TRANSFORMING THE EUKARYOTIC TREE OF LIFE THROUGH ISOLATION, CHARACTERISATION, AND PHYLOGENETICS OF NOVEL FREE-LIVING PROTISTS

by

Yana Eglit

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

at

Dalhousie University
Halifax, Nova Scotia
December 2022

Table of Contents

List of Tables

List of Figures

Abstract

The overwhelming majority of eukaryotic diversity is microbial and protistan. Major new groups are discovered nearly every year up to this day. I established several dozen cultures of heterotrophic protists comprised of two kinds of 'unknowns': 1. *Incertae sedis* organisms previously described using morphological approaches but devoid of sequence data and difficult to place in the tree of eukaryotes based on their current taxonomy alone; and 2. Previously unattested morphotypes. Particularly productive were samples from anaerobic, hypersaline, and alkaline environments, and combinations thereof—in those environments, eukaryote predators of other protists comprised a generous portion of hitherto uncharacterised diversity. I used light microscopy to carefully characterise their cell morphology and life histories, and, for some taxa, electron microscopy was used to examine their fine structure. Using ribosomal DNA sequencing, I obtained molecular data for 39 isolates for phylogenetic analyses. Among them were considerable contributions to the diversity of breviates, metamonads, and novel rhizarian lineages among Filosa and Endomyxa. The phylogenetic position of several of these isolates remained unresolved by ribosomal DNA phylogenies; thus, two of these taxa—the gliding marine cell *Meteora sporadica* and 'Protist X', an anaerobic flagellate that preys upon other microbial eukaryotes—were then subject to transcriptome sequencing and multigene phylogenomic analyses. Surprisingly, these analyses revealed that *Meteora sporadica* and 'Protist X' form a robust clade with the deep-branching Hemimastigophora. This hints at the existence of a new biologically diverse supergroup in the eukaryotic Tree of Life. Finally, a survey of available molecular environmental sequence data found that about half of the novel taxa have been virtually undetected by environmental methods to date. Altogether, this emphasises the importance of cultivation not only for downstream research, but in the exploration of biodiversity itself.

List of Abbreviations Used

3%LS	3% LB in Seawater
aa	amino acid
ASV	Amplicon Sequence Variant
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BS	Bootstrap Support
bp	base pair
bs	bootstrap support
BUSCO	Benchmarking Universal Single-Copy Orthologues
cDNA	complementary Deoxyribonucleic Acid
CLO	*Carpediemonas*-like Organism
CRuMs	Collodictyonid + *Rigifila* + *Mantamonas*
DIC	Differential Interference Contrast (microscopy)
DNA	Deoxyribonucleic Acid
EPA	Evolutionary Placement Analysis
FSR	Fast Site Removal
FW	Freshwater
GTR	Generalised Time-Reversible (model of DNA evolution)
gCF	gene Concordance Factor
HS5	Hypersaline Medium #5
LB	Lysogeny Broth
LBA	Long Branch Attraction
LECA	Last Eukaryotic Common Ancestor
LGT	Lateral Gene Transfer
LSU	Large Subunit (of ribosomal DNA/RNA)
LWR	Likelihood-to-Weight Ratio
MDA	Multiple Displacement Amplification
ML	Maximum Likelihood
mRNA	messenger RNA
MRO	Mitochondrion-related Organelle

NCBI	National Center for Biotechnology Information
NCBI nt	National Center for Biotechnology Information nucleotide databases
nt	nucleotide
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PMSF	Posterior Mean Site Frequency
pp	posterior probability
ppt	parts per thousand
RAxML	Randomised 'Axelerated' Maximum Likelihood
RNA	Ribonucleic Acid
rDNA	ribosomal Deoxyribonucleic Acid
rRNA	ribosomal Ribonucleic Acid
RT	Reverse Transcription
Sar	Stramenopiles, alveolates, and rhizaria
SEM	Scanning Electron Microscopy
SRA	Sequence Read Archive
SSU	Small Subunit (of ribosomal DNA/RNA)
SW	Seawater
Taq	(DNA polymerase of) *Thermus aquaticus*
TEM	Transmission Electron Microscopy
UFB	Ultra-Fast Bootstrap (also UFboot)
V_4	Hypervariable region V_4 (in SSU-rDNA)
V_9	Hypervariable region V_9 (in SSU-rDNA)
WGA	Whole Genome Amplification

Acknowledgements

It takes a village to raise a PhD. A village spanning extensive time and space, nationalities and identities, professions and personal roles. The 'lone genius' is a persistent myth, and most of us only have a chance at the 'lone' part anyway.

The governing body of this village, of course, responsible for its very existence, is my supervisory committee. It's hard to overstate how absurdly lucky I am with a committee who are not only amazing people and work well together, but have also become my friends. At times, I felt like I was 'cheating' at grad school, unscathed by a whole genre of doctoral nightmares. Every single one of you is a personal and scientific role model to me. Erin Bertrand, I wasn't sure at first how a polar marine microbe proteomics expert would find any interest in my obscure and very not-photosynthetic (and apparently not highly abundant) critters. You not only contributed much needed outside(-ish) perspective to my projects, but I'd outright forget you weren't even in my immediate field. I still can't follow your mass spec work a fraction as well as you understood my obscure little critters. Over a short time, I watched you cultivate a healthy, happy, inclusive lab that is as supportive as you are, and if anyone is crazy enough to hand me keys to a lab someday, I'll be asking for advice. Warm regards to Greg, Avery, and the menagerie: Simon, and the crazy pair "Fatty and the Other One", whatever their original names were.

Andrew Roger, obviously I would not have made a dent in all the phylogenomics without you. You have a gift for translating math to biologists. You and Udeni also have a gift for throwing epic parties, and you know precisely what cat gif to send to a struggling student or colleague. Your openness and humility are an inspiration, and you too have cultivated a supportive and accepting lab. Thank you for putting up with my whining about everything grad school and bioinformatics related.

No village can survive without members who do the work to support its activity, the adults in the building so to speak. I appreciate Ping Li and Patricia Scallion for running the TEM and SEM facilities and providing helpful advice for electron microscopy. Rod in stores has always added an upbeat, cheerful interaction to the day—on top of running watercolour tutorials and musical sessions in the department. Aileen, Julie, and Carolyn in the Biology Office for ensuring the place functions as well as anything can at Dal, but also for being fierce advocates for graduate students. And thanks to Louie and the rest of the custodial staff, with whom I wish my late

schedule overlapped a bit more.

The trade networks between villages are essential to a thriving research economy. I have been extremely privileged to work in a small close-knit field that generally celebrates collaboration over competition. ISoP and ISEP have been welcoming academic organisations, without which I probably would not have made it to graduate school. I have too many wonderful collaborators to name, but especially involved with the work in this thesis were Ivan Čepička, Takashi Shiratori, and Vasily Zlatogursky, Thank you to everyone who demonstrates that there is no conflict between good science and good people: science truly thrives in a low-ego environment.

A few poor souls had to work in the village workshop with me. Practically at my bench almost the entire time was my grad school partner-in-crime Gordon Lax. Thank you for all your patient help and advice around the lab, and an indelible friendship. You're a force of collected calm in chaotic moments, and seem to have an endless ability to fix things well beyond your specialisation: lab protocols, tech issues, and even relationships between other people. You also found Joana for us, and Chester, and Rosa. Lastly, thank you for all your help with typography and LaTeXtypesetting for this thesis. I would not have been able to tame the beast long enough to compile this thesis without you. Kiran More and Andrea Gigeroff: our ex-lab member pandemic-support-group-turned-weekly-ritual played a big role in rebuilding some of the science enthusiasm that fizzled out in the plague years. I'm excited to continue working with you as scientists, and as friends. Likewise for other past lab members and current friends and collaborators—Sebastian Hess, Anna Busch, Robyn Buchwald, Claire Burnard—I was lucky to share the village workshop with you all, even if the chisels and hammers would sometimes end up in unexpected locations. I never imagined how much I would miss that while working alone during the plague years. To the newest member of the Simpson Lab, Liz Weston: thank you for taking care of cultures during the chaos of the past couple months, and best of luck in your adventures in protistology—may your protists be both fascinatingly weird and kind to your research! There is no limit to the beasts you can slay, you are stronger than you think.

Those working in neighbouring workshops (guildhalls?) have been a tremendous source of support, encouragement, and inspiration. The Roger, Bertrand, Archibald, and Slamovits labs in particular, and CGEB/ICG in general, have all been wonderful for science and social life. I have received so much support from the permanent members of ICG/the 8th floor: John Archibald, Ford Doolittle and Mike Gray, Claudio Slamovits, Susana Breglia (hard to believe you TA'd my first ever protistology course in undergrad!), Wanda Danilchuk, Marlena Dlutek, and Bruce Curtis (thank you for all the BLAST and other computing help!). Additional thanks

to Sara and Loay (talk about taking mushrooming to a whole new level!), Cat, Megan, Catalina, Scott M. (I owe you for this LaTeXthesis template!), Noor, Scott P., Elden, Laura, Courtney, Kelsey, Morgan, Shelby, Lucie, and so, so many others. Of the 8th Floor Tupper denizens who have since embarked on further adventures, I am especially grateful to Michelle Leger and all her advice and support, and travel, and wildly inappropriate conversations, and hilarious protist-related sketches (and, of course, cat gifs). Jon Jerlström-Hultqvist (and Heather Topple) enriched our world with Swedish pastries, protist-themed homebrew, and MinION wizardry alike—your ability to face strange new lab techniques and master them is an inspiration. Jan and Sophie de Vries feasted at the village kitchen Gordon operated, and at many other kitchens too: thank you for all the delicious chats about science and life at large! Alicia and Jeremy: Alicia, thank you for arranging a fantastic outreach opportunity, and thanks to both of you for more feasting (and teaching us how to properly eat lobster). Lastly, thanks to fellow biology graduate students not mentioned by name, especially BOGS and everyone involved who made so many fun events happen! And thanks to the broader community beyond the ivory walls, from our trivia group to my Ukrainian friends to everyone who plays a role in making Halifax a welcoming city.

