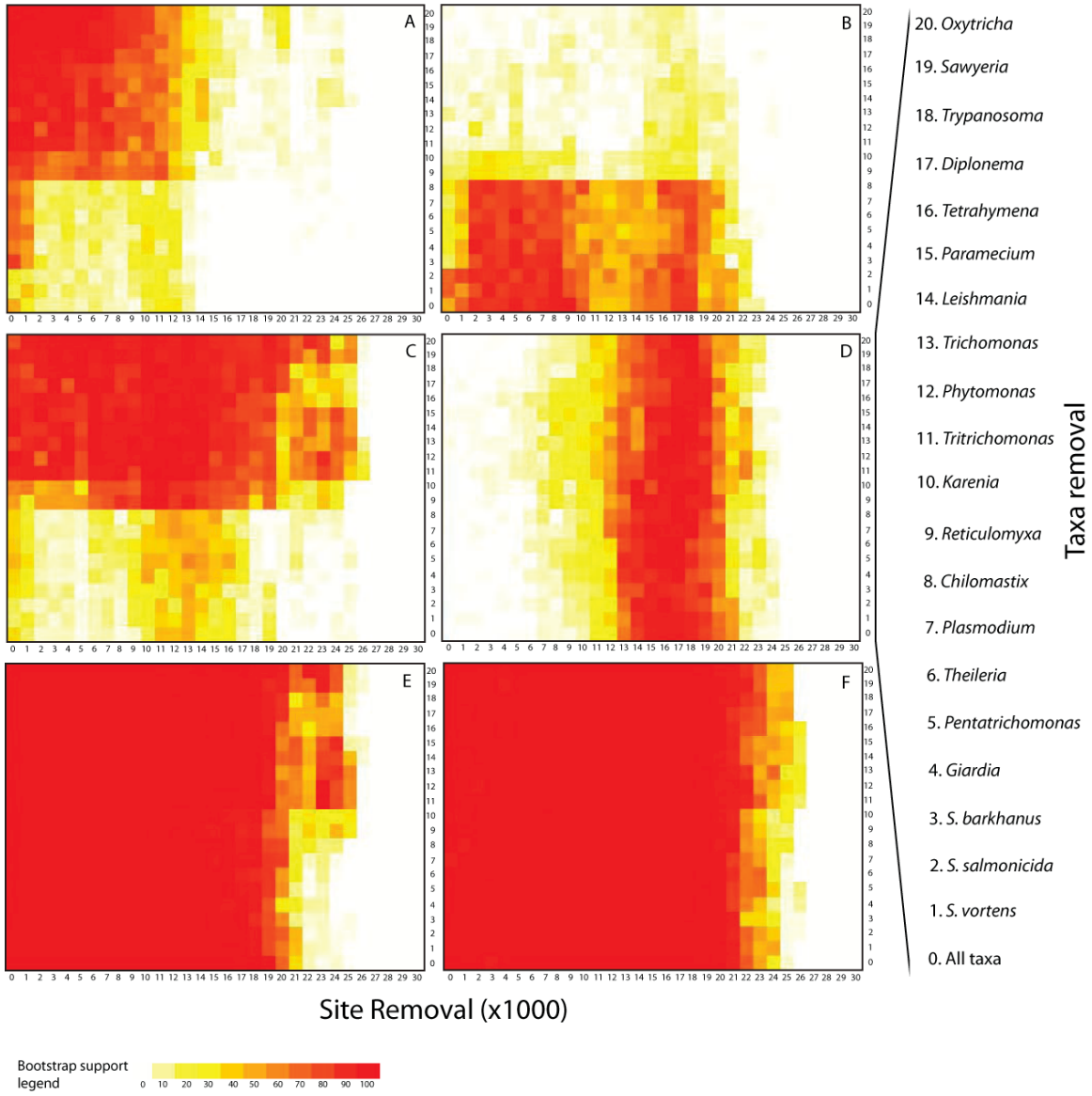
Appendix E.2. Bayesian phylogeny of 'Main' dataset reconstructed using the program PhyloBayes. Consensus based on three independent chains run for 40000 generations with burnin set to 4000. Numbers at nodes represent Bayesian posterior probabilisties.



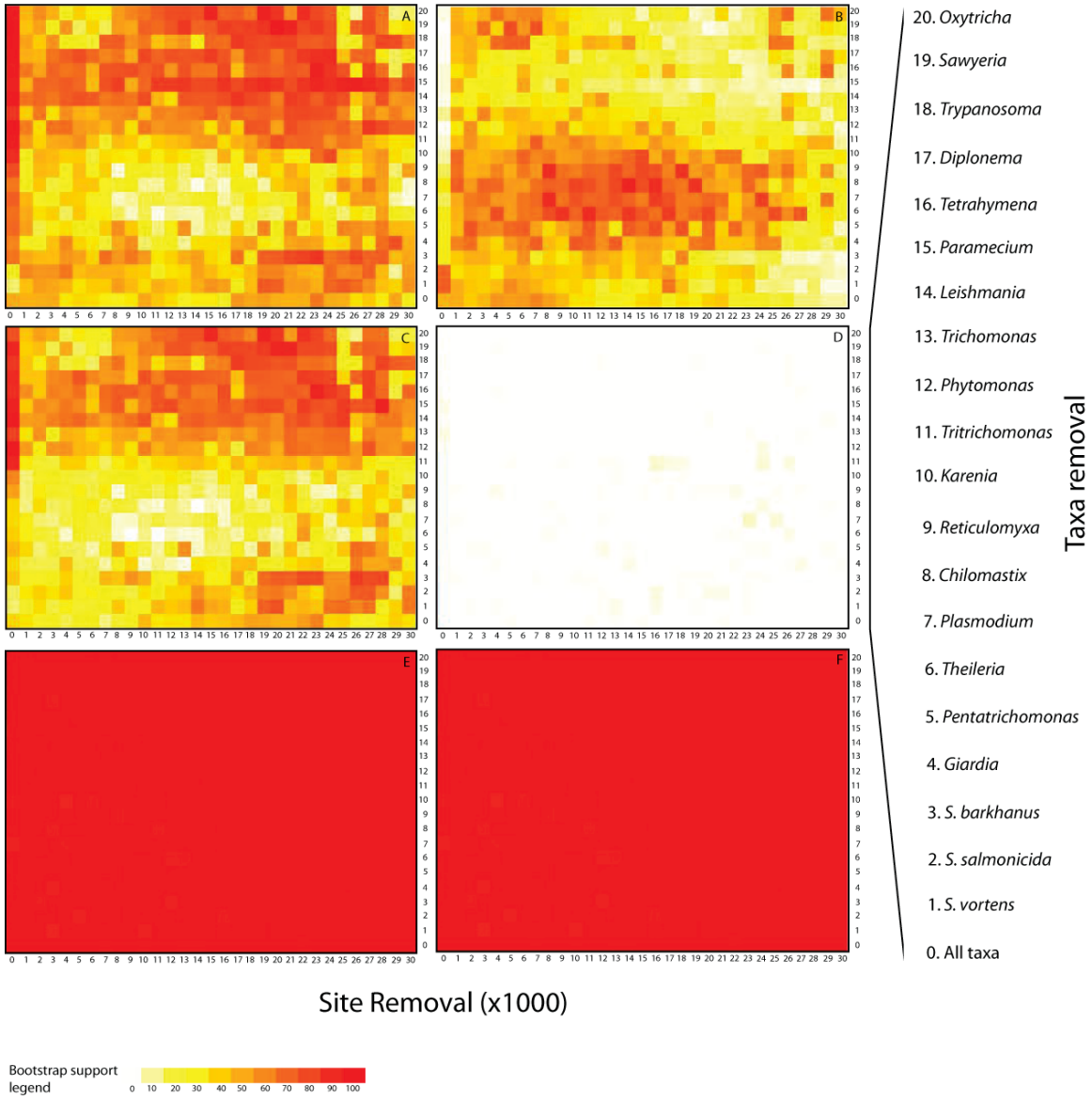
Appendix E.3. Bayesian phylogeny of 'Main' dataset reconstructed using program PhyloBayes. This tree is showing results of the 'rogue' chain that did not converge with other three chains (Appendix E.2). The trees are identical except for position of *Reticulomyxa filosa*. The chain was run for 40000 generations with burnin set to 4000. Numbers at nodes represent Bayesian posterior probabilisties.



Appendix E.4 Graph depicting the change of bootstrap support for different topologies as fast-evolving taxa are removed sequentially from the 'CLO_rep' dataset. After removal of 12 taxa the support for paraphyly of Excavata (i.e., a *Malawimonas*-unikonts grouping) is close to zero, and Excavata monophyly is strongly supported.



Appendix E.5 Heat map graphs depicting the bootstrap support for different topologies as fast-evolving sites are removed together with fast-evolving taxa from the 'CLO_rep' dataset. Numbers along the x axis represent the number of removed sites (x1000) and numbers on the y axis represent the number of removed taxa. The actual taxa removed are listed on the far right. A. Bootstrap support for monophyletic Excavata. B. Bootstrap support for paraphyletic Excavata (i.e., the ML topology in Figure 4.1). C. Bootstrap support for *Malawimonas*-Metamonada monophyly. D. Bootstrap support for *Malawimonas*-Metamonada-unikonts monophyly. E. Bootstrap support for Metamonada. F. Bootstrap support for unikont clade.



Appendix E.6 Heat map graphs depicting the bootstrap support for different topologies as sites are removed from fast-evolving taxa according to 'ratio of rates', together with removal of fast-evolving taxa from the 'CLO_rep' dataset. Numbers along the x axis represent the number of sites replaced with '-' character in fast-evolving taxa (x1000) and numbers on the y axis represent the number of removed taxa. The actual taxa removed are listed on the far right. A. Bootstrap support for monophyletic Excavata. B. Bootstrap support for paraphyletic Excavata (i.e., the ML topology). C. Bootstrap support for *Malawimonas*-Metamonada monophyly. D. Bootstrap support for *Malawimonas*-Metamonada-unikonts monophyly. E. Bootstrap support for Metamonada. F. Bootstrap support for unikont clade.

Appendix E.7. Table showing the bootstrap support for excavate paraphyly (i.e., a *Malawimonas*-unikont grouping) for the 'CLO_rep' dataset, for the 'CLO_rep' dataset when all three CLO taxa were excluded, and for ten 'jacknifed' datasets constructed by removing 10521 sites at random from the 'Main' dataset to give them the same number of sites as the 'CLO_rep' dataset (32995 sites). Support values estimated from 300 regular boostrap replicates.

Dataset	Bootstrap support	Confidence interval
CLO_rep	63	+/- 5.6
CLO_rep but no CLOS	60	+/- 5.7
Jacknifing 1	64	+/- 5.5
Jacknifing 2	65	+/- 5.5
Jacknifing 3	74	+ /-5
Jacknifing 4	93	+/- 3
Jacknifing 5	76	+/- 4.9
Jacknifing 6	71	+/- 5.2
Jacknifing 7	75	+/- 5
Jacknifing 8	66	+/- 5.5
Jacknifing 9	83	+/- 4.3
Jacknifing 10	92	+/- 3.1

Appendix E.8. Table of the used genes and gene sampling per taxon. '1' marks presence of particular gene for particular taxon in the dataset, while '0' marks missing data.

Taxon	Actin	Atub	Btub	EF2	Eif-5A	Gtub	HSP70-E	HSP70C	HSP90	Pace2C	Rad51A	Rpl13A	Rpl13e	Rpl14e	Rpl15	Rpl18	Rpl2	Rpl24A	Rpl3	Rpl4b	Rpl5	Rpl7a	Ubq	ar21	arc20	arf3	arpc1	atp6	calr	capz
Acanthamoeba castellani	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1
Alexandrium tamarense	1	1	1	1	1	0	1	1	1	0	0	1	0	1	1	1	0	1	1	0	0	1	1	0	1	1	0	1	0	0
Amphimedon queenslandica	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Andalucia incarcerata	1	1	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Arabidopsis thaliana	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
Carpediemonas membranifera	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1
Bigelowiella natans	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Blastocystis hominis	0	1	1	1	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1
Blastocladiella emersonii	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	0	1	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	1
Chilomastix caulleriy	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	1	1	1	0	0
Chlamydomonas reinhardtii	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Chondrus crispus	1	1	1	1	1	0	1	1	1	0	0	1	0	0	1	1	1	0	1	0	1	1	1	0	0	1	0	0	0	0
Ergobibamus cyprinoides	0	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1
Cryptosporidium sp.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0
Cyanidioschizon merolae	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0
Dictyostelium discoideum	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	0	1	0	1	1	0	1	0	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	1	0	1	0	0	0	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Euglena gracilis	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0
Giardia intestinalis	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0
Glaucocystis nostia	1	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0
Guillardia theta	1	1	1	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0
Histiona aroides	1	1	1	1	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isochrysis galbana	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	1	1	1	0	1	1	0	0	0	0	1	1	1	0	0
Jakoba bahamensis	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	0	0
Jakoba libera	1	1	1	1	1	0	0	0	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0
Karenia brevis	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1	0	1	1	1	0	0	1	1	0	1	0
Leishmania major	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	0	0	1	1	1	1	0
Malawimonas californiensis	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0

Taxon	Actin	Atub	Btub	EF2	Eif-5A	Gtub	HSP70-E	HSP70C	HSP90	Pace2C	Rad51A	Rpl13A	Rpl13e	Rpl14e	Rpl15	Rpl18	Rpl2	Rpl24A	Rpl3	Rpl4b	Rpl5	Rpl7a	Ubq	ar21	arc20	arf3	arpc1	atp6	calr	capz
Malawimonas jakobiformis	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Mastigamoeba balamuthi	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Micromonas sp.	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0
Monocercomonoides sp.	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Naegleria gruberi	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Nematostella vectensis	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1
Nuclearia simplex	1	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0
Oryza sativa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oxyrrhis marina	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0
Oxytricha trifallax	0	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0
Paramecium caudatum	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0
Pavlova lutheri	1	1	1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0
Pentatrichomonas hominis	1	0	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	0	0	1	0	1	1	1
Peranema trichophorum	1	1	1	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	1	0
Perkinsus marinus	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0
Phaeodactylum tricornutum	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
Physarym polycephalum	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	1	0	0	1	1	0	1	1
Phytophthora infestans	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
Phytomonas serpens	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0
Plasmodium falciparum	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	1	1	1	0	1
Polytomella parva	1	1	1	1	1	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0
Porphyra yezoensis	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0
Proterospongia sp.	1	1	1	1	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	1	1	1	0	1	1	1	0
Prymnesium parvum	1	1	1	1	1	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	0
Reclinomonas americana	1	1	1	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Reticulomyxa filosa	1	1	1	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	1
Rhizopus oryzae	0	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	0
Kipferlia bialata	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0
Sarcocystis sp.	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Sawyeria marylandensis	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1
Spironucleus barkhanus	0	1	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	0
Seculamonas ecuadoriensis	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	1	1	1

Taxon	Actin	Atub	Btub	EF2	Eif-5A	Gtub	HSP70-E	HSP70C	HSP90	Pace2C	Rad51A	Rpl13A	Rpl13e	Rpl14e	Rpl15	Rpl18	Rpl2	Rpl24A	Rpl3	Rpl4b	Rpl5	Rpl7a	Ubq	ar21	arc20	arf3	arpc1	atp6	calr	capz
Sphaerophorma arctica	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Spironucleus salmonicida	1	1	1	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0
Spironucleus vortens	1	0	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	0
Spizellomyces punctata	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0
Stachyamoeba lipophora	1	1	0	1	1	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0
Telonema subtilis	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0
Tetrahymena termophila	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1
Thalassiosira pseudonana	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0
Thecamonas trahens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1
Theileria parva	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0
Tsukubomonas globosa	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Toxoplasma gondii	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1
Trichoplax adhaerens	1	1	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	1
Trichomonas vaginalis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Trimastix pyriformis	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1
Tritrichomonas foetus	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
Trypanosoma brucei	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0
Ustilago maydis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
Taxon	cct-A	cct-B	cct-D	cct-E	cct-G	cct-N	cct-T	cct-Z	cpn60	crfg	ctsl1	ef1alpha	fh	fibri	fpps	gdi2	glcn	gnb2l	gnbpa	grc5	h3	h4	hla-B	hmt1	hsp70mt	if2b	if2g	if2p	if6	ino1
Acanthamoeba castellani	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	1	1	1
Alexandrium tamarense	1	1	0	1	1	0	1	1	1	1	0	1	0	1	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0
Amphimedon queenslandica	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	0
Andalucia incarcerata	1	1	0	0	1	0	1	1	1	0	1	1	0	0	1	1	0	1	1	1	1	1	0	0	1	0	1	1	0	1
Arabidopsis thaliana	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0
Carpediemonas membranifera	0	0	0	0	0	1	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Bigelowiella natans	1	0	1	1	1	0	1	0	1	0	1	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0	0	1	0
Blastocystis hominis	0	1	0	1	0	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1	0	0	1
Blastocladiella emersonii	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1

Taxon	cct-A	cct-B	cct-D	cct-E	cct-G	cct-N	cct-T	cct-Z	cpn60	crfg	ctsl1	ef1alpha	fh	fibri	fpps	gdi2	glcn	gnb2l	gnbpa	grc5	h3	h4	hla-B	hmt1	hsp70mt	if2b	if2g	if2p	if6	ino1
Chilomastix caulleriy	1	0	1	0	1	0	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1
Chlamydomonas reinhardtii	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Chondrus crispus	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0
Ergobibamus cyprinoides	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1
Cryptosporidium sp.	1	1	0	0	1	0	0	1	0	0	1	1	0	0	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	0
Cyanidioschizon merolae	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	0
Dictyostelium discoideum	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1
Drosophila melanogaster	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0
Euglena gracilis	0	1	0	1	1	1	1	0	1	0	1	1	0	1	1	1	0	1	0	1	1	0	1	0	1	0	1	1	1	0
Giardia intestinalis	1	1	1	0	1	1	1	0	1	1	0	1	0	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1
Glaucocystis nostia	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	1	1	1	1	0	0	1	0	1	0
Guillardia theta	0	1	0	0	1	0	1	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	0
Histiona aroides	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0	0	0	0	1	0	1	1
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isochrysis galbana	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	1	0
Jakoba bahamensis	1	0	1	0	0	1	0	1	0	0	1	1	1	1	0	1	0	1	1	1	0	1	1	0	0	0	1	1	0	1
Jakoba libera	0	1	0	1	0	1	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1
Karenia brevis	0	0	0	1	1	0	0	0	1	0	1	1	0	1	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0
Leishmania major	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
Malawimonas californiensis	0	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	0	1	0	1	1	0	0	1	0
Malawimonas jakobiformis	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	1	0	1	0	1	0	0	0	1	1	1	1	0
Mastigamoeba balamuthi	1	0	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
Micromonas sp.	1	1	1	1	1	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1	1	0	1	0	1	0	0	0	0	0
Monocercomonoides sp.	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	1
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Naegleria gruberi	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Nematostella vectensis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0
Nuclearia simplex	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
Oryza sativa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oxyrrhis marina	1	1	1	1	0	0	1	1	0	1	1	0	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1	0	1	0
Oxytricha trifallax	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0
Paramecium caudatum	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0
Pavlova lutheri	0	1	0	1	1	0	1	1	0	1	0	0	0	1	0	1	0	1	0	1	1	1	0	0	0	1	0	0	0	0

Taxon	cct-A	cct-B	cct-D	cct-E	cct-G	cct-N	cct-T	cct-Z	cpn60	crfg	ctsl1	ef1alpha	fh	fibri	fpps	gdi2	glcn	gnb2l	gnbpa	grc5	h3	h4	hla-B	hmt1	hsp70mt	if2b	if2g	if2p	if6	ino1
Pentatrichomonas hominis	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0
Peranema trichophorum	1	0	0	1	0	1	1	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	0	1	0	0
Perkinsus marinus	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
Phaeodactylum tricornutum	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0
Physarym polycephalum	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1
Phytophthora infestans	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1
Phytomonas serpens	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	1	0
Plasmodium falciparum	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1
Polytomella parva	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0
Porphyra yezoensis	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	1	0	1	1	1	1	0	1	0	1	1	1	1
Proterospongia sp.	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0
Prymnesium parvum	1	1	1	1	0	0	0	0	1	0	1	1	0	0	0	1	0	1	0	1	1	0	1	0	1	1	0	0	1	1
Reclinomonas americana	1	0	0	1	0	0	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1
Reticulomyxa filosa	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0	1	1	0	1	1
Rhizopus oryzae	1	1	1	0	1	1	1	1	0	0	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0	0	1	0	1	0
Kipferlia bialata	1	0	1	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	1	0
Sarcocystis sp.	0	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	1	0	1	1	0	1	0	0	1	1	0	1	0
Sawyeria marylandensis	1	1	1	0	1	0	0	1	0	1	1	1	0	1	0	1	0	1	1	1	1	1	1	0	1	0	1	0	1	1
Spironucleus barkhanus	0	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1	1	0	0	0	0	0	1	1	0
Seculamonas ecuadoriensis	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0	0	1	1	1	1	1	0	1	1
Sphaerophorma arctica	0	1	1	1	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0
Spironucleus salmonicida	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	0
Spironucleus vortens	0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0
Spizellomyces punctata	1	1	1	0	0	1	1	1	0	1	0	0	1	1	0	1	0	1	1	1	1	0	1	0	0	0	1	1	0	0
Stachyamoeba lipophora	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0
Telonema subtilis	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	0	1	0	1	1
Tetrahymena termophila	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0
Thalassiosira pseudonana	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1
Thecamonas trahens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Theileria parva	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	0
Tsukubomonas globosa	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Toxoplasma gondii	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	1	1	0

Taxon	cct-A	cct-B	cct-D	cct-E	cct-G	cct-N	cct-T	cct-Z	cpn60	crfg	ctsl1	ef1alpha	fh	fibri	fpps	gdi2	glcn	gnb2l	gnbpa	grc5	h3	h4	hla-B	hmt1	hsp70mt	if2b	if2g	if2p	if6	ino1
Trichoplax adhaerens	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0
Trichomonas vaginalis	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Trimastix pyriformis	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	1	1	0	0	0	0	1	1
Tritrichomonas foetus	1	1	1	0	1	0	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	1	0	1
Trypanosoma brucei	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
Ustilago maydis	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Taxon	l10a	mat	mcm-A	mcm-B	mcm-C	mcm-D	metap2	mra1	ndf1	nsf1-C	nsf1-E	nsf1-G	nsf1-I	nsf1-J	nsf1-K	nsf1-L	nsf1-M	nsf2-A	nsf2-F	orf2	osgep	pace2-A	pace2B	pace5	pp2A-b	psma-A	psma-B	psma-C	psma-E	psma-F
Acanthamoeba castellani	1	1	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1
Alexandrium tamarense	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Amphimedon queenslandica	1	1	1	0	0	0	1	0	1	1	1	1	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1
Andalucia incarcerata	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0
Arabidopsis thaliana	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Carpediemonas membranifera	1	1	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1
Bigelowiella natans	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1
Blastocystis hominis	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1
Blastocladiella emersonii	1	1	0	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Chilomastix caulleriy	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	1	1	0	1	0	1	0	1	1	0
Chlamydomonas reinhardtii	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Chondrus crispus	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	0
Ergobibamus cyprinoides	1	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Cryptosporidium sp.	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cyanidioschizon merolae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Dictyostelium discoideum	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1

Taxon	cct-A	cct-B	cct-D	cct-E	cct-G	cct-N	cct-T	cct-Z	cpn60	crfg	ctsl1	ef1alpha	fh	fibri	fpps	gdi2	glcn	gnb2l	gnbpa	grc5	h3	h4	hla-B	hmt1	hsp70mt	if2b	if2g	if2p	if6	ino1
Trichoplax adhaerens	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0
Trichomonas vaginalis	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Trimastix pyriformis	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	1	1	0	0	0	0	1	1
Tritrichomonas foetus	1	1	1	0	1	0	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	1	0	1
Trypanosoma brucei	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
Ustilago maydis	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Taxon	l10a	mat	mcm-A	mcm-B	mcm-C	mcm-D	metap2	mra1	ndf1	nsf1-C	nsf1-E	nsf1-G	nsf1-I	nsf1-J	nsf1-K	nsf1-L	nsf1-M	nsf2-A	nsf2-F	orf2	osgep	pace2-A	pace2B	pace5	pp2A-b	psma-A	psma-B	psma-C	psma-E	psma-F
Acanthamoeba castellani	1	1	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1
Alexandrium tamarense	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Amphimedon queenslandica	1	1	1	0	0	0	1	0	1	1	1	1	0	0	0	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1
Andalucia incarcerata	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0
Arabidopsis thaliana	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Carpediemonas membranifera	1	1	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1
Bigelowiella natans	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1
Blastocystis hominis	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1
Blastocladiella emersonii	1	1	0	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Chilomastix caulleriy	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	1	1	0	1	0	1	0	1	1	0
Chlamydomonas reinhardtii	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Chondrus crispus	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	0
Ergobibamus cyprinoides	1	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Cryptosporidium sp.	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cyanidioschizon merolae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Dictyostelium discoideum	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1

Taxon	l10a	mat	mcm-A	mcm-B	mcm-C	mcm-D	metap2	mra1	ndf1	nsf1-C	nsf1-E	nsf1-G	nsf1-I	nsf1-J	nsf1-K	nsf1-L	nsf1-M	nsf2-A	nsf2-F	orf2	osgep	pace2-A	pace2B	pace5	pp2A-b	psma-A	psma-B	psma-C	psma-E	psma-F
Euglena gracilis	1	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	1	1
Giardia intestinalis	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1
Glaucocystis nostia	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	0	1	0	1
Guillardia theta	1	0	1	0	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	0	1	0	1	1	1	0	1	1	1
Histiona aroides	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isochrysis galbana	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	1	0	0	1
Jakoba bahamensis	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0
Jakoba libera	1	1	0	1	1	0	1	0	1	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	1	1	1	1
Karenia brevis	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0
Leishmania major	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Malawimonas californiensis	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	0	1	0	1
Malawimonas jakobiformis	1	1	0	0	0	0	0	1	0	0	0	1	0	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1	0	1
Mastigamoeba balamuthi	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	0	1	1	1
Micromonas sp.	1	1	0	0	0	0	0	1	1	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0
Monocercomonoides sp.	1	1	0	0	1	0	1	1	0	0	0	1	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Naegleria gruberi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Nematostella vectensis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Nuclearia simplex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Oryza sativa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1
Oxyrrhis marina	0	1	1	0	1	0	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	1
Oxytricha trifallax	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0
Paramecium caudatum	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	0
Pavlova lutheri	1	0	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0
Pentatrichomonas hominis	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	1	0	1	1
Peranema trichophorum	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Perkinsus marinus	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Phaeodactylum tricornutum	0	1	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Physarym polycephalum	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0
Phytophthora infestans	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1
Phytomonas serpens	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
Plasmodium falciparum	0	1	1	0	0	1	1	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	1	1	0	0	1	0	0	1
Polytomella parva	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Porphyra yezoensis	1	1	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1
Proterospongia sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1
Prymnesium parvum	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0

Taxon	l10a	mat	mcm-A	mcm-B	mcm-C	mcm-D	metap2	mra1	ndf1	nsf1-C	nsf1-E	nsf1-G	nsf1-I	nsf1-J	nsf1-K	nsf1-L	nsf1-M	nsf2-A	nsf2-F	orf2	osgep	pace2-A	pace2B	pace5	pp2A-b	psma-A	psma-B	psma-C	psma-E	psma-F
Reclinomonas americana	1	1	0	0	0	0	1	1	1	1	1	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	1
Reticulomyxa filosa	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
Rhizopus oryzae	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1
Kipferlia bialata	1	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
Sarcocystis sp.	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0
Sawyeria marylandensis	1	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	1	0	1
Spironucleus barkhanus	1	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0
Seculamonas ecuadoriensis	1	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	1	1
Sphaerophorma arctica	1	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	1	0	1	0
Spironucleus salmonicida	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Spironucleus vortens	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0
Spizellomyces punctata	0	1	0	0	0	0	1	0	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1
Stachyamoeba lipophora	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Telonema subtilis	1	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	1	1	0	1	0	0	1	0	1	0	1	1
Tetrahymena termophila	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Thalassiosira pseudonana	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Thecamonas trahens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Theileria parva	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Tsukubomonas globosa	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1	1	1	0	1	1	0	0	1	0	1	1	1	1	1
Toxoplasma gondii	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	1
Trichoplax adhaerens	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
Trichomonas vaginalis	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Trimastix pyriformis	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	0	1
Tritrichomonas foetus	1	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1
Trypanosoma brucei	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Ustilago maydis	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Taxon	psma-G	psma-H	psma-J	psmb-K	psmb-L	psmb-M	psmb-N	psmd	rac	rad23	ran	rf1	rla2a	rla2b	rpl11	rpl12	rpl17	rpl19	rpl20	rpl21	rpl26	rpl27	rpl30	rpl31	rpl32	rpl33	rpl35	rpl43	rpl44	rpl6
Acanthamoeba castellani	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Alexandrium tamarense	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Amphimedon queenslandica	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Taxon	psma-G	psma-H	psma-J	psmb-K	psmb-L	psmb-M	psmb-N	psmd	rac	rad23	ran	rf1	rla2a	rla2b	rpl11	rpl12	rpl17	rpl19	rpl20	rpl21	rpl26	rpl27	rpl30	rpl31	rpl32	rpl33	rpl35	rpl43	rpl44	rpl6
Andalucia incarcerata	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Arabidopsis thaliana	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Carpediemonas membranifera	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Bigelowiella natans	0	1	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Blastocystis hominis	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Blastocladiella emersonii	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	0	0	1	0	0	0	1	0	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	1	1	0	0	1	1
Chilomastix caulleriy	1	1	1	1	1	0	1	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1
Chlamydomonas reinhardtii	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chondrus crispus	0	0	0	1	1	1	1	1	1	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1	1	1	0	0	0	1
Ergobibamus cyprinoides	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Cryptosporidium sp.	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cyanidioschizon merolae	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dictyostelium discoideum	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Diplonema papilatum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Euglena gracilis	0	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Giardia intestinalis	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0
Glaucocystis nostia	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Guillardia theta	1	1	1	0	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
Histiona aroides	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0	1	0	1	0	1	1
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Isochrysis galbana	0	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jakoba bahamensis	0	0	1	1	1	0	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	0	1	1	0	1	1	0	1
Jakoba libera	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0
Karenia brevis	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	1	0	1	1	0
Leishmania major	0	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Malawimonas californiensis	0	0	0	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1
Malawimonas jakobiformis	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Mastigamoeba balamuthi	0	0	0	0	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Micromonas sp.	1	1	0	1	0	1	0	0	0	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Monocercomonoides sp.	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1

Taxon	psma-G	psma-H	psma-J	psmb-K	psmb-L	psmb-M	psmb-N	psmd	rac	rad23	ran	rf1	rla2a	rla2b	rpl11	rpl12	rpl17	rpl19	rpl20	rpl21	rpl26	rpl27	rpl30	rpl31	rpl32	rpl33	rpl35	rpl43	rpl44	rpl6
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Naegleria gruberi	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1
Nematostella vectensis	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nuclearia simplex	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oryza sativa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oxyrrhis marina	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oxytricha trifallax	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
Paramecium caudatum	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pavlova lutheri	1	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	0	0	1	1
Pentatrichomonas hominis	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	1	1	1	1	1	0	0	0	0	1	0	0	1
Peranema trichophorum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Perkinsus marinus	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Phaeodactylum tricornutum	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Physarym polycephalum	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0
Phytophthora infestans	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Phytomonas serpens	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Plasmodium falciparum	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1
Polytomella parva	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Porphyra yezoensis	1	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Proterospongia sp.	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	1	0	0	0	1	1	1
Prymnesium parvum	1	1	1	0	1	0	1	0	0	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	1
Reclinomonas americana	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Reticulomyxa filosa	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Rhizopus oryzae	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kipferlia bialata	0	1	1	0	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Sarcocystis sp.	1	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sawyeria marylandensis	1	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Spironucleus barkhanus	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0
Seculamonas ecuadoriensis	1	0	0	0	0	0	1	1	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	1	1	1	0	1	1	1
Sphaerophorma arctica	1	1	0	0	0	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Spironucleus salmonicida	0	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0
Spironucleus vortens	0	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	1	1
Spizellomyces punctata	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Taxon	psma-G	psma-H	psma-J	psmb-K	psmb-L	psmb-M	psmb-N	psmd	rac	rad23	ran	rf1	rla2a	rla2b	rpl11	rpl12	rpl17	rpl19	rpl20	rpl21	rpl26	rpl27	rpl30	rpl31	rpl32	rpl33	rpl35	rpl43	rpl44	rpl6
Stachyamoeba lipophora	0	0	0	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1
Telonema subtilis	0	0	0	1	0	1	0	1	0	0	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0
Tetrahymena termophila	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1
Thalassiosira pseudonana	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Thecamonas trahens	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Theileria parva	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tsukubomonas globosa	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Toxoplasma gondii	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Trichoplax adhaerens	1	0	0	0	1	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1
Trichomonas vaginalis	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0
Trimastix pyriformis	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Tritrichomonas foetus	1	1	0	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Trypanosoma brucei	0	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
Ustilago maydis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Taxon	rpl9	rpo-A	rpo-B	rpo-C	rppO	rps10	rps11	rps12	rps14	rps15	rps16	rps17	rps18	rps2	rps20	rps23	rps26	rps27	rps3	rps4	rps5	rps6	rps8	s15a	s15p	sap40	sra	srp54	srs	suca
Acanthamoeba castellani	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Alexandrium tamarense	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	0
Amphimedon queenslandica	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Andalucia incarcerata	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1
Arabidopsis thaliana	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1
Carpediemonas membranifera	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Bigelowiella natans	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	1	1	1	1	0	0	0	1
Blastocystis hominis	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Blastocladiella emersonii	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	0	0	0	0	0	0	1	1	0	1	1	0	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0	0	0
Chilomastix caulleriy	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Chlamydomonas reinhardtii	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Chondrus crispus	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	1	1	1	1	0	1	1	0	1	0	1	1	0	0
Ergobibamus cyprinoides	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cryptosporidium sp.	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	0

Taxon	rpl9	rpo-A	rpo-B	rpo-C	rppO	rps10	rps11	rps12	rps14	rps15	rps16	rps17	rps18	rps2	rps20	rps23	rps26	rps27	rps3	rps4	rps5	rps6	rps8	s15a	s15p	sap40	sra	srp54	srs	suca
Cyanidioschizon merolae	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	1	0	0	1	1	1	1	1
Dictyostelium discoideum	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Euglena gracilis	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
Giardia intestinalis	1	0	0	1	1	0	1	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0
Glaucocystis nostia	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0
Guillardia theta	1	1	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1
Histiona aroides	0	0	0	0	1	0	1	0	1	1	1	1	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0	1	1	0
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isochrysis galbana	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
Jakoba bahamensis	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	1	0	1	1	1	0
Jakoba libera	0	0	0	0	1	0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	0	0	1	1	0	0	1	0
Karenia brevis	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0
Leishmania major	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Malawimonas californiensis	1	0	0	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1
Malawimonas jakobiformis	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1
Mastigamoeba balamuthi	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Micromonas sp.	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Monocercomonoides sp.	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Naegleria gruberi	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Nematostella vectensis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	1	0	0	1	1	1	1	0
Nuclearia simplex	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	0	0	1
Oryza sativa	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oxyrrhis marina	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	1	1	1
Oxytricha trifallax	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1
Paramecium caudatum	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1
Pavlova lutheri	1	0	0	0	0	1	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	1	0
Pentatrichomonas hominis	1	0	0	0	1	0	0	1	1	0	1	0	1	1	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	1
Peranema trichophorum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	1	1
Perkinsus marinus	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Phaeodactylum tricornutum	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1
Physarym polycephalum	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1

Taxon	rpl9	rpo-A	rpo-B	rpo-C	rppO	rps10	rps11	rps12	rps14	rps15	rps16	rps17	rps18	rps2	rps20	rps23	rps26	rps27	rps3	rps4	rps5	rps6	rps8	s15a	s15p	sap40	sra	srp54	srs	suca
Phytophthora infestans	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Phytomonas serpens	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
Plasmodium falciparum	1	0	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Polytomella parva	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	0	0	1
Porphyra yezoensis	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0
Proterospongia sp.	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0
Prymnesium parvum	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1
Reclinomonas americana	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Reticulomyxa filosa	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Rhizopus oryzae	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	0
Kipferlia bialata	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Sarcocystis sp.	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	1	1
Sawyeria marylandensis	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Spironucleus barkhanus	1	1	0	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Seculamonas ecuadoriensis	1	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1
Sphaerophorma arctica	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
Spironucleus salmonicida	1	0	0	0	1	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0
Spironucleus vortens	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Spizellomyces punctata	0	1	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	1	0	1
Stachyamoeba lipophora	1	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Telonema subtilis	1	0	0	0	0	1	0	1	1	1	1	0	1	1	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1
Tetrahymena termophila	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Thalassiosira pseudonana	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Thecamonas trahens	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Theileria parva	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1
Tsukubomonas globosa	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toxoplasma gondii	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Trichoplax adhaerens	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Trichomonas vaginalis	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Trimastix pyriformis	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Tritrichomonas foetus	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
Trypanosoma brucei	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ustilago maydis	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Taxon	tfiid	topo1	trs	ubc	vata	vatb	vatc	vate	wd	wrs	xpb
Acanthamoeba castellani	0	1	1	0	1	1	1	1	0	1	1
Alexandrium tamarense	0	0	0	1	1	0	0	1	0	1	0
Amphimedon queenslandica	0	1	0	1	1	1	1	1	1	1	1
Andalucia incarcerata	1	0	1	1	1	1	0	1	1	0	0
Arabidopsis thaliana	1	1	1	0	1	1	1	1	1	1	1
Carpediemonas membranifera	0	0	0	1	1	1	0	0	1	0	0
Bigelowiella natans	0	0	1	0	1	0	0	1	1	0	0
Blastocystis hominis	0	0	1	0	1	1	0	1	0	1	1
Blastocladiella emersonii	1	0	1	1	1	1	1	0	0	1	1
Capsaspora owczarzaki	1	1	1	1	1	1	1	1	1	1	1
Cercomonas longicauda	0	1	0	1	0	0	1	0	0	0	0
Chilomastix caulleriy	0	1	1	1	1	1	0	0	0	1	0
Chlamydomonas reinhardtii	1	0	1	1	1	1	1	1	0	1	0
Chondrus crispus	0	1	0	1	0	0	0	0	0	0	0
Ergobibamus cyprinoides	1	1	1	1	1	1	1	1	1	1	1
Cryptosporidium sp.	1	1	1	1	1	1	1	1	0	1	1
Cyanidioschizon merolae	1	1	1	1	1	1	1	1	0	0	1
Dictyostelium discoideum	1	1	1	1	1	1	1	1	1	1	1
Diplonema papilatum	0	0	0	1	0	0	1	0	0	0	0
Drosophila melanogaster	1	1	1	1	1	1	1	1	1	1	1
Emilliania huxleyi	0	1	1	1	1	1	1	0	1	1	1
Euglena gracilis	0	1	1	1	1	1	1	1	1	0	0
Giardia intestinalis	0	0	1	1	1	1	0	0	1	1	1
Glaucocystis nostia	0	1	1	1	0	1	1	1	0	0	0
Guillardia theta	1	0	0	1	1	1	0	1	0	0	1
Histiona aroides	0	1	0	0	1	1	1	1	0	1	0
Homo sapiens	1	1	1	1	1	1	1	1	1	1	1
Isochrysis galbana	1	0	0	0	1	1	1	0	1	0	0
Jakoba bahamensis	0	0	1	1	0	0	1	1	1	0	0
Jakoba libera	0	0	1	1	1	1	1	1	1	1	0
Karenia brevis	0	0	1	1	1	1	0	1	0	1	1
Leishmania major	0	1	1	1	1	1	1	1	0	1	1
Malawimonas californiensis	0	0	1	1	0	1	0	1	0	1	0
Malawimonas jakobiformis	1	1	0	1	1	1	1	1	1	0	0

Taxon	tfiid	topo1	trs	ubc	vata	vatb	vatc	vate	wd	wrs	xpb
Mastigamoeba balamuthi	0	0	1	1	1	1	1	1	1	1	1
Micromonas sp.	0	0	1	0	0	1	0	0	0	0	0
Monocercomonoides sp.	1	1	1	1	1	1	1	1	0	1	0
Monosiga brevicollis	1	1	1	1	1	1	1	1	1	1	1
Naegleria gruberi	1	1	1	1	1	1	1	1	1	1	1
Nematostella vectensis	1	1	1	1	1	1	1	1	1	1	1
Neurospora crassa	1	1	1	0	1	1	1	1	0	1	1
Nuclearia simplex	0	0	0	1	0	1	0	1	1	0	0
Oryza sativa	1	1	1	1	1	1	1	1	1	1	1
Oxyrrhis marina	0	0	0	1	1	1	1	1	0	0	0
Oxytricha trifallax	0	0	0	1	0	1	0	0	0	0	0
Paramecium caudatum	1	1	1	0	1	1	1	1	0	1	1
Pavlova lutheri	0	1	0	1	1	1	1	1	1	1	0
Pentatrichomonas hominis	0	0	1	0	1	0	0	1	0	0	0
Peranema trichophorum	0	0	1	1	1	1	0	0	0	0	1
Perkinsus marinus	1	1	1	1	1	1	1	1	1	1	1
Phaeodactylum tricornutum	1	1	1	1	1	1	1	1	1	1	1
Physarym polycephalum	0	0	0	0	1	1	1	0	0	1	0
Phytophthora infestans	1	1	1	1	1	1	1	1	0	1	1
Phytomonas serpens	0	0	0	0	0	0	0	0	0	0	0
Plasmodium falciparum	0	1	1	1	1	1	1	0	1	0	1
Polytomella parva	0	0	0	1	1	1	1	0	0	0	0
Porphyra yezoensis	0	1	1	1	1	1	1	1	0	1	1
Proterospongia sp.	0	0	0	0	1	0	0	0	0	0	0
Prymnesium parvum	0	0	0	1	1	1	1	1	1	0	0
Reclinomonas americana	0	0	1	1	1	0	0	1	1	1	0
Reticulomyxa filosa	0	0	0	1	1	1	0	0	1	0	0
Rhizopus oryzae	1	0	0	1	1	1	1	1	0	0	0
Kipferlia bialata	1	1	1	1	1	1	0	1	0	1	0
Sarcocystis sp.	1	0	1	1	1	0	0	0	0	0	0
Sawyeria marylandensis	0	0	1	1	1	1	0	0	0	1	0
Spironucleus barkhanus	0	0	1	1	1	1	0	0	0	1	0
Seculamonas ecuadoriensis	1	1	1	1	1	1	1	0	0	0	0
Sphaerophorma arctica	0	0	0	1	1	1	1	1	0	0	0

Taxon	tfiid	topo1	trs	ubc	vata	vatb	vatc	vate	wd	wrs	xpb
Spironucleus salmonicida	0	0	0	0	1	1	0	0	0	1	0
Spironucleus vortens	0	0	1	1	1	1	0	0	1	1	1
Spizellomyces punctata	0	0	0	1	0	0	0	0	0	0	1
Stachyamoeba lipophora	0	0	1	0	1	1	1	0	0	0	0
Telonema subtilis	0	0	1	1	1	1	0	0	0	0	0
Tetrahymena termophila	1	1	1	1	1	1	1	1	1	1	1
Thalassiosira pseudonana	1	1	1	1	1	1	1	1	0	1	1
Thecamonas trahens	1	1	1	1	1	1	1	1	1	1	1
Theileria parva	1	1	1	1	1	1	1	1	1	1	1
Tsukubomonas globosa	1	0	1	1	1	1	1	1	1	1	0
Toxoplasma gondii	1	0	1	0	1	1	0	1	0	1	1
Trichoplax adhaerens	0	1	0	1	1	1	1	0	0	0	0
Trichomonas vaginalis	1	0	1	1	1	1	1	1	1	1	1
Trimastix pyriformis	1	0	1	1	1	1	1	1	0	0	0
Tritrichomonas foetus	0	0	1	1	1	1	0	1	0	0	1
Trypanosoma brucei	1	1	1	1	1	1	1	1	0	1	1
Ustilago maydis	1	1	1	1	1	1	1	1	0	1	1

Appendix E. 9. Table of data sources, that were added or modified compared to (Hampl et al. 2009).

Organism	Source of Data	Comment
Amphimedon queenslandica	NCBI	
Blastocladiella emerosnii	NCBI	
Capsaspora owczarzaki	Broad institute	Origins of Multicellularity Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)
Cyanidioschizon merolae	NCBI	
Diplonema papilatum	TBestDB	
Emilliania huxleyi	JGI	These sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community
Guallardia theta	TBestDB	
Karenia brevis	NCBI	
Leishmania major	TriTrypDB	
Monosiga brevicollis	JGI	King *et al*. 2008. *Nature* **451:** 783-788
Naegleria gruberi	JGI	Jillian *et al.* 2010. *Cell* **140:** 631-642
Nematostella vectensis	JGI	Putnam *et al.* 2007. *Science* **317:** 86-94
Neurospora crassa	NCBI	
Pentatrichomonas hominis	NCBI	
Perkinsus marinus	NCBI	
Phaeodactylum tricornutum	NCBI	
Phytomonas serpens	NCBI	
Phytophthora infestans	Broad institute	"*Phytophthora infestans* Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)
Plasmodium falciparum	PlasmoDB	
Polytomella parva	NCBI	
Porphyra yezoensis	NCBI	
Prymnesium parvum	NCBI	
Reticulomyxa filosa	NCBI	
Rhizopus orzyae	NCBI	
Sphaeroforma arctica	TBestDB	
Telonema subtilis	NCBI	
Tetrahymena thermophila	Tetrahymena genome database	
Thecamonas trahens	Broad institute	Origins of Multicellularity Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)
Trichoplax adhaerens	NCBI	
Tritrichomonas foetus	NCBI	
Trypanosoma brucei	TriTrypDB	

Appendix F

Supplementary Materials for Chapter 5

Appendix F.1. Table showing detailed information about targeting predictions. 'n/a' signs that protein was not found in paerticular ogranims. 'n' signs non-mitochondrial prediction and numbers sign the probability of the particular protein to be targeted to mitochodnria. 'no/yes' sign the absence/presence of targeting peptides.

Organism	***Carpediemonas***					***Ergobibamus***					***Kipferlia***					***Chilomastix***				
Prediction software	**Mitoprot**	**Mitopred**	**Predotar**	**TargetP**	**Targetin pep.**	**Mitoprot**	**Mitopred**	**Predotar**	**TargetP**	**Targetin pep.**	**Mitoprot**	**Mitopred**	**Predotar**	**TargetP**	**Targetin pep.**	**Mitoprot**	**Mitopred**	**Predotar**	**TargetP**	**Targetin pep.**
Energy metabolism																				
Pyruvate:ferredoxin oxydoreductase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no
[FeFe- hydrogenase	n	n	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Malic enzyme	n	n	n	n	no	n	n	n	n	no	n	99	n	n	no	n	n	n	n	no
NADH dehydrogenase 51Kd	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n	85	n	n	no	n/a	n/a	n/a	n/a	n/a
NADH dehydrogenase 24 Kd	n	85	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Adenylate kinase	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no
Succinyl-CoA:acetate CoA transferase	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
Succinate CoA synthetase alpha	n	84.6	n	n	no	n	99	n	0.57	no	0.98	85	n	0.6	yes	n/a	n/a	n/a	n/a	n/a
Succinate CoA synthetase beta	n	99	n	n	yes	n	99	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
Acetyl-CoA synthetase	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no
Hydrogenase maturase protein F	n/a	n/a	n/a	n/a	n/a	n	92	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
Hydrogenase maturase protein E	n	n	n	n	n	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hydrogenase maturase protein G	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Appendix F.1. continued

Amino acid metabolism																				
Serine hydroxymethyltransferase	0.74	n	n	0.57	yes	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
T-protein	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
P-protein	n/a	n/a	n/a	n/a	n/a	0.84	n	0.7	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
L-protein	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
H-protein	n	n	n	n	no	0.92	85	0.45	0.56	yes	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
L-protein (Spirochaete version)	n	n	n	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Alanine transaminase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n	n	n	n	no	0.79	n	0.43	n	no
Branched chain amino acid aminotransferase	n/a	n/a	n/a	n/a	n/a	n	92	0.83	n	no	n	n	n	n	no	n	n	n	n	no
Phosphatidylserine decarboxylase	n/a	n/a	n/a	n/a	n/a	n	92	0.59	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Aspartate aminotransferase mitochondrial	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n	n	n	n	no	n	99	n	n	no
Glutamate dehydrogenase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n	99	n	n	no	n	99	n	n	no
Ornithine aminotransferase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Ornithine carbamoyltransferase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n	n	n	n	no	0.62	n	n	n	no

Appendix F.1. continued

Iron-Sulfur cluster synthesis																				
Scaffold protein IscU	0.90	99	0.76	0.3	yes	0.85	92	0.8	0.87	yes	0.92	99	0.7	0.58	yes	0.93	99	0.83	0.5	yes
Cystein desulfurase IscS	n	n	n	n	n	0.59	85	0.63	n	no	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no
Ferredoxin	0.83	n	0.32	n	yes	n/a	n/a	n/a	n/a	n/a	0.93	n	0.57	0.2	yes	n	n	n	n	no
Isa1/Isa2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
ISD11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Frataxin	n	92	0.52	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Glutaredoxin	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no
DNAJ Jac1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Iron-sulfur cluster co-chaperone HscB	n/a	n/a	n/a	n/a	n/a	0.87	n	n	0.79	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Oxygen scavenging system																				
Hydroperoxide reductase ruberythrin	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Thioredoxin	0.55	n	n	n	yes	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Thioredoxin peroxidase	n	n	n	n	no	n	92	0.61	0.51	no	n	n	n	n	no	n	n	n	n	no
Thioredoxin reductase	n/a	n/a	n/a	n/a	n/a	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n	99	0.56	n	no
Super oxide dismutase	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no	n	n	n	n	no

Appendix F.1. continued

Import machinery																				
Inner membrane translocase 23 (Tim23)	0.59	85	0.44	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Outer membrane translocase 40 (Tom40)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Inner membrane translocase Pam18	0.55	99	0.61	0.63	yes	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.82	92	0.43	0.25	yes
Mdj11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Mge	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Sam50	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hydrogenosomal integral membrane prote	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hydrogenosomal integral membrane prote	0.75	21	n	n	no	0.52	92	n	n	no	n/a	n/a	n/a	n/a	n/a	0.75	n	0.92	n	no
Mitochondrial carrier protein	0.67	99	0.42	n	no	n	85	n	n	no	n	99	n	n	no	n/a	n/a	n/a	n/a	n/a
Mitochondrial folate transporter	0.64	85	0.42	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Mitochondrial processing peptidase beta-s	0.74	92	0.35	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Heat shock protein 70	n	99	0.78	n	no	n	n	0.92	n	no	n	n	n	n	no	n/a	n/a	n/a	n/a	n/a
Heat shock protein CPN10	0.86	n	0.56	0.21	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Heat shock protein CPN60	n/a	n/a	n/a	n/a	n/a	0.77	n	0.99	n	no	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Fatty acid metabolism																				
Long chain fatty acid CoA ligase	n	n	n	n	no	0.79	99	n	n	no	n	n	n	n	no	n	n	n	n	no

14. Appendix G

Article describing *Hicanonectes teleskopos*

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Sep 07, 2011

This is a License Agreement between Martin Kolisko ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2743730728489
License date	Sep 07, 2011
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Eukaryotic Microbiology
Licensed content title	Light Microscopic Observations, Ultrastructure, and Molecular Phylogeny of Hicanonectes teleskopos n. g., n. sp., a Deep-Branching Relative of Diplomonads
Licensed content author	JONG SOO PARK,MARTIN KOLISKO,AARON A. HEISS,ALASTAIR G.B. SIMPSON
Licensed content date	Jul 1, 2009
Start page	373
End page	384
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Total	0.00 USD

J. Eukaryot. Microbiol., 56(4), 2009 pp. 373–384

DOI: 10.1111/j.1550-7408.2009.00412.x

Light Microscopic Observations, Ultrastructure, and Molecular Phylogeny of *Hicanonectes teleskopos* n. g., n. sp., a Deep-Branching Relative of Diplomonads

JONG SOO PARK, MARTIN KOLISKO, AARON A. HEISS and ALASTAIR G.B. SIMPSON

Canadian Institute for Advanced Research, Program in Integrated Microbial Diversity, and Department of Biology, Dalhousie University, Halifax, B3H 4J1 Canada

ABSTRACT. We describe *Hicanonectes teleskopos* n. g., n. sp., a heterotrophic flagellate isolated from low-oxygen marine sediment. *Hicanonectes teleskopos* has a ventral groove and two unequal flagella, and rapidly rotates during swimming. At the ultrastructural level *H. teleskopos* is a ''typical excavate'': it displays flagellar vanes, a split right microtubular root, ''I,'' ''B,'' and ''C'' fibres, a singlet microtubular root, and a possible composite fibre. Small subunit rRNA (SSU rRNA) gene phylogenies and an ''arched'' B fibre demonstrate that *H. teleskopos* belongs to Fornicata (i.e. diplomonads, retortamonads, and relatives). It forms a clade with the deep-branching fornicate *Carpediemonas*, with moderate-to-strong bootstrap support, although their SSU rRNA gene sequences are quite dissimilar. *Hicanonectes* differs from *Carpediemonas* in cell shape, swimming behaviour, number of basal bodies (i.e. 4 vs. 3), number of flagellar vanes (i.e. 2 vs. 3), anterior root organization, and by having a cytopharynx. Like *Carpediemonas* and *Dysnectes*, *Hicanonectes* has conspicuous mitochondrion-like organelles that lack cristae and superficially resemble the hydrogenosomes of parabasalids, rather than the mitosomes of their closer relatives the diplomonads (e.g. *Giardia*).

Key Words. Anaerobe, basal eukaryote, excavate, *Giardia*, hydrogenosome, metamonad, microaerophile, protist, Protozoa.

FORNICATA is a recently established taxon within Excavata that houses diplomonads, retortamonads, *Carpediemonas*, and the newly described *Dysnectes* (Simpson 2003; Yubuki et al. 2007). Diplomonads are by far the best known of these groups, and include free-living, commensal, and parasitic species, for example, the well-studied human parasite *Giardia intestinalis*, also known as *Giardia lamblia* (Kulda and Nohýnková 1978). With one exception retortamonads are commensals or parasites, but have not been definitively connected to any human or livestock diseases (Kulda and Nohýnková 1978). *Carpediemonas* and *Dysnectes* are small, free-living, and slowly swimming biflagellated cells that inhabit oxygen-poor marine sediments (Ekebom, Patterson, and Vørs 1996; Lee and Patterson 2000; Simpson and Patterson 1999; Yubuki et al. 2007). There are just two nominal species of *Carpediemonas*, while *Dysnectes* is monospecific. Retortamonads, *Carpediemonas*, and *Dysnectes* have a ''typical excavate'' morphology, meaning that they possess a longitudinal feeding groove supported by a particular organization of cytoskeletal elements, and associated with a vane-bearing posterior flagellum (Simpson 2003; Yubuki et al. 2007). These three groups also display a unique organization of one cytoskeletal element, the B fibre/arched fibre, and this organization is the defining synapomorphy of Fornicata (Simpson 2003; Yubuki et al. 2007). The monophyly of Fornicata is well supported by small subunit ribosomal RNA (SSU rRNA) gene phylogenies (Keeling and Brugerolle 2006; Kolisko et al. 2008; Simpson 2003; Simpson et al. 2002b; Yubuki et al. 2007), and a *Carpediemonas*+diplomonads clade is recovered in protein phylogenies, in the absence of data from retortamonads and *Dysnectes* (Simpson, Inagaki, and Roger 2006; Simpson, MacQuarrie, and Roger 2002a; Simpson et al. 2002b). Small subunit rRNA gene trees indicate that diplomonads and retortamonads are specifically related, although a recent study recovers retortamonads as a paraphyletic group (Cepicka et al. 2008). *Carpediemonas* and *Dysnectes* are recovered as successive basal branches within Fornicata (Yubuki et al. 2007).

Fornicata is of great importance to researchers interested in eukaryotic cell evolution, for two main reasons: Firstly, molecular phylogenies that include diplomonads have tended to place them at or near the base of the eukaryotic tree (Ciccarelli et al. 2006; Hashimoto et al. 1994, 1995; Sogin et al. 1989); consequently diplomonads have often been considered to be relatively underived representatives of the first eukaryotic cell (e.g. Cavalier-Smith 1995). Secondly, these organisms do not possess classical mitochondria and were for a long time considered to be amitochondrial. However, genes of mitochondrial origin have been discovered in diplomonad nuclear genomes (Horner and Embley 2001; Roger et al. 1998; Tachezy, Sánchez, and Müller 2001) and extremely small ($\sim$ 50 nm across) remnant mitochondrial organelles, called mitosomes, have been identified in *G. intestinalis* (Tovar et al. 2003). The only known function of the *Giardia* mitosomes is iron–sulfur cluster assembly. In contrast to *Giardia*, *Carpediemonas membranifera* and *Dysnectes brevis* both possess rather large double membrane-bounded organelles with no cristae (Simpson and Patterson 1999; Yubuki et al. 2007). The larger biovolume of these organelles hints that they might be involved in more metabolic processes than the mitosomes of *Giardia*.

From a sample of low-oxygen marine sediment we encountered and cultured a biflagellated protist with a relatively inconspicuous longitudinal groove, which rotated during swimming and did not closely resemble any described organism that we were aware of. Surprisingly, ultrastructural study indicated that this organism is a ''typical excavate,'' and that it bears conspicuous mitochondrion-like organelles that lack cristae. Phylogenetic analyses of its SSU rRNA genes, and the presence of the proposed structural synapomorphy, demonstrate that the organism is a member of Fornicata. It is sufficiently distinct in ultrastructure and at the sequence level to merit description as a new taxon, *Hicanonectes teleskopos* n. g., n. sp.

MATERIALS AND METHODS

Isolation and culture. *Hicanonectes teleskopos* was isolated from anoxic marine intertidal sediments from a sheltered embayment on Salt Spring Island, BC, Canada (48°46′N, 123°28′W; sampled October 2007). Crude cultures were established by dispersing a sediment sample in TYSGM media (Diamond 1982), prepared without bovine serum, diluted 1:1 with sterile seawater, and supplemented with 30 ml horse serum/L. A xenic monoeukaryotic culture was established through one round of single cell isolation and several rapid transfers of the culture to fresh media. Cultures were maintained in this medium in 15 ml polypropylene tubes with 10–12 ml of media per tube. Cultures were grown at 21 °C and subculturing was performed every 4 d.

Light microscopy. Live *H. teleskopos* cells were observed with phase contrast microscopy using a Zeiss Axiovert 200 M mi-

Corresponding Author: A.G.B. Simpson, Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia, Canada B3H 4J1—Telephone number: 902 494 1247; FAX number: 902 494 3736; e-mail: alastair.simpson@dal.ca

croscope equipped with an Axiocam HR digital camera. Lengths and widths of live *H. teleskopos* cells ($n = 17$) were determined from digital micrographs using the camera software (Axiovision 4.6).

Electron microscopy. For freeze substitution, approximately 1.5 ml of cell culture was pelleted at 1,000 *g* for 30 min in a microcentrifuge tube. The supernatant was aspirated and material from the pellet loaded into hexadecene-coated 200 μm-deep gold-plated planchettes for high-pressure freezing using liquid nitrogen. The frozen specimens were transferred, under liquid nitrogen, to sample tubes containing an anhydrous solution of 2.0% (w/v) OsO_4 and 0.1% (v/v) glutaraldehyde in HPLC-grade acetone. The temperature was raised from $-160\,^{\circ}C$ at a rate of $1\,^{\circ}C$/h, holding the temperature steady for ~ 24 h at $-90\,^{\circ}C$ and for ~ 12 h at $-60\,^{\circ}C$. During the $-60\,^{\circ}C$ period, the specimens were removed from the fixation cocktail and placed in pure acetone. Subsequently, the temperature was raised by $2\,^{\circ}C$/h, with a 24-h holding period at $-30\,^{\circ}C$, until the specimens were at $-20\,^{\circ}C$. They were then transferred to $-20\,^{\circ}C$ for ~ 24 h, and then to $4\,^{\circ}C$ for ~ 12 h, before being allowed to warm to room temperature. Specimens were suspended in pure acetone and agitated to remove them from the planchettes; all subsequent handling was done using 1,000 μl micropipettes with cut-off tips. Specimens were transferred through a series of Spurr's resin mixtures (one-third resin, two-thirds resin, and three changes of full resin), before final embedding.

For conventional chemical fixation, cells were centrifuged at 8,000 *g* for 3 min and fixed for 30 min at room temperature in a cocktail containing 1% (v/v) glutaraldehyde, 5% (w/v) sucrose, and 0.1 M cacodylate buffer (pH 7.4). After rinsing the cells 3 times in 0.1 M cacodylate buffer with 5% (w/v) sucrose, cells were post-fixed for 1 h in 0.8% (w/v) OsO_4 and 5% (w/v) sucrose in 0.1 M cacodylate. After being rinsed free of post-fixative, cells were concentrated by centrifugation and trapped in 1.5% (w/v) agarose. Agarose blocks were dehydrated by applying a graded series of ethanols, and then embedded in Spurr's resin.

Serial sections (~ 70 nm) were cut with a diamond knife on a Leica UC6 ultramicrotome (Leica, Knowlhill, Wetzlar, Germany) and were subsequently stained with saturated uranyl acetate in 50% ethanol and with lead citrate. Sections were observed using a Tecnai 12 transmission electron microscope (FEI, Hillsboro, OR, Philips) fitted with a goniometer stage.

SSU rRNA gene sequencing. DNA was isolated using a hexadecyltrimethyl ammonium bromide (CTAB) protocol (Clark 1992). The SSU rRNA gene was amplified using universal eukaryotic primers (Medlin et al. 1988). A fragment of expected size was then purified from the gel and subcloned into pCR4 TOPO vector (Invitrogen, Carlsbad, CA). Four clones were partially sequenced and identity of the sequences was confirmed using BLAST (Altschul et al. 1990). Subsequently, two identical clones were fully bidirectionally sequenced by primer walking. This sequence has been deposited in GenBank as Accession number FJ628363.

Phylogenetic analyses. The master alignment used for this study was constructed using CLUSTALX 1.83 (Thompson et al. 1997). Some sequences were realigned to the main alignment using the program MAFFT (Katoh et al. 2005) with the EINSII algorithm. The alignment was then edited manually in BioEdit 7.0.5.3 (Hall 1999). Ambiguously aligned regions were excluded from analysis; the final trimmed dataset was 827-bp long. The alignment is available upon request.

The main dataset included a wide taxon sampling of diplomonads, representatives of the retortamonad genera *Retortamonas* and *Chilomastix*, plus *C. membranifera*, *D. brevis*, and some uncultured eukaryote sequences similar to Fornicata, including CPS-GM-5 and its relatives (see Takishita et al. 2007). A variety of other eukaryotes was used as an outgroup, including representatives of all other major groups of excavates. We also analysed a second dataset, identical to the first except that the genus *Chilomastix* was excluded. The reason for this was the placement of *Chilomastix* in SSU rRNA gene trees: *Chilomastix* SSU rRNA gene sequences do not form a clade with those from *Retortamonas*, the other member of Retortamonadida, but instead fall at the base of the diplomonad+retortamonad clade (Cepicka et al. 2008). While this pattern could represent the true phylogeny it is also consistent with the *Chilomastix* lineage being more rapidly evolving than *Retortamonas*, and erroneously placed towards the base of the tree as a long-branch attraction artifact. The exact placement of *Chilomastix* within the basal part of the tree is also relatively unstable (Cepicka et al. 2008), which is consistent with this possibility.

Phylogenetic trees were reconstructed using maximum likelihood (ML) and Bayesian methods. For both datasets, the model of sequence evolution (GTR+Γ+I) was selected using the Akaike information criterion, as implemented in the program Modeltest (Posada and Crandall 1998). The ML trees were estimated using PAUP* 4b10 with 10 random taxon additions followed by tree-bisection-reconnection (Swofford 2003), and subsequently bootstrapped with 500 replicates using the program IQPNNI 3.2 (Vinh and von Haeseler 2004). The Bayesian analysis was carried out in MrBayes 3.2 (Huelsenbeck 2000) with two independent runs, each with four independent chains running for 1.4×10^7 generations (a burn-in of 2.5×10^5 generations was used), with default heating parameter (0.2) and sampling frequency (0.01).

RESULTS

Light microscopy. Cells are biflagellated and broadly oval shaped with a longitudinal groove (Fig. 1–4). The groove is difficult to see while the cell is swimming. Cells are 6.5–10.0 μm long (average ± SD 8.5 ± 0.9 μm, $n = 17$) and 4.5–8.0 μm wide (5.8 ± 0.9 μm). The posterior flagellum is 2.5–3.5 times the length of the cell body. It runs through the groove and then trails freely behind the cell (Fig. 1, 2). The second flagellum is directed leftwards and slightly anteriorly, and is approximately 1–1.5 times the cell length (Fig. 3). In phase contrast microscopy the nucleus can be observed in the anterior end of the cell (Fig. 1). Most cells contain vacuoles with prokaryotic contents (Fig. 1, 4). The groove runs along the whole length of the cell body. In the anterior third of the cell the right wall of the groove is sharply defined, and contains a thick cytoskeletal element (presumably, the right root [RR] and B fibre—see ''ultrastructure'' below) that can be observed readily using phase contrast (Fig. 1, 2). The margin then undergoes a bend, causing the groove gradually to narrow posteriorly, and the cytoskeletal element becomes less conspicuous. At the posterior end of the cell, the groove deflects to the left where it continues into the cell as a narrow, difficult-to-observe cytopharynx (Fig. 1, 2). When moving, the cell swims in more or less straight lines, with a rapid rotation about its longitudinal axis.

Ultrastructure. Terms for ultrastructural components are as used by Simpson and Patterson (1999) and Yubuki et al. (2007), except that we identified the flagella and basal bodies numerically, as per O'Kelly (1993) and Simpson (2003). The general appearance of the cells under light microscopy was also observed by transmission electron microscopy. The nucleus and flagellar apparatus are in the anterior part of the cell, with the flagellar apparatus located subapically (Fig. 5, 8). Posterior to the flagellar apparatus is the ventral groove. The edges of the groove are supported by microtubules originating from basal body 1 (Fig. 6–8). The most sharply defined part of the groove is immediately next to its right margin (Fig. 8). We observed one encysted cell in our chemical fixation (Fig. 9). This was approximately rounded with

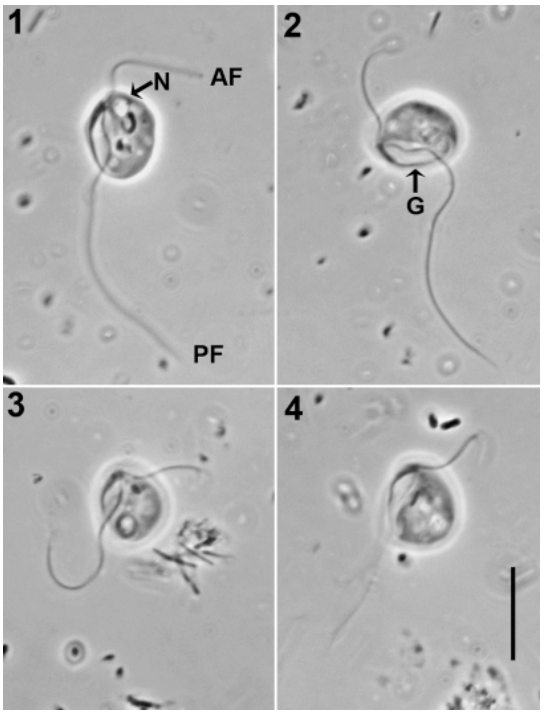


Fig. **1–4.** Phase contrast light micrographs of living cells of *Hicanonectes teleskopos* n. g., n. sp. Note the delicate curving cytopharynx at the posterior end of the groove (**1**), and the flagellar insertion and flagellar lengths (**2, 3**). AF, anterior flagellum; G, groove; N, nucleus; PF, posterior flagellum. Scale bar = 10 μm.

internalized flagellar axonemes and a loose-fitting, relatively thin cyst wall without scales or fibrous materials.

In trophic cells the nucleus is very closely associated with the basal bodies. It is rounded and lacks a central nucleolus; instead, densely staining material is seen concentrated at one end (Fig. 5). No Golgi apparatus was observed. Rounded mitochondrion-like organelles up to 300 nm in diameter were observed throughout the cells (Fig. 5, 8). These lack cristae. However, the organelles were not well preserved, and we could not visualize clearly the bounding membranes (Fig. 10, 11). Some mitochondrion-like organelles appeared to be undergoing division (Fig. 11). Faint parallel striations were sometimes observed in grazing sections of the surface of these organelles (Fig. 12). Food vacuoles with prokaryotic contents were seen in the middle and posterior of the cell (Fig. 5, 8). In some cells we observed small prokaryotes coated with numerous filaments ($\sim$ 300 nm diameter, excepting the filaments), which were free within the cytoplasm and might be endosymbionts (Fig. 13). These prokaryotes were not associated closely with the mitochondrion-like organelles.

Flagellum 1 (i.e. the posterior flagellum) and flagellum 2 (i.e. the anterior flagellum) each have a normal 9+2 axoneme (Fig. 14, 29). The transition between the 9+2 structure and the basal body typically occurs within the cytoplasm, 200–300 nm below the level of insertion (Fig. 17, 19, 20). Flagellum 1 has two vanes, each supported by a fine paraxonemal lamellum (Fig. 14). One vane is very conspicuous and broad (maximum breadth $\sim$ 600 nm), and is located on the ventral side of the axoneme, while the second vane is much smaller and is located on the dorsal side (Fig. 14, 15). The ventral vane originates near the flagellar insertion (Fig. 15). The dorsal vane originates more distally. The outer edge of the ventral vane has a striated appearance in grazing section, with the striations arranged perpendicular to the axoneme (Fig. 16). There are normally four basal bodies in the flagellar apparatus (Fig. 17, 39), each with a normal triplet structure (Fig. 18, 25, 33). Basal bodies 1 and 2 give rise to flagellum 1 and 2, respectively, and are about 330 nm in length (Fig. 17, 19, 20). Basal bodies 3 and 4 are normally non-flagellated, and they are sometimes shorter than the flagellated basal bodies. All basal bodies have a short ($\sim$ 70 nm) cartwheel structure (e.g. Fig. 33). The flagellated basal bodies 1 and 2 are arranged at a slightly obtuse angle (Fig. 19, 39) and are separated by $\sim$ 50 nm at their closest (Fig. 18). Basal body 2 is directed leftwards and somewhat anteriorly. Non-flagellated basal body 3 is situated close and nearly parallel to basal body 2, but is located more ventrally (Fig. 18, 39). Non-flagellated basal body 4 lies to the left of basal body 1 and is directed leftwards (Fig. 17, 39).

There is one microtubular root, the anterior root (AR), associated specifically with basal body 2. The AR originates adjacent to the right/anterior side of basal body 2, in association with fibrous and dense material on its interior side (Fig. 18, 21, 39). Additional fibrous material also connects basal body 2 to basal body 1 (Fig. 21). The AR extends to the left-dorsal side of basal body 2 together with the dense material (Fig. 21, 39). The number of microtubules in the AR gradually increases at a rate of roughly one per transverse section ($\sim$ 70 nm), up to nine microtubules (Fig. 22). There is a dorsal fan of individual microtubules associated with the AR (Fig. 19, 20, 39). These appear to originate alongside and parallel to the AR, and support the cell membrane outside the confines of the groove. Two clusters of individual internal microtubules (IMt), IMt1 and IMt2, originate from different microtubular organizing centres (MTOCs), which are closely associated with the proximal ends of basal bodies 3 and 4, respectively (Fig. 23, 27, 28, 39).