I am grateful to everyone who spent time in the scriptorium with me, sharing coffee and pastries as you watched me stare in frustration at the mostly blank parchment, or grinding pigment for the next batch of Python code. Catalina, Lucie, Megan—you've been excellent writing buddies! Any self-proclaimed aspiring writer or artist seems to have a 'retreat in the desert' phase, and my dear friends from our Indiana days, Sam Miller and Nathan Taylor, arranged the perfect thesis writing retreat, with tacos, saguaros, and the oppressive heat of Arizona in August to make sure you stay inside and write. Additionally, thank you for coming all the way to Halifax to attend my defense, and even doing some fencing with my supervisor. Nerd Club lives on!

This village has a rich chronicle with all who helped shape this path, in particular past teachers and mentors. Thank you to everyone in my past who pushed and motivated me to put in the work towards becoming a scientist as a reality. I thank my parents for valuing education. I was particularly lucky to discover research near the start of my undergraduate degree, and probably would not have been able to go down this path if a few individuals did not take a chance on a naïve undergrad with terrible grades. I hope to someday pass on this opportunity to the next generation of research enthusiasts.

Finally, my supervisor, mentor, and dear friend, Alastair. Of course, without you, I would

not have entered graduate school in the first place, but that was only the beginning. You are the most patient and caring mentor anyone could ask for. Thoughtful, careful and meticulous to an almost irritating level, generous, humble, bad at doors, understanding, stubborn enough to successfully supervise this misfit to the bitter end. When the Great Plague descended upon the land, in the chaos of First Lockdown, you and Gina made sure I was not to face the crisis alone. And Gina, thank you for all the supporting words and food(!) and hiking breaks, insect hunts, garden tours, and scribal evenings. Good luck on your new horticultural adventure and may your plants thrive freely, if only slightly out of control. I owe both of you a great debt.

To everyone: thank you for taking part in my PhD village.

Typeset in XeLaTeXusing the EB Garamond font family for the main body text.

Chapter 1

Introduction

Much of the central philosophy in cell biology and biochemistry is driven by the idea of unity, encapsulated by a quotation attributed to microbial biochemist Jacques Monod: "...anything found to be true of *E. coli* must also be true of Elephants" (Friedmann 2004). In other words, findings in one organism should be expected to be relevant to another. Furthermore, we study (ideally, convenient) model systems in attempt to learn about The Cell or The Organism, particularly as it relates to human biology. Yet, incorporating the evolutionary context complicates the idea of unified biology of all life. That is not to say that general principles cannot be found across the tree of eukaryotes or tree of life, but rather that the extrapolation of biological principles from one or a few select organisms must be done with care.

New discoveries in biology inherently fall into a phylogenetic and evolutionary context. Not only does this context impact the interpretation of the discovery, but also how the discovery is applied to other organisms. Furthermore, hypotheses and generalisations are made from the patterns inferred by examining a sampling of total diversity from the perspective of a specific system. All of these considerations demand not only a *correct* phylogeny, but also adequate sampling of organisms from this phylogeny, or taxon sampling. This is particularly true for fields like cell biology, where evolutionary approaches have become more common only relatively recently—unlike, for example, ecology or physiology—and significant difficulties are still presented by incomplete or suboptimal sampling.

Incorrect phylogenetic inferences or assumptions can lead to incorrect interpretations of new data. The common assumption that morphologically simpler life forms 'predate' or diverge before their more complex counterpart has diminished in light of evidence building up

showing small and/or parasitic animals to be derived rather than ancestral forms (Garey and Schmidt-Rhaesa 1998), or the collapse of "crown eukaryotes" (*e.g.* Baldauf 2003). Incorrect phylogenies involving key model systems have also led to mistaken interpretations (*e.g.* Hughes and Friedman 2004a). Even with a correct phylogeny, skewed sampling across taxa can likewise lead to inaccuracies, to date a relatively common problem in animal-model-centric studies of protein evolution (*e.g.* Hehenberger et al. 2020 *vs.* Cahill 2017), or systems as a whole (*e.g.* Corfield and Berry 2015). Or a chance absence of a feature in a particular chosen representative for a clade can be inferred, incorrectly, as an absence for an entire group (discussed with respect to choanoflagellates in (Richter et al. 2018).

In this Introduction, I will discuss taxon sampling and its relevance. I will then provide a brief overview of where improvements in taxon sampling among microbial eukaryotes (protists) have come from in light of recent technological and methodological innovations.

1.1 What is taxon sampling?

Taxon sampling is the selection of organisms examined in a particular study. Usually, it is a subsample of known, or available, taxon sampling, tailored to a specific group and question. Typically, taxon sampling is discussed in the context of inferring the correct phylogenetic tree (*e.g.* Rosenberg and Kumar 2001; Heath et al. 2008); however, here we will focus on the role of taxon sampling in *applying* the inferred phylogenetic tree to broader questions. Both the correct tree and appropriate taxon sampling are crucial for quality ancestral state reconstructions and evolution-informed hypothesis generation. The selection of taxa would ideally strive to be as complete, balanced, and appropriate for the question on hand as possible (Plazzi et al. 2010) for applying phylogenies as well as constructing them, but in practice, usually remains a subjective process. Optimal taxon sampling for one question may be inappropriate for another. It is important to note here that the concept of 'basal lineages' is misleading, and that a 'species-poor' deeply diverging lineage is not necessarily primitive or slow-evolving for a trait of interest, or, in fact, most traits (for a thorough discussion, see Crisp and Cook 2005).

Novel taxa and expanded taxon sampling can lead to major paradigm shifts and rule-breaking—regardless of phylogenetic placement. For example, a bacterial phagocytosis-like process in the planctomycete Uab was found via new taxon discovery (Shiratori et al. 2019), as well as the presence of a mixed 'archaeal' and 'bacterial' type phospholipid membranes in a novel bacterium

(Villanueva et al. 2020), and the first rotary flagellum known in any eukaryote along with the first example of a toroidal swimmer, in amoebozoan *Idionectes* (Hess et al. 2019). In another example, total mitochondrial organelle loss had been considered implausible until the discovery that the oxymonad *Monocercomonoides* lacks any relic of the mitochondrion (Karnkowska et al. 2016).

Alternatively, novel taxa can change our understanding of the nature or evolutionary history of a specific phylogenetic group. The new archaean sister lineages to eukaryotes, Asgard Archaea (Spang et al. 2015 Spang et al. 2018; Zaremba-Niedzwiedzka et al. 2017), are in the process of imposing a dramatic overhaul of our understanding of eukaryogenesis (Zaremba-Niedzwiedzka et al. 2017; Akıl and Robinson 2018; Lu et al. 2020; Akıl et al. 2020) and the evolution of features previously considered as unique to eukaryotes. The conclusion that the apicomplexan apicoplast was a relict plastid of a previously photosynthetic ancestor was confirmed by the discovery of its algal relative *Chromera* (Moore et al. 2008); similarly, the predicted flagellate ancestry of red algae was supported by the discovery of a novel deep-branching predatory flagellate, *Rhodelphis* (Gawryluk et al. 2019). The examples above all concern the discovery of novel sister groups to the taxon of interest; however, reconstruction of evolutionary history can also be enhanced by adding more nodes to a lineage of interest: for example adding barthelonids to the "grade" of *Carpediemonas*-like organisms contributing to the reconstruction of fornicate mitochondrial evolution (Yazaki et al. 2020).

Taxon sampling can be said to come in *shallow* or *deep* forms: the former can be seen as an expansion of sampling in a known clade (fig. 1.1; labels s, d), for example by sequencing more *Carpediemonas*-like organisms among fornicates (Kolisko et al. 2010b); the latter would be introducing deep lineages of a clade—for example, the discovery of Asgard Archaea mentioned above, or finding *Hemimastix* (Lax et al. 2018) or *Rhodelphis* (Gawryluk et al. 2019), in the context of deep eukaryote phylogeny. This distinction is fundamentally arbitrary and depends on the question being asked: expanding the sampling of *Carpediemonas*-like organisms is shallow from the perspective of the tree of eukaryotes or the supergroup Metamonada, but deep with respect to diplomonads or diversity of *Giardia* species. Both forms of sampling are important in reconstructing evolutionary history: while it is obvious that deep sampling contributes new nodes (and, thus, last common ancestor (LCA) reconstructions) along the lineage of interest, shallow sampling also improves accuracy in reconstructing the subsequent evolutionary history of a group.