Basal body 1 is linked to two major microtubular roots, the left root (LR) and the right root (RR), as well as a singlet microtubular root, and some non-microtubular fibres (e.g. B, I, A, and C fibres; Fig. 24, 25, 39). The RR originates aligned to the right side of basal body 1 (Fig. 24–26, 39), although its proximal (anterior) end is also closely associated with basal body 2 and the proximal end of the AR (Fig. 19). It expands rapidly to include about 20 microtubules in a single curved row (Fig. 24). Slightly posterior to the opening of the groove, the RR splits into an inner portion of seven microtubules (IRR) and an outer portion of 13 microtubules (ORR, Fig. 29, 30, 39). The ORR curves around to form part of the support for the right wall of the groove, while the IRR remains associated with the floor of the groove (Fig. 29, 31, 39). More distal sections of the right wall of the groove show up to 24 microtubules, representing a combination of ORR-derived microtubules, IMt1, and possibly some dorsal fan microtubules (Fig. 32). The material that we identify as the A fibre is closely associated with the dorsal side of the RR, was seen only at the proximal end of the RR, and is quite indistinct (Fig. 25). The I fibre is closely associated with the ventral side of the RR (Fig. 24–27, 39). It is composed of a double-leaved sheet connected to the ventral face of the RR by a latticework structure, and has a total thickness of $\sim$ 40 nm (Fig. 24). Posterior to the split of the RR, the I fibre continues only with the outer portion of the ORR (Fig. 28–31). The B fibre material has two distinct elements. The first element has a laminate appearance in transverse section. This element originates from the right side of the LR, then arches across the ventral side of basal body 1 to become closely associated with the RR (Fig. 24, 26, 39). The second element appears as a dense amorphous sheet with a relatively lucent field associated with its

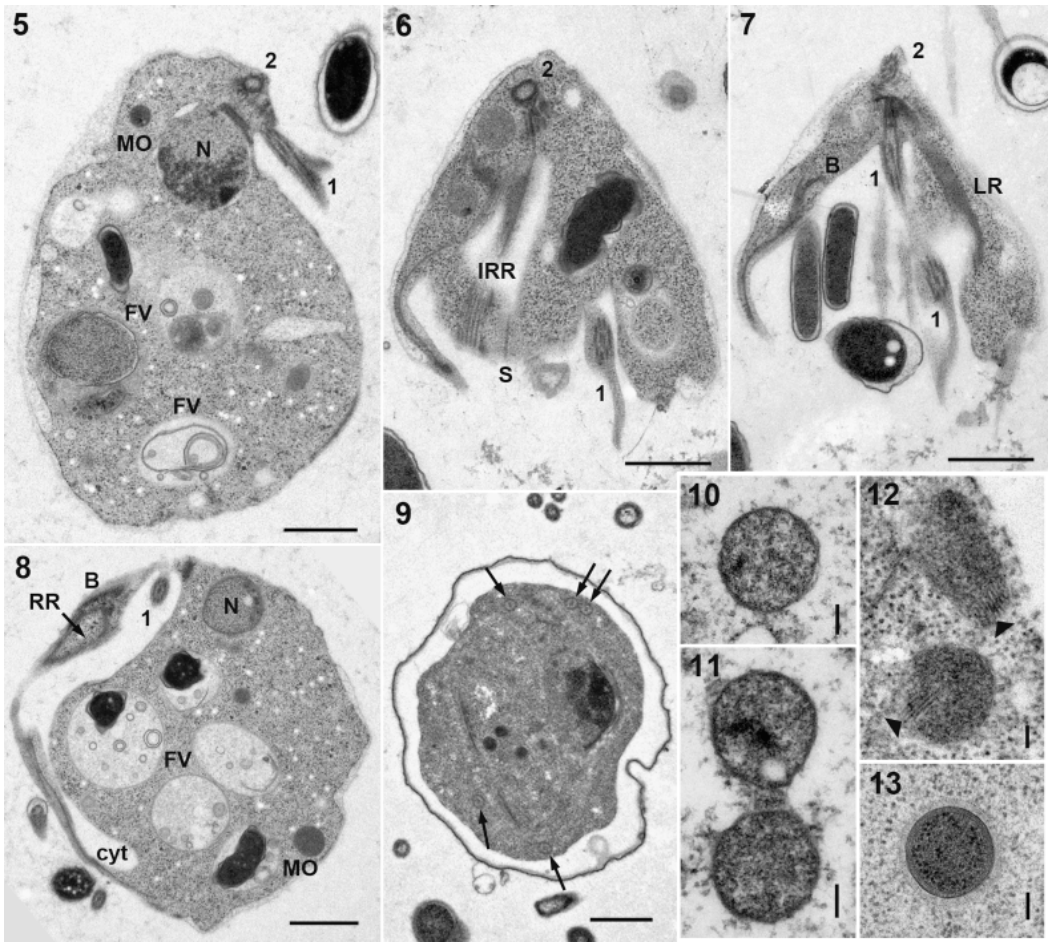


Fig. **5–13.** Transmission electron micrographs of *Hicanonectes teleskopos* n. g., n. sp., ultra-thin sections. **5–8.** Cells in longitudinal sections. 5. Lateral view; 6–8. Ventral view. **9.** Cyst showing flagellar axonemes (arrows) within the cytoplasm (chemical fixation). **10–12.** Mitochondrion-like organelles (10, 11: chemical fixation). Arrowheads in 12 indicate faint parallel striations in the surface of the organelles. **13.** Intracytoplasmic prokaryote surrounded by numerous filaments. 1, flagellum/basal body 1 (= posterior flagellum/basal body); 2, flagellum/basal body 2 (= anterior flagellum/basal body); B, B fibre, cyt: cytopharynx; FV, food vacuole; IRR, inner portion of right root; LR, left root; MO, mitochondrion-like organelle; N, nucleus; RR, right root; S, singlet microtubular root. Scale bars for 5–9 = 1 μm; for 10–13 = 100 nm.

ventral face (Fig. 28–31, 39). This second element projects orthogonally from the ventral face of the laminate element (Fig. 33, 39). The laminate element ends near the opening of the groove, while the dense element runs posteriorly immediately under the cell membrane of the right margin of the groove (Fig. 27–29, 31). Initially the dense element of the B fibre runs parallel to the RR and the I fibre (Fig. 27, 28). The subsequent curvature of the RR means that the RR and the I fibre end up arranged perpendicular to the B fibre (Fig. 31). The singlet microtubular root (S) originates close to basal body 1 and the dorsal side of the RR (Fig. 24, 26, 34, 39). It continues posteriorly, close to but separate from the left side of the RR (Fig. 27–29).

The LR is composed of a single row of 10 microtubules that are closely linked to the non-microtubular C fibre (Fig. 27–29, 39). The C fibre adheres to the dorsal side of the LR (Fig. 24, 25, 27, 39). It appears as one dense and one fine sheet in transverse section, and is about 40 nm thick (Fig. 27). The C fibre ends at the level of the opening of the groove. Slightly posterior to this, an additional microtubule originates close to the dorsal side of the LR, and runs parallel to the main LR microtubules (Fig. 28).

As the groove broadens, its right margin is supported primarily by the B fibre and the ORR (Fig. 7, 8, 29, 39). The outer side of this groove wall is also supported by IMt1 microtubules (Fig. 37). The left wall of the groove is supported by the LR (Fig. 7, 39). The right half of the floor of the groove is supported by the singlet root, the IRR, and a few microtubules that originate from the ORR (Fig. 6, 29). The left half of the groove is supported by sparse microtubules derived from the LR and IMt2. Farther down the groove, the I fibre and B fibre are reduced and then lost (Fig. 32). Halfway down the groove, there is a bend in the right wall (Fig. 6–8).

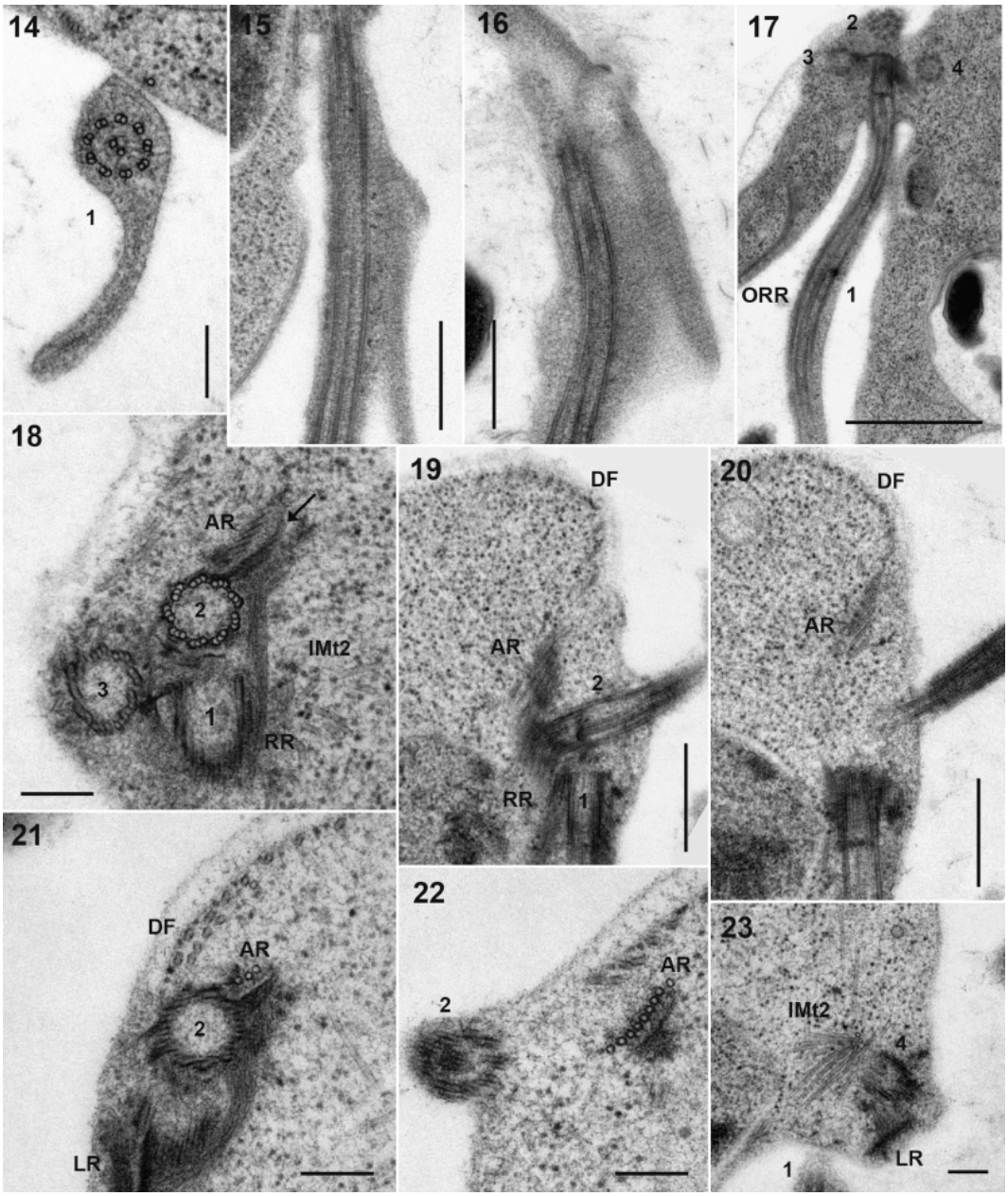


Fig. **14–23.** Transmission electron micrographs of *Hicanonectes teleskopos* n. g., n. sp., ultra-thin sections. **14.** Posterior flagellum in transverse section showing normal 9+2 axoneme structure and two flagellar vanes (ventral much larger than dorsal). **15, 16.** Posterior flagellum in longitudinal section, showing the origin of the ventral vane shortly after emergence (15), and the striated appearance of the outer portion of the ventral vane (16). **17.** Longitudinal section through the anterior portion of the cell, showing four basal bodies (i.e. 1–4). **18.** Section showing triplet structure of basal bodies 2 and 3. Note the anterior root (AR) and electron-dense material originating adjacent to basal body 2. **19, 20.** Near-consecutive sections showing the AR associated with the dorsal fan. Note the anterior end of the right root (RR) located close to basal body 2 and the AR. **21, 22.** Near-consecutive transverse sections of AR, showing origin of the root and number of microtubules, and electron-dense material around the basal bodies. **23.** IMt2 originating from an MTOC near basal body 4. 1, flagellum/basal body 1 (= posterior flagellum/basal body); 2, flagellum/basal body 2 (= anterior flagellum/basal body); 3, non-flagellated basal body 3; 4, non-flagellated basal body 4; DF, dorsal fan; IMt 2, radiation of internal microtubules; LR, left root; ORR, outer portion of right root; MTOC, microtubular organising centre. Scale bars for 14, 18, 21, 22, and 23 = 200 nm; for 15, 16, 19, and for 20 = 500 nm; for 17 = 1 μm.

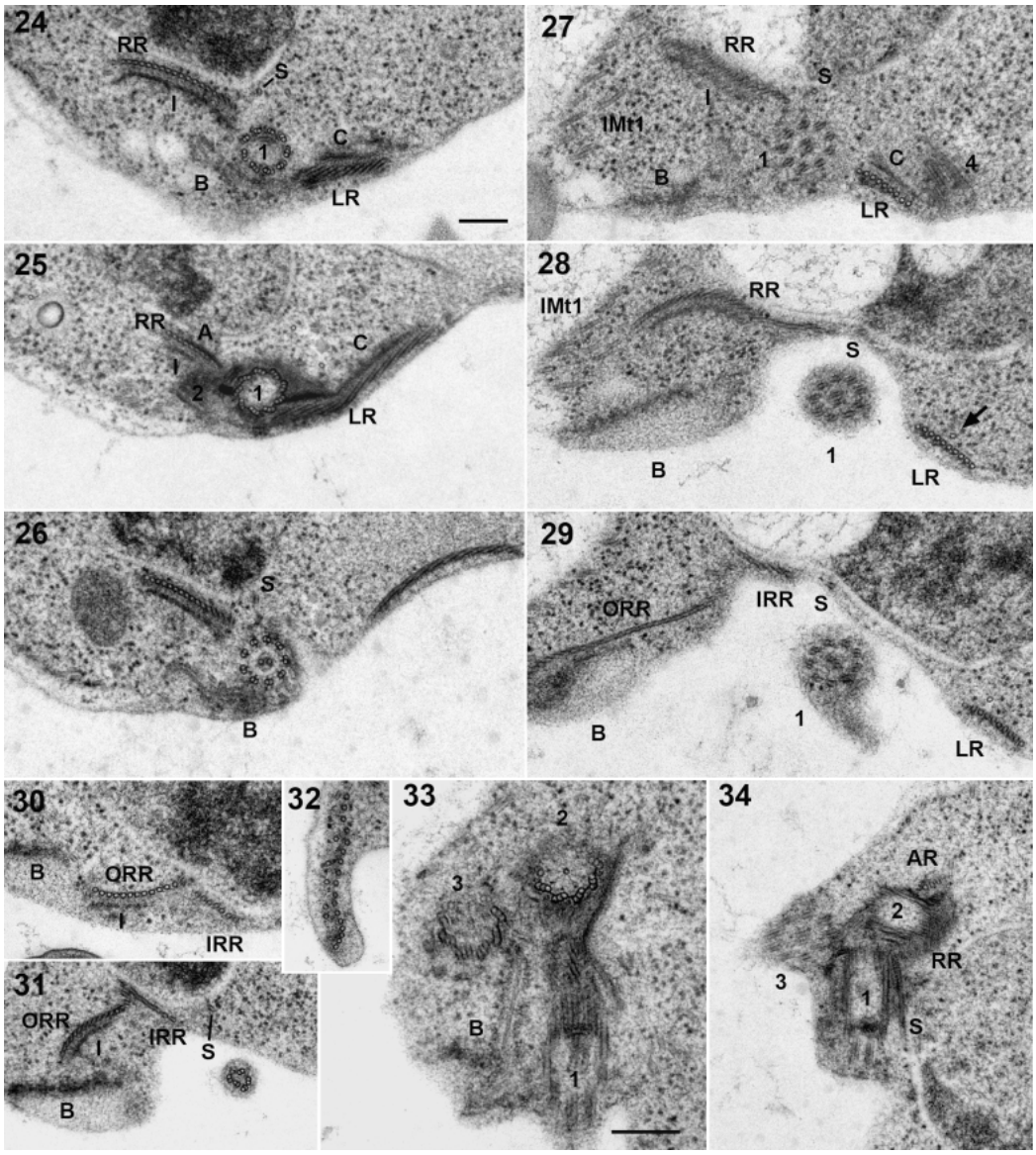


Fig. **24–34.** Transmission electron micrographs of *Hicanonectes teleskopos* n. g., n. sp., ultra-thin sections. **24.** Transverse section of basal body 1 showing RR and LR, singlet microtubule, and associated non-microtubular fibres. Note the origin of the lamellate element of the B fibre against the LR. **25, 26.** Non-consecutive serial sections showing basal body 1 and associated fibres. Note the A fibre (in 25), indistinct by the level of the transition zone (26). **27–29.** Non-consecutive serial sections showing the LR in transverse section, as well as the outer portion of the RR and anterior end of the groove. The RR separates into outer and inner portions near the opening of the groove. Arrow in 28 indicates an additional single microtubule parallel to dorsal side of the left root, posterior to the termination of the C fibre. **30.** Transverse section of the IRR and ORR, showing the maximal number of microtubules. **31.** Section showing the angle between the IRR and ORR, and the normal position of the amorphous element of the B fibre. **32.** Transverse section showing right side of the groove, after the termination of the B fibre. The groove wall is supported by microtubules originating from the ORR, as well as IMt1. **33.** Section showing the relationship between the two distinct elements of the B fibre (i.e. one with laminate appearance and the other including electron-dense sheet and amorphous material). Note also the cartwheel structures in basal bodies 2 and 3. **34.** Section showing the origin of the singlet microtubular root. 1, flagellum/basal body 1 (= posterior flagellum/basal body); 2, basal body 2 (= anterior basal body); 3, non-flagellated basal body 3; 4, non-flagellated basal body 4; A, A fibre; AR, anterior root; B, B fibre; C, C fibre; I, I fibre; IMt1, radiation of internal microtubules; IRR, inner portion of right root; LR, left root; ORR, outer portion of right root; RR, right root; S, singlet microtubular root. Scale bars (in 24) = 200 nm for all figures except 33, for in 33 = 200 nm.

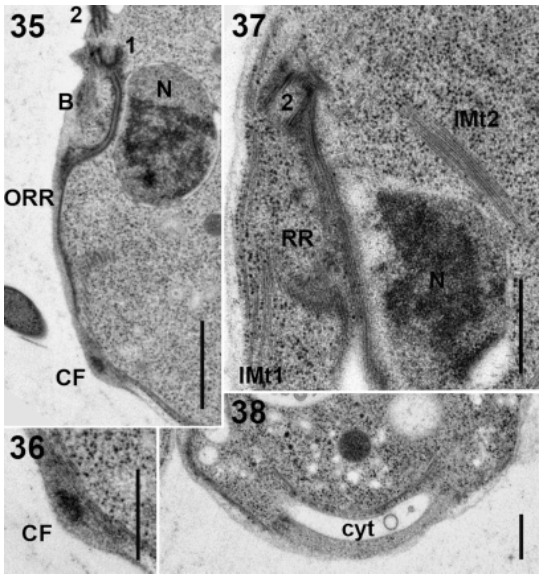


Fig. **35–38.** Transmission electron micrographs of *Hicanonectes teleskopos* n. g., n. sp., ultra-thin sections. **35.** Longitudinal section showing the right wall of the groove. Note the position of the possible composite fibre. **36.** Higher magnification view of the possible composite fibre. **37.** Section showing elements of IMt1 and IMt2. **38.** Posterior end of the cell, showing the cytopharynx. 1, basal body 1 (= posterior basal body); 2, flagellum/basal body 2 (= anterior flagellum/basal body); B, B fibre; CF, composite fibre; cyt, cytopharynx; IMt1 and IMt2, radiations of internal microtubules; N, nucleus; ORR, outer portion of right root; RR, right root. Scale bars for 35 = 1 μm; for 36–38 = 500 nm.

Posterior to this point, the right wall includes a small non-microtubular fibre, which we interpret to be the composite fibre (Fig. 35, 36). The cytopharynx is located in the posterior portion of the cell, and is supported by several microtubules originating ultimately from the RR (Fig. 8, 38).

Molecular phylogeny. In our analyses we included 36 sequences from Fornicata (including seven environmental sequences) and an outgroup of 32 other eukaryotes (Fig. 40a). The overall topology of Fornicata is mostly consistent with recently published results (e.g. Cepicka et al. 2008; Keeling and Brugerolle 2006; Kolisko et al. 2005, 2008; Yubuki et al. 2007). The monophyly of Fornicata, including our new isolate, was highly supported, with 96% bootstrap support (BS) and posterior probability (PP) of 1. Within Fornicata a major clade was recovered that included all diplomonads and the retortamonad *Retortamonas*, but not the retortamonad *Chilomastix*. The monophyly of diplomonads+*Retortamonas* was highly supported (99% BS; 1 PP), as was the monophyly of hexamitine diplomonads, including enteromonads (100% BS; 1 PP). The genus *Retortamonas* branches weakly with the giardiine diplomonads *Giardia* and *Octomitus* (46% BS; 0.77 PP).

The other fornicates, including our new isolate, formed a series of branches attached to the base of the diplomonads+*Retortamonas* clade. In order, these were (i) *Chilomastix* spp., (ii) *D. brevis*, (iii) a clade comprising *C. membranifera*, our new isolate, and uncultured eukaryote D4P08A09, and (iv) a tight clade consisting of several sequences from uncultured eukaryotes, including CPS-GM5, whose close relative was identified as a *Carpediemonas/Dysnectes*-like excavate using light microscopy (Takishita et al. 2007). The branching pattern among these clades, however, received low or very low BS.

Our new isolate formed a highly supported clade with uncultured eukaryote sequence D4P08A09 (100% BP: 1 PP). These two sequences in turn formed a clade with *C. membranifera* that received moderate BS (70%), and PP 1. It is noteworthy that sequence D4P08A09 was obtained from anoxic intertidal marine sediments, a similar environment to that from which both our new isolate and *C. membranifera* were originally isolated.

We repeated the analysis without the sequences from the retortamonad *Chilomastix* (Fig. 40b). A similar ML topology was recovered, and most important nodes received similar statistical support, with two exceptions. Firstly, the BS for the monophyly of *Carpediemonas*, our new isolate, and uncultured eukaryote D4P08A09 increased to 86% (PP remained at 1). Secondly, support for the placement of *D. brevis* as the sister group to diplomonads+retortamonads increased to 80% BS (up from 46%), and PP 1 (up from 0.94).

DISCUSSION

The affinities and assignment of *Hicanonectes*. Simpson (2003) recognized an assemblage of flagellates called ''typical excavates'' that share eight distinctive morphological characters (see also O'Kelly 1993, 1997; O'Kelly and Nerad 1999; Simpson and Patterson 1999; Yubuki et al. 2007). These features are (1) a ventral groove used for suspension feeding, (2) flagellar vanes, (3) a splitting of the RR, (4) an I fibre, (5) a B fibre, (6) a singlet microtubular root associated with basal body 1, (7) a C fibre, and (8) a composite fibre. Our new isolate displays at least seven of these characters (Table 1). The sole uncertainty concerns the composite fibre: we identified a delicate non-microtubular element in the same position as the composite fibre of typical excavates, but did not determine its substructure. There are also no data as to whether the typical excavate *Malawimonas* has a composite fibre (O'Kelly and Nerad 1999; Simpson 2003). Other aspects of the ultrastructure of our new isolate are similar to previously studied typical excavates: for example, the arrangement of the left, right, and singlet roots in their support of the feeding groove, and the presence of a detectable A fibre (Simpson 2003; Simpson and Patterson 1999). Based on morphology we consider our new isolate to be a typical excavate, along with jakobids, retortamonads, *Trimastix*, *Malawimonas*, *Carpediemonas*, and *Dysnectes* (Simpson 2003; Yubuki et al. 2007). However, assignation of an organism as a typical excavate does not resolve its phylogenetic position, because molecular phylogenies demonstrate clearly that typical excavates are not a monophyletic group (Dacks et al. 2001; Rodriguez-Ezpeleta et al. 2007; Simpson et al. 2002b, 2006).

Our SSU rRNA gene phylogenies place our new isolate in the clade Fornicata, which includes the typical excavates Retortamonadida, *Carpediemonas*, and *Dysnectes*, as well as the non-typical excavate group Diplomonadida (Cepicka et al. 2008; Simpson et al. 2002b, 2006; Yubuki et al. 2007). Statistical support for this position is strong irrespective of phylogenetic method, and withstands minor changes in taxon sampling. Furthermore the B fibre complex in our new isolate originates against the LR and then arches across the ventral face of basal body 1 to associate with the RR. This arrangement is characteristic of *Carpediemonas*, *Dysnectes*, and retortamonads, and is the proposed synapomorphy that defines the taxon Fornicata (Simpson 2003; Simpson and Patterson 1999; Yubuki et al. 2007).

Some other conspicuous ultrastructural features are consistent with, although not diagnostic of, placement with Fornicata. Our new isolate has two opposed vanes, dorsal and ventral, on its

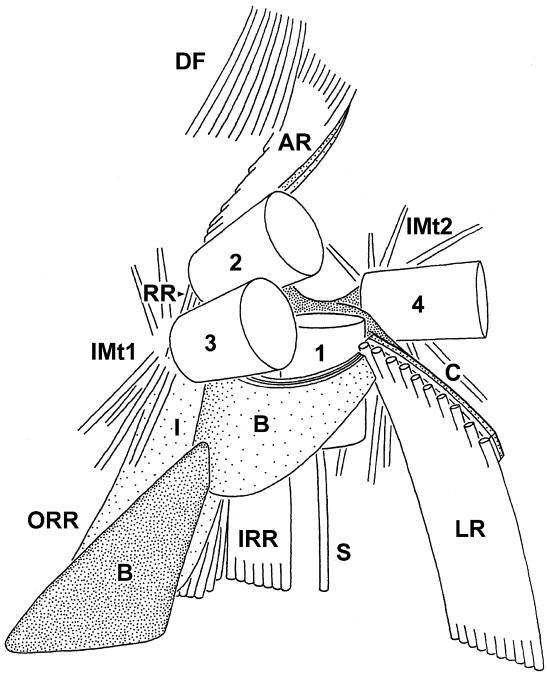


Fig. **39.** Diagrammatic representation of the proximal flagellar apparatus of *Hicanonectes teleskopos* n. g., n. sp., viewed from the ventral side. Basal bodies are depicted as large cylinders and the flagella themselves are not shown. The lucent field associated with the ventral side of the amorphous element of the B fibre has been omitted. All major microtubular roots continue beyond the confines of the diagram. 1, basal body 1 (= posterior basal body); 2, basal body 2 (= anterior basal body); 3, non-flagellated basal body 3; 4, non-flagellated basal body 4; AR, anterior root; B, B fibre; C, C fibre; DF, dorsal fan; I, I fibre; IMt1 and IMt 2, radiations of internal microtubules; IRR, inner portion of right root; LR, left root; ORR, outer portion of right root; RR, right root; S, singlet microtubular root.

posterior flagellum, although the dorsal vane is poorly developed. This is similar to all typical excavate groups except jakobids and *Malawimonas jakobiformis*, which each have only one of the vanes (Lara, Chatzinotas, and Simpson 2006; O'Kelly 1997; O'Kelly and Nerad 1999; Simpson and Patterson 2001). In place of classical-looking mitochondria, our new isolate has rounded organelles that lack cristae. This is similar to other Fornicata, but also to Preaxostyla and Parabasala, none of which have classical crista-bearing mitochondrial organelles (Brugerolle and Patterson 1997; Carpenter, Waller, and Keeling 2008; Hampl and Simpson 2008; O'Kelly, Farmer, and Nerad 1999; Simpson and Patterson 1999; Yubuki et al. 2007). In summary, both molecular phylogenies and easy-to-interpret morphological characters indicate that our new isolate belongs to the taxon Fornicata.

Of the previously described groups within Fornicata, our new isolate is most similar at the ultrastructural level to *Carpediemonas* and *Dysnectes*. It shares with *C. membranifera* 11 of the 13 ultrastructural features highlighted in Table 1 (see also Fig. 41). Like *C. membranifera* our new isolate has an I fibre that includes a double-leaved sheet, while the A fibre is indistinct. However, there are several differences. Our new isolate has four basal bodies, two flagellar vanes, and conspicuous radiations of IMt, whereas *C. membranifera* has three basal bodies, three vanes, and few IMt (Simpson and Patterson 1999). The AR of *C. membranifera* is more delicate, and has a different orientation and relationship to the dorsal fan. The RR and LR of *C. membranifera* consist of 16 and six microtubules, respectively, while those of our new isolate comprise 20 and 10 microtubules, respectively. In *C. membranifera*, most of the width of the groove is supported by microtubules derived ultimately from the RR, whereas microtubules from the RR support only about half the width of the groove in our new isolate. *Carpediemonas membranifera* does not have a distinct cytopharnyx at the posterior end of the groove, while our new isolate has no observable dictyosome.

Our new isolate and *D. brevis* share 10 of the 13 ultrastructural features of excavates (Table 1, see also Fig. 41). The number of microtubules in the RR is similar (18 vs. 20), as is the proportion of the groove supported by microtubules derived from the RR (about half). Both species have two flagellar vanes, while dictyosomes have not been found in either. As with both our new isolate and *C. membranifera*, the A fibre of *D. brevis* is indistinct at best (Yubuki et al. 2007 report it as absent). However, *D. brevis* has two basal bodies in interphase, not four as in our new isolate. The LR of *D. brevis* is more strongly developed, with more microtubules (17 vs. 10), and a thicker C fibre, as well as an extension of the B fibre down its ventral face. *Dysnectes brevis* lacks a distinct cytopharynx. Furthermore, *D. brevis* has a much less substantial AR, and apparently no dorsal fan at all (Yubuki et al. 2007).

As indicated by our SSU rRNA gene phylogenies, our new isolate is markedly dissimilar at the sequence level to other described fornicate taxa. Our ML and Bayesian trees place it as the sister group of *C. membranifera*, along with a related though distinct environmental sequence. Bootstrap support is moderate or strong, depending on taxon sampling. Tentatively, we consider it most likely that *C. membranifera* is the closest formally described relative of our new isolate. Nonetheless, in light of the considerable morphological differences and SSU rRNA gene sequence dissimilarity between our new isolate and *C. membranifera*, we propose that it represents a distinct genus from *Carpediemonas* (and *Dysnectes*). Because we are not aware of any previously described organism with which our isolate can be identified, we here introduce the new genus and new species *H. teleskopos* n. g., n. sp. Formal diagnoses are given at the end of the discussion.

The anterior root and dorsal fan. The AR of *H. teleskopos* n. g., n. sp. is well developed, with nine microtubules attached to a supporting non-microtubular fibre. This root travels anteriorly, and the dorsal fan appears to originate alongside it, in parallel. This organization differs markedly from that in most typical excavates including other fornicates. In *Carpediemonas* and *Dysnectes*, as well as in *Malawimonas* and *Trimastix pyriformis*, the AR curves to run posteriorly down the left side of the cell, contains only one to four microtubules, and is, at most, lightly reinforced with non-microtubular material (Brugerolle and Patterson 1997; O'Kelly and Nerad 1999; O'Kelly et al. 1999; Simpson and Patterson 1999; Yubuki et al. 2007). The dorsal fan, where present, originates along the length of the AR in these taxa. In retortamonads and jakobids, there is no true AR. The dorsal fan originates in close association with basal body 2 in jakobids (Lara et al. 2006; O'Kelly 1997; Patterson 1990), and with the sheet-like non-microtubular ''lapel'' in retortamonads (Bernard, Simpson, and Patterson 1997; Brugerolle 1973). There are some parallels between *H. teleskopos* and *Trimastix marina*. In *T. marina* the AR is also large (15–16 microtubules), is directed anteriorly, and is associated with non-microtubular material (Simpson, Bernard, and Patterson 2000). However, the non-microtubular material is associated with the exterior-most face of the AR, rather than the interior-most face as in *H. teleskopos*, and the dorsal fan is most closely associated with the face of the AR, and not the edge of the root as in *H. teleskopos*. Most likely the AR of ancestral typical

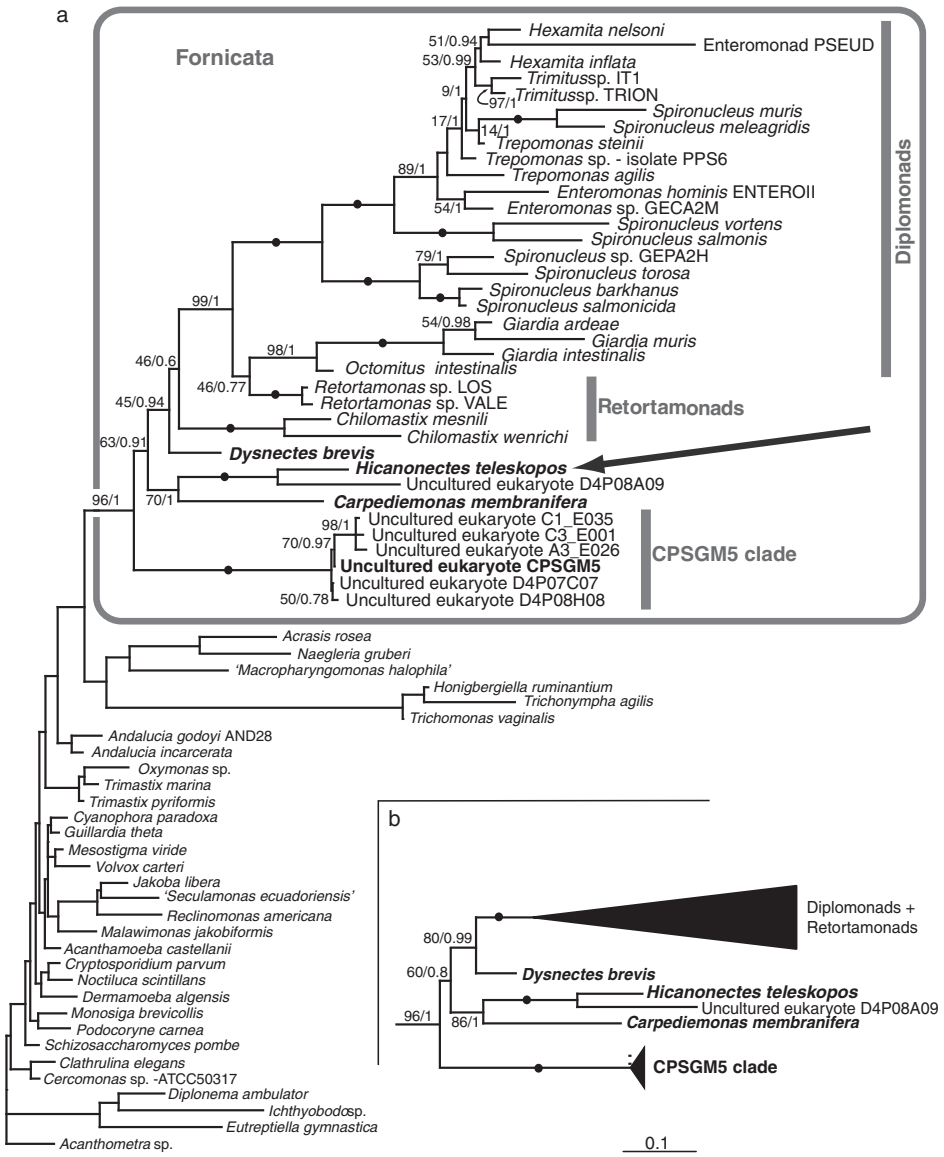


Fig. **40.** Molecular phylogeny of Fornicata showing the position of *Hicanonectes teleskopos* n. g., n. sp. **a.** Maximum likelihood (ML) tree based on the full dataset of SSU rRNA gene sequences (GTR+Γ+I model). **b.** Maximum likelihood topology within Fornicata when *Chilomastix* was excluded from the analysis. Numbers along branches show ML IQPNNI bootstrap percentages and Bayesian posterior probabilities. Support values for nodes outside Fornicata are not shown. Statistical support is not shown for nodes within diplomonads or the CPSGM5 clade when bootstrap support is $<50\%$ and posterior probability is <0.8. Black circles indicate bootstrap support of 100% and posterior probability of 1.

excavates was small and directed leftwards and posteriorly, with *H. teleskopos* and *T. marina* representing two independent lineages in which the AR convergently expanded in size, become more anteriorly directed, and associated differently with the dorsal fan.

The B fibre. In *H. teleskopos* n. g., n. sp., there are two distinct components to the B fibre complex: a laminate sheet-like element, which might be considered the B fibre sensu stricto, and a second, more diffuse element consisting of a dense non-laminate sheet and a relatively lucent zone. It is this second element that extends further posteriorly and associates most closely with the right margin of the groove. The B fibre of *D. brevis* appears also to comprise two distinct elements (fig. 16, 17, 24, and 25 in Yubuki et al. 2007). Although not explicitly identified previously, the second

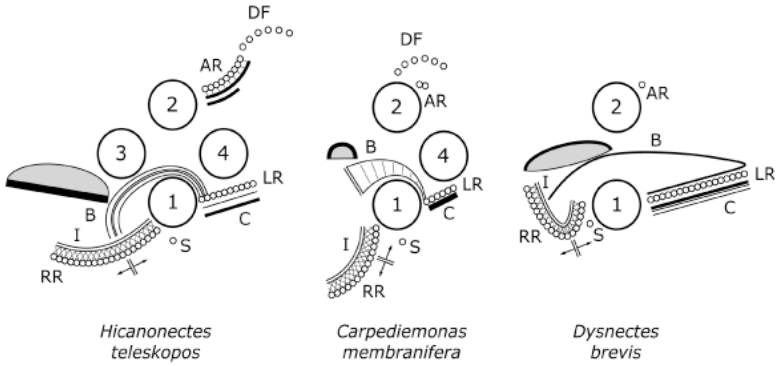


Fig. **41.** Diagrams of the flagellar apparatus of basal fornicates, using the system of Sleigh (1988), as per Simpson (2003). Each basal body is represented in tip-to-base orientation, and its associated microtubules and non-microtubular structures are aligned with it. Not all non-microtubular structures are shown (e.g. possible A fibres are omitted). Note that not all structures shown are present at the same level of sectioning; microtubular roots are shown at their fullest extent, although they frequently gain or lose microtubules along their length. Note also that non-microtubular elements, especially the B fibre, may not be present in the shape indicated in all planes of section; in *Dysnectes* it runs along the outermost six or so microtubules of LR as that root increases from 1 to 17, and is not present at all along more-inward microtubules. *Dysnectes* is depicted from information in Yubuki et al. 2007; *Carpediemonas* after Simpson 2003, with reference to original micrographs (Simpson and Patterson 1999). 1, basal body 1 (= posterior basal body); 2, basal body 2 (= anterior basal body); 3, non-flagellated basal body 3; 4, non-flagellated basal body 4; B, B fibre; C, C fibre; DF, dorsal fan; I, I fibre; LR, left root; RR, right root; S, singlet microtubular root.

element may also be present in *C. membranifera* (fig. 3f, g and 4b–f in Simpson and Patterson 1999). The precise appearance of this element differs in all three species, although some of this variation might be due to different fixation protocols. The presumed homologue of the B fibre in retortamonads is the structure originally referred to as the arched fibre (Simpson 2003; Simpson and Patterson 1999). This fibre has a second discrete element of amorphous material that is associated with its ventral face, and which lies within the right margin of the anterior part of the groove (Bernard et al. 1997; Brugerolle 1977). This extra material might be homologous to the second element of the B fibre complex of *Hicanonectes*, *Carpediemonas*, and *Dysnectes*. We have not identified a positionally equivalent element in other typical excavates. There is an amorphous element associated with the B fibre in the jakobid *Andalucia incarcerata*, but this is sandwiched between the I fibre and B fibre, rather than being at the right margin of the groove wall (Simpson and Patterson 2001), and is unlikely to be homologous.

Mitochondrion-like organelles. The mitochondrion-like organelles of *H. teleskopos* n. g., n. sp., are rounded and of moderate size, $\sim$ 300 nm across, and they are common within the cell. This is similar to the mitochondrion-like organelles of *D. brevis* (Yubuki et al. 2007). In *C. membranifera* the mitochondrion-like organelles are elongate rather than rounded, and may be connected as a network, but are otherwise similar (Simpson and Patterson 1999). The organelles in all three taxa are much larger than the mitosomes of the diplomonad *G. intestinalis*, which are apparently not involved in ATP generation (Tovar et al. 2003). In biovolume, the organelles of *H. teleskopos*, *D. brevis*, and *C. membranifera* are more similar to the hydrogenosomes of

Table **1.** Summary of the structural features of *Hicanonectes teleskopos* n. g., n. sp. and other excavates.

Taxon	Ventral groove	Flagellar vanes (#)	Split RR	I fibre	B fibre (origin)	SR	C fibre (# sheets)	CF	A fibre	Number of basal bodies	AR	Dorsal fan	Mitochondrial organelle
Hicanonectes	+	+ (2)	+	+	+ (LR)	+	+ (2)	+	+	4	+	+	NC
Carpediemonas	+	+ (3)	+	+	+ (LR)	+	+ (2)	+	+	3	+	+	NC
Dysnectes	+	+ (2)	+	+	+ (LR)	+	+ (3)	+	?[a]	2	+	–	NC
Retortamonads	+	+ (2/3)	+	+	+ (LR)	+	+ (2)	+	+	4	–	+	NC
Trimastix	+	+ (2)	+	+	+ (RR)	+	+ (2)	+	+	4	+	+	NC
Malawimonas	+	+ (1)	+	+	+ (BB)	+	+ (4)	+	+	2	+	+	C
Jakobids	+	+ (1)	+	+	+ (BB)	+	+ (1)	+	+	2	–	+	C (1 NC)
Diplomonads	+	–[b]	+	+	–	?	–	–	?	4	+	–	NC
Parabasalids	–	–	–	–	–	–	?	–	?	Varies	–	+	NC
Oxymonads[c]	–	–	–	+	+ (RR)	+	+ (2)	–	?	4	+	+	NC
Euglenozoa	–	–	–	–	–	–	–	–	?	2	+	+	C (mostly)
Heterolobosea	±	–[b]	+	+	–	–	–	–	?	2/4	?	+	C (mostly)

''Typical excavates'' are shaded grey.

[a]Recorded as absent by Yubuki et al. (2007), but see Figure 16 in Yubuki et al. (2007) for a possible delicate A fibre.

[b]Probably non-homologous flagellar vanes in one isolated subgroup – see O'Kelly (1997), Simpson (2003).

[c]Based on the underived oxymonad *Monocercomonoides*—see Simpson et al. (2002b).

+, present; –, absent; ?, uncertain; AR, anterior root; BB, basal body; C, cristae in mitochondrion-like organelles; CF, composite fibre; LR, left root; NC, no cristae in mitochondrion-like organelles; RR, right root; SR, singlet root.

parabasalids, but are somewhat smaller than the typical ~ 500 nm diameter (e.g. see Benchimol 2008). They also resemble the mitochondrion-related organelles of the preaxostylan excavate *T. pyriformis* (Brugerolle and Patterson 1997; O'Kelly et al. 1999), which might also perform hydrogenosomal ATP generation (based on the presence of transcripts for PFO (pyruvate: ferredoxin oxidoreductase and [FeFe]-hydrogenase, Hampl et al. 2008). Overall, this study confirms that Fornicata is a lineage in which classical mitochondria are universally absent, yet contains a series of organisms with mitochondrion-like organelles that differ in appearance from the mitosomes of *Giardia*. Further studies of these organisms may help clarify how the *Giardia* mitosome evolved from presumably less-reduced mitochondrion-like organelles.

The relationships among Fornicata. A well-sampled and well-resolved phylogenetic tree of Fornicata is required for understanding important aspects of the evolution of this group (e.g. possible transitions between different types of mitochondrion-like organelles), and developing a rational higher taxonomy of its more recently discovered major lineages. The deep portions of the Fornicata tree are presently not well-resolved using SSU rRNA gene data. We recovered a *Hicanonectes*+*Carpediemonas* clade with reasonable statistical support, but found mostly poor support for the relationships among (i) the *Hicanonectes*+*Carpediemonas* clade, (ii) *Dysnectes*, (iii) the clade including CPSGM-5, (iv) a diplomonad+*Retortamonas* clade, and (v) *Chilomastix*, where included. We did recover a *Dysnectes*+diplomonad+retortamonad clade with high PP, and with reasonably strong BS when *Chilomastix* was excluded. We suggest that an unstable position of *Chilomastix* was masking otherwise moderate-to-strong support for the *Dysnectes*+diplomonad+retortamonad clade. However, our tree topology conflicts with the analysis of Yubuki et al. (2007) where *Dysnectes* was recovered as the deepest branch within Fornicata (i.e. *Carpediemonas* was more closely related to diplomonads and retortamonads). The poor resolution of the tree of Fornicata based on SSU rRNA gene sequence might be an effect of poor taxon sampling in deep lineages, and/or variable rates of sequence evolution for this gene in these taxa. We anticipate that analyses that include additional deeply branching fornicates and examine multi-gene datasets will be required to resolve the deep relationships among Fornicata.

Taxonomic Summary

Assignment: Eukaryota; Excavata; Fornicata

Hicanonectes n. g.

Diagnosis. Free-living, biflagellated, and colourless cells bearing a longitudinal groove with a sharply defined right wall. The posterior flagellum beats within the groove, and bears vanes. The posterior end of the groove forms a curved cytopharynx. Mitochondrion-like organelles lack cristae. The anterior microtubular root is well developed and directed laterally and anteriorly (rather than curving to be directed leftwards and posteriorly). The cell rotates rapidly while swimming.

Type species. *Hicanonectes teleskopos* Park, Kolisko, Heiss & Simpson

Etymology. *Hicanonectes* = ''adequate swimmer'' (Greek; masculine). This organism is a more conventional and effective swimmer than its most similar relatives, *Carpediemonas* and *Dysnectes* (the latter name meaning ''bad swimmer''—Yubuki et al. 2007). It is, however, of only moderate abilities when compared with many flagellates from other taxonomic groups.

Hicanonectes teleskopos n. sp.

Diagnosis. Cells oval-shaped and 6.5–10.0 μm long. The posterior flagellum is 2.5–3.5 times as long as the cell; the anterior flagellum is directed sharply leftwards and is 1.0–1.5 times as long as the cell.

Type material. Block of resin-embedded cells for electron microscopy deposited with the Protist Type Specimen Slide Collection, U.S. Natural History Museum, Washington, DC, as USNM 1122785. This material constitutes the name-bearing hapantotype for the species.

Type locality. Anoxic layer of intertidal sediment, Salt Spring Island, BC, Canada (48°46′N and 123°28′W).

Etymology. *teleskopos* = ''Far see-er'' (Greek), recognizes the Canadian Institute for Advanced Research (CIfAR, pronounced ''see-far'') for long-standing support of microbial evolution research in Canada, and commemorates the isolation of this species immediately after the first full meeting of the CIfAR Program in Integrated Microbial Biodiversity, in October 2007.

ACKNOWLEDGMENTS

Thanks to David Walsh (University of British Columbia) for sampling assistance, Zhiyuan Lu (Dalhousie University) for assistance with high-pressure freezing fixation, and Harold Tarrant (Newcastle University, Australia) for advice on ancient Greek. This work is supported by NSERC grant 298366-04 to AGBS, and the Canadian Institute for Advanced Research (CIfAR) Program in Integrated Microbial Biodiversity. J.S.P. is partly supported by a Korea Research Foundation Grant funded by the Korean Government (MOEHRD, No. KRF-2007-357-C00119). M.K. is supported by the Nova Scotia Health Research Foundation. A.A.H. is partly supported by a Dalhousie University graduate student scholarship.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, **215**:403–410.

Benchimol, M. 2008. Structure of the hydrogenosome. *In*: Tachezy, J. (ed.), Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer-Verlag, Berlin. p. 75–97.

Bernard, C., Simpson, A. G. B. & Patterson, D. J. 1997. An ultrastructural study of a free-living retortamonad, *Chilomastix cuspidata* (Larsen and Patterson, 1990) n. comb. (Retortamonadida, Protista). *Eur. J. Protistol.*, **33**:254–265.

Brugerolle, G. 1973. Etude ultrastucturale du trophozoite du kyste chez le genre *Chilomastix* Aléxéieff, 1910 (Zoomastigophorea, Retortamonadida Grassé, 1952). *J. Protozool.*, **20**:574–585.

Brugerolle, G. 1977. Ultrastructure du genre *Retortamonas* Grassi 1879 (Zoomastigophorea, Retortamonadida Wenrich 1931). *Protistologica*, **13**:233–240.

Brugerolle, G. & Patterson, D. J. 1997. Ultrastructure of *Trimastix convexa* Hollande, an amitochondriate anaerobic flagellate with a previously undescribed organisation. *Eur. J. Protistol.*, **33**:121–130.

Carpenter, K. J., Waller, R. F. & Keeling, P. J. 2008. Surface morphology of *Saccinobaculus* (Oxymonadida): implications for character evolution and function in oxymonads. *Protist*, **159**:209–221.

Cavalier-Smith, T. 1995. Cell cycles, diplokaryosis and the archezoan origin of sex. *Arch. Protistenkd.*, **145**:189–207.

Cepicka, I., Kostka, M., Uzlikova, M., Kulda, J. & Flegr, J. 2008. Non-monophyly of Retortamonadida and high genetic diversity of the genus *Chilomastix* suggested by analysis of SSU rDNA. *Mol. Phylogenet. Evol.*, **48**:770–775.

Ciccarelli, D. F., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311**:1283–1287.

Clark, C. G. 1992. DNA purification from polysaccharide-rich cells. *In*: Lee, J. J. & Soldo, A. T. (eds.), Protocols in Protozoology. Allen Press, Lawrence, KS. **1**:D-3.1–D-3.2.

Dacks, J. B., Silberman, J. D., Simpson, A. G. B., Moryia, S., Kudo, T., Ohkuma, M. & Redfield, R. J. 2001. Oxymonads are closely related to the excavate taxon *Trimastix*. *Mol. Biol. Evol.*, **18**:1034–1044.

Diamond, L. S. 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other lumen-dwelling protozoa. *J. Parasitol.*, **68**:958–959.

Ekebom, J., Patterson, D. J. & Vørs, N. 1996. Heterotrophic flagellates from coral reef sediments (Great Barrier Reef, Australia). *Arch. Protistenkd.*, **146**:251–272.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*, **41**:95–98.

Hampl, V. & Simpson, A. G. B. 2008. Possible mitochondria-related organelles in poorly-studied ''amitochondriate'' eukaryotes. *In*: Tachezy, J. (ed.), Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer-Verlag, Berlin. p. 265–283.

Hampl, V., Silberman, J. D., Stechmann, A., Diaz-Triviño, S., Johnson, P. J. & Roger, A. J. 2008. Genetic evidence for a mitochondriate ancestry in the 'amitochondriate' flagellate *Trimastix pyriformis*. *PLoS ONE*, **3**, e1383.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K-I. & Hasegawa, M. 1995. Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. *Mol. Biol. Evol.*, **12**:782–793.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K-I. & Hasegawa, M. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.*, **11**:65–71.

Horner, D. S. & Embley, T. M. 2001. Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes. *Mol. Biol. Evol.*, **18**:1970–1975.

Huelsenbeck, J. P. 2000. MrBayes: Bayesian inference of phylogeny. Distributed by the author. Department of Biology, University of Rochester.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acid Res.*, **33**:511–518.

Keeling, P. J. & Brugerolle, G. 2006. Evidence from SSU rRNA phylogeny that *Octomitus* is a sister lineage of *Giardia*. *Protist*, **157**:205–212.

Kolisko, M., Cepicka, I., Hampl, V., Kulda, J. & Flegr, J. 2005. The phylogenetic position of enteromonads: a challenge for the present models of diplomonad evolution. *Int. J. Syst. Evol. Microbiol.*, **55**:1729–1733.

Kolisko, M., Cepicka, I., Hampl, V., Leigh, J., Roger, A. J., Kulda, J., Simpson, A. G. B. & Flegr, J. 2008. Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, α-tubulin and HSP90 genes: implications for the evolutionary history of the double karyomastigont of diplomonads. *BMC. Evol. Biol.*, **8**:e205.

Kulda, J. & Nohýnková, E. 1978. Flagellates of the human intestine and of intestines of other species. *In*: Kreier, J. P. (ed.), Parasitic Protozoa: Volume II: Intestinal Flagellates, Histomonads, Trichomonads, Amoeba, Opalinids, and Ciliates. Academic Press, San Diego. p. 1–128.

Lara, E., Chatzinotas, A. & Simpson, A. G. B. 2006. *Andalucia* (n. gen.) – The deepest branch within jakobids (Jakobida, Excavata), based on morphological and molecular study of a new flagellate from soil. *J. Eukaryot. Microbiol.*, **53**:112–120.

Lee, W. J. & Patterson, D. J. 2000. Heterotrophic flagellates (Protista) from marine sediments of Botany Bay, Australia. *J. Nat. Hist.*, **34**:483–562.

Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene*, **71**:491–499.

O'Kelly, C. J. 1993. The jakobid flagellates: structural features of *Jakoba*, *Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes. *J. Eukaryot. Microbiol.*, **40**:627–636.

O'Kelly, C. J. 1997. Ultrastructure of trophozoites, zoospores and cysts of *Reclinomonas americana* Flavin & Nerad, 1993 (Protista *incertae sedis*: Histionidae). *Eur. J. Protistol.*, **33**:337–348.

O'Kelly, C. J. & Nerad, T. A. 1999. *Malawimonas jakobiformis* n. gen., n. sp. (Malawimonadidae n. fam.): a *Jakoba*-like heterotrophic nanoflagellate with discoidal mitochondrial cristae. *J. Eukaryot. Microbiol.*, **46**:522–531.

O'Kelly, C. J., Farmer, M. A. & Nerad, T. A. 1999. Ultrastructure of *Trimastix pyriformis* (Klebs) Bernard et al.: similarities of *Trimastix* species with retortamonad and jakobid flagellates. *Protist*, **150**:149–162.

Patterson, D. J. 1990. *Jakoba libera* (Ruinen, 1938), a heterotrophic flagellate from deep ocean sediments. *J. Mar. Biol. Assoc. UK*, **70**:381–393.

Posada, D. & Crandall, C. A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**:817–818.

Rodriguez-Ezpeleta, N., Brinkman, H., Burger, G., Roger, A. J., Gray, M. W. & Lang, B. F. 2007. Toward resolving the eukaryotic tree: the phylogenetic position of jakobids and cercozoans. *Curr. Biol.*, **17**:1420–1425.

Roger, A. J., Svärd, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. & Sogin, M. L. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harboured an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. USA*, **95**:229–234.

Simpson, A. G. B. 2003. Cytoskeletal organisation phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int. J. Syst. Evol. Microbiol.*, **53**:1759–1777.

Simpson, A. G. B. & Patterson, D. J. 1999. The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the ''excavate hypothesis''. *Eur. J. Protistol.*, **35**:353–370.

Simpson, A. G. B. & Patterson, D. J. 2001. On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*. *J. Eukaryot. Microbiol.*, **48**:480–492.

Simpson, A. G. B., Bernard, C. & Patterson, D. J. 2000. The ultrastructure of *Trimastix marina* Kent 1880 (Eukaryota), an excavate flagellate. *Eur. J. Protistol.*, **36**:229–252.

Simpson, A. G. B., Inagaki, Y. & Roger, A. J. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary position of ''primitive'' eukaryotes. *Mol. Biol. Evol.*, **23**:615–625.

Simpson, A. G. B., MacQuarrie, E. K. & Roger, A. J. 2002a. Eukaryotic evolution: early origin of canonical introns. *Nature*, **419**:270.

Simpson, A. G. B., Roger, A. J., Silberman, J. D., Leipe, D. D., Edgcomb, V. P., Jermiin, L. S., Patterson, D. J. & Sogin, M. L. 2002b. Evolutionary history of ''early diverging'' eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol. Biol. Evol.*, **19**: 1782–1791.

Sleigh, M. A. 1988. Flagellar root maps allow speculative comparisons of root patterns and of their ontogeny. *BioSystems*, **21**:277–282.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**:75–77.

Swofford, D. L. 2003. PAUP: Phylogenetic Analyses Using Parsimony (*and other Methods). Sinauer Associates, Sunderland, MA.

Tachezy, J., Sánchez, L. B. & Müller, M. 2001. Mitochondrial type iron–sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Mol. Biol. Evol.*, **18**:1919–1928.

Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y. & Maruyama, T. 2007. Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles*, **11**:563–576.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **24**:4876–4882.

Tovar, J., León-Avilla, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., Hernandez, M., Müller, M. & Lucocq, J. M. 2003. Mitochondrial remnant organelles of *Giardia* function in iron–sulfur protein maturation. *Nature*, **426**:172–176.

Vinh, L. S. & von Haeseler, A. 2004. IQPNNI: moving fast through tree space and stopping in time. *Mol. Biol. Evol.*, **8**:1565–1571.

Yubuki, N., Inagaki, Y., Nakayama, T. & Inouye, I. 2007. Ultrastructure and ribosomal RNA phylogeny of the free-living heterotrophic flagellate *Dysnectes brevis* n. gen, n. sp., a new member of Fornicata. *J. Eukaryot. Microbiol.*, **54**:191–200.

Received: 11/05/08, 02/27/09; accepted: 03/03/09

15. Appendix H

Article describing *Ergobibamus teleskopos* that I co-authored

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Sep 07, 2011

This is a License Agreement between Martin Kolisko ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2743730973896
License date	Sep 07, 2011
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Eukaryotic Microbiology
Licensed content title	Cell Morphology and Formal Description of Ergobibamus cyprinoides n. g., n. sp., Another Carpediemonas-Like Relative of Diplomonads
Licensed content author	JONG SOO PARK,MARTIN KOLISKO,ALASTAIR G.B. SIMPSON
Licensed content date	Nov 1, 2010
Start page	520
End page	528
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Total	0.00 USD

J. Eukaryot. Microbiol., 57(6), 2010 pp. 520–528

DOI: 10.1111/j.1550-7408.2010.00506.x

Cell Morphology and Formal Description of *Ergobibamus cyprinoides* n. g., n. sp., Another *Carpediemonas*-Like Relative of Diplomonads

JONG SOO PARK, MARTIN KOLISKO and ALASTAIR G.B. SIMPSON

Department of Biology, Canadian Institute for Advanced Research, Program in Integrated Microbial Diversity, Dalhousie University, Oxford Street, Halifax, Nova Scotia, Canada B3H 4J1

ABSTRACT. About 20 new isolates of *Carpediemonas*-like organisms (CLOs) have been reported since 2006. Small subunit rRNA gene phylogenies divide CLOs into six major clades: four contain described exemplars (i.e. *Carpediemonas*, *Dysnectes*, *Hicanonectes*, and *Kipferlia*), but two include only undescribed organisms. Here we describe a representative of one of these latter clades as *Ergobibamus cyprinoides* n. g., n. sp., and catalogue its ultrastructure. *Ergobibamus cyprinoides* is a bean-shaped biflagellated cell, 7–11.5 μm long, with a conspicuous groove. Instead of classical mitochondria there are cristae-lacking rounded organelles 300–400 nm in diameter. The posterior flagellum has a broad ventral vane and small dorsal vane. There are normally four basal bodies, two non-flagellated. There is one anterior root (AR), containing six microtubules. The posterior flagellar apparatus follows the ''typical excavate'' pattern of a splitting right root supported by fibres ''I,'' ''B,'' and ''A,'' a ''composite'' fibre, a singlet root, and a left root (LR) with a ''C'' fibre. The B fibre originates against the LR—a synapomorphy of the taxon Fornicata—supporting the assignation of *Ergobibamus* to Fornicata, along with diplomonads, retortamonads, and other CLOs. Distinctive features of *E. cyprinoides* include the complexity of the AR, which is intermediate between *Hicanonectes*, and *Carpediemonas* and *Dysnectes*, and a dorsal extension of the C fibre.

Key Words. Anaerobe, basal eukaryote, diplomonad, excavate, *Giardia*, hydrogenosome, microaerophile, protist, protozoa, ultrastructure.

DIPLOMONADS and retortamonads are mostly parasitic/commensal protozoa that are well known for their peculiar cell biology, especially the diminutive, biochemically reduced ''mitosomes'' that diplomonads possess instead of classical mitochondria (Morrison et al. 2007; Nohýnková, Tůmová, and Kulda 2006; Tovar et al. 2003). These protists are also notable because of the very deep-branching position of diplomonads in most molecular phylogenies of eukaryotes that are rooted with prokaryote outgroups (Ciccarelli et al. 2006; Hashimoto et al. 1994, 1995; Morrison et al. 2007; Sogin et al. 1989), although there is strong evidence that these results are due to a long branch attraction artefact rather than reflecting historical signal (Embley and Martin 2006; Hampl et al. 2009; Philippe et al. 2005).

For a long time it was unclear which protists, if any, were closely related to diplomonads and retortamonads. A little under a decade ago, however, small subunit (SSU) rRNA and protein-coding gene phylogenies established that *Carpediemonas membranifera* was a specific relative (Simpson, MacQuarrie, and Roger 2002a; Simpson et al. 2002b). *Carpediemonas membranifera* is a small free-living flagellate observed in preparations from marine sediments under suboxic conditions (Bernard, Simpson, and Patterson 2000; Ekebom, Patterson, and Vørs 1996; Kolisko et al. 2010; Larsen and Patterson 1990; Lee and Patterson 2000). Its gross morphology and cytoskeleton closely resemble those of mitochondrial excavate taxa, especially *Malawimonas*, and it possesses cristae-lacking mitochondrial-like organelles that are much larger and more conspicuous than the mitosomes of diplomonads (Simpson and Patterson 1999).

Recent studies have shown that *Carpediemonas* is not phylogenetically isolated, as a substantial novel diversity of *Carpediemonas*-like organisms (CLOs) it has been cultured from marine/saline sediment material from diverse locations around the world (Kolisko et al. 2010; Park et al. 2009; Yubuki et al. 2007). The CLOs cultured to date represent at least six very distinct clades referred to as CL1–CL6 by Kolisko et al. (2010), with *C. membranifera* itself representing CL4 (Fig. 1). In SSU rRNA gene trees these six clades form a poorly resolved cloud at the base of diplomonads and retortamonads (Kolisko et al. 2010). By virtue of their phylogenetic position CLOs are important organisms for understanding the evolutionary origins of diplomonads and retortamonads, and their mitochondria-related organelles. Because of the continued interest in diplomonads as possible primitive and/or early diverging eukaryotes, the study of CLOs also contributes to our understanding of early eukaryote cell evolution.

Representatives of clades CL1 and CL3 have recently been characterized by transmission electron microscopy (TEM), and described formally as *Dysnectes brevis* and *Hicanonectes teleskopos* (Park et al. 2009; Yubuki et al. 2007). Both broadly resemble *Carpediemonas* in their cytoskeletal organization, and also have relatively large, cristae-lacking mitochondrial organelles (Park et al. 2009; Yubuki et al. 2007). A third organism, which represents CL6, was studied previously under the name *Carpediemonas bialata* (Lee and Patterson 2000) but has been assigned a new generic vehicle mainly on phylogenetic grounds and is now called *Kipferlia bialata* (Kolisko et al. 2010). As yet there are few TEM data for *Kipferlia* (Kolisko et al. 2010). To date there are neither TEM data nor formal descriptions of any organisms belonging to clades CL2 or CL5.

In this work we characterize a representative of clade CL5 using TEM, and formally describe it as a new genus, *Ergobibamus* n. g., with *Ergobibamus cyprinoides* n. sp. as its type.

MATERIALS AND METHODS

Isolation and culturing. The organism studied here Isolate CL was cultured from intertidal anoxic sediments sampled near Mahone Bay, Nova Scotia, Canada (44°26′N, 64°21′W). Approximately 1 ml of the sample was inoculated into 15-ml conical tubes containing 12 ml of modified Tryptone-yeast extract-serum-gastric mucin media (Diamond 1982), prepared without bovine serum or mucin, diluted 1:1 with sterile seawater, and supplemented with 30 ml/L of horse serum (Sigma, St. Louis, MO). The xenic, monoeukaryotic culture was established via filtering through a 5-μm filter, followed by a sequence of rapid serial transfers. Low oxygen conditions were maintained in the sealed culture tubes by the metabolic activity of the accompanying prokaryotes.

Light microscopy. Live cells were observed using phase contrast on a Zeiss Axiovert 200 M microscope (Carl Zeiss, Jena, Germany) equipped with an Axiocam HR digital camera. Cell sizes ($n = 30$) were determined using microphotographs and the camera software (Axiovision 4.6).

Corresponding Author: A. Simpson, Department of Biology, Canadian Institute for Advanced Research, Program in Integrated Microbial Diversity, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia, Canada B3H 4J1—Telephone number: +902 494 1247; FAX number: +902 494 3736; e-mail: alastair.simpson@dal.ca

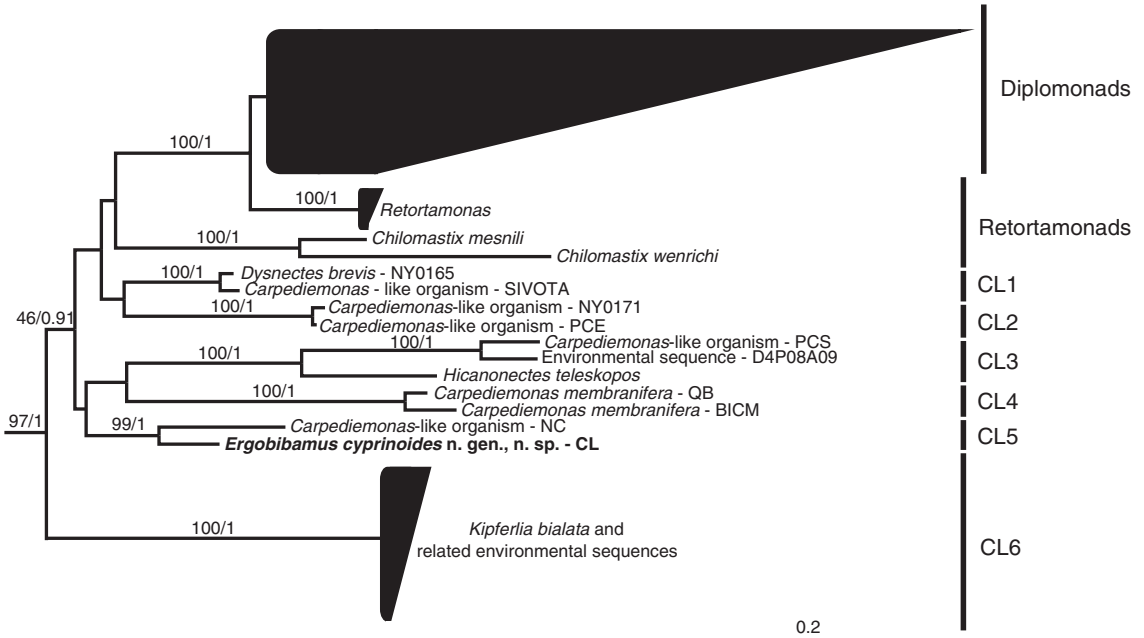


Fig. **1.** Phylogenetic tree depicting the major clades of *Carpediemonas*-like organisms (clades CL1–CL6), after the analysis by Kolisko et al. (2010). The clades of diplomonads, retortamonads and CL6 are collapsed, with the horizontal edges approximating the longest and shortest branch within each group. The outgroup of 31 eukaryote sequences is not shown. The statistical support is as follows: % bootstrap support based on 10,000 bootstrap replicates/MrBayes posterior probability values.

Electron microscopy. For freeze substitution, specimens were treated according to the protocol described in Park et al. (2009). Briefly, approximately 1.5 ml of cell culture were pelleted at 1,000 *g* for 30 min in a microcentrifuge tube. The supernatant was aspirated, and pellet material was loaded into hexadecene-coated 200-μm-deep gold-plated planchettes for high-pressure freezing using liquid nitrogen. The frozen specimens were transferred, under liquid nitrogen, into an anhydrous solution of 2.0% (w/v) OsO_4 and 0.1% (v/v) glutaraldehyde in HPLC-grade acetone. The temperature was raised from $-160\,^{\circ}C$ at a rate of 1 °C/h, holding the temperature steady for ~ 24 h at $-90\,^{\circ}C$ and for ~ 12 h at $-60\,^{\circ}C$. During the $-60\,^{\circ}C$ period, the specimens were transferred to pure acetone. Subsequently, the temperature was raised by 2 °C/h, with a 24-h holding period at $-30\,^{\circ}C$ until the specimens were at $-20\,^{\circ}C$. They were then transferred to $-20\,^{\circ}C$ for ~ 24 h, then to 4 °C for ~ 12 h, then to room temperature. After removal from the planchettes, specimens were transferred through a series of Spurr's resin mixtures, before final embedding.

For standard ''chemical'' fixation for TEM, cells were centrifuged at 8,000 *g* for 3 min and fixed for 30 min at room temperature in a cocktail containing 1% (v/v) glutaraldehyde and 5% (w/v) sucrose in 0.1 M cacodylate buffer (pH 7.4). After rinsing the cells 3 times in 0.1 M cacodylate buffer with 5% (w/v) sucrose, cells were post-fixed for 1 h in 0.8% (w/v) OsO_4 and 5% (w/v) sucrose in 0.1 M cacodylate. After being rinsed free of post-fixative, cells were concentrated by centrifugation and trapped in 1.5% (w/v) agarose. Agarose blocks were dehydrated by applying a graded series of ethanols, and then embedded in Spurr's resin.

Serial sections (~ 70 nm) were cut with a diamond knife on a Leica UC6 ultramicrotome (Leica, Wetzlar, Germany), then stained with saturated uranyl acetate in 50% ethanol and with lead citrate. Sections were observed using a Tecnai 12 TEM (FEI^{TM} Company, Hillsboro, OR) fitted with a goniometer stage and a 1-megapixel digital camera. In general, the freeze-substitution fixation gave better results, however usable material was sparse, and some structures were more difficult to visualize in this material.

RESULTS

Light microscopy. Cells of Isolate CL are bean-shaped, more-or-less inflexible, biflagellated, and possess a large, easily visible longitudinal groove (Fig. 2–4). The cells are 7–11.5 μm long (mean = 9 μm, SD ± 1 μm, $n = 30$) and 3.5–6.5 μm wide (5 ± 0.7 μm). The posterior flagellum is $\sim$ 2–2.5 cell lengths long, runs through the groove, and continues behind the cell (Fig. 2–4). The portion of the flagellum within the groove beats with high amplitude waves (Fig. 4), while the distal half of the flagellum is relatively passive. The anterior flagellum is about the same length as the cell body, inserts almost apically, and beats ahead of the cell as well as dorsally and ventrally (Fig. 2–4). A small bulge appears between the points of insertion of the two flagella (Fig. 2, 4). The ovate nucleus is located in the anterior part of the cell (Fig. 2), and vacuoles containing prokaryotic food contents are usually visible within the cell (Fig. 2–4). The phagocytosis of bacteria occurs at the posterior end of the groove. The cells swim slowly in relatively straight lines with occasional slow wobbling.