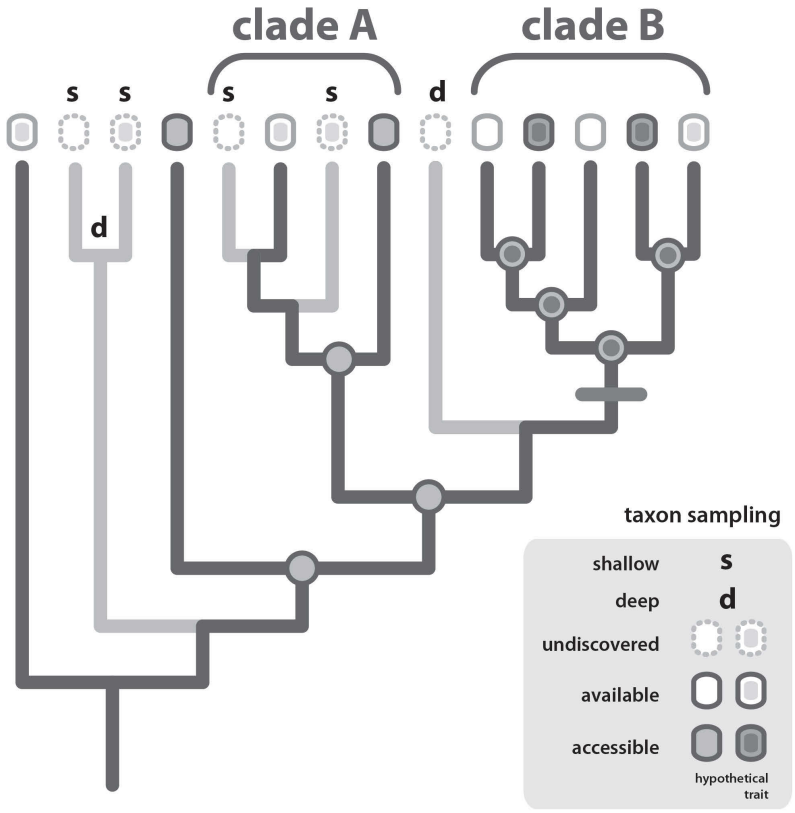


Figure 1.1: Types of taxon sampling. SHALLOW taxon sampling is expanding the sampling within a clade, whereas DEEP taxon sampling adds new branches (and a last common ancestor (LCA)) outside the clade of interest; note that the distinction between the two is subjective and context-dependent. While the hypothetical new lineage between clades A and B would represent deep sampling pertaining to the evolution of clade B, it would be shallow sampling within clade A+B and arguably still shallow sampling with respect to clade A, as it would contribute to reconstructing an extant LCA in the lineage leading to A. The novel clade outside A+B itself represents deep sampling, whereas expanding the taxa within it is shallow.

AVAILABLE taxon sampling is a taxon that is known and characterised—in the modern age, usually molecularly, though placement suspected by morphology can be considered tentatively available as well. ACCESSIBLE taxon sampling, on the other hand, is whether one is able to include the taxon in a particular analysis; for example, if a genome of a known organism has not yet been sequenced, that organism is available but not accessible to sampling in the given context. The same taxon can be accessible or inaccessible depending on the question and methods used, whereas a taxon is either available or not, regardless of context. Due to differential loss, retention, gain, and derivation of characters (shown here as filled in ovals), an undersampled clade may lead to improper assumptions about its LCA, as in this diagram where a trait (filled-in circle) would be assumed to evolve much more recently in the tree (bar) than it actually did, due to chance available and accessible sampling of taxa that do not have this trait.

Furthermore, taxon sampling can be contrasted between *available*—the mere knowledge that an organism branches at a particular node in the tree of life, in this era at least via some molecular data-and *accessible*-organisms that can be interrogated in a given context or by a given technology of interest (fig. 1.1, ovals). For accessing molecular cell biology, simply knowing that *Hemimastix* exists and where it goes in the tree of eukaryotes is not enough: one must also have appropriate tools and prior research to interrogate the organism; *Hemimastix* is not yet, as of this writing, accessible for molecular cell biology, or comparative genomics (although a culture exists). In a similar vein, heliozoans have been known for over a century, but did not become truly available until ribosomal small subunit rDNA separated them into four distantly related groups (Nikolaev et al. 2004), with the emergence of centrohelids as a separate lineage further supported by phylogenomics to be fairly distinct from the major supergroups (Burki et al. 2016). Adding more choanoflagellate transcriptomes (Richter et al. 2018) would be an example of expanded *accessible shallow* taxon sampling from the perspective of molecular biology. Often, a trade-off exists between advanced or expensive technology and the taxon sampling accessible to it, particularly if involves a reduction in throughput.

1.2 Why does taxon sampling matter?

To illustrate the importance of accurate trees and appropriate taxon sampling, I will use a charismatic example from animal phylogeny featuring groups containing two important classical model systems: arthropods and nematodes. Nematodes, being small worms with a simplified body cavity, or coelom, were considered a 'simpler' animal (then, pseudocoelomates) diverging later than, say, flatworms[1] (acoelomates) and expected to branch outside arthropods+chordates[2] (coelomates) (Adoutte et al. 1999). This view was in line with the prevalent concept of evolutionary grades of complexity at the time (Adoutte et al. 1999). A competing hypothesis was that the shared moulting, or ecdysis, unites arthropods+nematodes as Ecdysozoa (first proposed in Barnes et al. 1993 *The Invertebrates, a New Synthesis*, 2nd Ed.; Valentine and Collins 2000). Early ribosomal RNA gene phylogenies supported the coelomate idea (Ohama et al. 1984), but

[1] Namely, platyhelminths; flatworms are polyphyletic (Cannon et al. 2016).

[2] For the sake of simplicity, I will focus specifically on nematodes, arthropods, and chordates in this discussion.

this was later shown to be a result of long branch attraction resulting from the elevated evolutionary rate of *Caenorhabditis elegans* (*C. elegans*) specifically, and thus alleviated by its replacement with a slower evolving taxon (Aguinaldo et al. 1997; Adoutte et al. 1999). Early protein phylogeny work, however, contradicted the ribosomal inferences (Sidow and WK Thomas 1994), contributing to a heated debate in animal phylogenetics (Wägele and Misof 2001; Wolf et al. 2004) that eventually settled down by 2010 (Holton and Pisani 2010) in favour of Ecdysozoa.

The story of nematodes and arthropods matters here for several reasons: first, it is a striking example of how taxon sampling impacts inference (*e.g.* acknowledged in Holton and Pisani 2010) and how an early drawback of new technology is often the insufficient taxon sampling available to its nascent form. The reversal to coelomates in the early days of protein phylogenies was, in part, a consequence of the available nematode sequences coming from one species, the fast-evolving *C. elegans*. This has continued on with the addition of 'minor', small and 'simple' taxa like bryozoans, tardigrades, and rotifers, altering our understanding of animal evolution (Garey and Schmidt-Rhaesa 1998; Laumer et al. 2015). Second, it is a case where the traditional view of evolution by gradual gain in complexity was eventually interrupted by improved molecular phylogenies. In addition to the collapse of the pseudocoelomates, the acoelomates (flatworms) were later shown to fall within the massive clade containing Arthropods, molluscs, and chordates (reviewed in Telford et al. 2015), supporting that simplification of the coelom has occurred several times independently. Lastly, arthropods and nematodes contain two of the most important model systems in cell biology: *Drosophila melanogaster* and *C. elegans*, respectively. Extrapolating any organismal biology from those systems to general animal evolution and applying these ideas to hypotheses in chordate evolution relies upon correct assumptions about their phylogenetic relationships.

The protein phylogeny confusion coincided with an explosion of developmental and molecular cell biology work on the model systems *Drosophila* and *C. elegans*, forcing authors doing ancestral reconstruction to pick one of the hypotheses (Friedman and Hughes 2001; Hughes and Friedman 2004b *vs.* Nicholson et al. 2005), or agnostically include both (Bertrand et al. 2004; Kim and Ausubel 2005). This affected studies like those mapping ancestral genome duplications (Friedman and Hughes 2001) or losses (Hughes and Friedman 2004b); for example, in the latter incorrectly assuming losses in *Drosophila* and *C. elegans* as parallel rather than shared. (Lees-Miller et al. 2020) point out an example where a gene was absent in both invertebrate models and thus assumed to be a later, vertebrate, innovation, but then found not only in insects

but even outside animals. Genes found in presumed deep-branching organisms were assumed on that basis to be ancient innovations, for example HOX genes in a flatworm (Balavoine and Telford 1995). Thus, referring to the correct phylogeny is extremely important when making evolutionary inferences.