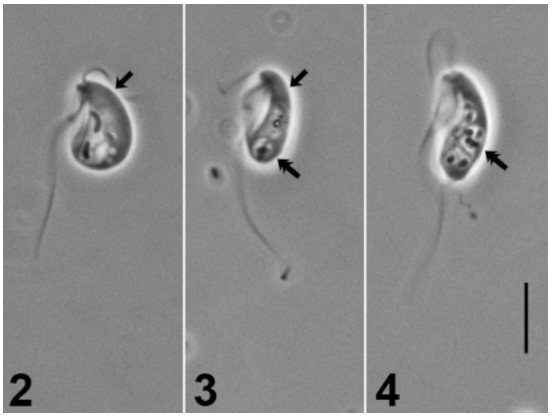


Fig. **2–4.** Phase-contrast light microphotographs of cells of *Ergobibamus cyprinoides* n. g., n. sp. Arrows in Fig. **2** and **3** denote nucleus. Double-headed arrows in Fig. **3** and **4** indicate food vacuoles. Scale bar: 5 μm.

TEM. The nucleus is located in the anterior region of the cell, close to the basal bodies and has an ovoid shape (Fig. 5, 6). It lacks a central nucleolus, although small dense condensations were observed around the periphery (Fig. 5). Mitochondrion-like organelles, several per cell, are concentrated in the general vicinity of the nucleus (Fig. 6). These are rounded, usually between 300 and 400 nm in diameter, and have two very closely adpressed bounding membranes, which are difficult to distinguish clearly in our preparations (Fig. 6–10). The matrix appears homogeneous except for a small diffuse region of increased density (Fig. 7, 8). No cristae-like structures were observed inside the organelles (Fig. 7–10). Stacked endoplasmic reticulum was sometimes observed in the cytoplasm (Fig. 11), but no discrete Golgi apparatus was observed. Cells include vacuoles containing presumed remnants of digested prey, and sometimes apparently empty vacuoles as well (Fig. 5).

There are normally four basal bodies per cell (Fig. 12). Basal body 1 gives rise to the posterior flagellum (Fig. 13, 14), and basal body 2 to the anterior flagellum, while basal bodies 3 and 4 are non-flagellated. In one cell microtubular elements of an additional flagellar organelle were observed (Fig. 15, asterisk). Basal bodies 1 and 2 are about 350 nm long (Fig. 5, 12). Basal bodies 3 and 4 are somewhat shorter (Fig. 12). Basal bodies 1 and 2 are arranged at a slightly obtuse angle, and are separated by $\sim$ 100 nm at their closest (Fig. 12, 15, 17). Basal body 2 originates to the right side of basal body 1, and is directed leftwards (Fig. 15, 18). Basal bodies 3 and 4 lie to the right and left sides of basal body 1, respectively (Fig. 12, 15, 19, 20).

The posterior flagellum (1) and the anterior flagellum (2) each have a normal 9+2 axoneme (Fig. 13, 14, 17). The transition zone between the 9+2 structure and the basal body is slightly below the actual level of insertion into the cell membrane (Fig. 5, 15). The posterior flagellum (1) has two vanes, one ventral and one dorsal (Fig. 13). The vane on the ventral side of the axoneme is broader than the axoneme (Fig. 13). It originates within 500 nm of flagellar emergence (Fig. 6), in association with a circular reinforcing structure that is connected to its inner edge (Fig. 14, arrow). The dorsal vane is much narrower (Fig. 13), originates after the ventral vane, and terminates sooner.

Basal body 2 connects to a single anterior root (AR) of about six microtubules, which is supported on the anterior side by thin dense material, and runs close to the cell membrane (Fig. 17, 18). The AR is closely associated with microtubules of a dorsal fan (Fig. 17, 18), which presumably support the dorsal cell membrane. Some of the dorsal fan microtubules run between the AR and the cell membrane, roughly parallel to the AR microtubules, and may originate there (Fig. 17, 18). The AR curves over the anterior of the cell and then begins to travel down the left side of the cell (Fig. 15, 18).

Basal body 1 is associated with the major structures that support the ventral groove, namely the right root (RR), the left root (LR), and four non-microtubular fibres (i.e. B, I, A, and C fibres), as well as the singlet microtubular root. The RR and its associated B, I, and A fibres, and the singlet microtubular root support the right side and margin of the groove, while the LR and C fibre support the left side of the groove, with its less well-defined margin (Fig. 15, 16, 19, 21, 22). The RR originates on the right side of basal body 1 (Fig. 19, 20), but is also very close to the base of basal body 2, though not aligned with it (Fig. 18). The RR rapidly grows to $\sim$ 18 microtubules that form a single curved row, and almost immediately splits into an inner portion of six microtubules (IRR) and an outer portion of $\sim$ 12 microtubules (ORR, Fig. 19–21). Close to the opening of the groove, the number of microtubules of the ORR increases, with $>$30 microtubules seen in some sections, but the IRR remains as about six microtubules (Fig. 25). The A fibre is a very thin element closely associated with the dorsal side of the RR, in particular connected to the IRR (Fig. 21). The I fibre is closely associated with the ventral side of the RR. It has a total thickness of $\sim$ 65 nm and has a fine latticework appearance, though with a more dense outer sheet that may appear double leaved (Fig. 20–22). After the splitting of the RR into IRR and ORR, the I fibre is associated only with the outer portion of the ORR (Fig. 19–21). The B fibre is a complex of elements that originates against the ventral side of the LR (Fig. 19–21). The B fibre runs across the ventral side of basal body 1, left to right (Fig. 21, 22), then descends down the right side of the groove (Fig. 23). Near its origin it is up to $\sim$ 150 nm wide, appearing indistinct but with more dense margins, at least in our freeze-substitution fixes (Fig 20, 21). Longitudinal sections show that the B-fibre complex includes an element with conspicuous lateral striations every $\sim$ 25 nm (Fig. 16). The singlet microtubular root originates near basal body 1 and the dorsal side of the RR, and runs parallel to, and to the left of, the IRR (Fig. 19–21). As the groove opens and broadens ventrally the B fibre supports the right margin of the groove, with the ORR initially lying closer to the base of the right margin, while the IRR and singlet microtubule associate with the floor of the groove, with the singlet near the midline of the groove (Fig. 23). The groove floor between the IRR and ORR is supported by microtubules that diverge individually from the left side of the ORR.

The LR originates near the left side of basal body 1 (Fig. 19–21) and expands rapidly to be composed of six microtubules in a single row (Fig. 24). It is supported on its dorsal side by a complex C fibre (Fig. 22, 24) and on its ventral side by multilayered material that we regard as part of the B fibre, in other words the B-fibre complex extends distally to support at least the anterior portion of the LR (Fig. 20–22, 24). The C fibre is about 150 nm thick in total, and consists of two main components—first a series of vanes connecting to each microtubule and projecting dorsally, and second, dorsal to that, a more dense multilayered structure that is narrow left to right (Fig. 22, 24). The C fibre continues to support the LR as the groove opens and appears to support the origins of additional microtubules to the outer edge of the LR (Fig. 15). We infer that microtubules are lost from the inner edge of the LR at the same time, with at least some of them becoming individual microtubules that support the left half of the floor of the groove. The LR and C fibre continue to support the left margin of the groove, which is less well marked than the right margin (Fig. 23).

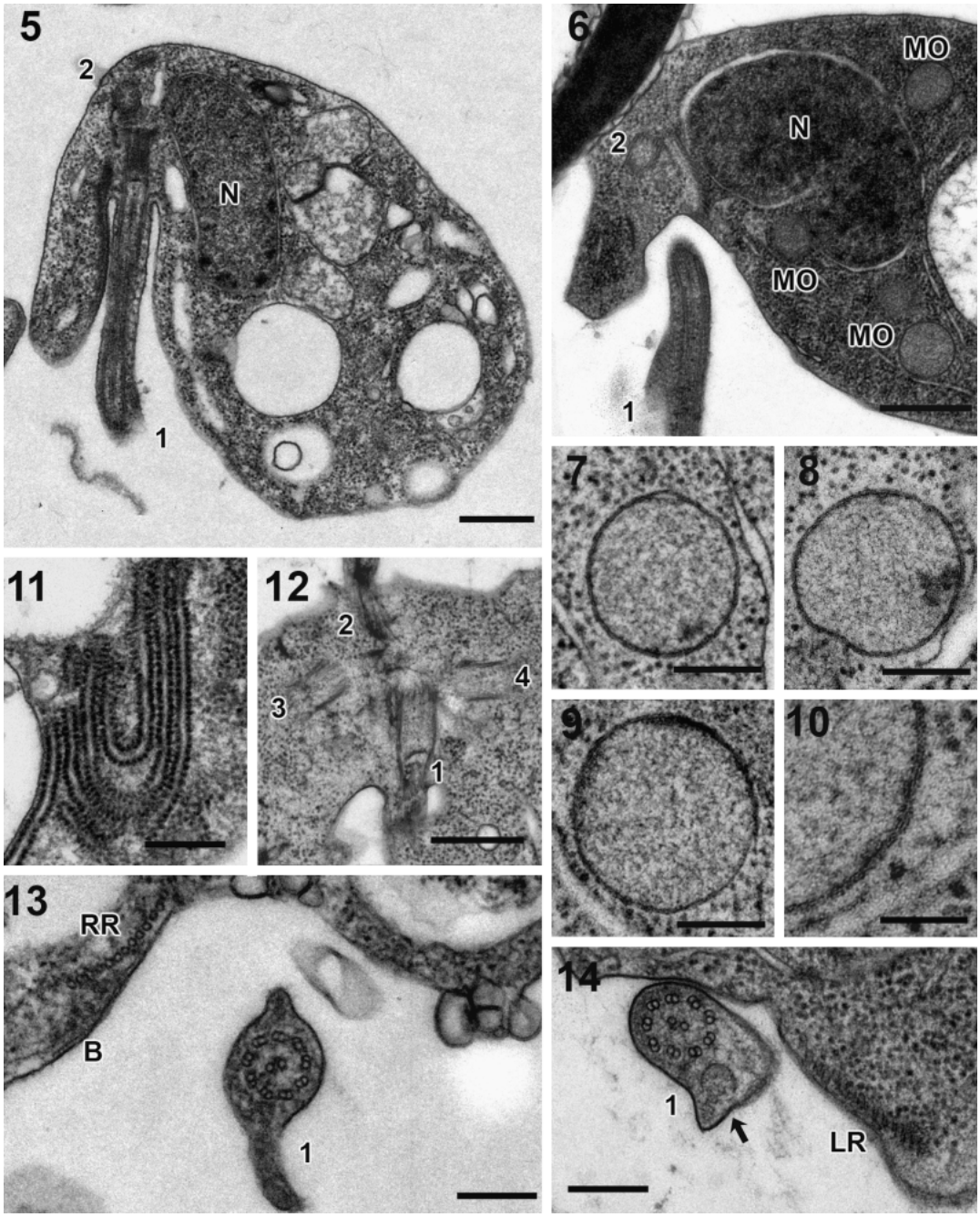


Fig. **5–14.** Transmission electron micrographs of *Ergobibamus cyprinoides* n. g., n. sp., ultrathin sections. **5–6.** Cells in oblique sections showing the anterior portion of the cell. **7–9.** Mitochondrion-like organelles (MO). **10.** High magnification view of the two closely adpressed bounding membranes of an MO. **11.** Stacked endoplasmic reticulum in the cytoplasm. **12.** Section showing the four basal bodies. **13.** Transverse section of the anterior portion of the groove showing the posterior flagellum with two flagellar vanes, ventral much broader than dorsal. **14.** Posterior flagellum in transverse section showing the origin of the ventral vane in association with a circular structure (arrow). 1, flagellum/basal body 1 (= posterior flagellum/basal body); 2, flagellum/basal body 2 (= anterior flagellum/basal body); 3 and 4, non-flagellated basal bodies; B, B fibre; LR, left root; N, nucleus; RR, right root. Scale bars for Fig. 5, 6, 12: 500 nm; scale bars for Fig. 7–9, 11 and 13–14: 200 nm; scale bar for Fig. 10: 100 nm. Figures 6–10, 12, and 14 show material fixed by freeze-substitution fixation; all other micrographs are of material fixed using conventional chemical fixation.

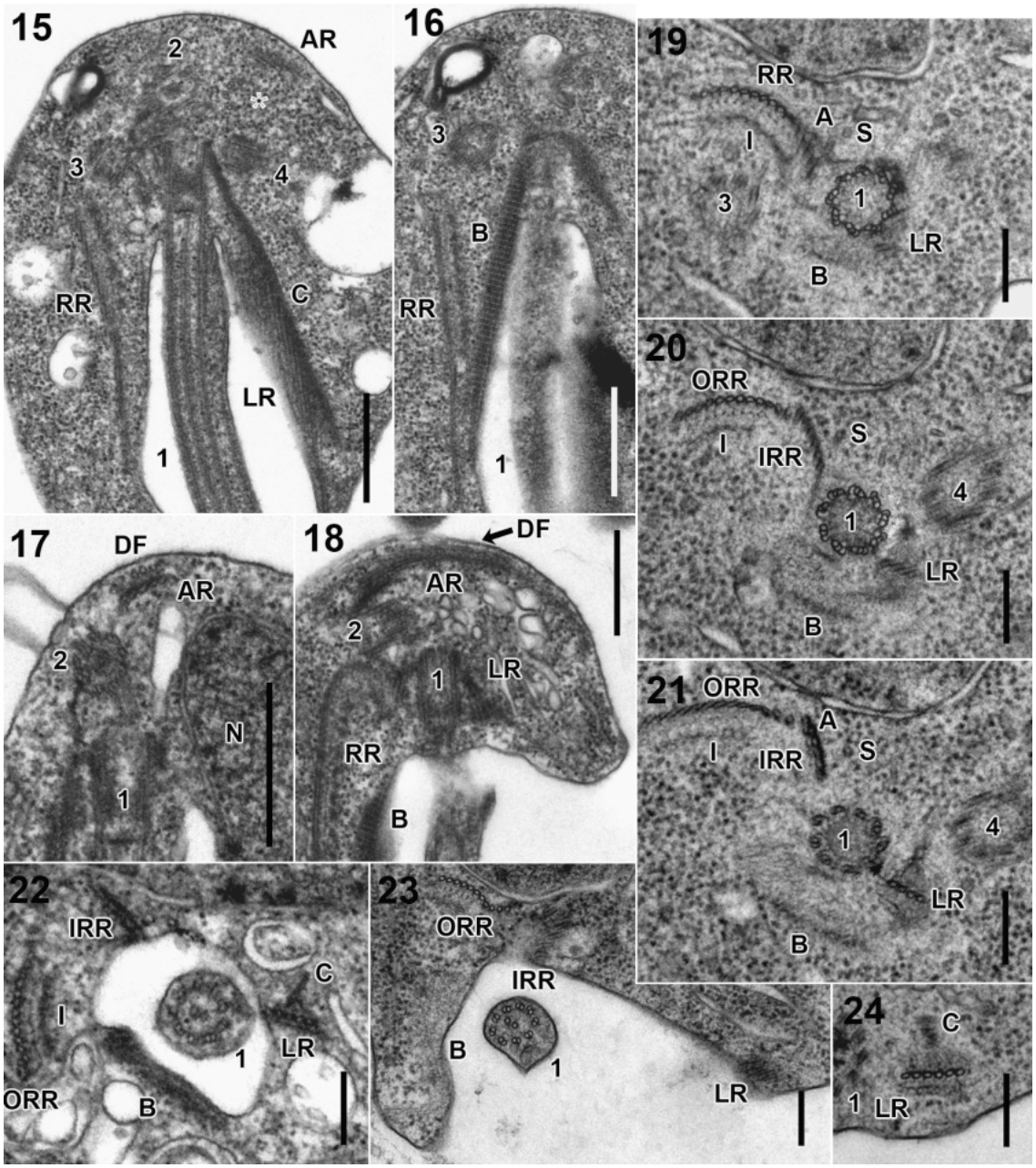


Fig. **15–24.** Transmission electron micrographs of *Ergobibamus cyprinoides* n. g., n. sp., ultrathin sections. **15–16.** Ventral view of the cell. Asterisk: partial elements of another non-flagellated basal body. **17.** Transverse section of the anterior root (AR), showing proximity to the dorsal fan (DF). **18.** Near-longitudinal section of the AR. **19–21.** Non-consecutive serial sections of basal body 1, right root (RR), and left root (LR), singlet microtubular root (S), and the several non-microtubular fibres associated mainly with the right root, namely the A fibre (A), B fibre (B), and I fibre (I). **22.** Extreme anterior end of the groove (i.e. just below the insertion of flagellum 1), showing location of the LR and the inner and outer portions of the right root (IRR, ORR), as well as the associated non-microtubular fibres. **23.** Transverse section of the anterior end of the groove. **24.** Transverse section of the LR and associated C fibre (C). Note also the material associated with the ventral side of the LR 1, flagellum/basal body 1 (= posterior flagellum/basal body); 2, flagellum/basal body 2 (= anterior flagellum/basal body); 3 and 4, non-flagellated basal bodies; N, nucleus. Scale bars for Fig. 15–18: 500 nm; scale bars for Fig. 19–24: 200 nm. Fig. 19–21, 23, and 24 show material fixed by freeze-substitution fixation; other micrographs are of material fixed using conventional chemical fixation.

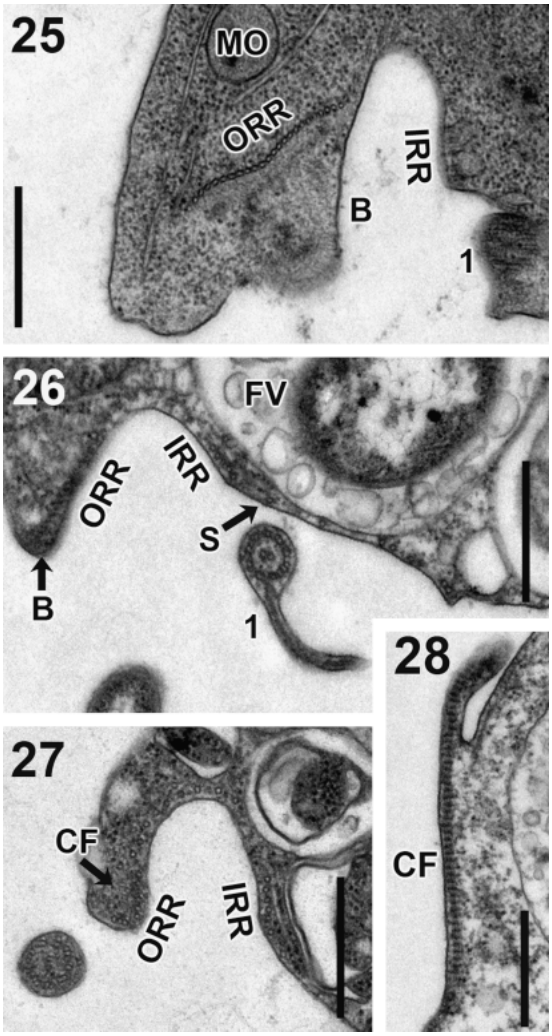


Fig. **25–28.** Transmission electron micrographs of *Ergobibamus cyprinoides* n. g., n. sp., ultrathin sections. **25.** Transverse section of the outer portion of the right root (ORR) relatively close to the anterior end of the groove, showing ~ 30 microtubules. **26.** Right margin of the groove more posteriorly, showing the ORR closely associated with the B fibre (B). **27.** Transverse section of the right margin of the groove in the posterior portion of the cell. Note the absence of the B fibre, but presence of a difficult-to-see composite fibre (CF). **28.** Longitudinal section of CF showing the striations about 35 nm apart. 1, flagellum 1 (= posterior flagellum); FV, food vacuole; IRR, inner portion of the right root; MO, mitochondrion-like organelle. Scale bars: 500 nm. Fig. 25 shows material fixed by freeze-substitution fixation.

Further down the cell the LR is greatly reduced or lost altogether, although its precise termination was not observed in this study, while the ORR and B fibre become more closely associated under the right margin of the groove, and the B fibre is gradually reduced (Fig. 26). In the posterior portion of the groove, the B fibre terminates, but a composite fibre (CF) originates against the cytoplasmic side of the ORR microtubules (Fig. 27). The CF is indistinct in transverse section, at least with freeze substitution fixation (Fig. 27). Transverse striations with a period of ~ 35 nm can be observed in longitudinal section (Fig. 28).

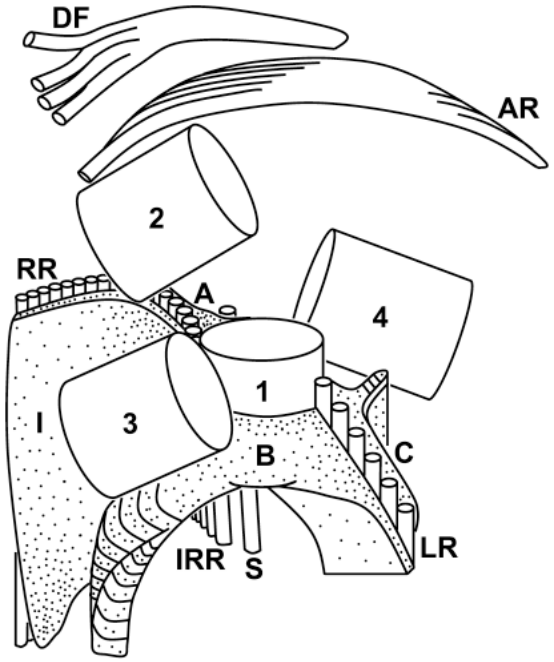


Fig. **29.** Diagram illustrating the proximal flagellar apparatus of *Ergobibamus cyprinoides* n. g., n. sp., as seen from the ventral side. 1–4, Basal bodies 1–4; A, A fibre; AR, anterior root; B, B-fibre complex; C, C fibre; DF, dorsal fan; I, I fibre; IRR, inner portion of the right root; LR, left root; RR, right root; S, singlet root.

A diagrammatic representation of the proximal flagellar apparatus is shown in Fig. 29.

DISCUSSION

At present the genera *Carpediemonas*, *Dysnectes*, *Hicanonectes*, and *Kipferlia* are each monospecific. For brevity, this discussion will often use the genus name alone when describing the characteristics of the sole species in that genus.

At the level of light microscopy Isolate CL is very similar to *C. membranifera*, *D. brevis*, and *K. bialata*, which are all bean-shaped cells with conspicuous grooves and which usually swim slowly with little or no rotation (Table 1; Kolisko et al. 2010; Yubuki et al. 2007). Isolate CL is readily distinguished from *H. teleskopos*, which is more ovoid, has a more subtle groove terminating in a cytopharynx visible by light microscopy, and swims with rapid rotation (Park et al. 2009). Recent SSU rRNA gene data show that Isolate CL and the similar Isolate NC form a clade that is very distinct from all other clades of CLOs (<75% sequence similarity; Kolisko et al. 2010). In addition, this lineage appears distinct in comparative analyses of nuclear protein-coding genes (Kolisko, unpubl. data). Phylogenies of SSU rRNA genes do not support a specific relationship with any one other lineage in particular (Fig. 1; Kolisko et al. 2010). Therefore, assigning our new isolate to any existing genus would unite organisms that are quite dissimilar genetically, and would almost certainly result in a paraphyletic or polyphyletic taxon. As discussed below, Isolate

Table 1. Morphological features of *Ergobibamus cyprinoides* n. gen., n. sp. compared with other *Carpediemonas*-like organisms.

	Cell shape	Discrete cytopharynx	Swimming pattern	MO shape	Dictyosome observed	# Basal bodies	# Flagellar vanes	Dorsal vane	B fibre down LR	Maximum # ORR mts	Size of LR	Dorsal extension to C fibre	AR, # mts	Dorsal fan
Ergobibamus cyprinoides	Bean-shaped	Absent	Slow, little rotation	Rounded	–	4[a]	2	Narrow	+	~ 30	Large	+	6	+
Carpediemonas membranifera	Bean-shaped	Absent	Slow, little rotation	Elongate	+	3	3	Broad	–	11	Small	–	2	+
Dysnectes brevis	Bean-shaped	Absent	Very slow, nodding	Rounded	–	2	2	Narrow	+	~ 23	Large	–[b]	1	–
Hicanonectes teleskopos	Ovoid	Present	With rapid rotation	Rounded	–	4	2	Narrow	–	> 13 (~ 20?)	Large	–	9	+

[a]Additional doublet elements seen in one cell (see Fig. 15, asterisk).
[b]No narrow extension, but C fibre somewhat thickened.
#, number of; +, present; –, absent; AR, anterior root; LR, left root; MO, mitochondrion-like organelle; mts, microtubules; ORR, outer portion of right root.

CL differs somewhat at the level of ultrastructure from *Dysnectes* and *Carpediemonas*. There are few ultrastructural data yet from *Kipferlia*, but Isolate CL lacks the features held to be diagnostic for this taxon (Kolisko et al. 2010). Bearing in mind the level of sequence and ultrastructural dissimilarity from these other taxa, and especially, the risk of creating paraphyletic or polyphyletic taxa, the most appropriate course of action is to assign Isolate CL to a new genus as well as a new species. We here propose the name *E. cyprinoides* n. g. n. sp. (formal description below).

The sub-cellular morphology of *E. cyprinoides* Isolate CL is broadly similar to that of *Carpediemonas*, *Dysnectes*, and *Hicanonectes* (Park et al. 2009; Simpson and Patterson 1999; Yubuki et al. 2007). Like these other taxa our new isolate has the full suite of ''typical excavate'' characters: a ventral suspension-feeding groove, flagellar vanes on the posterior flagellum (flagellum 1), a split RR, a ''singlet'' root, non-microtubular ''B,'' ''C,'' and ''I'' fibres, and a conspicuous CF (see Simpson 2003, noting that the identification of CF of *Hicanonectes* is tentative; Park et al. 2009).

Like *Carpediemonas*, *Dysnectes*, and *Hicanonectes*, our new isolate has relatively large, cristae-lacking organelles in place of classical mitochondria, has both dorsal and ventral flagellar vanes, and has a B-fibre complex that connects to the LR. An origin of the B-fibre complex against the LR is the proposed synapomorphy for the taxon Fornicata, to which *Carpediemonas*, *Dysnectes*, and *Hicanonectes* are assigned, along with retortamonads and diplomonads (Park et al. 2009; Simpson 2003; Simpson and Patterson 1999; Yubuki et al. 2007). Its presence therefore supports the molecular phylogenetic evidence that *Ergobibamus* belongs to Fornicata (Fig. 1; Kolisko et al. 2010). The first two features are probably plesiomorphic for Fornicata: both are present also in the preaxostylan *Trimastix* (Brugerolle and Patterson 1997; Simpson, Bernard, and Patterson 2000) while parabasalids also have cristae-lacking hydrogenosomes but lack typical excavate flagellar vanes. Preaxostyla and parabasalids are likely to be the closest relatives of Fornicata, and the three clades together may descend from a common ancestor that lacked classical mitochondria (Cavalier-Smith 2003; Hampl et al. 2005, 2009; Simpson and Roger 2004).

Ergobibamus shares various other similarities with some, but not all other CLOs, as well as displaying a few unique features. Overall *Ergobibamus* is perhaps most similar to *Hicanonectes*, in spite of their differing appearance by light microscopy, and least similar to *Carpediemonas* (Table 1). *Ergobibamus*, like *Hicanonectes* and *Dysnectes*, has rounded mitochondrion-like organelles, whereas *C. membranifera* has elongate organelle profiles. Our inability to locate a Golgi body in *Ergobibamus* is typical for Fornicata as a whole, but one was observed in *Carpediemonas* (Simpson and Patterson 1999). The routine presence of two additional non-flagellated basal bodies in interphase was reported previously in *Hicanonectes* (Park et al. 2009), but not in *Dysnectes* (Yubuki et al. 2007), while *Carpediemonas* has a single non-flagellated basal body during interphase (Simpson and Patterson 1999). The relative narrowness of the dorsal flagellar vane of *Ergobibamus* is similar to that of *Hicanonectes* and *Dysnectes* (Park et al. 2009; Yubuki et al. 2007) and contrasts with *Carpediemonas* and retortamonads in which both dorsal and ventral vanes are broad (Simpson and Patterson 1999). A dorsal vane might be altogether absent in *K. bialata* (Kolisko et al. 2010), but more data are required to confirm this. A third, lateral vane is present in *Carpediemonas* and in the retortamonad *Retortamonas* (Brugerolle 1977; Simpson and Patterson 1999) but appears to be absent in all other Fornicata, including *Ergobibamus*. The marked extension of the B-fibre complex down the LR is shared by *Ergobibamus* and *Dysnectes* but not *Carpediemonas* or *Hicanonectes*. Note that a similar extension of the arched fibre is seen in some retortamonads (see Brugerolle 1973, 1977). The

outer portion of the RR in *Ergobibamus* contains a large number of microtubules at its maximum (~ 30), more similar to *Dysnectes* (23) and *Hicanonectes* (at least 13, perhaps ~ 20) than to *Carpediemonas* (11). *Ergobibamus* is broadly similar to *Hicanonectes* and *Dysnectes* in that there is a substantial number of microtubules in, or derived from, the LR, and only about half of the width of the groove is supported by microtubules derived from the RR (Park et al. 2009; Yubuki et al. 2007). This contrasts with *Carpediemonas* where there are only six LR microtubules in total, and almost all of the width of the groove is supported by microtubules derived from the RR (Simpson and Patterson 1999). The narrow-but-deep multilayered ''extension'' of the C fibre on its dorsal side is not found in other CLOs studied to date, although a broader multilayered C fibre is present in *Dysnectes*, as well as in retortamonads (Bernard, Simpson, and Patterson 1997; Yubuki et al. 2007). The general location and path of the AR in *Ergobibamus* is most similar to *Dysnectes* and *Carpediemonas* (Simpson and Patterson 1999; Yubuki et al. 2007). The AR is larger than in *Dysnectes* and *Carpediemonas* with ~ six microtubules vs. one to two, but less extensive than the AR of ~ nine reinforced microtubules seen in *Hicanonectes* (Park et al. 2009; Simpson and Patterson 1999; Yubuki et al. 2007). *Ergobibamus* differs from *Dysnectes* in that the latter has little or no dorsal fan (Yubuki et al. 2007). *Kipferlia bialata* has a distinctive membrane-like extension of the right wall of the groove (Kolisko et al. 2010), which is absent from *Ergobibamus*, as well as from *Carpediemonas*, *Dysnectes*, and *Hicanonectes*.

The cristae-lacking mitochondrion-like organelles of *E. cyprinoides* resemble closely the organelles of *Dysnectes* and *Hicanonectes* in being rounded, relatively large, and present in numbers of several per cell (Park et al. 2009; Yubuki et al. 2007). The organelles of *C. membranifera* are narrower, but more elongate, and perhaps ramifying (Simpson and Patterson 1999). In size and proportion of cell volume the organelles of all of these CLOs differ markedly from the mitosomes of their relatives, the diplomonads (Tovar et al. 2003). The small size of diplomonad mitosomes (i.e. ~ 150 nm spherical equivalent in *Giardia intestinalis*) mirrors their relatively limited functionality—those of *Giardia* appear to have no role in ATP synthesis and their only known function is iron–sulfur cluster synthesis (Hjort et al. 2010; Tachezy and Smid 2008; Tovar et al. 2003). Although somewhat smaller in general, the organelles of CLOs more closely resemble the hydrogenosomes of parabasalids, which are anaerobic and fermentative energy-generating organelles (Benchimol 2008; Hrdý et al. 2004; Hrdý, Tachezý, and Müller 2008). The organelles of CLOs might be more similar biochemically to parabasalid hydrogenosomes than to mitosomes, and for example, have a role in ATP synthesis. Information on inferred and actual biochemistry of a range of CLOs is eagerly awaited.

The comparative biology of CLOs promises to shed much light on the evolutionary origin of diplomonads. However, such data will be much more useful in the light of phylogenies that accurately describe the relationships among the several main lineages of CLOs, diplomonads, and retortamonads, which might themselves constitute more than one clade (Cepicka et al. 2008). At present the only phylogenies available encompassing all of these organisms are based on SSU rRNA genes and do not clearly resolve these relationships (Kolisko et al. 2010). It is hoped that well-sampled multigene phylogenies will provide a better estimate of this phylogenetic framework.

Taxonomic Summary

Assignment: Eukaryota; Excavata; Fornicata

Ergobibamus n. g.

Diagnosis. Free-living, biflagellated, colourless cells bearing a longitudinal groove. The posterior flagellum beats within the groove, and bears vanes—a broader vane on the ventral side and narrower vane on the dorsal side. With a substantial anterior microtubular root (~ six microtubules), and a narrow multilayered element in the ''C'' fibre. Mitochondrion-like organelles lack cristae.

Type species. *Ergobibamus cyprinoides* Park, Kolisko, and Simpson

Etymology

''Ergo bibamus''—''therefore, let us drink'' (Latin, masculine.). The title of a poem by Goethe. This poem expresses sentiments consistent with those of Horace's Ode I.11, the original source of the phrase ''Carpe diem''—seize the day.

Ergobibamus cyprinoides n. sp.

Diagnosis. *Ergobibamus* cells, bean-shaped and 7–11.5 um long, with a posterior flagellum ~ 2–2.5 times the length of the cell body.

Type material. Block of resin-embedded cells for electron microscopy deposited with the Protist Type Specimen Slide Collection, US Natural History Museum, Washington, DC as 2054434. This material constitutes the name-bearing hapantotype for the species.

Type locality. Anoxic layer of marine intertidal sediment, Mahone Bay, Nova Scotia, Canada (44°26′N, 64°21′W).

Etymology. *cyprinoides*, ''Carp-like'' (Latin). When first isolated, the type strain was referred to as ''Carp-like'' (CL), owing to its similarity to *Carpediemonas* when viewed by light microscopy.

Gene sequence. The SSU rRNA gene sequence from *E. cyprinoides*, Isolate CL, was reported by Kolisko et al. (2010) and has the GenBank Accession Number GU827592.

ACKNOWLEDGMENTS

Thanks to Aaron Heiss and Zhiyuan Lu (Dalhousie University) for assistance with high-pressure freezing fixation, and to Ping Li (Dalhousie University) for technical assistance with TEM. Thanks to Jeffrey Silberman (University of Arkansas) for valuable comments on the manuscript. This work is supported by NSERC grant 298366-04 to A.G.B.S., and the Tula Foundation through the Centre for Comparative Genomics and Evolutionary Bioinformatics (CGEB) at Dalhousie University. J.S.P. is supported by a CGEB postdoctoral fellowship. M.K. is supported by the Nova Scotia Health Research Foundation.

LITERATURE CITED

Benchimol, M. 2008. Structure of the hydrogenosome. *In*: Tachezy, J. (ed.), Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer-Verlag, Berlin Heidelberg. p. 75–97.

Bernard, C., Simpson, A. G. B. & Patterson, D. J. 1997. An ultrastructural study of a free-living retortamonad, *Chilomastix cuspidata* (Larsen and Patterson, 1990) n. comb. (Retortamonadida, Protista). *Eur. J. Protistol.*, **33**:254–265.

Bernard, C., Simpson, A. G. B. & Patterson, D. J. 2000. Some free-living flagellates (Protista) from anoxic habitats. *Ophelia*, **52**:113–142.

Brugerolle, G. 1973. Etude ultrastucturale du trophozoite du kyste chez le genre *Chilomastix* Aléxéieff, 1910 (Zoomastigophorea, Retortamonadida Grassé, 1952). *J. Protozool.*, **20**:574–585.

Brugerolle, G. 1977. Ultrastructure du genre *Retortamonas* Grassi 1879 (Zoomastigophorea, Retortamonadida Wenrich 1931). *Protistologica*, **13**:233–240.