1.3 Where are new protist taxa coming from?

Discovery and cultivation of microbes has gone on for a few centuries, but molecular biology has thoroughly changed what the act of microbial sampling entails. Morphological data are now routinely accompanied by sequence information (or, in some cases, sequence information is completely unassociated with any morphology), and previously-described organisms are 're-discovered' from this molecular perspective. Next-generation sequencing has been an important driver behind this 'molecular revolution', providing an accessible tool for objectively assigning microbes to their evolutionary places and at the same time empowering researchers to ask more complex questions about non-model systems. Previously undescribed taxa can now be characterised by molecular approaches without the need to culture, via single-cell ribosomal DNA/RNA sequencing (Lax et al. 2018; Lax and Simpson 2020), transcriptomics (Kolisko et al. 2014; Krabberød et al. 2017; Lax et al. 2018) and genomics (Gawryluk et al. 2016a; Strassert et al. 2018; Wideman et al. 2019), in addition to environmental metagenomics (West et al. 2018; Obiol et al. 2020). Furthermore, the ease and cost of next-generation sequencing unleashed a revolution in bioinformatic tools that has relaxed requirements for the purity of cultures, which no longer need be axenic nor even mono-eukaryotic for targeted molecular work or even whole genome sequencing to be practical (Leger et al. 2017; Monteil et al. 2019; Gawryluk et al. 2019). As a result, organisms that were previously missed due to dependencies on other prokaryotic or eukaryotic microbes (*e.g.* as food sources, symbionts or oxygen scavengers), are now more accessible to sequencing and/or cultivation. This has led to a major influx of a variety of protistan heterotrophs (especially eukaryotrophs) and anaerobes into research. Similar parallels are noted on the prokaryote side (Lewis et al. 2020).

The combination of better sampling and improved phylogenetic techniques has led to a richer and more stable tree of eukaryotes, with a variety of new groupings and deep-branching lineages added in the last 15 years (for review, see Burki et al. 2020; also refer to Figure 1.2). In

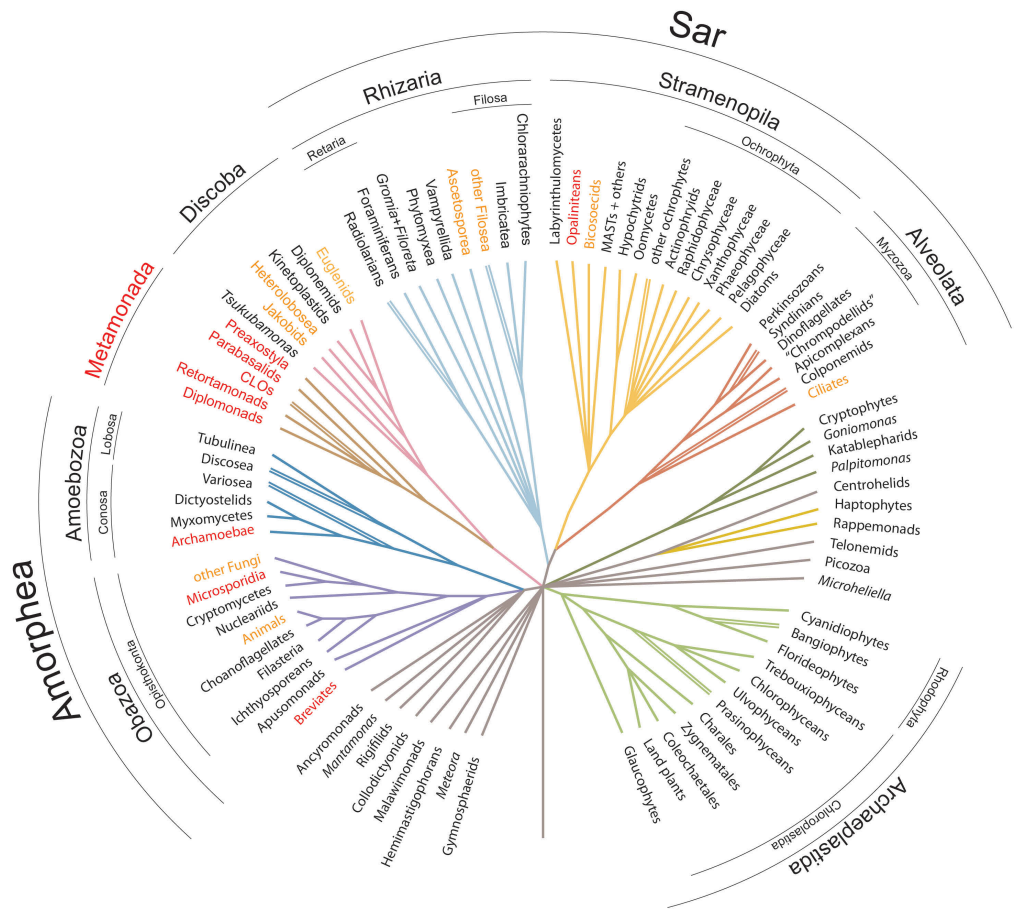


Figure 1.2: Reference tree of eukaryotes. Modified from (Simpson and Eglit 2016) with some updates based on (Burki et al. 2020) and own unpublished phylogenomic data. Fully anaerobic lineages are in red, while lineages known to contain anaerobes are in orange.

many ways, these developments mirror those in the earlier animal phylogenies discussed above: lower complexity taxa often turned out to be secondarily reduced—*e.g.* the complete demise of 'archezoa' as an assemblage of ancestrally mitochondrion-lacking extant taxa (Embley 2006)—and numerous prior inferences based on incorrect phylogenies were overturned.

In this thesis, I markedly increase the deep taxon sampling of heterotrophic protists from the perspective of resolving the tree of eukaryotes. I prioritised cultivation-based approaches, since, in addition to expanding the available taxon sampling by contributing new molecular sequences, I specifically aimed to broaden the collection of taxa accessible to further downstream studies—for example, genome sequencing or cell biology investigations. In chapter 2, I describe establishing several dozen cultures of organisms that are distinctive on the SSU rDNA (DNA

encoding small subunit ribosomal RNA) sequence level. In the case of some particularly phylogenetically distinct isolates, I took advantage of the cultivation-based approach to examine their fine structure using electron microscopy. I also used rDNA sequence data to probe a selection of publicly available environmental sequence datasets to explore the distribution of the newly discovered groups across habitat types. In chapters 3 and 4, select organisms whose position in the tree of eukaryotes was not resolved by single gene (*i.e.* SSU rDNA) molecular phylogenies were subject to multi-gene (phylogenomic) analyses. Sequences for protein-coding genes were obtained from either bulk-culture- or single-cell-transcriptome sequencing. In the course of this work, we improved the shallow taxon sampling of several major clades and revealed the existence of a candidate novel supergroup.

Chapter 2

Isolation and characterisation of novel heterotrophic protists

2.1 Introduction

Despite its over 350 year history (Lane 2015), the age of discovery of new eukaryotic microbes has yet to taper down (Burki et al. 2020). A half century of molecular phylogenetics studies (since Ishikawa 1977) has subtantially transformed our understanding of eukaryote evolution. An important part of this has been environmental sequencing approaches, which have revealed entire novel clades known only from molecular sequence data (eg. Takishita et al. 2007, López-García et al. 2003). Some of these clades, like the conspicuously diverse environmental clade of diplonemids from deep sea (López-García et al. 2007, Lara et al. 2009, Scheckenbach et al. 2010) and planktonic environments (de Vargas et al. 2015a), have subsequently had some of their members characterised by morphology (Gawryluk et al. 2016b). In a similar vein, the Marine Alveolate 1 (MALV-1) clade was found to be an assemblage of parasites of other protists (Bråte et al. 2012). Other environmental clades, however, remain undersampled or altogether uncharacterised. For example, to date, neither Novel Clade 12 (NC-12) (Bass et al. 2018) in Rhizaria nor the phylogenetically ambiguous NAMAKO-1 and NAMAKO-2 lineages (Takishita et al. 2007) have representatives with known cell morphology.

While some claims have been made about researchers getting closer to exhausting the microbial eukaryote diversity of some types of planktonic habitats (Vargas et al. 2015b), overall

it is clear that uncharacterised biodiversity remains considerable (Forster et al. 2016). A majority of eukaryotic microbial ecology studies to date have focused on planktonic habitats (Vaulot et al. 2022a, Forster et al. 2016), which are more amenable to DNA and RNA extraction and sequencing. A much smaller abundance of samples exists for soils (approx. 800 in MetaPR2, compared with 3259 for water column; Vaulot et al. 2022a), and a tiny handful for marine and freshwater sediments (<20 in MetaPR2). Among sediment samples, a substantially higher portion of sequences cannot be classified (Forster et al. 2016). In addition, the claims that the discovery of biodiversity in some areas is saturating are based on short read amplicon sequencing. Meanwhile, metagenomic, metatranscriptomic, and especially, single-cell approaches are contributing increasingly more to the sampling of existing clades of microbial eukaryotes (Lax and Simpson 2020, Lax et al. 2019), and, importantly, to finding new lineages (Lax et al. 2018, Wideman et al. 2019).

Although the discovery of novel biodiversity in microbial ecology is typically equated with environmental sequencing, cultivation remains an important exploratory tool. Cultivation not only presents another set of techniques for expanding our sampling of novel diversity, but brings a considerable advantage by enabling a much greater range of research on the organism in question. Going back to diplonemids, since the initial discovery of the 'environmental clade' and subsequent characterisation of a few representatives' morphology using single cell methods, a considerable number of additional diplonemid cultures have been established (*e.g.* Tashyreva et al. 2018a), on top of sparking development of transformation techniques in existing strains (Kaur et al. 2018). The cultures facilitated ultrastructure studies, allowed the demonstration of flexibility in transition between osmotrophy and bacterivory on an organismal level (Prokopchuk et al. 2022), and the discovery of exceptional accumulations of barite and celestite crystals in diplonemids (Pilátová et al. 2022). Thus, the establishment, molecular characterisation, and first-pass morphological description of a cultivated organism is only a starting point as long as the culture remains viable.

In this work, I report efforts to cultivate novel diversity from across the tree of eukaryotes, focused substantially but not exclusively on selected poorly sampled environments (anaerobic and/or hypersaline and/or alkaline) and understudied life histories—specifically, eukaryotrophy, or eukaryotes feeding on other eukaryotes.

2.1.1 Habitats

2.1.1.1 Anoxic environments

Low oxygen environments are naturally present in a variety of habitats, from marine and freshwater sediments to oxygen minimum zones in the water column and anoxic layers of meromictic lakes, to digestive tracts and other microhabitats on and inside animal bodies. Typical sediments contain a redoxcline from oxygenated surface layers to anoxic depths, where a gradient of optimal electron acceptors supports stratified layers of differing microbial communities, for example denitrifiers closer to the oxic layer and sulfate reducers farther away, with a methanogen zone at the very bottom (Treude 2012). Conventionally, focus on anaerobic microbes has largely been on their prokaryotic fraction, as well as parasitic eukaryotes of medical or economic importance. In fact, microbial eukaryote diversity was expected to be fairly low in anoxic layers, in part due to sulfide toxicity (Triadó-Margarit and Casamayor 2015).

By contrast, molecular sequencing and cultivation efforts suggest there remains much diversity of anaerobic microbes to be characterised (Zuendorf et al. 2006, Takishita et al. 2007, Epstein and López-García 2008, Stoeck et al. 2010, Edgcomb et al. 2011, Orsi et al. 2012, Triadó-Margarit and Casamayor 2015). Polyxenic culturing approaches and expansion of interests beyond the classical medically relevant protistan parasites have yielded a number of novel anaerobic lineages within the past 20 years (Kolisko et al. 2010a, Táborský et al. 2017, Gawryluk et al. 2016c, Burki et al. 2013, Walker et al. 2006, Yubuki et al. 2015a). Many of these lineages represent novel adaptations to low oxygen environments, including resistance to toxic effects of anaerobiosis in key metabolic pathways (Stairs et al. 2018). Furthermore, recent works show that facultative anaerobes and those not far along the path of adaptation are widespread throughout the tree of eukaryotes (Gawryluk et al. 2016c, Roger et al. 2017, Leger et al. 2019). Lastly, anaerobic eukaryotrophs are a particularly undersampled assemblage (see discussion in Chapter 4). Thus, I set out to explore some of the undercharacterised diversity of anaerobic communities in sediments through a combination of sampling low oxygen environments and enrichments with high organic content promoting sufficient prokaryotic growth to suppress oxygen levels.

2.1.1.2 Saline lakes

High salt environments (with salinity substantially higher than that of seawater) include human made solar salterns (Ventosa et al. 2014), brine channels in sea ice (DN Thomas and Dieckmann 2009), deep hypersaline anoxic basins (DHABs) (Stock et al. 2012), and athallasic salt lakes (Harding and Simpson 2018) – to name a few. In this chapter, the focus will be on salterns and thalassic salt lakes, with the chemical and physical diversity of these environmnets simplified to two types: salt lakes and, with high pH, soda (alkaline) lakes. Saline lakes of both types usually vary considerably in salinity throughout time, in some cases subjecting their organismal communities not only to low water availability and an extreme osmotic environment, but also challenging rapid changes in conditions. My search for hypersaline protists from salt lakes has mostly been focused on habitats with at least 150ppt salinity, whereas soda lake samples with salinity as low as 30ppt (but pH>9) were still considered in this chapter.

2.1.1.3 Soda lakes

Soda lakes are usually found as endorrheic bodies of water in regions with little exposed sedimentary rock (Schagerl and Renaut 2016, Jones et al. 1998); without the calcium and magnesium from deposits of shells in ancient aquatic sediments to sequester carbonate ions from water, sodium carbonate and sodium bicarbonate ions accumulate to dominate the lake chemistry, up to high salinities and pH. They can be chemically diverse (Sorokin et al. 2015, Boros and Kolpakova 2018). Classic examples are Rift Valley lakes in Eastern Africa, and lakes in the mountains along the west coast of the Americas, such as Mono Lake in California. A population of such lakes exists in the Pacific Northwest region in both US and Canada, such as Soap Lake in Washington (Sorokin et al. 2007), or Goodenough and Last Chance Lakes in interior BC (Zorz et al. 2019). Despite being among the more productive lake ecosystems (Pirlot et al. 2005), the eukaryotic microbiology of alkaline lakes has received relatively little attention (Pirlot et al. 2005, Yasindi and Taylor 2006, Zorz et al. 2019), particularly from a morphological perspective (but see Yasindi and Taylor 2016, though mostly focusing on ciliates)

2.1.1.4 Eukaryotrophs

In the early part of the molecular era, the majority of focus on protists has been on autotrophs, medically-relevant parasites, and bacterivorous heterotrophs. Technological limitations at the time severely restricted molecular analyses of cultures containing more than one eukaryote.

Over the past two decades, this has been turned around and the majority of major novel heterotrophic protistan lineages discovered in recent years have been eukaryotrophic (Burki et al. 2020; Tikhonenkov 2020). This reflects a "catching up" once better sequencing and bioinformatic techniques have enabled transcriptomic and now genomic study of mixtures of eukaryotes. For example, non-metazoan holozoans were until recently primarily bacterivorous or parasitic, comprised of groups like choanoflagellates and ichthyosporea. In the past decade alone, three distinct novel lineages of eukaryotrophs were added to the holozoan tree (Hehenberger et al. 2017, Tikhonenkov et al. 2020a). Furthermore, *Ancoracysta* was recently found as new distinct lineage outside known supergroups (Janouškovec et al. 2017). Even the newest major lineage in Archaeplastida turned out to be eukaryotrophic (Gawryluk et al. 2019)!

2.1.2 Summary

This chapter summarises a "low-throughput" approach using light microscopy and classical cultivation techniques to isolate morphologically-distinctive protists that represent candidate new major lineages or taxa whose phylogenetic affinities are not immediately obvious. It is important to note that my sampling approaches do not attempt to capture a snapshot of diversity at the time, nor quantify relative or absolute abundances of any organism therein. The aim is cherry-pick potentially novel or otherwise interesting organisms for further study. Results include microscopy images and phylogenies showing isolated organisms loosely grouped by taxonomic affiliation; residual organisms that escaped molecular characterisation by SSU rDNA (DNA encoding small subunit ribosomal rDNA) are included at the end. The SSU rDNA sequences are also used to attempt to detect probable similar taxa in public short read environmental sequence data.

2.2 Materials and Methods

2.2.1 Overview

Samples, usually containing sediment, were obtained in 15 mL or 50 mL conical centrifuge tubes, and typically enriched within 2-3 days with suitable media (Table 2.3). Enrichments were monitored every 1 or 2 days on an inverted microscope with phase contrast optics for protists with morphology resembling either previously described *incertae sedis* taxa ('known unknowns') or

organisms we could not find similar descriptions to in the literature ('unknown unknowns'). Candidates were imaged further with DIC optics on a Zeiss Axiovert 200M and its AxioCam M5 (Carl Zeiss AG) camera to confirm morphology as well as attempt to predict culturing approaches most likely to be successful, *e.g.* by searching for evidence of eukaryotrophy. Some organisms were incubated overnight under a petroleum jelly sealed coverslip prior to imaging. Further details on isolation and cultivation provided below.

For molecular study, approximately 10mL volumes (typically) of established cultures were pelleted by centrifugation and subjected to DNA extraction using the QIAGEN DNEasy Blood and Tissue Kit (QIAGEN 567 N.V., Hilden, Germany), with elution in 30-50uL RO water. In cases where cultivation appeared unlikely to be successful, or where the presence of a second (prey) eukaryote in culture was unavoidable, single cells were picked with a glass pipette drawn out over a flame and placed in 5uL sterile molecular grade or RO water in a 200 µL PCR tubes. In some cases, high magnification images for each picked cell were obtained on the Zeiss Axiovert. Sometimes multiple cells were pooled. The cells were subject to Whole Genome Amplification using a Multiple Displacement Amplification (MDA) kit (GE Health). The amplified product was then used directly as input for PCR.

Name	Fwd or rev	Sequence (5'-3')
25F	F	CATATGCTTGTCTCAAAGATTAAGCCA
42F	F	CTCAARGAYTAAGCCATGCA
82F	F	GAAACTGCGAATGGCTC
EukA	F	AACCTGGTTGATCCTGCCAGT
Hemi2-342F	F	ACTTTCGATTGTAGGATAGA
1492R	R	AAGTCGTAACAAGGT
1498R	R	CACCTACGGAAACCTTGTTA
EukB	R	TGATCCTTCTGCAGGTTCACCTAC
Hemi2-1103R	R	AAAACTTGCGATTTCTCTGG
SSF-1345R	R	TAATCTAGCCCCATCACGTTGCA

Table 2.1: Primers used for obtaining SSU rDNA sequences

PCR with universal eukaryotic primers (Table 2.1); many of the universal primer sequences were obtained from Adl et al. 2014) was used to amplify partial SSU rDNA (Table 2.2.1 for Sanger sequencing (Génome Québec). Upon succesful sequencing, the Sanger reads were manually trimmed and assembled in Geneious R10 (Kearse et al. 2012), then used to search NCBI nt (nucleotide) via BLASTn for matches with significant (usually >97%) sequence identity. The sequences were then added to a pan-eukaryotic SSU rDNA alignment and phylogeny inferred

using maximum likelihood under the GTR+I+Γ model in RAxML (Stamatakis 2014). Organisms whose evolutionary position was unresolvable with SSU rDNA phylogenies were then candidates for multi-gene phylogenomic analyses using transcriptome sequencing (for examples, see chapters 3 and 4).