Brugerolle, G. & Patterson, D. J. 1997. Ultrastructure of *Trimastix convexa* Hollande, an amitochondriate anaerobic flagellate with a previously undescribed organization. *Eur. J. Protistol.*, **33**:121–130.

Cavalier-Smith, T. 2003. The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, *Carpediemonas*, Eophar yngia) and Loukozoa emend. (Jakobea, *Malawimonas*): their evolutionary

affinities and new higher taxa. *Int. J. Syst. Evol. Microbiol.*, **53**:1741–1758.

Cepicka, I., Kostka, M., Uzlikova, M., Kulda, J. & Flegr, J. 2008. Non-monophyly of Retortamonadida and high genetic diversity of the genus *Chilomastix* suggested by analysis of SSU rDNA. *Mol. Phylogenet. Evol.*, **48**:770–775.

Ciccarelli, D. F., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311**:1283–1287.

Diamond, L. S. 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other lumen-dwelling protozoa. *J. Parasitol.*, **66**:958–959.

Ekebom, J., Patterson, D. J. & Vørs, N. 1996. Heterotrophic flagellates from coral reef sediments (Great Barrier Reef, Australia). *Arch. Protistenkd.*, **146**:251–272.

Embley, T. M. & Martin, W. 2006. Eukaryote evolution, changes and challenges. *Nature*, **440**:623–630.

Hampl, V., Horner, D. S., Dyal, P., Kulda, J., Flegr, J., Foster, P. G. & Embley, T. M. 2005. Inference of the phylogenetic position of oxymonads based on nine genes: support for Metamonada and Excavata. *Mol. Biol. Evol.*, **22**:2508–2518.

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G. B. & Roger, A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic ''supergroups''. *Proc. Natl. Acad. Sci.*, **106**:3859–3864.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K-I. & Hasegawa, M. 1995. Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. *Mol. Biol. Evol.*, **12**:782–793.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K-I. & Hasegawa, M. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.*, **11**:65–71.

Hjort, K., Goldberg, A. V., Tsaousis, A. D., Hirt, R. P. & Embley, T. M. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Phil. Trans. R. Soc. B*, **365**:713–727.

Hrdý, I., Tachezý, J. & Müller, M. 2008. Metabolism of trichomonad hydrogenosomes. *In*: Tachezy, J. (ed.), Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer-Verlag, Berlin Heidelberg. p. 113–145.

Hrdý, I., Hirt, P. R., Dolezal, P., Bardonová, L., Foster, P. G., Tachezy, J. & Embley, T. M. 2004. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature*, **432**:618–622.

Kolisko, M., Silberman, J. D., Cepicka, I., Yubuki, N., Takishita, K., Yabuki, A., Leander, B. S., Inouye, I., Inagaki, Y., Roger, A. J. & Simpson, A. G. B. 2010. A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats. *Environ. Microbiol.*, published online 28 June 2010, doi: 10.1111/j.1462-2920.2010.02239.x.

Larsen, J. & Patterson, D. J. 1990. Some flagellates (Protista) from tropical marine sediments. *J. Nat. Hist.*, **24**:801–937.

Lee, W. J. & Patterson, D. J. 2000. Heterotrophic flagellates (Protista) from marine sediments of Botany Bay, Australia. *J. Nat. Hist.*, **34**:483–562.

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aleym, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E. J., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, G., Svard, S. G. & Sogin, M. L. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*, **317**:1921–1926.

Nohýnková, E., Tůmová, P. & Kulda, J. 2006. Cell division of *Giardia intestinalis*: flagellar developmental cycle involves transformation and exchange of flagella between mastigonts of diplomonad cell. *Eukaryot. Cell*, **5**:753–761.

Park, J. S., Kolisko, M., Heiss, A. A. & Simpson, A. G. B. 2009. Light microscopic observations, ultrastructure, and molecular phylogeny of *Hicanonectes teleskopos* n. g., n. sp., a deep-branching relative of diplomonads. *J. Eukaryot. Microbiol.*, **56**:373–384.

Philippe, H., Zhou, Y., Brinkmann, H., Rodriguez, N. & Delsuc, F. 2005. Heterotachy and long-branch attraction in phylogenetics. *BMC. Evol. Biol.*, **5**, art. 50.

Simpson, A. G. B. 2003. Cytoskeletal organisation phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int. J. Syst. Evol. Microbiol.*, **53**:1759–1777.

Simpson, A. G. B. & Patterson, D. J. 1999. The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the ''excavate hypothesis.'' *Eur. J. Protistol.*, **35**:353–370.

Simpson, A. G. B. & Roger, A. J. 2004. Excavata and the origin of amitochondriate eukaryotes. *In*: Hirt, R. P. & Horner, D. S. (ed.), Organelles, Genomes, and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. CRC Press, Boca Raton. p. 27–53.

Simpson, A. G. B., Bernard, C. & Patterson, D. J. 2000. The ultrastructure of *Trimastix marina* Kent 1880 (Eukaryota), an excavate flagellate. *Eur. J. Protistol.*, **36**:229–252.

Simpson, A. G. B., MacQuarrie, E. K. & Roger, A. J. 2002a. Eukaryotic evolution: early origin of canonical introns. *Nature*, **419**:270.

Simpson, A. G. B., Roger, A. J., Silberman, J. D., Leipe, D. D., Edgcomb, V. P., Jermiin, L. S., Patterson, D. J. & Sogin, M. L. 2002b. ''Evolutionary history of ''early diverging'' eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol. Biol. Evol.*, **19**:1782–1791.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**:75–77.

Tachezy, J. & Smid, O. 2008. Mitosomes in parasitic protists. *In*: Tachezy, J. (ed.), Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer-Verlag, Berlin Heidelberg. p. 201–230.

Tovar, J., León-Avilla, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., Hernandez, M., Müller, M. & Lucocq, J. M. 2003. Mitochondria remnant organelles of *Giardia* function in iron-sulfur protein maturation. *Nature*, **426**:172–176.

Yubuki, N., Inagaki, Y., Nakayama, T. & Inouye, I. 2007. Ultrastructure and ribosomal RNA phylogeny of the free-living heterotrophic flagellate *Dysnectes brevis* n. gen, n. sp., a new member of Fornicata. *J. Eukaryot. Microbiol.*, **54**:191–200.

Received: 03/18/10, 07/31/10; accepted: 08/04/10

16. Appendix I

Article analyzing phylogeny of *Carpediemonas*-like organisms using 7 genes that I co-authored

Multigene phylogenies of diverse *Carpediemonas*-like organisms identify the closest relatives of 'amitochondriate' diplomonads and retortamonads

Kiyotaka Takishita[a,2], Martin Kolisko[b,c,2], Hiroshi Komatsuzaki[d], Akinori Yabuki[d], Yuji Inagaki[d,e], Ivan Cepicka[f], Pavla Smejkalová[f,g], Jeffrey D. Silberman[h], Tetsuo Hashimoto[d,e], Andrew J. Roger[b] and Alastair G.B. Simpson[c,1]

[a]Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Kanagawa, 237-0061, Japan
[b]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada
[c]Department of Biology, Dalhousie University, Halifax, Nova Scotia, B3H 1X5, Canada
[d]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8572, Japan
[e]Center for Computational Sciences and Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8577, Japan
[f]Department of Zoology, Faculty of Science, Charles University in Prague, Prague, 128 44, Czech Republic
[g]Department of Parasitology, Faculty of Science, Charles University in Prague, Prague, 128 44, Czech Republic
[h]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, 72701, USA
[2]Equally contributing authors

Corresponding author: *E-mail alastair.simpson@dal.ca; Tel. (+1) 902 494 1247; Fax (+1) 902 494 3756.

Running head: Multigene phylogenies of Fornicata

Abstract

Diplomonads, retortamonads, and "*Carpediemonas*-like" organisms (CLOs) are a monophyletic group of protists that are microaerophilic/anaerobic and lack typical mitochondria. Most diplomonads and retortamonads are parasites, and the pathogen *Giardia intestinalis* is known to possess reduced mitochondrion-related organelles (mitosomes) that do not synthesize ATP. By contrast, free-living CLOs have larger organelles that superficially resemble some hydrogenosomes, organelles that in other protists are known to synthesize ATP anaerobically. This group represents an excellent system for studying the evolution of parasitism and anaerobic, mitochondrion-related organelles. Understanding these evolutionary transitions requires a well-resolved phylogeny of diplomonads, retortamonads and CLOs. Unfortunately, until now the deep relationships amongst these taxa were unresolved due to limited data for almost all of the CLO lineages. To address this, we assembled a dataset of up to six protein-coding genes that includes representatives from all six CLO lineages, and complements existing rRNA datasets. Multigene phylogenetic analyses place CLOs as well as the retortamonad *Chilomastix* as a paraphyletic basal assemblage to the lineage comprising diplomonads and the retortamonad *Retortamonas.* In particular, the CLO *Dysnectes* was shown to be the closest relative of the diplomonads + *Retortamonas* clade with strong support. This phylogeny is consistent with a drastic degeneration of mitochondrion-related organelles during the evolution from a free-living organism resembling extant CLOs to a probable parasite/commensal common ancestor of diplomonads and *Retortamonas.*

Key words: *Carpediemonas*-like organisms; diplomonads; Excavata; hydrogenosomes; mitochondria; mitosomes.

Introduction

Diplomonads (e.g., the human pathogen *Giardia intestinalis*) and their close relatives retortamonads are a collection of heterotrophic flagellates that includes many commensal and parasitic species, and some free-living forms (Brugerolle and Lee 2000). They are microaerophiles that lack typical mitochondria, and which tend to branch at very deep positions among eukaryotes in molecular phylogenetic analyses rooted with prokaryotic outgroups (Ciccarelli et al. 2006; Hashimoto et al. 1994, 1995; Leipe et al. 1993; Sogin 1989; Sogin et al. 1989). It was once proposed that these protists were primitive eukaryotes that had diverged before the acquisition of the mitochondrion (Cavalier-Smith 1983; Sogin et al. 1989). However, it is now widely believed that the 'long-branch attraction artifact' is responsible for the basal position of diplomonads, as they tend to form extremely long branches in phylogenetic trees (e.g., Philippe et al. 2000; Philippe and Germot 2000). Further, extremely reduced and tiny (~50-150 nm across) mitochondrion-related organelles, called mitosomes, have been identified in *Giardia* (Tovar et al. 2003). The mitosomes of *Giardia* do not synthesise ATP and their only known function is in iron-sulfur cluster synthesis (Hjort et al. 2010; Tovar et al. 2003). The characteristics of mitosomes contrast sharply with those of classical mitochondria, and also with the mitochondrion-related organelles of the best-studied relatives of diplomonads, the parabasalids (e.g., the human pathogen *Trichomonas vaginalis)*. The hydrogenosomes of parabasalids are much larger (~500 nm across - Benchimol 2001; Clemens and Johnson 2000), and they produce ATP by substrate-level phosphorylation through an anaerobic pathway that yields hydrogen as a waste product (Lindmark and Müller 1973; Müller 1993; Steinbüchel and Müller 1986).

A specific relationship between diplomonads and parabasalids is supported by various molecular phylogenetic data (e.g., Andersson et al. 2005; Arisue et al. 2005; Hampl et al. 2005, 2009; Henze et al. 2001), however the split between these lineages is likely ancient. A decade ago it was shown that diplomonads and retortamonads are actually more closely related to an obscure free-living microaerophilic heterotrophic flagellate called *Carpediemonas membranifera* (Simpson et al. 2002, 2006; Simpson and Patterson 1999), with the diplomonad + retortamonad + *Carpediemonas* clade becoming known as 'Fornicata' (Simpson 2003). Recently, some five additional major lineages of "*Carpediemonas*-like" organisms (CLOs) have been discovered, four represented by the species *Dysnectes brevis*, *Hicanonectes teleskopos*, *Kipferlia bialata,* and *Ergobibamus cyprinoides* (Kolisko et al. 2010; Park et al. 2009, 2010; Takishita et al. 2007; Yubuki et al. 2007) and one lineage with no described members, which was labeled 'CL2' by Kolisko et al. (2010). Small subunit rRNA (SSU rRNA) gene phylogenies suggest that CLOs (including *Carpediemonas*) form a series of branches at the base of the diplomonad + retortamonad lineage. Interestingly, all CLOs studied in detail by transmission electron microscopy (TEM) harbor mitochondrion-like organelles that are much larger than the mitosomes of *Giardia* and superficially resemble parabasalid hydrogenosomes (Park et al. 2009, 2010; Simpson and Patterson 1999; Yubuki et al. 2007). These findings make CLOs key to understanding the evolutionary origins of diplomonad cells, especially the extreme reductive evolution of mitochondrion-related organelles that occurred in the diplomonad lineage.

In order to correctly describe the evolutionary history within Fornicata it is necessary to have an accurate picture of the phylogenetic relationships among diplomonads, retortamonads, and the major CLO lineages. These relationships are poorly

resolved by SSU rRNA gene phylogenies, with differing results in different analyses, and generally low support for most deep nodes within Fornicata (Kolisko et al. 2010). Such lack of phylogenetic resolution in SSU rRNA gene trees may be attributed to limited informative phylogenetic signal. To date, however, multigene analyses of Fornicata have included at most one CLO and/or one retortamonad, and therefore say nothing about interrelationships amongst the major CLO clades (Kolisko et al. 2008; Simpson et al. 2002, 2006).

In the present study we conducted phylogenetic analyses based on up to seven genes that include representatives of all six major CLO lineages known to date. Several deep branches within the taxon Fornicata were resolved with strong statistical support, with CLOs representing a paraphyletic assemblage at the base of Fornicata, and retortamonads not recovered as a monophyletic group. In particular, we found strong evidence that *Dysnectes* is a close relative of diplomonads (and *Retortamonas*).

Results

Single gene phylogenies

Global eukaryotic phylogenies for the single gene datasets of SSU rRNA, α-tubulin, β-tubulin, the cytosolic isoforms of the 70 kDa and 90 kDa heat shock proteins (HSP70 and HSP90), and translation elongation factors 1α and 2 (EF-1α and EF2) were estimated with maximum likelihood (ML) and Bayesian methods (supplementary materials: Figs. S1-S7). Diplomonads, retortamonads and CLOs (i.e., the taxon Fornicata) formed a monophyletic lineage in the phylogenetic trees of SSU rRNA genes, α-tubulin, HSP90, HSP70, and EF-1α. This grouping received moderate-to-strong statistical support - 79%-93% bootstrap proportion (BP) and 0.99-1.00 Bayesian posterior probabilities (PP) - in the case of SSU rRNA, HSP90, and α-tubulin, respectively, and low statistical support in the case of HSP70 and EF-1α. In the ML and Bayesian trees of EF2, members of the Fornicata other than *Carpediemonas* were weakly clustered, and *Carpediemonas* branched with the heterolobosean *Naegleria gruberi* with low statistical support. In the ML and Bayesian trees of β-tubulin, the parabasalid clade was positioned cladistically within Fornicata. On face value this might suggest a possible lateral gene transfer, however the exact position of the parabasalid lineage within the Fornicata radiation is poorly supported, and a sister relationship between parabasalids and Fornicata cannot be ruled out.

As in the case of previous well-sampled SSU rRNA gene phylogenies (Kolisko et al. 2010), all single gene phylogenies weakly resolved many of the relationships amongst diplomonads, retortamonads, and major clades of CLOs. Nonetheless, the phylogenies of α-tubulin, EF-1α, EF2, and HSP90 all strongly suggested an evolutionary affinity between *Dysnectes* and diplomonads, or diplomonads plus the retortamonad taxon *Retortamonas* (86-100% BP and 0.99-1.00 PP).

Multiple gene phylogenies

In order to more robustly resolve the deep relationships within the taxon Fornicata we assembled several multigene datasets: These consisted of 4-5 genes (HSP70, HSP90, EF-1a and EF2, with or without SSU rRNA) and a diverse outgroup, or 6-7 genes (all six protein coding genes, with or without SSU rRNA) and a limited outgroup. These sets of protein-coding genes were assessed as potentially congruent when using these particular outgroups according to an analysis with the program Concaterpillar (see methods).

The analyses of the multigene datasets resolved the taxon Fornicata as a monophyletic lineage. This clade received 99% ML bootstrap support in the 4-gene and 5-gene analyses, and a posterior probability of 1.00, estimated using the CAT model implemented in PhyloBayes, in the Bayesian analysis of the 4-gene dataset (supplementary materials Figs. S8 and S9). Unsurprisingly, the Fornicata clade was also strongly supported in 6-gene and 7-gene analyses, with a small outgroup.

With respect to the ingroup, analyses of the 6-gene dataset and 7-gene dataset both yielded overall topologies that were essentially identical (Fig. 1 & 2). The diplomonad lineage and *Retortamonas* formed a strongly supported clade (100% BP with 6-gene and 7-gene ML analyses, and 1.00 PP with 6-gene Bayesian analysis). The other Fornicata members formed a series of branches attached sequentially to the base of the diplomonads + *Retortamonas* clade. In order, these were (i) *Dysnectes* (CL1), (ii) *Kipferlia* (CL6), (iii) *Chilomastix* (retortamonad), (iv) a clade comprising *Ergobibamus* (CL5), *Hicanonectes* (CL3) and the unnamed clade CL2, and (v) *Carpediemonas* (CL4). Statistical support measures for two nodes - one uniting *Dysnectes* and the diplomonads+*Retortamonas* clade and the other uniting *Dysnectes*, *Kipferlia*, *Chilomastix* and the diplomonads+*Retortamonas* clade - were 100% BP in the 6-gene and 7-gene ML analyses and 1.00 PP in the 6-gene Bayesian analysis (these strongly supported nodes are designated as nodes A and B in Figs. 1 & 2), while other deep branching points received low or very low support. *Hicanonectes* and CL2 were clustered with each other with 100% BP in the 6-gene and 7-gene ML analyses and 1.00 PP in the 6-gene Bayesian analysis. This grouping was also recovered with 100% BP if the relatively data-poor CL2 isolate PCE was excluded from the 6-gene dataset (tree not shown; no Bayesian analysis performed). The clade of CL2 and CL3 with CL5 received low ML bootstrap support, but had a high posterior probability in the Bayesian analysis (6-gene dataset; Fig. 2). The results for the AU tests were in good agreement with bootstrap analyses: All topologies that were not rejected at the 5% level contained nodes A and B (data not shown).

The results of the 4-gene and 5-gene analyses (Figs S8, S9) are congruent with the 6-gene and 7-gene analyses with the exception of the position of *Hicanonectes*. Instead of *Hicanonectes* forming a clade with CL2, *Hicanonectes* branched as the sister taxon to the *Ergobibamus* clade (CL5) with 75% BP and 1.00 PP in the 4-gene analyses, and with less than 50% BP in the 5-gene analyses. This incongruent result may reflect two factors, firstly the absence of α-tubulin and β-tubulin data from the 4-5 genes analyses (both genes individually provide strong support for a *Hicanonectes*-CL2 clade - supplementary materials Figs. S2 and S3), and secondly, a positive phylogenetic signal in the HSP90 dataset: Phylogenies for HSP90 support a clade of *Hicanonectes* with CL5, and statistical support is high (91% BP) when a large outgroup is used (supplementary material Fig. S5). Interestingly, however, support is lower (63% BP) when the small outgroup is used, as per the 6-gene and 7-gene analysis (not shown). On balance we tentatively favor the topology shown in the 6-gene and 7-gene analyses - *Hicanonectes* specifically related to CL2 - since it reflects strong support from two different gene partitions, and the conflicting signal in the HSP90 data is somewhat sensitive to taxon sampling.

Discussion

CLOs and *Chilomastix* are a basal grade within Fornicata; *Dysnectes* is the CLO closest to diplomonads

Several previously published SSU rRNA gene phylogenies have shown that the diplomonads, retortamonads, and CLOs form a robust monophyletic grouping among excavates or eukaryotes (Keeling and Brugerolle 2006; Kolisko et al. 2008, 2010; Simpson et al. 2002; Yubuki et al. 2007), but the deep relationships within this Fornicata lineage remained unclear. This poor and/or inconsistent resolution of relationships amongst CLO lineages left open the possibility that CLOs were actually a clade. Our results instead strongly indicate that CLOs are not a monophyletic group.

Most strikingly, clade CL1, represented by *Dysnectes*, was found to be the closest relative of the diplomonads + *Retortamonas* clade among the CLOs examined here. This position was strongly supported in all multigene analyses. Interestingly, it is mostly or entirely consistent with some previous SSU rRNA gene phylogenies (Kolisko et al. 2010; Park et al. 2009), although the most taxon-rich analysis published to date actually placed both *Dysnectes* and unnamed clade CL2 as the sistergroup to diplomonads and retortamonads (*Retortamonas* and *Chilomastix*), with poor support (Kolisko et al. 2010). Despite the availability of a detailed ultrastructural study of *Dysnectes brevis* (Yubuki et al. 2007), we are unable, however, to nominate any obvious potential apomorphies shared by *Dysnectes* and diplomonads to the exclusion of other CLOs. It would be interesting to compare to the cell structure of *Dysnectes* to that of the *Retortamonas* isolates for which there are SSU rRNA gene sequences, once these data are available (see below).

Further, a clade of diplomonads, *Retortamonas*, *Dysnectes*, *Chilomastix* and *Kipferlia* (CL6) also obtained maximal statistical support in our multigene phylogenies, even though the precise positions of *Chilomastix* and *Kipferlia* within this clade were not resolved. Meanwhile, the remaining CLO lineages were not monophyletic either, usually forming two distinct clades attached sequentially to the base of Fornicata. In particular, the CL4 clade, represented by *Carpediemonas,* branched in the deepest position within the Fornicata lineage, albeit with low statistical support. In all, our analysis strongly suggests that CLOs (and retortamonads) collectively represent several lineages - each at least as phylogenetically distinct as diplomonads - that connect as a series to the stem of the diplomonads.

The evolution of diplomonads and their mitosomes

The now-strong inference that CLOs represent several successive branches at the base of the diplomonad + *Retortamonas* lineage suggests that CLOs are paraphyletic, in other words, that diplomonads are descended from organisms similar to extant CLOs. Widespread features of extant CLOs were probably also features of these direct diplomonad ancestors; For example, they are likely to have been small free-living biflagellated cells with a typical excavate feeding groove. Further, since all extant CLOs discovered to date live in saline habitats it is possible that saline water represents the ancestral habitat for diplomonads and retortamonads. Concerted searches for CLOs in freshwater habitats should be performed to test this hypothesis.

This phylogenetic scheme is helpful in understanding the evolution of the mitochondrion-related organelles in the Fornicata lineage. *Giardia* has extremely small and functionally reduced mitochondrion-derived organelles, mitosomes, that do not produce ATP (Tovar et al. 2003), and to date there is little evidence of large and/or ATP-generating mitochondrial organelles in other diplomonads. Very little at all is known about the organelles in retortamonads. By contrast, all CLOs examined to date by TEM

unambiguously possess double-membrane bounded mitochondrion-like organelles that are substantially larger than *Giardia* mitosomes, and relatively abundant (Park et al. 2009, 2010; Simpson and Patterson 1999; Yubuki et al. 2007; see Fig. 3). This includes *Dysnectes,* which is most closely related to diplomonads, as well as *Hicanonectes, Ergobibamus* and *Carpediemonas,* which appear to represent at least two more distantly related lineages in our analyses (see previous section). The metabolic capacity of mitochondrion-related organelles in CLOs remains unknown, but the larger size of these organelles suggests that they might be involved in more diverse metabolic processes than the mitosomes of *Giardia*. Their presence in all studied CLO lineages suggests that ancestors of diplomonads had larger mitochondrion-derived organelles, and that much of the reductive evolution experienced by the mitochondrion-related organelles in diplomonads might have happened only after the divergence of the *Dysnectes* lineage. Given that most known diplomonads and retortamonads are parasites or commensals of vertebrates or invertebrates (Brugerolle and Lee 2000), the last common ancestor of all diplomonads and *Retortamonas* is likely to have been parasitic or commensal. It is possible that the shift from a free-living lifestyle to parasitism/commenism is somehow associated with the reductive evolution of mitosomes.

A possible "free-swimmer" clade

Our most gene-rich analyses recover a strongly supported relationship between CL2, represented by strains NY0171 and PCE, and CL3 represented by *Hicanonectes*. This phylogenetic affinity has not been recovered by previous SSU rRNA gene analyses, in which both CL2 and CL3 have unstable phylogenetic positions (Kolisko et al. 2010). These organisms differ morphologically from other CLOs studied to date. Studied CLOs from clades CL1, CL4, CL5, and CL6 are bean-shaped cells that rotate little or slowly while swimming, often swim quite slowly, and tend to associate with surfaces (Kolisko et al. 2010; Yubuki et al. 2007). Isolates NY0171 and PCE (CL2) and *Hicanonectes* (CL3) by contrast are oval-shaped cells that rapidly rotate during swimming, and tend not to associate specifically with surfaces (Park et al. 2009; Kolisko et al. 2010; see Fig. 3). The feeding groove is less conspicuous than in other CLOs, but terminates in a curving cytopharynx that can be visualized readily by light microscopy. These members of clades CL2 and CL3 could be adapted to a more free-swimming niche than other CLOs, and they may well share these adaptations due to descent from a common ancestor. There are detailed electron microscopical data from *Hicanonectes* (Park et al. 2009), but not yet for any strain from clade CL2. It will be important to compare the cytoskeletal organization of a CL2 representative to that of *Hicanonectes* to test the possibility of special homology.

In addition to *Hicanonectes*, clade CL3 contains isolate PCS, which differs from all other CLOs, being an elongate cell with a single flagellum (not two), and with a very reduced groove (Kolisko et al. 2010). Given the similarity between *Hicanonectes* and CL2, we infer tentatively that the PCS lineage evolved by descent from a free-swimming *Hicanonectes*-like ancestor, but further data on isolate PCS would be valuable.

Uncertain phylogenetic position of *Chilomastix*

All our analyses recovered retortamonads as a non-monophyletic group, with *Chilomastix caulleryi* always branching in a deeper position than *Retortamonas,* with two strongly supported nodes separating them. Previous SSU rRNA gene analyses placed *Chilomastix*

(represented by different species, namely *C. mesnili* and *C. wenrichi*) as sister to the diplomomad + *Retortamonas* clade to the exclusion of *Carpediemonas* and *Dysnectes*, suggesting that retortamonads were paraphyletic (Cepicka et al. 2008). Our analyses of a multigene dataset with comprehensive CLO sampling are more consistent with retortamonad polyphyly than with paraphyly. The apparent non-monophyly of retortamonads, especially polyphyly, would seem to conflict with previous TEM studies that show considerable morphological similarity between *Retortamonas agilis* and *Chilomastix* spp. (Bernard et al. 1997; Brugerolle 1973, 1977; see Simpson and Patterson 1999 for possible synapomorphies). At present, however, there is no overlap in the retortamonad species for which there are published TEM data, and those studied by molecular methods (Cepicka et al. 2008; Silberman et al. 2002). This might explain some of the discrepancy if either *Retortamonas* or *Chilomastix* were not monophyletic. Once the required morphological/sequence data are available, it is very likely that the taxonomy of retortamonads will have to be reconsidered.

Perspectives

The availability of a diversity of CLOs in culture will allow more direct examinations of the evolution of mitochondrion-related organelles in the Fornicata lineage. As a next step, the biochemical functions of the mitochondrion-related organelles in several CLOs should be examined, with *Dysnectes* being of particular interest. One powerful strategy to begin such a program is large-scale expressed sequence tags (EST) analyses. In fact, it has been inferred that the reduced mitochondrion-like organelles in the free-living preaxostylan excavate *Trimastix* and the parasitic stramenopile *Blastocystis* may perform hydrogenosome-like anaerobic ATP generation based on the presence of transcripts for pyruvate: ferredoxin oxidoreductase and [FeFe]-hydrogenase in EST data (Hampl et al. 2008; Stechmann et al. 2008). If the acquisition, retention and loss of various biochemical functions of organelles can be traced through the deepest branches within Fornicata, we will have a much better understanding of when and how the *Giardia intestinalis* mitosome came to be so simple, and exactly what sort of organelle it evolved from.

Methods

Total RNA/DNA extractions and cDNA synthesis

In this study we obtained new sequence data from organisms from all six clades of *Carpediemonas*-like organisms (CLOs) (labeled CL1-CL6 by Kolisko et al. 2010), plus the retortamonad *Chilomastix caulleryi.*

Genomic DNA was isolated from all CLO cultures investigated in this study as reported by Kolisko et al. (2010). *Chilomastix caulleryi* was isolated from a frog *Kassina senegalensis,* which was kept at the Department of Parasitology, University of Veterinary Sciences, Brno, Czech Republic. The frog was sacrificed and dissected, and 0.25 ml of the intestinal contents was pipetted into Dobell and Leidlaw's biphasic medium (Dobell and Leidlaw 1926). The original culture (KAS5) contained *Trichomitus batrachorum* (Parabasalia), *Chilomastix caulleryi* and unidentified bacteria. It was maintained at room temperature in Dobell and Leidlaw's medium, and transferred once a week. Later, the monoeukaryotic strain CHILO1 of *Chilomastix caulleryi* was created from KAS5 by dilution. Originally, it was cultured as per KAS5; later on it was transferred into TYSGM-9 medium (Diamond 1982) without mucin and Tween 80. It was maintained at room temperature and was transferred once a week. The strains KAS5 and CHILO1 have been

deposited in the collection of the Department of Parasitology, Charles University in Prague, Prague, Czech Republic. Genomic DNA was extracted from *C. caulleryi* as reported by Cepicka et al. (2008) for other *Chilomastix* species.

Total RNA was isolated from *Dysnectes brevis* (strain NY0165), *Kipferlia bialata* (strain NY0173), and unnamed strain NY0171 with the RNeasy Plant Mini Kit (Qiagen, USA), and synthesis of cDNA from total RNA was performed using the 3' RACE System (Invitrogen, USA). In addition, total RNA was also isolated from *Carpediemonas membranifera* (strain BICM), *Kipferlia bialata* (strain NY0173), *Chilomastix caulleryi*, and *Ergobibamus cyprinoides* (strain CL) using Trizol (Invitrogen, USA) for expressed sequence tags (EST) analysis (see below).

Gene/transcript amplification, cloning, and sequencing

Some fragments of genes encoding the nucleus-encoded proteins α-tubulin, β-tubulin, HSP70, HSP90, EF-1α, and EF2 as well as the SSU rRNA gene were amplified from genomic DNA or cDNA by PCR with various combinations of degenerate primers (supplementary material Table S1). A referenced list of the broad-range primers used, including newly designed primers, is presented in supplementary material Table S2. HSP90 transcripts from *Kipferlia bialata* (strain NY0173) and unnamed strain NY0171, and EF-1α transcripts from *Dysnectes brevis* (NY0165) and strain NY0171 were amplified using the 3' RACE System (Invitrogen) with exact-match primers designed on the initially amplified DNA fragments, following the manufacturer's instruction. Thermocycling was run for 35-40 cycles in all cases, with annealing temperatures of 48-55°C (except for the SSU rRNA gene of *Chilomastix caulleryi*, where an annealing temperature of 58°C was used). The amplified products were visualized by agarose gel electrophoresis. Amplicons of expected sizes were cloned using the pCR2.1 vector with the TOPO TA Cloning Kit (Invitrogen) or the pGem T-easy vector cloning kit (Promega, USA). The plasmid DNA was extracted from several positive clones grown overnight in liquid LB medium using the Nucleospin Plasmid Kit (Macherey-Nagel, Germany) or QIAprep Miniprep Kit (Qiagen), and inserts were bidirectionally sequenced. Other gene sequences were retrieved from EST data from *Carpediemonas membranifera* (BICM), *Kipferlia bialata* (NY0173), *Chilomastix caulleryi*, and *Ergobibamus cyprinoides* (CL) generated by 454 sequencing (see supplementary material Table S1). The transcripts were assembled into contigs with MIRA 3 (Chevreux et al. 1999) and Newbler (454 Life Sciences, USA), and the corresponding amino acid sequences were then deduced from the contig sequences. The sequences were analyzed using Sequencher version 4.2 (Gene Codes Corporation, USA), Genetyx-Mac version 14 (Software Development, Japan), or Geneious 4.8 (Biomatters Limited, New Zealand). New sequences obtained in the present study were deposited in GenBank (AB600279-AB600326).

Phylogenetic analyses

The nucleotide sequences of SSU rRNA genes and the deduced amino acid sequences of α-tubulin, β-tubulin, HSP70, HSP90, EF-1α, and EF2 from diplomonads, retortamonads, and CLOs were separately aligned with the corresponding sequences from various eukaryotic groups using ClustalX version 2.0 (Thompson et al. 1997). While β-tubulin gene sequences were obtained from all CLOs as well as *Chilomastix*, the datasets for the other five genes contained missing data (see supplementary material Table S1). The alignments were inspected by eye and manually edited, and ambiguously aligned sites

were removed from the datasets prior to the phylogenetic analyses. The analyzed datasets had the following dimensions: SSU rRNA gene: 77 taxa, 875 sites; α-tubulin: 76 taxa, 381sites; β-tubulin: 72 taxa, 400 sites; HSP70: 184 taxa, 508 sites; HSP90: 154 taxa, 559 sites; EF-1α: 88 taxa, 406 sites; EF2: 54 taxa, 778 sites. We also analyzed an HSP90 dataset with only a small outgroup (19 taxa total, 588 sites). The alignment data are deposited in Treebase.

For each single-gene dataset the maximum-likelihood (ML) phylogenetic tree and corresponding bootstrap support values (100 replicates) were calculated using RAxML 7.2.1 (Stamatakis 2006) under the GTRGAMMA model (for the SSU rRNA dataset) or PROTGAMMALGF model (for the protein datasets), with 4 categories of rate variation. PhyloBayes 2.3 (Lartillot and Philippe 2004) was also run on each single gene dataset. PhyloBayes analyses were run for 1,000 or greater generations with the LG + Γ + F model (for the protein datasets) and the GTR + Γ model (for the SSU rRNA gene dataset).

Two different outgroups were selected for the multigene analyses. The first was a small outgroup consisting only of taxa that were both relatively short-branching and thought to be closely related to Fornicata, for which two *Trimastix* species and *Malawimonas jakobiformis* were selected. The second was a larger and phylogenetically broad sample of relatively short-branching eukaryotes (47 outgroup taxa). Neither set included parabasalids, because they formed relatively long branches in the single gene phylogenies, and actually branched within the ingroup in the β-tubulin phylogeny (see results). The program Concaterpillar (Leigh et al. 2008) was used to test the topological congruence of the six protein coding genes considered for both taxon sets. When two *Trimastix* species and *Malawimonas* were used as outgroup taxa, the null hypothesis that the phylogenetic signals among the six genes are concordant was not rejected at the 5% level. Thus, the alignment including the six protein coding genes ("6 gene dataset"), as well as that including the six protein coding genes and the SSU rRNA gene ("7 gene dataset"), were phylogenetically analyzed along with this small outgroup sampling. With the larger outgroup, the Concaterpillar analysis rejected the null hypothesis at the 5% level, but suggested at least that the signals in EF-1α, EF2, HSP70, and HSP90 genes were not significantly incongruent to each other. Thus, we excluded both α- and β-tubulins from a concatenated dataset with the large outgroup sampling, resulting in "4-gene dataset" composed of EF-1α, EF2, HSP70, and HSP90 and "5-gene dataset" composed of EF-1α, EF2, HSP70, HSP90, and the SSU rRNA gene. In the 4-7 gene datasets some protein coding genes from Fornicata taxa were missing, because their sequences are not available (see supplementary material Table S1).