SSU rDNA sequences were also used to search publicly available V4 and V9 short read environmental SSU rDNA sequence datasets (see Table 2.6) for high probability matches in other locations and environments. Additionally, some organisms were subject to preliminary transmission electron microscopy (TEM) and/or scanning electron microscopy (SEM) imaging to examine ultrastructure and surface morphology, respectively. Further details of these procedures are provided below.

Table 2.2: Methods for extracting and amplifying SSU rDNA sequences for novel isolates. (continues on the next page)

Isolate ID	Organism or taxon	Isolated by	SSU?	whole culture or MDA?	Fwd Primer	Rev Primer	Temp	nested?	Fwd nest primer	Rev nest primer	Nest temp
LRS2 LÖ	colponemid	YE	Andrea Gigeroff								
F2 LÖ	colponemid	YE	Andrea Gigeroff								
GEM Colp	colponemid	YE	Andrea Gigeroff								
PSL3 Pal	colponemid	YE	Andrea Gigeroff								
TZLM1-NF	colpodellid	YE	full	MDA	EukA	EukB	63.0	Y	82F	1498R	58.0
SoapColp	colponemid-like	YE	partial	MDA	EukA	EukB	60.0	N			
PCE barth	barthelonid	YE	full	whole culture	82F	1498R	58.0	N			
LRM2 barth	barthelonid	YE	collaborator: Yazaki et al. 2020								
FB11 barth	barthelonid	YE	full	whole culture	82F	1498R	58.0	N			
TZLM1 RC	retortacarp	YE	full	whole culture	EukA	EukB	60.0	N			
TZLM3 RCL	retortacarp	YE	partial	whole culture	82F	1498R	55.0	N			
GEM RC	retortacarp	YE	partial	whole culture	EukA	EukB	58.0	N			
Soap18RC (2018)	retortacarp	YE		whole culture	EukA	EukB	60.0	N			
Soap20A RC (2020)	retortacarp	YE	partial	whole culture	EukA	EukB	58.0	N			
FB10N	breviate	YE	full	whole culture	EukA	EukB	60.0	N			
AuB brev	breviate	YE	full	whole culture	EukA	EukB	60.0	N			
LRM1b	breviate	YE	partial	whole culture	EukA	EukB	58.0	N			
PCE brev	breviate	YE	full	whole culture	EukA	EukB	55.0	N			
Saa brev	breviate	YE	full	whole culture	EukA	EukB	60.0	N			
LRM2N6brev	breviate	YE	full	whole culture	EukA	EukB	60.0	N			
(halobrev)	breviate	YE	none								
CARMGS BLO	BLO (breviate)	YE	full	whole culture	25F	EukB	61.0	N			
SnP L3C	Lithocolla	YE	collaborator: Galindo et al. 2019								
LAR Mayo	amoeba (paramoeba)	YE	partial		82F	1498R	60.0	N			
(LAR MayoP)	parasome	YE	none								
(LAR MayoP)	parasite	YE	none								
PCE SSF	squishy flagellate	YE	partial	MDA	25F	SSF-1345R	55.0	N			
Saa SSF	squishy flagellate	YE	partial	MDA	82F	1498R	60.0	Y	82F	1492R	55.0
TBB1 PG	pyriform glider	YE	partial	MDA	EukA	EukB	53.0	N			
QSI PG	pyriform glider	YE	full	MDA	EukA	EukB	54.0	N			
DM2	Heliomorpha (=Dimorpha)	Gordon Lax	full	MDA	EukA	EukB	55.0	Y	25F	EukB	60.0
BoP4.1 Q	Q	YE	partial	whole culture	EukA	EukB	60.0	Y	EukA	1498R	58.0
BoP4.1 HT	Discocelis	YE	full	whole culture	EukA	EukB	53.0	Y	EukA	1498R	53.0
BoP4.1 RotF	rotating flagellate	YE	partial	whole culture	25F	EukB	63.0	N			
BWH RotFL	rotating flagellate like	YE	partial	MDA	EukA	EukB	60.0	N			
(LIS MC)	Multicilia	YE	none								
(PSL3 MC)	Multicilia (halophile)	YE	none	MDA							
LRM3-1F	uniflagellate amoebozoan	YE	partial	MDA	25F	EukB	63.0	N			

For sampling location information, see Table A.1. Isolates in brackets have not yet been sequenced and are included to clarify that. Names under "SSU?" are of collaborators who obtained the sequences.

Table 2.2: Methods for extracting and amplifying SSU rDNA sequences for novel isolates. (continued from the previous page)

Isolate ID	Organism or taxon	Isolated by	SSU?	whole culture or MDA?	Fwd Primer	Rev Primer	Temp	nested?	Fwd nest primer	Rev nest primer	Nest temp
KSS Trichosph.	Trichosphaerium	YE	collaborator: Matthew Brown								
TZLM1-YAF	Dactylomonas / Selenaion sp.	YE	full	whole culture	25F	EukB	55.0	Y	25F	1498R	55.0
TZLM3-JC	Jakobacarp	YE	full	whole culture	EukA	1498R	55.0	Y	25F	1498R	60.0
(Soap20 RF)	(Ridgy Flagellate)	YE	none								
CuF1-MH	centrohelid	YE	partial	MDA	EukA	EukB	58.0	N			
CuSP2-3 BLOL	palpitomonad	YE	partial	MDA	25F	EukB	55.0	N			
(CuF1-Anc)	(Ancyromonas melba)	YE	none								
BW2 Spiro	spironemid	YE	full	MDA	82F	1498R	52.0	N			
BW2 Hemi	Hemimastix	YE	full	whole culture	Hemi2-342F	Hemi2-1103R	55.0	N			
LBC3 Met	Meteora	YE	full	whole culture	EukA	1498R	63.0	Y	82F	1498R	58.0
MSW3C	Protist X	YE	full	MDA	EukA	EukB	55.0	N			
CuCandC4	Protist X	YE	partial	MDA	EukA	EukB	63.0	N			
FBttCande	Protist X	YE	full	MDA	EukA	EukB	60.0	N			
LRM3-SF	small flagellate /amoeba	YE	full	whole culture	25F	EukB	61.0	Y	42F	1498R	58.0
SJB2	excavate-like' flagellate	YE	full	whole culture	EukA	EukB	58.0	N			

For sampling location information, see Table A.1. Isolates in brackets have not yet been sequenced and are included to clarify that. Names under "SSU?" are of collaborators who obtained the sequences.

2.2.2 Sampling, enrichment, and cultivation

Samples were kept sealed and 0.5-1 mL aliquots including sediment, if present, were added to enrichment media, typically a couple days after sampling. One sample, LRM3, was re-enriched two years after sampling date. Marine enrichment media were based on autoclaved natural seawater from the Northwest Arm, Halifax, while hypersaline samples were added to medium "HS#5" (NaCl 137.6, KCl 3.8, $MgCl_2 \cdot 6H_2O$ 13.45 $MgSO_4 \cdot 7H_2O$ 1.65, $CaCl_2 \cdot 2H_2O$ 0.65 $g \cdot L^{-1}$; Park 2012), diluted to similar salinities to those of the source samples as determined by a high salinity refractometer. Alkaline lake samples were enriched using medium "TZ" (KH_2PO_4 0.2, $MgCl_2$ 0.1, KCl 0.2, NH_4Cl 0.5, NaCl 100, Na_2CO_3 68, $NaHCO_3$ 38 $g \cdot L^{-1}$; adapted from Mesbah et al. 2007), likewise diluted to appropriate concentrations determined by a high salinity refractometer. Aerobes were usually enriched for by supplementing the medium with a single autoclaved wheat grain or with 0.1% v/v LB medium in an unvented tissue culture flask stored horizontally. Anaerobes were enriched in 12mL volumes in a sealed 15mL conical centrufuge tube sealed and stored upright, with 3% v/v LB (eg. in seawater, "3%LS") and/or four autoclaved grains added to provide organic enrichment for the prokaryotic community. In some cases, resazurin was used as a chemical indicator to assess oxygen levels and detect anoxic layers in enrichments (Karakashev et al. 2003).

The enrichment cultures were monitored by light microscopy every 1-2 days. Cells of interest were either directed to cultivation efforts or, in some cases, picked for single-cell molecular methods (see above). Cultures were established either by serial dilutions or long-term maintenance leading to decline in contaminants, or via single cell isolation. For eukaryotrophs and anaerobes, prey eukaryote and bacterial communities were supplied or established first. Upon establishment, most culture were passaged at 1, 2, or 4 week intervals, with the exception of slower-growing halophiles passaged every several months.

2.2.3 Electron microscopy

By default, cells were harvested by centrifugation. More fragile organisms, especially anaerobes, were instead left undisturbed in 15mL conical centrifuge tubes as the top fluid was removed until the last 2 mL, the bottom of which was gently collected and added to fixative (Table A.2). Cells were spun between washes, and then dehydrated through an ethanol or acetone series and infiltrated and embedded in EPON 812 resin. The resin blocks were sectioned in series using a Leica UC6 ultramicrotome. Preliminary examination skipped staining, but sections of adequately

fixed preparations were stained with lead citrate and either uranyl acetate or Uranyless™ (Benmeradi et al. 2015). Sections were imaged on a Tecnai-12 transmission electron microscope using a FEI Gatan camera.

SEM fixation was done on cells added to poly-L-lysine coated coverslips in multiwell plates (Table A.3). Then, exchanges of increasing concentrations of ethanol were performed for dehydration. Cells attached to coverslips were then dried by a critical point dryer (Leica EM CPD300) and coated with a gold-palladium alloy in a Leica EM ACE200 sputter coater, both at the EM Core Facility in the Life Sciences Centre. Imaging was done at the DalTech SEM-FIB facility, Faculty of Mechanical Engineering. Downstream image processing for both electron and light microscopy was done using FIJI (Rasband 1997; Schneider et al. 2012).

2.2.4 Phylogenetic analyses and environmental sequence placements

The reference global alignment for SSU rDNA from the *Meteora* project (Eglit *et al.* in prep; chapter 3), was expanded considerably in its taxon sampling by adding sequences of organisms characterised here, as well as environmental sequences from NCBI and published environmental datasets (Takishita et al. 2007, Jamy et al. 2020a). The alignment was trimmed with a mask derived from gblocks (Castresana 2000) readjusted for a final matrix of 1176 sites and 250 taxa. A maximum likelihood phylogeny was inferred in RAxML-NG (Kozlov et al. 2019) under the GTR+Γ+I model, and support values provided by 200 non-parametric bootstrap replicates. SSU rDNA sequences of several isolates ended up branching within Rhizaria, so to make a second alignment focused on that group, an original alignment kindly provided by Sebastian Hess was expanded with additional relevant taxa plus known environmental lineages from NCBI nt, as well as the literature (Takishita et al. 2007, Bass et al. 2018), particularly for endomyxea. The taxon sampling of Filosa was intentionally kept sparse as resolving branching orders within filosan groups is difficult using SSU rDNA alone. The alignment mask was started from gblocks and curated manually for a final matrix of 1192 sites across 113 taxa. The phylogeny was inferred using the same parameters as above.

The available V4 and V9 regions of the 39 new SSU rDNA sequences reported here (38 and 36 regions, respectively) were used as queries for a greedy BLASTn search of a custom combined public environmental sequences database (14 million sequences, see 2.3) with the settings `-perc_identity 80`, `-max_target_seqs 500`, and `-qcov_hsp_perc 70`. The setting

`-qcov_hsp_perc 70` was included to reduce spurious high percent identity matches over ultra-short portions of the sequence, while allowing for partial sequence matches. The corresponding matching sequences were then aligned against the untrimmed reference global eukaryotic SSU rDNA alignment using PaPaRa (SA Berger and Stamatakis 2011) and short sequence placements inferred via RAxML-EPA (SA Berger et al. 2011) using the above reference phylogeny. Matches among Rhizaria with likelihood weight ratios ≥ 0.5 were further refined by repeating the alignment and placement procedures described above with the densely sampled Rhizaria SSU rDNA alignment and corresponding reference phylogeny. This analysis is outlined in fig. 2.1.

Shorthand	Reference	Type	Size	Habitat types	Acknowledgements
VAMPS2021	Huse et al. 2014	V4, V9	1106510	various	
TARAv9	de Vargas et al. 2015a	V9	2306025	sunlit open ocean	
DeepSeaV9	Schoenle et al. 2021	V9	472664	deep sea sediment (bathyal, abyssal, hadal)	Alexandra Schoenle
NeoMetaTv4	Mahé et al. 2017	V4	75613	neotropical soils: Costa Rica, Panama, Ecuador	Frédéric Mahé
MalaspinaV4	Obiol et al. 2020	V4	25843	marine water column	
BCSodaLakesV4	Zorz et al. 2019	V4	587	interior BC soda lake mats (and sediment)	Jackie Zorz
GuaymasV4	Pasulka et al. 2019	V4	11129	hydrothermal vent, surface layers of sediment	
Cariaco	Suter et al. 2021	V4	14176	marine anoxic water column, Cariaco basin	Elizabeth Suter
Biomarks	Vaulot et al. 2022a	V4	9097	metanalysis: TARA, Malaspina, polar, other	Daniel Vaulot
SoilMetaT	Geisen et al. 2015	V4	199275	temperate soils	
GordaRidge	Hu et al. 2021	V4	20399	10-80C vent fluids above hydrothermal vent	
RiaFormosa	Filker et al. 2015	V4	1515	saltern water column, 3 ponds	Sabine Filker
Filker2017	Filker et al. 2017	V4	251807	39-440ppt salinity, salterns, multiple locations	Sabine Filker
TOTAL:			1445324 0		

Table 2.3: Public environmental sequence datasets surveyed in Chapter 2

Sequences corresponding to placements at edges of interest with likelihood weight ratios ≥ 0.9 were extracted using an R script and then used for reciprocal BLASTn against NCBI nt (downloaded 08 Aug 2022), with Viruses, Bacteria, Bilateria, Tracheophytes, and Dikarya removed from the search for computational simplification, combined with full length and near-full length private SSU rDNA sequences for queried organisms using `blastdb_aliastool`.

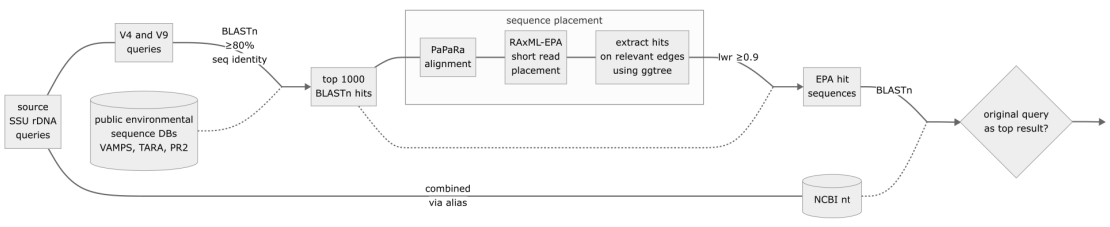


Figure 2.1: Overview of the environmental sequence survey analysis.

2.3 Results

Over 100 cultures were established, of which 66 are still viable as of August 2022 (Table 2.3). From these, I determined 39 SSU rDNA sequences, with all but the phenotypically distinctive palpitomonad 'BLOL' had <97% sequence identity to anything in the NCBI nucleotide database, thus representing novel species or genera of eukaryotes. Of these SSU rDNA sequences, only half had any probable matches in the public short read environmental sequence data surveyed in this study. Of the major novel lineages, *Meteora sporadica* and Protist X are examined in greater detail in chapters 3 and 4 of this thesis, respectively, while Hemimastigophora have been published as: Lax, Eglit, Eme *et al* (2018).

Table 2.4: Isolates cultivated in Chapter 2. (continues on the next page)

Isolate name	Description	Still viable? (2022)	DNA extracted?	SSU rDNA sequence?	Source	Habitat	(p)redator: of, (b)acterivore
LRM1B	breviate (sister to pygsuia)	Y	Y	Y	Delta de l'Ebre, Spain	M, A	b
FB10N2-Lnbrev	long-necked breviate – *Subulatomonas*	Y	Y	Y	San Juan Is, WA	M, A	b
FB5Mast	mastigella-like – pelobiont	Y	Y	N	San Juan Is, WA	M, A	b
SaaNSSF	seriously squishy flagellate – endomyxean	N	Y	Y	Saanich inlet, BC	M, A	p: SaaB, LRM1b
PCE SSF	seriously squishy flagellate – endomyxean	Y	Y	Y	PEI-Cavendish-Eelgrass mud	M,A	p:PCEBrev (also SaaBrev)
SaaN"brev"	breviate (basal to pygsuia branch)	Y	Y	Y	Saanich inlet, BC	M, A	b
LRM2N4-Barth	*Barthelona* – fornicate	Y	Y	N	Delta de l'Ebre, Spain	M, A	b
FB11N4-Barth	*Barthelona* – fornicate	N	Y	Y (p)	San Juan Is, WA	M, A	b
Cu11-3S_SB	*Barthelona*-like – fornicate	Y	Y	N	Curaçao	H, A	b
LRS2-WGF/LO	colponemid	Y			Delta de l'Ebre, Spain	H	p:LRS3Perc
FS2SB	*Barthelona*-like, smaller than Cu11 – fornicate	N	N	N	Faro, Portugal	H, A	b
LRS2-SB	*Barthelona*-like, smaller than Cu11 – fornicate	Y	N	N	Delta de l'Ebre, Spain	H, A	b
PSL3-SB	*Barthelona*-like, smaller than Cu11 – fornicate	N	N	N	Cuba	H,A	b
FS2-Hlobo	poss. *Euplaisiobystra*?	Y	N	N	Faro, Portugal	H	b
TZLM5-Ph	*Pharyngomonas* sp.	Y	Y	N	Lake Manyara, Tanzania	H, B	b
TZLM1-T	trichomonad	Y	Y	N	Lake Manyara, Tanzania	H, B, A	b?
TZLM1-RC	Retortacarp, fornicate	Y	Y	Y	Lake Manyara, Tanzania	H, B, A	b
TZLM1-NF	Colpodellid "narrow flagellate"	Y	Y	Y	Lake Manyara, Tanzania	H, B	p:TZLM1-Ph
TZLM1-YAF	yet another flagellate' - *Dactylomonas*-like	Y	Y	Y (p)	Lake Manyara, Tanzania	H, B, A	b?
TZLM3-RCL	Retortacarp-like (confirmed)	Y	Y	Y (p)	Lake Manyara, Tanzania	F, B, A	b
TZLM5-Cerco	Cercomonad (anaerobe?)	Y	Y	N	Lake Manyara, Tanzania	F, B, A/O	b
TZLM3-PGL	Pyriform glider like	N	N	N	Lake Manyara, Tanzania	F, B,A?	p:TZLM3Cerco
TZLM3-Jakobacarp	Heterolobosean	Y	Y	Y	Lake Manyara, Tanzania	F,B	b
MSW3C	"Protist X"	Y	Y	Y	McNab's Is, Halifax	M, A	p: BICM, MB1
TBB1-PG	pyriform glider	N	Y	N	Boundary Bay	M, A	p: LRM1b
LIS-MC	*Multicilia*	Y			Luke Island sediment	M	p: Hak-Van
SnP-MC	*Multicilia*	N	N	N	Split Nose Point, sediment	M	p:Hak-Van