For each concatenated alignment the ML tree was estimated and bootstrapping with 500 replicates was performed using the program RAxML 7.2.6 with the LG + Γ model (for amino acid sequences) and the GTR + Γ model (for nucleotide sequences). For Bayesian analyses the program PhyloBayes 2.3 was run on the 4-gene and 6-gene datasets. PhyloBayes analyses were run with the CAT + Γ model, for 55,000 generations for the 4-gene dataset, and 20,000 generations for the 6-gene dataset. In both RAxML and PhyloBayes analyses, the model parameters and branch lengths were optimized separately for each gene partition as suggested by the Concaterpillar analyses (data not shown). To test for the possibility of missing data from isolate PCE influencing the analyses of the 6-gene dataset (see results), the ML analysis of this dataset was repeated with PCE excluded,

The 6-gene dataset was used for the approximately unbiased (AU) test (Shimodaira 2002). We tested the topologies describing all possible combinations of relationships between the following major clades of Fornicata: the diplomonads + *Retortamonas* clade, *Dysnectes* (CL1), a clade of CL2 + CL3 (which receives 100% bootstrap support (BP) in the original 6-gene sequence analysis), *Carpediemonas* (CL4), *Ergobibamus* (CL5), and *Kipferlia* (CL6) (the CLO lineages CL1-6 were labeled as per Kolisko et al. 2010). The relationships within each of these clades and within the outgroup were constrained to those seen in the ML tree. This resulted in 10395 possible topologies that were tested by AU test. The log-likelihoods at sites were computed in RAxML with the LG + Γ model, and used as the input for CONSEL (Shimodaira and Hasegawa 2001).

Acknowledgements
This work is supported by NSERC Discovery grants numbers 298366-04 to AGBS and 227085-05 to AJR, by the Canadian Institute for Advanced Research (CIfAR), Program in Integrated Microbial Biodiversity, by Japanese Society for the Promotion of Science (JSPS) grants numbers 21370031, 22657025 and 20570219 supporting YI and TH, and by the Ministry of Education, Youth and Sport of the Czech Republic (project MSM0021620828). AK was supported by a JSPS Research Fellowship for Young Scientists (No. 201242).

References

Andersson JO, Sarchfield SW, Roger AJ (2005) Gene transfers from Nanoarchaeota to an ancestor of diplomonads and parabasalids. Mol Biol Evol **22**:85–90

Arisue N, Hasegawa M, Hashimoto T (2005) Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol Biol Evol **22**:409–420

Benchimol M (2001) Hydrogenosome morphological variation induced by fibronectin and other drugs in *Trichomonas vaginalis* and *Tritrichomonas foetus*. Parasitol Res **87:**215–222

Bernard C, Simpson AGB, Patterson DJ (1997) An ultrastructural study of a free-living retortamonad, *Chilomastix cuspidata* (Larsen & Patterson 1990) n. comb. (Retortamonadida, Protista). Eur J Protistol **33:**254–265

Brugerolle G (1973) Etude ultrastructurale du trophozoite et du kyste chez le genre *Chilomastix* Alexeieff 1910 (Zoomastigophorea, Retortamonadida Grasse 1952). J Protozool **20:**574–575

Brugerolle G (1977) Ultrastructure du genre *Retortamonas* Grassi 1879 (Zoomastigophorea, Retortamonadida, Wenrich 1932). Protistologica **13:**233–240

Brugerolle G, Lee JJ (2000) Order Diplomonadida. In Lee JJ, Leedale GF, Bradbury P (eds) The Illustrated Guide to The Protozoa, 2nd edn., vol. 2. Society of Protozoologists, Lawrence, pp 1125–1135

Cavalier-Smith T (1983) A six kingdom classification and a unified phylogeny. In Schwemmler W, Schenk HEA (eds) Endocytobiology II. De Gruyter, Berlin, pp 1027–1034

Cepicka I, Kostka M, Uzlíková M, Kulda J, Flegr J (2008) Non-monophyly of Retortamonadida and high genetic diversity of the genus *Chilomastix* suggested by analysis of SSU rDNA. Mol Phylogenet Evol **48:**770–775

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. Comput Sci Biol: Proc German Conference on Bioinformatics (GCB) 99, pp. 45–56

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science **311:**1283–1287

Clemens DL, Johnson PJ (2000) Failure to detect DNA in hydrogenosomes of *Trichomonas vaginalis* by nick translation and immunomicroscopy. Mol Biochem Parasitol **106:**307–313

Dobell C, Leidlaw PP (1926) On the cultivation of *Entamoeba histolytica* and some other entozoic amoebae. Parasitology **18**:283-318

Hampl V, Horner DS, Dyal P, Kulda J, Flegr J, Foster P, Embley TM (2005) Inference of the phylogenetic position of oxymonads based on nine genes: support for Metamonada and Excavata. Mol Biol Evol **22**:2508–2518

Hampl V, Silberman JD, Stechmann A, Diaz-Trivino S, Johnson PJ, Roger AJ (2008) Genetic evidence for a mitochondriate ancestry in the amitochondriate flagellate *Trimastix pyriformis*. PLoS One **3:**e1383

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic 'supergroups'. Proc Natl Acad Sci USA **106:**3859–3864

Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamoto K, Hasegawa M (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia.* Mol Biol Evol **11:**65–71

Hashimoto T, Nakamura Y, Kamaishi T, Nakamura F, Adachi J, Okamoto K, Hasegawa M (1995) Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. Mol Biol Evol **12:**782–793

Henze K, Horner DS, Suguri S, Moore DV, Sanchez LB, Müller M, Embley TM (2001) Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis.* Gene **281**:123–131

Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM (2010) Diversity and reductive evolution of mitochondria among microbial eukaryotes. Philos Trans R Soc Lond B Biol Sci **365:**713–727

Keeling PJ, Brugerolle G (2006) Evidence from SSU rRNA phylogeny that *Octomitus* is a sister lineage to *Giardia*. Protist **157:**205–212

Kolisko M, Cepicka I, Hampl V, Leigh J, Roger AJ, Kulda J, Simpson AG, Flegr J (2008) Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: implications for the evolutionary history of the double karyomastigont of diplomonads. BMC Evol Biol **8:**205

Kolisko M, Silberman JD, Cepicka I, Yubuki N, Takishita K, Yabuki A, Leander BS, Inouye I, Inagaki Y, Roger AJ, Simpson AGB (2010) A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats. Environ Microbiol **12:**2700–2710

Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol **21**:1095–1109

Leigh JW, Susko E, Baumgartner M, Roger AJ (2008) Testing congruence in phylogenomic analysis. Syst Biol **57:**104–115

Leipe DD, Gunderson JH, Nerad TA, Sogin ML (1993) Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol Biochem Parasitol **59:**41–48

Lindmark DG, Müller M (1973) Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. J Biol Chem **248:**7724–7728

Müller M (1993) The hydrogenosome. J Gen Microbiol **139:**2879–2889

Park JS, Kolisko M, Heiss AA, Simpson AGB (2009) Light microscopic observations, ultrastructure, and molecular phylogeny of *Hicanonectes teleskopos* n. gen., n. sp., a deep-branching relative of diplomonads. J Eukaryot Microbiol **56:**373–384

Park JS, Kolisko M, Simpson AGB (2010) Cell morphology and formal description of *Ergobibamus cyprinoides* n. g., n. sp., another *Carpediemonas*-like relative of diplomonads. J Eukaryot Microbiol **57:**520–528

Philippe H, Germot A (2000) Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol Biol Evol **17**:830–834

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Müller M, Le Guyader H (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc Biol Sci **267:**1213–1221

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol **51:**492–508

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17:**1246–1247

Silberman JD, Simpson AGB, Kulda J, Cepicka I, Hampl V, Johnson PJ, Roger AJ (2002) Retortamonad flagellates are closely related to diplomonads: implications for the history of mitochondrial function in eukaryote evolution. Mol Biol Evol **19:**777-786

Simpson AGB (2003) Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol **53**:1759–1777

Simpson AGB, Patterson DJ (1999) The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the excavate hypothesis. Europ J Protistol **35:**353–370

Simpson AGB, Roger AJ, Silberman JD, Leipe D, Edgcomb VP, Jermiin LS, Patterson DJ, Sogin ML (2002) Evolutionary history of 'early diverging' eukaryotes: the excavate taxon *Carpediemonas* is closely related to *Giardia.* Mol Biol Evol **19**:1782–1791

Simpson AGB, Inagaki Y, Roger AJ (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol **23:**615–625

Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. Am Zool **29:**487–499

Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA (1989) Phylogenetic significance of the kingdom concept: an unusual eukaryotic 16S-like ribosomal RNA from *Giardia lamblia*. Science **243:**75–77

Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688–2690

Stechmann A, Hamblin K, Perez-Brocal V, Gaston D, Richmond GS, Van Der Giezen M, Clark CG, Roger AJ (2008) Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes. Curr Biol **18:**580–585

Steinbüchel A, Müller M (1986) Anaerobic pyruvate metabolism of *Tritrichomonas foetus* and *Trichomonas vaginalis* hydrogenosomes. Mol Biochem Parasitol **20:**57–65

Tachezy J, Sanchez LB, Müller M (2001) Mitochondrial type iron–sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. Mol Biol Evol **18:**1919–1928

Takishita K, Yubuki N, Kakizoe N, Inagaki Y, Maruyama T (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. Extremophiles **11:**563–576

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **24:**4876–4882

Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, Van Der Giezen M, Hernández M, Müller M, Lucocq JM (2003) Mitochondrial remnant organelles of *Giardia* function in iron–sulphur protein maturation. Nature **426:**172–176

Yubuki N, Inagaki Y, Nakayama T, Inouye I (2007) Ultrastructure and ribosomal RNA phylogeny of the free-living heterotrophic flagellate *Dysnectes brevis* n. gen., n. sp., a new member of the Fornicata. J Eukaryot Microbiol **54:**191–200

Figure legends

Figure 1. Maximum-likelihood phylogeny of Fornicata based on concatenated SSU rRNA gene and six proteins (α-tubulin, β-tubulin, HSP70, HSP90, EF-1α, and EF2). Three excavate species, *Trimastix pyriformis*, *Trimastix marina*, and *Malawimonas jakobiformis,* were used to root the tree. Bootstrap probabilities are shown for nodes with support over 50%.

Figure 2. Maximum-likelihood phylogeny of Fornicata based on concatenated six proteins (α-tubulin, β-tubulin, HSP70, HSP90, EF-1α, and EF2). Three excavate species, *Trimastix pyriformis*, *Trimastix marina*, and *Malawimonas jakobiformis,* were used to root the tree. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure 3. Diagram summarizing the well-supported phylogenetic relationships within Fornicata, showing the general appearance of representative cells, and the relative sizes of the mitochondria-related organelles. Note that the representative diplomonad depicted is *Hexamita,* but the size of the mitochondrial organelle is based on *Giardia* (Tovar et al. 2003). The length of the long axis of the *Carpediemonas* organelle is arbitrary. *Kipferlia* organelle size from N, Yubuki, pers. comm.). Scale bar represents 10 μm for all organisms, except 15 μm for *Chilomastix* (*caulleryi*), and represents approximately 500 nm for the diagrams of mitochondria-related organelles.

Figure S1. Maximum-likelihood phylogeny of SSU rRNA gene sequences, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S2. Maximum-likelihood phylogeny of α-tubulin, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S3. Maximum-likelihood phylogeny of β-tubulin, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S4. Maximum-likelihood phylogeny of HSP70, including a broad range of eukaryotes, rooted with the endoplasmic reticulum isoform. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S5. Maximum-likelihood phylogeny of HSP90, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S6. Maximum-likelihood phylogeny of EF-1α, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S7. Maximum-likelihood phylogeny of EF2, including a broad range of eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S8. Maximum-likelihood phylogeny of Fornicata estimated from four concatenated proteins (EF-1α, EF2, HSP70, and HSP90), with a large outgroup of relatively short-branching eukaryotes. The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%. Thick branches represent relationships with over 0.90 Bayesian posterior probabilities.

Figure S9. Maximum-likelihood phylogeny of Fornicata with a large outgroup of relatively short-branching eukaryotes, based on concatenation of the SSU rRNA gene and four proteins (EF1, EF2, HSP70, and HSP90). The Fornicata members are shaded. Bootstrap probabilities are shown for nodes with support over 50%.

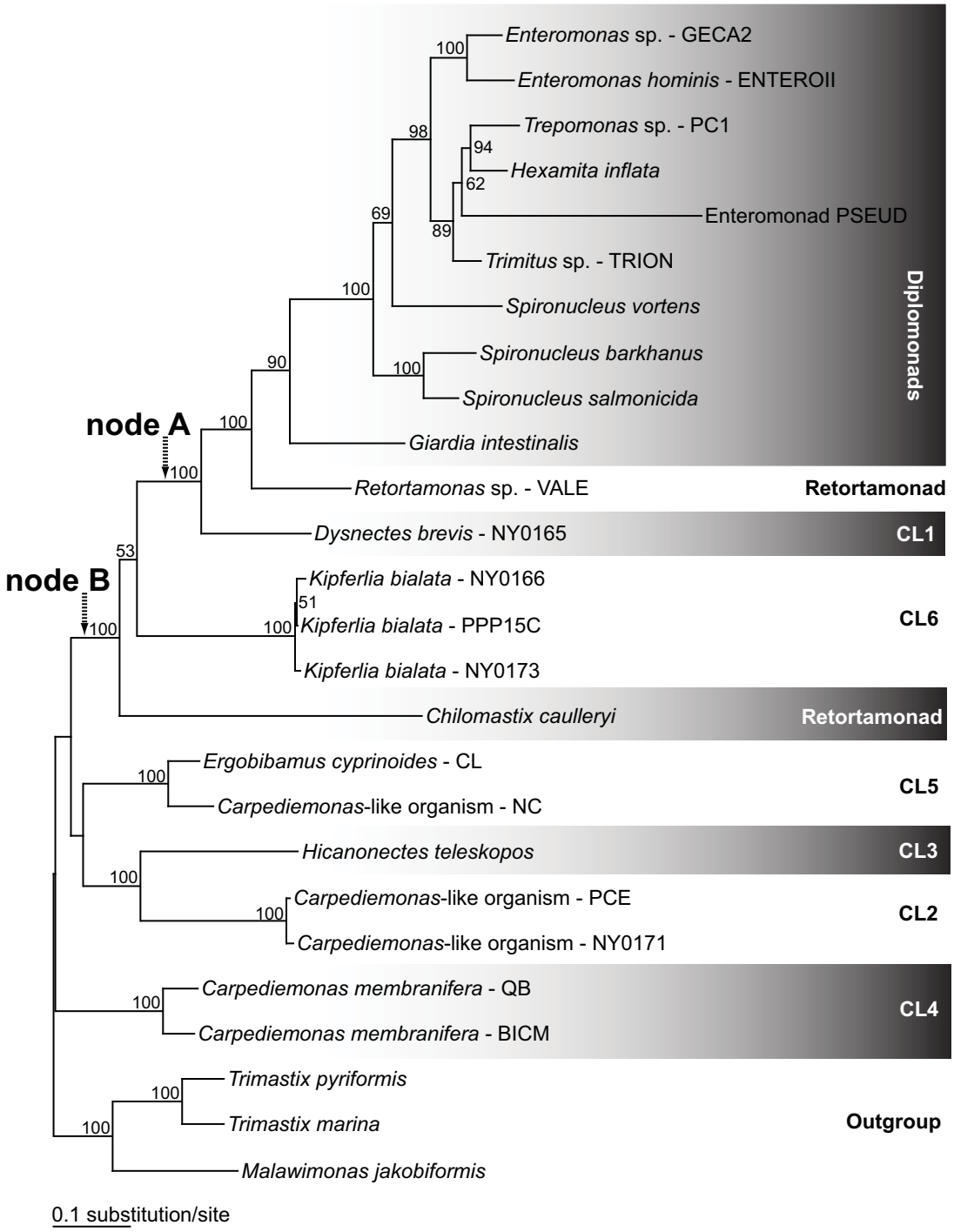

Enteromonas sp. - GECA2
Enteromonas hominis - ENTEROII
Trepomonas sp. - PC1
Hexamita inflata
Enteromonad PSEUD
Trimitus sp. - TRION
Spironucleus vortens
Spironucleus barkhanus
Spironucleus salmonicida
Giardia intestinalis
Diplomonads
node A
Retortamonas sp. - VALE
Retortamonad
Dysnectes brevis - NY0165
CL1
node B
Kipferlia bialata - NY0166
Kipferlia bialata - PPP15C
Kipferlia bialata - NY0173
CL6
Chilomastix caulleryi
Retortamonad
Ergobibamus cyprinoides - CL
Carpediemonas-like organism - NC
CL5
Hicanonectes teleskopos
CL3
Carpediemonas-like organism - PCE
Carpediemonas-like organism - NY0171
CL2
Carpediemonas membranifera - QB
Carpediemonas membranifera - BICM
CL4
Trimastix pyriformis
Trimastix marina
Malawimonas jakobiformis
Outgroup
0.1 substitution/site

Figure 1

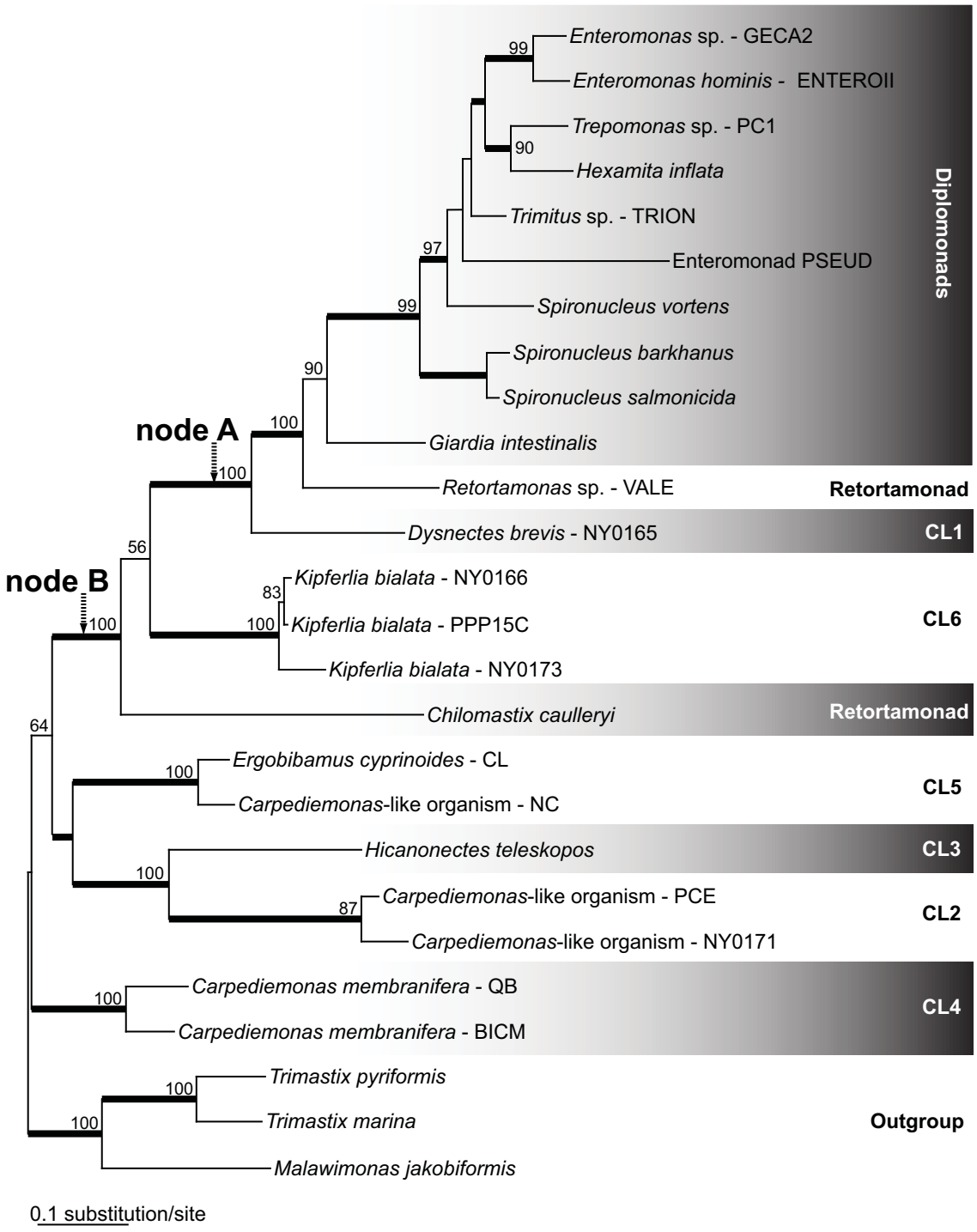
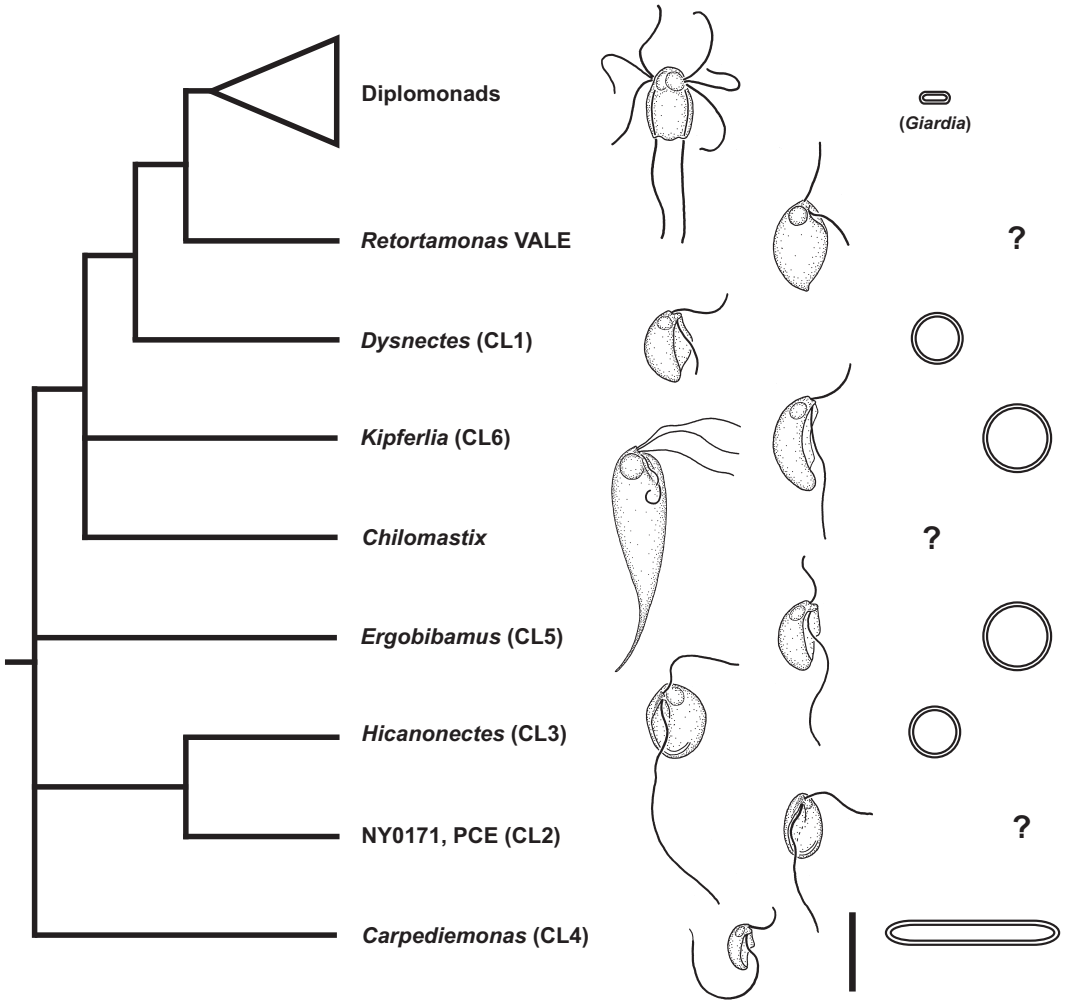

Enteromonas sp. - GECA2
Enteromonas hominis - ENTEROII
Trepomonas sp. - PC1
Hexamita inflata
Trimitus sp. - TRION
Enteromonad PSEUD
Spironucleus vortens
Spironucleus barkhanus
Spironucleus salmonicida
Giardia intestinalis
Diplomonads
node A
Retortamonas sp. - VALE
Retortamonad
Dysnectes brevis - NY0165
CL1
node B
Kipferlia bialata - NY0166
Kipferlia bialata - PPP15C
Kipferlia bialata - NY0173
CL6
Chilomastix caulleryi
Retortamonad
Ergobibamus cyprinoides - CL
Carpediemonas-like organism - NC
CL5
Hicanonectes teleskopos
CL3
Carpediemonas-like organism - PCE
Carpediemonas-like organism - NY0171
CL2
Carpediemonas membranifera - QB
Carpediemonas membranifera - BICM
CL4
Trimastix pyriformis
Trimastix marina
Malawimonas jakobiformis
Outgroup
0.1 substitution/site

Figure 2


Diplomonads
(Giardia)
Retortamonas VALE
?
Dysnectes (CL1)
Kipferlia (CL6)
Chilomastix
?
Ergobibamus (CL5)
Hicanonectes (CL3)
NY0171, PCE (CL2)
?
Carpediemonas (CL4)

SSU rRNA

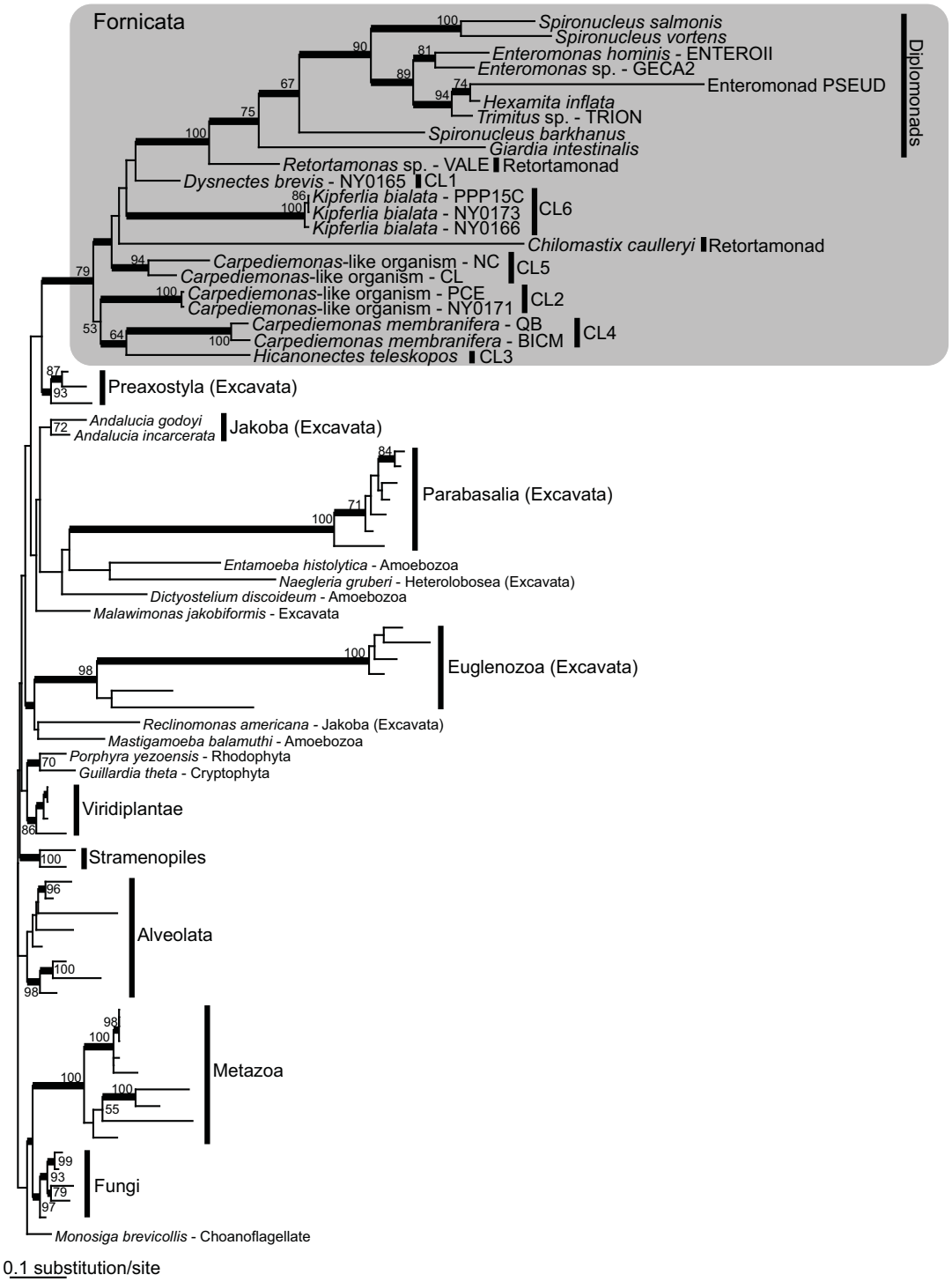
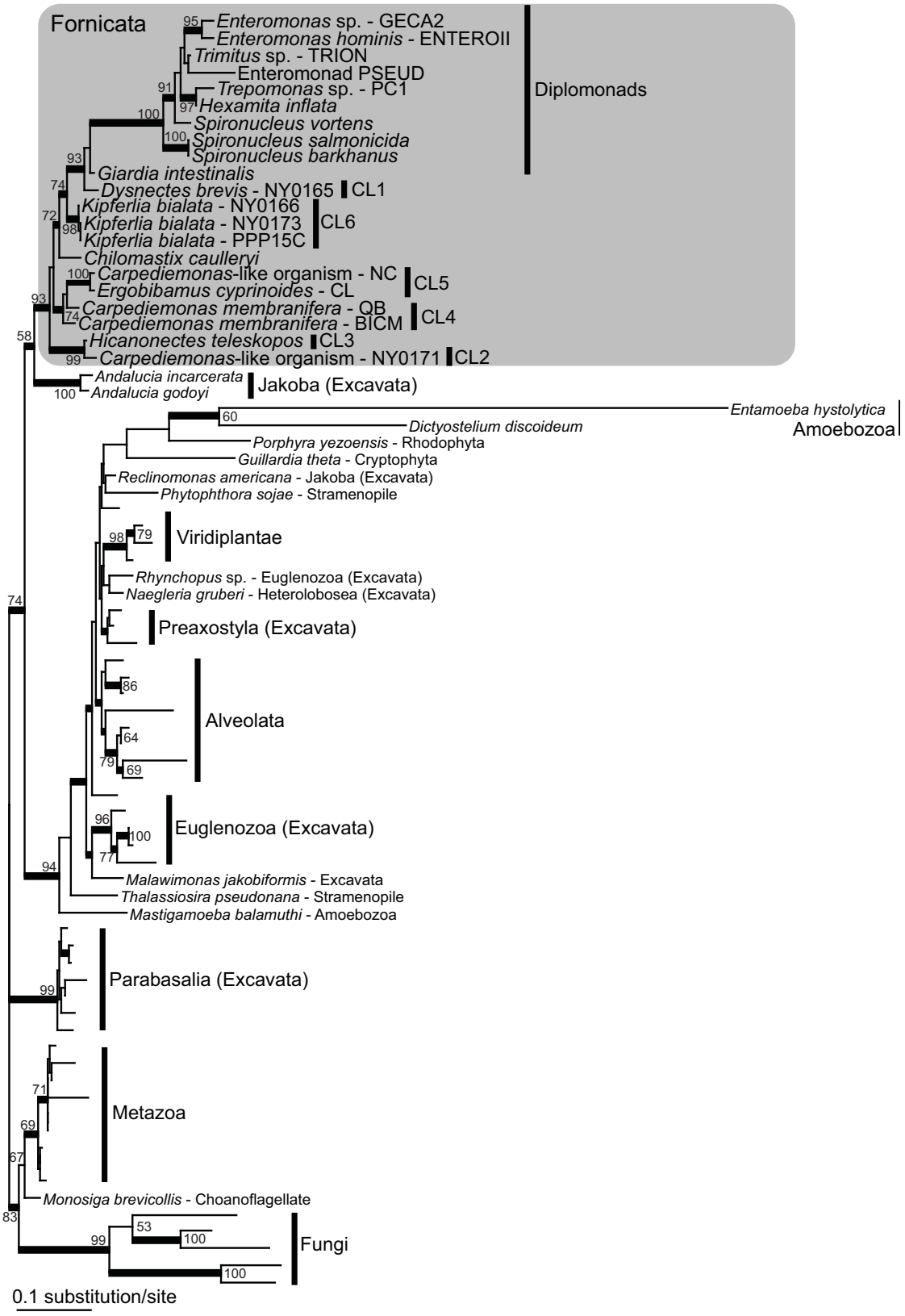
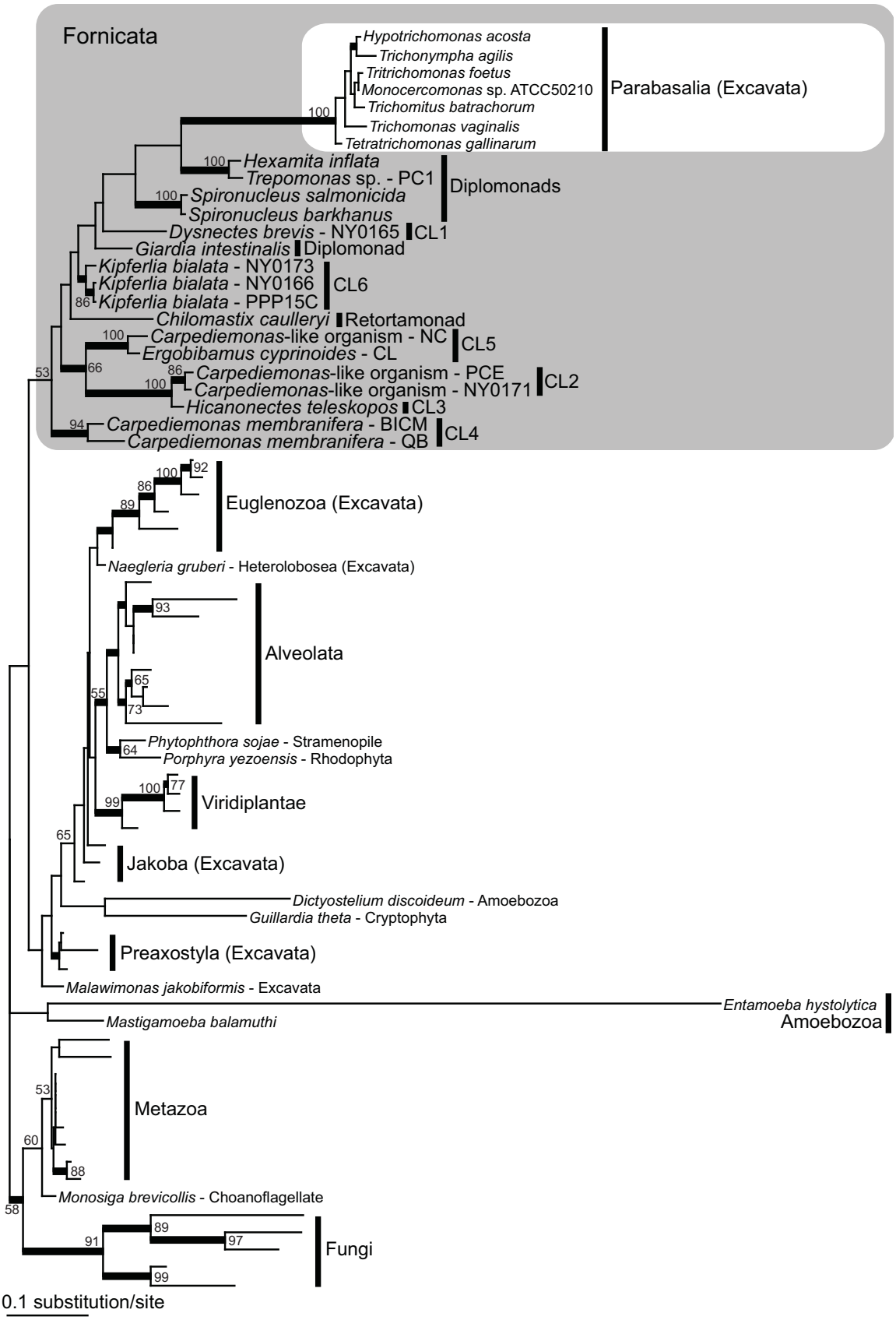


Figure S1

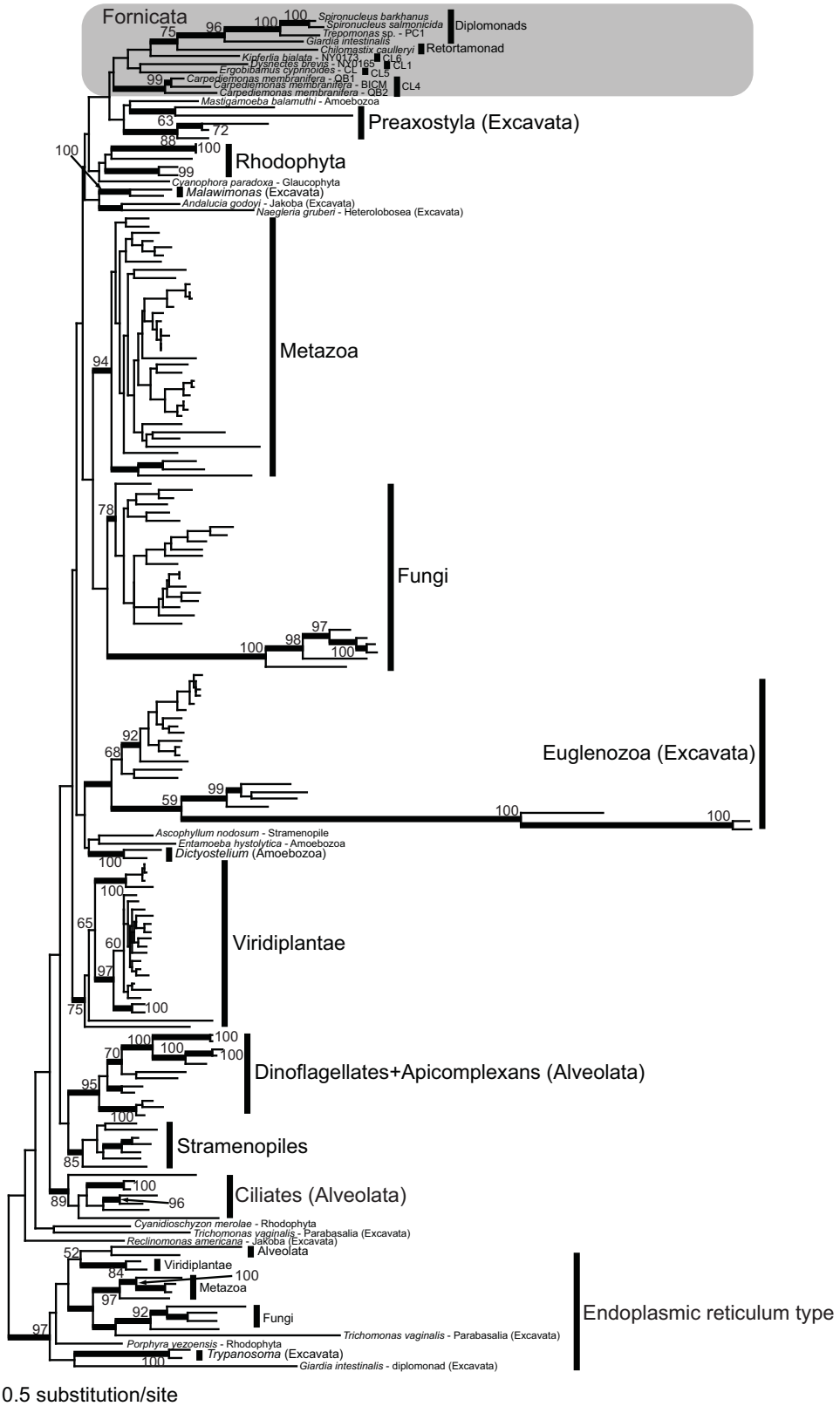
α-tubulin


Fornicata
Enteromonas sp. - GECA2
Enteromonas hominis - ENTEROII
Trimitus sp. - TRION
Enteromonad PSEUD
Trepomonas sp. - PC1
Hexamita inflata
Spironucleus vortens
Spironucleus salmonicida
Spironucleus barkhanus
Diplomonads
Giardia intestinalis
Dysnectes brevis - NY0165
CL1
Kipferlia bialata - NY0166
Kipferlia bialata - NY0173
Kipferlia bialata - PPP15C
CL6
Chilomastix caulleryi
Carpediemonas-like organism - NC
Ergobibamus cyprinoides - CL
CL5
Carpediemonas membranifera - QB
Carpediemonas membranifera - BICM
CL4
Hicanonectes teleskopos
CL3
Carpediemonas-like organism - NY0171
CL2
Andalucia incarcerata
Andalucia godoyi
Jakoba (Excavata)
Entamoeba hystolytica
Dictyostelium discoideum
Amoebozoa
Porphyra yezoensis - Rhodophyta
Guillardia theta - Cryptophyta
Reclinomonas americana - Jakoba (Excavata)
Phytophthora sojae - Stramenopile
Viridiplantae
Rhynchopus sp. - Euglenozoa (Excavata)
Naegleria gruberi - Heterolobosea (Excavata)
Preaxostyla (Excavata)
Alveolata
Euglenozoa (Excavata)
Malawimonas jakobiformis - Excavata
Thalassiosira pseudonana - Stramenopile
Mastigamoeba balamuthi - Amoebozoa
Parabasalia (Excavata)
Metazoa
Monosiga brevicollis - Choanoflagellate
Fungi
0.1 substitution/site

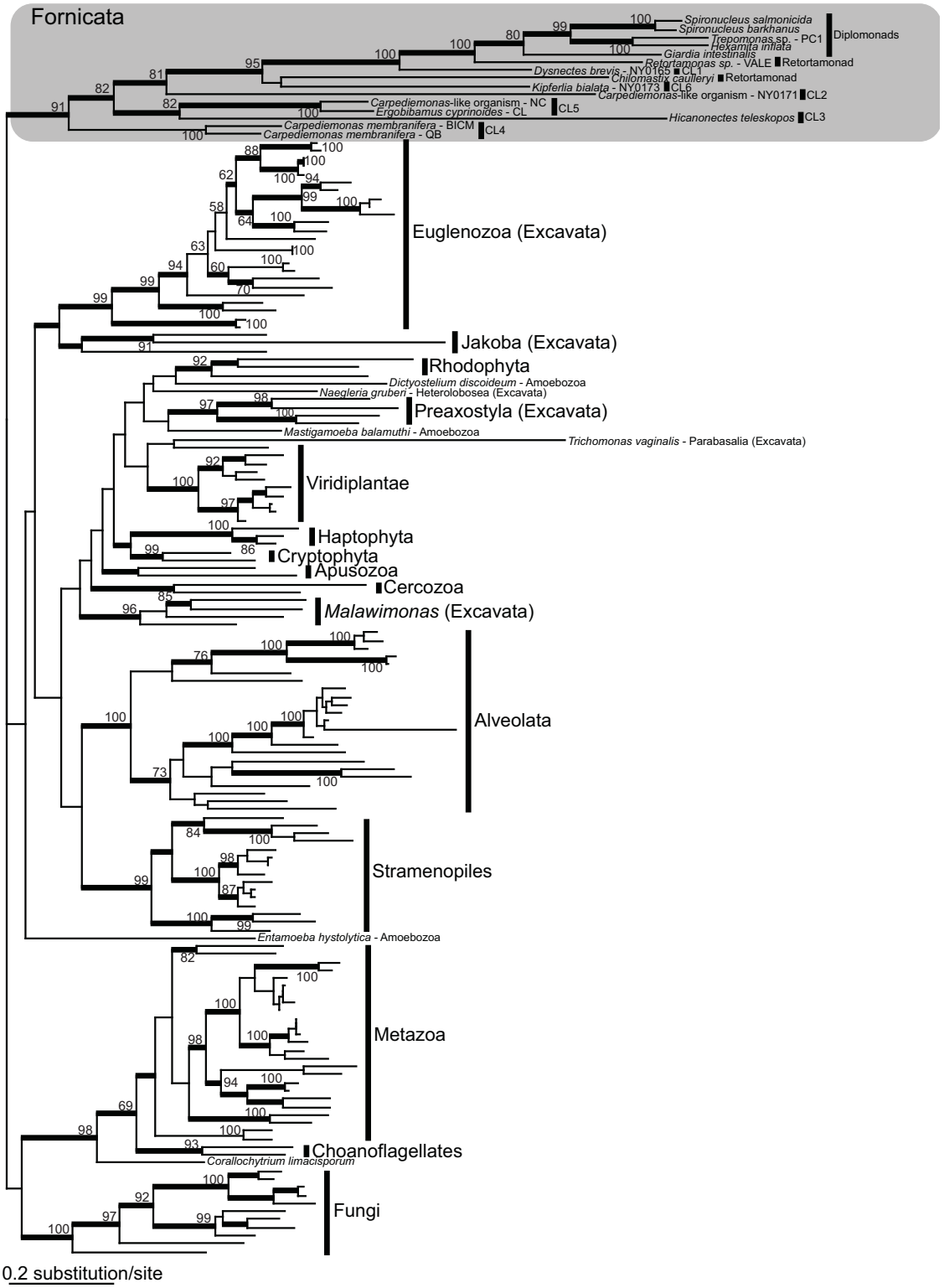
β-tubulin

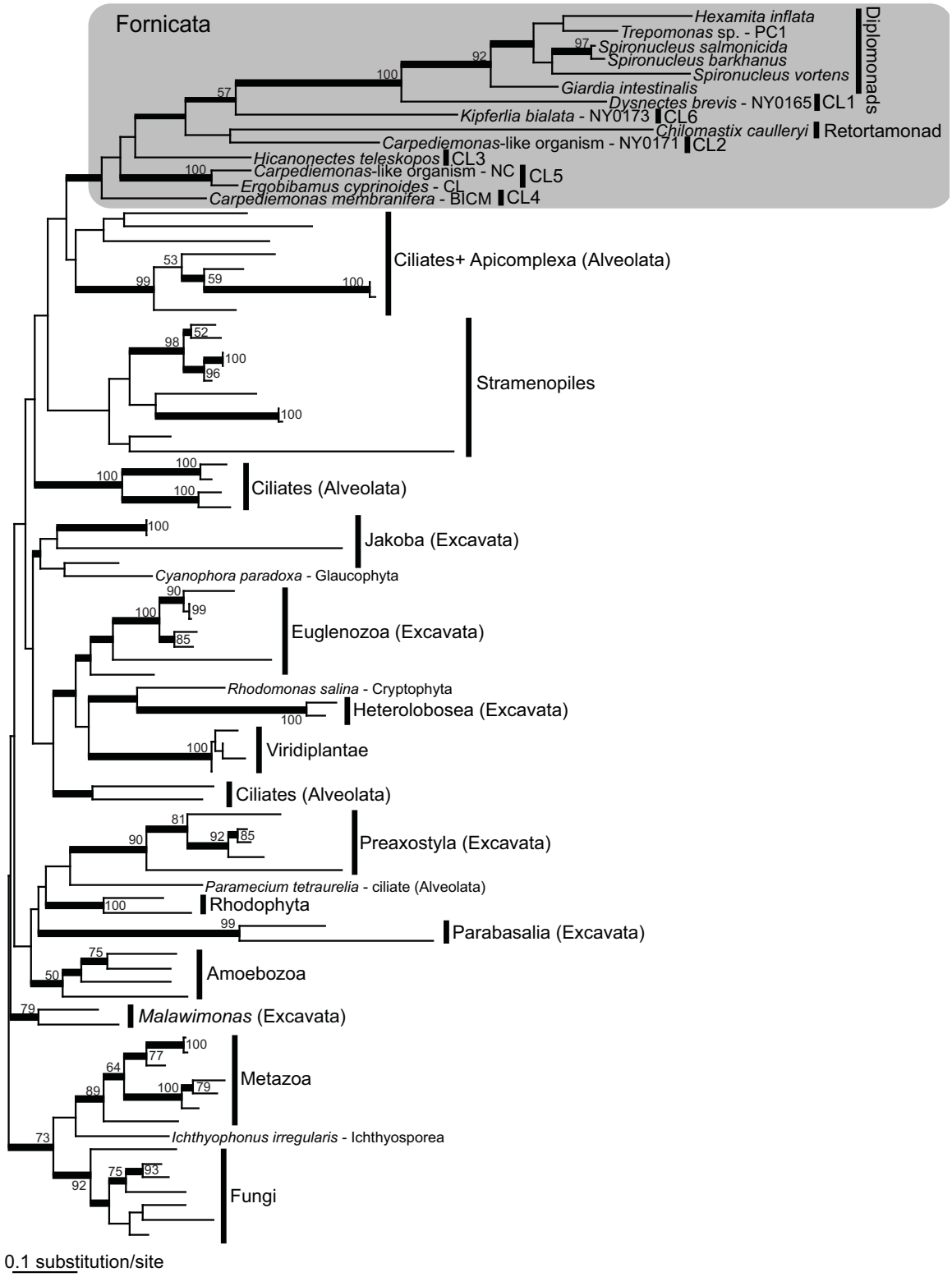

Fornicata
Hypotrichomonas acosta
Trichonympha agilis
Tritrichomonas foetus
Monocercomonas sp. ATCC50210
Trichomitus batrachorum
Trichomonas vaginalis
Tetratrichomonas gallinarum
Parabasalia (Excavata)
Hexamita inflata
Trepomonas sp. - PC1
Spironucleus salmonicida
Spironucleus barkhanus
Diplomonads
Dysnectes brevis - NY0165
CL1
Giardia intestinalis
Diplomonad
Kipferlia bialata - NY0173
Kipferlia bialata - NY0166
Kipferlia bialata - PPP15C
CL6
Chilomastix caulleryi
Retortamonad
Carpediemonas-like organism - NC
Ergobibamus cyprinoides - CL
CL5
Carpediemonas-like organism - PCE
Carpediemonas-like organism - NY0171
CL2
Hicanonectes teleskopos
CL3
Carpediemonas membranifera - BICM
Carpediemonas membranifera - QB
CL4
Euglenozoa (Excavata)
Naegleria gruberi - Heterolobosea (Excavata)
Alveolata
Phytophthora sojae - Stramenopile
Porphyra yezoensis - Rhodophyta
Viridiplantae
Jakoba (Excavata)
Dictyostelium discoideum - Amoebozoa
Guillardia theta - Cryptophyta
Preaxostyla (Excavata)
Malawimonas jakobiformis - Excavata
Entamoeba hystolytica
Mastigamoeba balamuthi
Amoebozoa
Metazoa
Monosiga brevicollis - Choanoflagellate
Fungi
0.1 substitution/site

HSP70

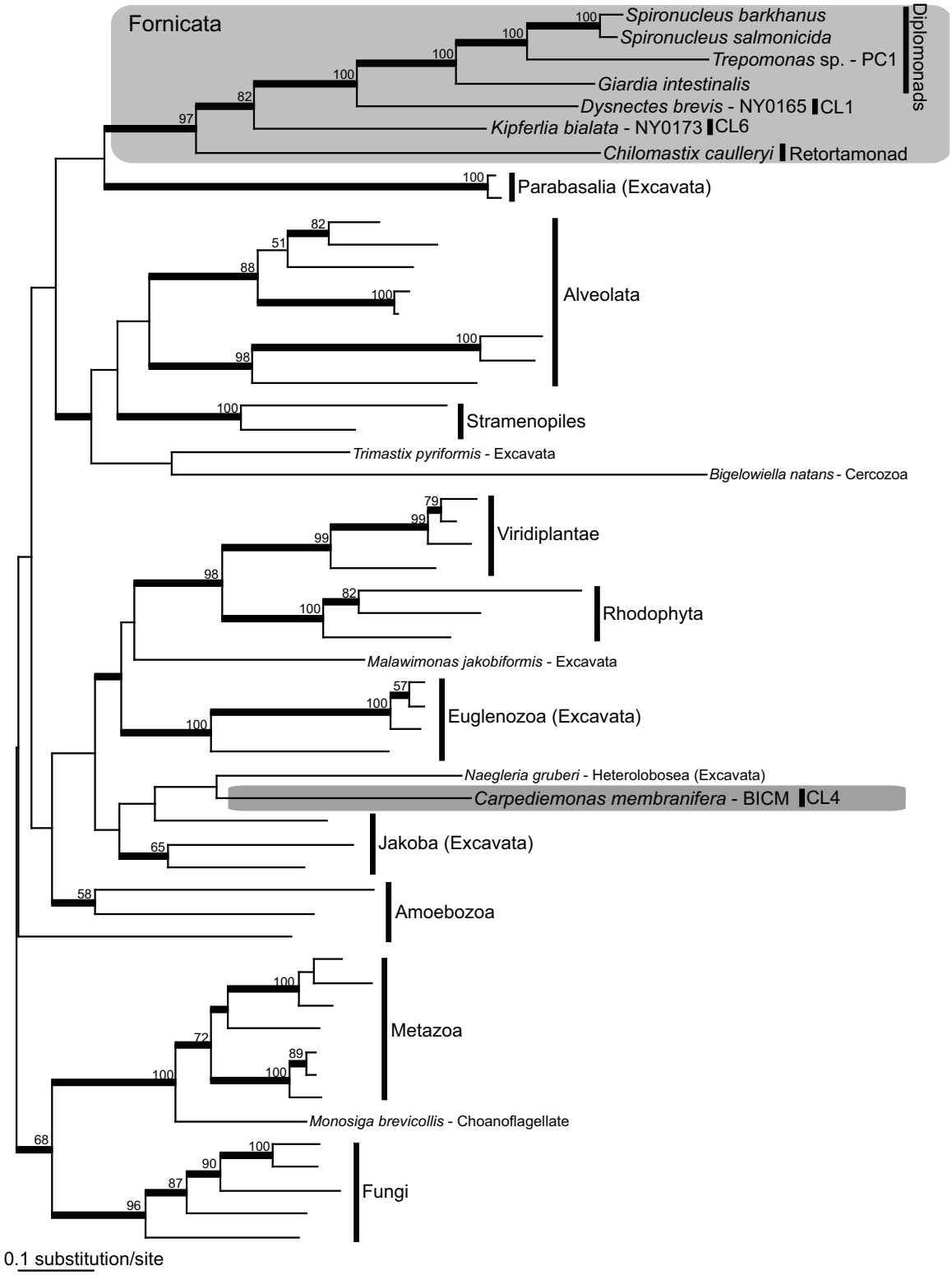

Fornicata
Spironucleus barkhanus
Spironucleus salmonicida
Trepomonas sp. - PC1
Giardia intestinalis
Diplomonads
Chilomastix caulleryi
Retortamonad
Kipferlia bialata - NY0173
CL6
Paranema trichophorum - NY0165
CL1
Ergobibamus cyprinoides - CL
CL5
Carpediemonas membranifera - QB1
Carpediemonas membranifera - BICM
CL4
Carpediemonas membranifera - QB2
Mastigamoeba balamuthi - Amoebozoa
Preaxostyla (Excavata)
Rhodophyta
Cyanophora paradoxa - Glaucophyta
Malawimonas (Excavata)
Andalucia godoyi - Jakoba (Excavata)
Naegleria gruberi - Heterolobosea (Excavata)
Metazoa
Fungi
Euglenozoa (Excavata)
Ascophyllum nodosum - Stramenopile
Entamoeba hystolytica - Amoebozoa
Dictyostelium (Amoebozoa)
Viridiplantae
Dinoflagellates+Apicomplexans (Alveolata)
Stramenopiles
Ciliates (Alveolata)
Cyanidioschyzon merolae - Rhodophyta
Trichomonas vaginalis - Parabasalia (Excavata)
Reclinomonas americana - Jakoba (Excavata)
Alveolata
Viridiplantae
Metazoa
Fungi
Endoplasmic reticulum type
Trichomonas vaginalis - Parabasalia (Excavata)
Porphyra yezoensis - Rhodophyta
Trypanosoma (Excavata)
Giardia intestinalis - diplomonad (Excavata)
0.5 substitution/site

HSP90


Fornicata
Spironucleus salmonicida
Spironucleus barkhanus
Trepomonas sp. - PC1
Hexamita inflata
Giardia intestinalis
Diplomonads
Retortamonas sp. - VALE
Retortamonad
Dysnectes brevis - NY0165
CL1
Chilomastix caulleryi
Retortamonad
Kipferlia bialata - NY0173
CL6
Carpediemonas-like organism - NY0171
CL2
Carpediemonas-like organism - NC
CL5
Ergobibamus cyprinoides - CL
Hicanonectes teleskopos
CL3
Carpediemonas membranifera - BICM
Carpediemonas membranifera - QB
CL4
Euglenozoa (Excavata)
Jakoba (Excavata)
Rhodophyta
Dictyostelium discoideum - Amoebozoa
Naegleria gruberi - Heterolobosea (Excavata)
Preaxostyla (Excavata)
Mastigamoeba balamuthi - Amoebozoa
Trichomonas vaginalis - Parabasalia (Excavata)
Viridiplantae
Haptophyta
Cryptophyta
Apusozoa
Cercozoa
Malawimonas (Excavata)
Alveolata
Stramenopiles
Entamoeba hystolytica - Amoebozoa
Metazoa
Choanoflagellates
Corallochytrium limacisporum
Fungi
0.2 substitution/site


EF-1α
Fornicata
Hexamita inflata
Trepomonas sp. - PC1
Spironucleus salmonicida
Spironucleus barkhanus
Spironucleus vortens
Giardia intestinalis
Diplomonads
Dysnectes brevis - NY0165 CL1
Kipferlia bialata - NY0173 CL6
Chilomastix caulleryi Retortamonad
Carpediemonas-like organism - NY0171 CL2
Hicanonectes teleskopos CL3
Carpediemonas-like organism - NC CL5
Ergobibamus cyprinoides - CL
Carpediemonas membranifera - BICM CL4
Ciliates+ Apicomplexa (Alveolata)
Stramenopiles
Ciliates (Alveolata)
Jakoba (Excavata)
Cyanophora paradoxa - Glaucophyta
Euglenozoa (Excavata)
Rhodomonas salina - Cryptophyta
Heterolobosea (Excavata)
Viridiplantae
Ciliates (Alveolata)
Preaxostyla (Excavata)
Paramecium tetraurelia - ciliate (Alveolata)
Rhodophyta
Parabasalia (Excavata)
Amoebozoa
Malawimonas (Excavata)
Metazoa
Ichthyophonus irregularis - Ichthyosporea
Fungi
0.1 substitution/site

EF2


Fornicata
Spironucleus barkhanus
Spironucleus salmonicida
Trepomonas sp. - PC1
Giardia intestinalis
Dysnectes brevis - NY0165 CL1
Kipferlia bialata - NY0173 CL6
Chilomastix caulleryi Retortamonad
Diplomonads
Parabasalia (Excavata)
Alveolata
Stramenopiles
Trimastix pyriformis - Excavata
Bigelowiella natans - Cercozoa
Viridiplantae
Rhodophyta
Malawimonas jakobiformis - Excavata
Euglenozoa (Excavata)
Naegleria gruberi - Heterolobosea (Excavata)
Carpediemonas membranifera - BICM CL4
Jakoba (Excavata)
Amoebozoa
Metazoa
Monosiga brevicollis - Choanoflagellate
Fungi
0.1 substitution/site

Four gene (EF-1α, EF2, HSP70, and HSP90) dataset

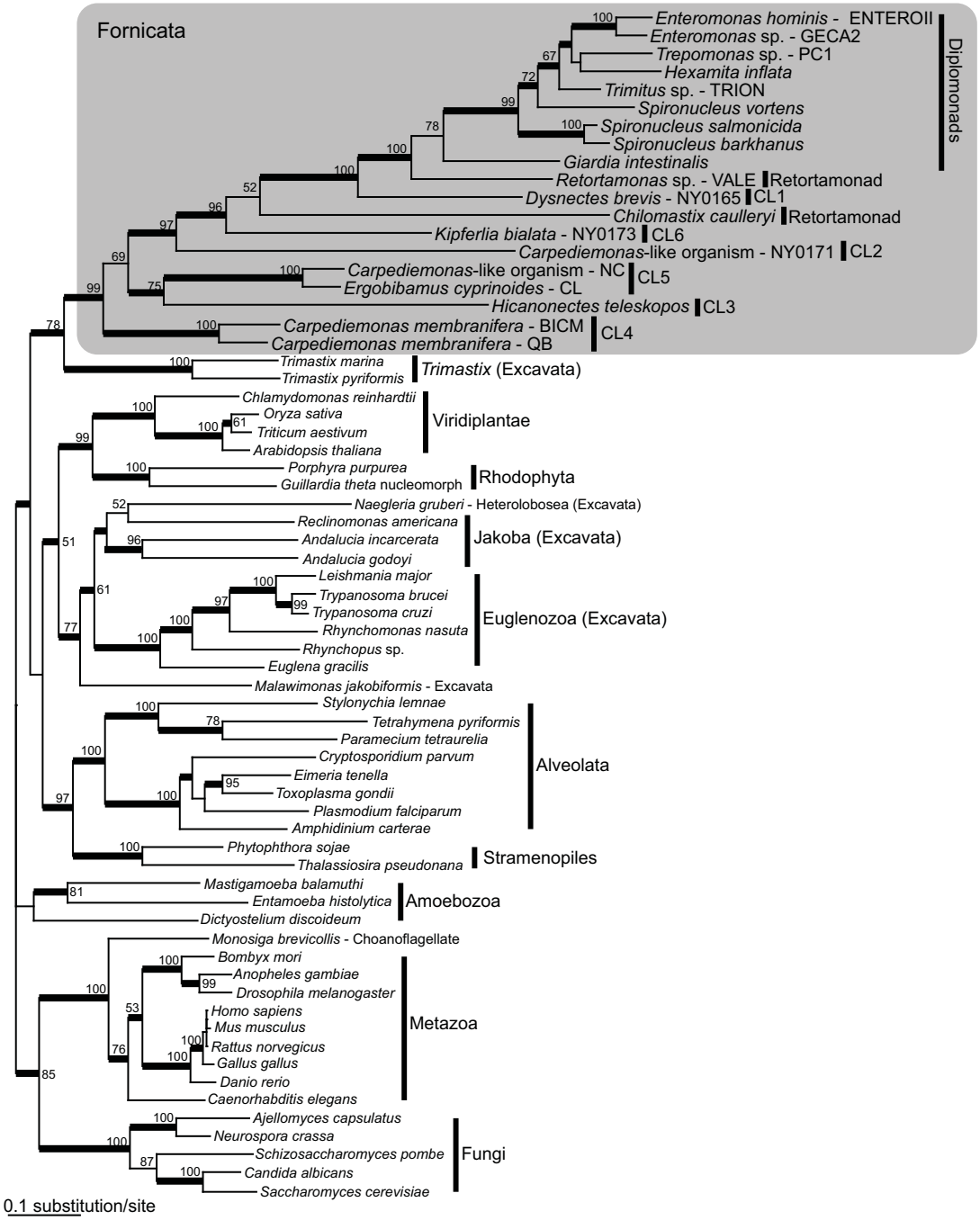


Figure S8

Five gene (SSU rRNA, EF-1α, EF2, HSP70, and HSP90) dataset

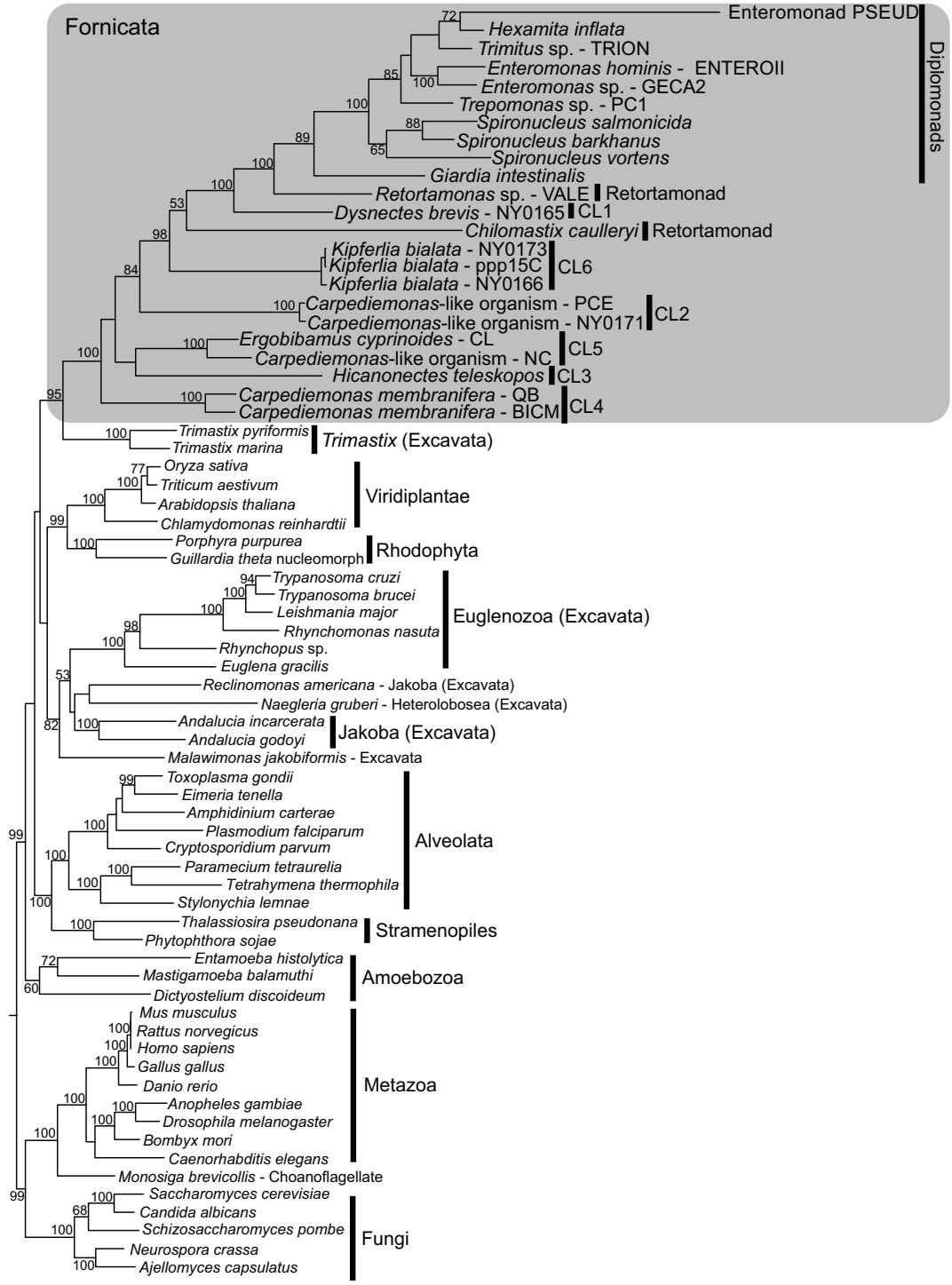


Figure S9

Table S1. Genes from deep-branching Fornicata lineages included in the phylogenetic analyses (indicated by 'X'). A red colored 'X' indicates a sequence newly obtained in this study using PCR or extracted from EST datasets. Names of PCR primers used are in parentheses.

Affiliation	Organism	SSU rRNA	α-tubulin	β-tubulin	HSP70	HSP90	EF-1α	EF2
CLO (CL1)	*Dysnectes brevis* (NY0165)	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)	X PCR (HSP70_88,HSP70_498r)	X PCR (100XF,910X1R)	X PCR* (HSP90_210z,HSP90_447zr)	X PCR (EF2-A1, EF2-A2)
CLO (CL2)	NY0171	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)		X PCR (110U2,910X1R)	X PCR (EF1AF1A,EF1AR1A)	
CLO (CL2)	PCE	X		X PCR (BtubA,BtubB)				
CLO (CL3)	*Hicanonectes teleskopos* (SB)	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)		X PCR (100XF,880XR)	X PCR (1XF,8XR)	
CLO (CL4)	*Carpediemonas membranifera* (BICM)	X	X EST	X PCR (BtubA,BtubB)	X EST	X PCR (100XF,880XR)	X PCR (1XF,8XR)	X EST
CLO (CL4)	*Carpediemonas membranifera* (QB)	X	X	X	X	X		
CLO (CL5)	*Ergobibamus cyprinoides* (CL)	X	X EST	X PCR (BtubA,BtubB)	X EST	X EST	X EST	

CLO (CL5)	NC	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)		X PCR (100XF,910X1R)	X PCR (1XF,8XR)	
CLO (CL6)	*Kipferlia bialata* (NY0173)	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)	X EST	X EST	X PCR (EF1A-B1, SaEF1aRr)	X PCR (EF2-C1, SaEF2-A1Rr2)
CLO (CL6)	*Kipferlia bialata* (PPP15c)	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)				
CLO (CL6)	*Kipferlia bialata* (NY0166)	X	X PCR (AtubA,AtubB)	X PCR (BtubA,BtubB)				
Retortamonad	*Chilomastix caulleryi*	X (EukA, EukB)	X EST	X EST	X EST	X EST	X EST	X EST

*The HSP90 gene from *Dysnectes brevis* was accidentally amplified with the primer set designed for the EF-1α gene (HSP90_210z and HSP90_447zr).

Table S2. List of PCR primers used in this study.

Gene	Primer direction	Primer name	Reference or sequence
SSU rRNA	Forward	EukA	Medlin et al. 1988
	Reverse	EukB	Medlin et al. 1988
α-tubulin	Forward	AtubA	Edgcomb et al. 2001
	Reverse	AtubB	Edgcomb et al. 2001
β-tubulin	Forward	BtubA	Edgcomb et al. 2001
	Reverse	BtubB	Edgcomb et al. 2001
HSP70	Forward	HSP70_88	5'-GCNRTNGGNATHGAYYTNGG-3'
		HSP70_95	5'-TAGGCGATGGCCGCAGCAGTGGG-3'
	Reverse	HSP70_498r	5'-TANGCNACNGCYTCRTCNGGRTT-3'
HSP90	Forward	100XF	Simpson et al. 2002
		110U2	Simpson et al. 2006
	Reverse	880XR	Simpson et al. 2006
		910X1R	Simpson et al. 2006
EF-1α	Forward	1XF	Simpson et al. 2008
		EF1AF1A	5'-GTGGACGCCGGNAARTCNACNACNAC-3'
		HSP90_210z	5'-TNATHGGNCARTTYGGNGTNGG-3'
		EF1A-B1	5'-GTNATHGGNCAYGTNGAYWSNGGNAA-3'

	Reverse	8XR	5'-CCGACAGCGACGGTCTGNYKCATRTC-3'
		EF1AR1A	5'-TCGGCCTGGGANGTNCCNGTNATCAT-3'
		HSP90_447zr	5'-TARAANTCNKYRTAYTCYTCYTCNGT-3'
		SaEF1aRr	5'-CATCTTGTTGATTCCGACGATCATCTGC-3'
		SaEF1aRr-2	5'-GCTTGACTCCCAGAGTGTAAGC-3'
EF2	Forward	EF2-A1	5'-GGNGCNGGNGARYTNCAYYTNGA-3'
		EF2-2	5'-GANGCNYTNYTNGARATGATNAT-3'
		EF2-B1	Hashimoto et al. 1995
		EF2-C1	Hashimoto et al. 1995
		SaEF2-3	5'-TACCTGCCCTCCCCCAGGCAGGCCCA-3'
	Reverse	EF2-A2	Hashimoto et al. 1995
		DbEF2-A1Rr	5'-CCTGGTAGTTGACGATAGGGG-3'
		DbEF2-2Rr	5'-GATGGCGTTGTAGGCCTCATCATC-3'
		DbEF2-B1Rr	5'-CCTTGTTGAGCATGAGCACGG-3'
		SaEF2-A1Rr2	5'-GAGGGGGAATCCGCAGAAATCCTCC-3'