SSU rDNA sequence: (Y)es, (N)o, (p)artial, from single (c)ell; Habitat types: (M)arine, (H)ypersaline, (B)asic (alkaline), (F)reshwater, (S)oil, (A)noxic, (O)xic. For sampling location information, see Table A.1.

Table 2.4: Isolates cultivated in Chapter 2. (continued from the previous page)

Isolate name	Description	Still viable? (2022)	DNA extracted?	SSU rDNA sequence?	Source	Habitat	(p)redator: of, (b)acterivore
SnP-Lithocolla	*Lithocolla* (nucleariid)	Y			Split Nose Point, sediment	M	p:ISO
HAK-MF-VAN	Vannellid (for *Multicilia*)	Y	N	N	Hakai (mud flat)	M	b
LRM2N6-brev	micro breviate, SaaBrev-like	Y	Y	Y	Delta de l'Ebre, Spain	M, A	b
AbB-Lnbrev	probable *Subulatomonas*	N	Y	Y	Bermagui, Aus	M,A	b
LBC3 Met	*Meteora*!	Y	Y	Y	La Boca, Cuba	M	b
BoP4.1 RotF	Rotating flagellate *Abollifer*-like (90%)	Y	Y	Y	Bay of Pigs, Cuba	M	p:GEMKIN, *Isochrysis*
BoP4.1 Kin	kinetoplastid (*Neobodo* sp.?)	Y	N	N	Bay of Pigs, Cuba	M	b
LIS-Tel	*Telonema* (to AMNH)	Y			Luke Island sediment	M	p: stramenopile(co-cultured)
LIS-Man	Mantamonad (to AMNH)	Y			Luke Island sediment	M	b
BoP4.1 HT	*Discocelia* sp., Cercozoa	Y	Y	Y	Bay of Pigs, Cuba	M	b
BoP4.1 Q	gliding organism "Q", Filosan	Y	Y	Y(p)	Bay of Pigs, Cuba	M	p:LARKIN,QSS-stram
BW1-Auri	*Aurigamonas*	Y			Bluff Wilderness Trail, HRM	F	p:BW2ss Spumella sp.
BW2-Hemi	*Hemimastix*	Y	Y	Y		F(S)	p:BW2ss Spumella sp.
BW2-SS	small stramenopile (prey) *Spumella* sp.	Y	Y	Y		F(S)	b
PSL3-Halobrev	halophilic "LNbrev"	N	N	N		H,A	b
PSL3-HaloMC	halophilic *Multicilia*	Y				H	p:small amoebozoan
CuF2-Ph	trilobed pharyngomonas-like	N	N	N		H,A?	b
CuJT2-HaloMC	halophilic *Multicilia*	Y	N	N		H	p:small amoebozoan
PSL3-Colp	*Palustrimonas* (colponemid)	Y			p:pharyngomonads		p:LRS3 Pharyngomonad
BWH-RotF-L	RotF-like (Filosan)	Y	Y(c)	Y(p)	Baddeck, Bras d'Or (by Kira)	M	p:diatoms, kineto
PCE Microbrev	tiny breviate	Y	Y	Y	PEI-Cavendish-Eelgrass mud	M,A	b
PCE barth-like	barthelonid (small!)	Y	Y	Y	PEI-Cavendish-Eelgrass mud	M,A	b
PCE-PCS-like	narrow uniflagellate CLO ("PCS")	N	N	N	PEI-Cavendish-Eelgrass mud	M,A	b
BCWL Pb	large pelobiont (for teaching)	N	N	N			b
KSS-Trich-5C	*Trichosphaerium*	Y			Keji SeaSide high tidal green pool	M	p:synechocystis; b (fac. eukaryotroph)
BoP3.4 BLOL	Palpitomonad	Y	N	N	Bay of Pigs, Cuba	M	b

SSU rDNA sequence: (Y)es, (N)o, (p)artial, from single (c)ell; Habitat types: (M)arine, (H)ypersaline, (B)asic (alkaline), (F)reshwater, (S)oil, (A)noxic, (O)xic. For sampling location information, see Table A.1.

Table 2.4: Isolates cultivated in Chapter 2. (continued from the previous page)

Isolate name	Description	Still viable? (2022)	DNA extracted?	SSU rDNA sequence?	Source	Habitat	(p)redator: of, (b)acterivore
CuCAR17MG BLO	Barthelonid-like organism	Y	Y	Y	Piscadera mangrove mud (~45ppt)	M,A	b (on flocs)
CuSP2-3 BLO-L	Palpitomonad	Y	Y	Y	sulfidic mud, SP2-3 70ppt	M,A?	p:BICM; b (fac. eukaryotroph)
CuSP2-1 Candelabra	Candelabra sp. 2	N	Y	Y	SP2-1 100ppt	M,A (also H)	p:CuSP2-2 halocolos?
CuFort1 MicroH	tiny centrohelid	Y	Y	Y	pond in front of fort	H	p:C3C choano
LRM3-1F	uniflagellate/amoeba	Y			Delta de l'Ebre, Spain (old sample)	M	b
LRM3-sf	tiny amoeba and flagellate forms	Y	Y	Y(p)	Delta de l'Ebre, Spain (old sample)	M	b
SJB2 "Excavate"	excavate-like new lineage	Y	Y	Y	Curaçao, San Juan Baai 2	M	b
CuFort1 Anc	*Ancyromonas melba*?	Y	N	N	pond in front of fort	H	b?
CuJT2017	*Pleurostomum*	?	N	N	Saliña Jan Thiel (saturated pond)	H	b
BC-GEM RCL	Retortacarp-like	Y	Y	Y (p)	Good Enough Lake mat (Zorz)	B,A	b
BC-LCM RCL	Retortacarp-like	N	N	N	Last Chance Lake mat (Zorz)	B,A	b
BC-LCS RCL	Retortacarp-like	N	N	N	Last Chance Lake mud (Zorz)	B,A	b
BC-GEM kineto	kinetoplastid; prey for GEM colp	Y	Y	Y(p)	Good Enough Lake mat (Zorz)	B (also in SW)	b
BC-GEM Colp	Colponemid	Y			Good Enough Lake mat (Zorz)		p:kinetoplastids
LAR-MAYO	*Paramoeba*	Y	Y	Y(p)	Larnaka beach by castle		p:ISO
LAR-MayoP	parasite of *Paramoeba* (cryptomycete?)	Y	Y	N	Larnaka beach by castle	M	p:Mayorellid
LAR-Kin	kinetoplastid (HMM prey?)	Y	N		Larnaka beach by castle	M	b
Soap2 RC	Retortacarp	N	Y	N	Soap Lake 2, WA (2018)	B,A	b
Soap2 Colp	Colponemid? Stramenopile???	Y	Y	Y	Soap Lake 2, WA (2018)	B,M	p:ISO
Soap20A RC	Retortacarp	Y	Y	Y	Soap Lake 2020, WA		b
Soap20B RC	Retortacarp	Y	redo		Soap Lake 2020, WA	B	b
Soap20A RF	ridgy flagellate; deep h-lobosean?	Y	Y			B	b
Soap20A Jakobacarp-like	Looks like Jakobacarp (TZLM3 JC)	Y	N			B	b
FHL Paulinellid	Paulinellid (non-ps)	Y			Friday Harbor (dock)	M	p: co-cultured stramenopiles? b?

SSU rDNA sequence: (Y)es, (N)o, (p)artial, from single (c)ell; Habitat types: (M)arine, (H)ypersaline, (B)asic (alkaline), (F)reshwater, (S)oil, (A)noxic, (O)xic. For sampling location information, see Table A.1.

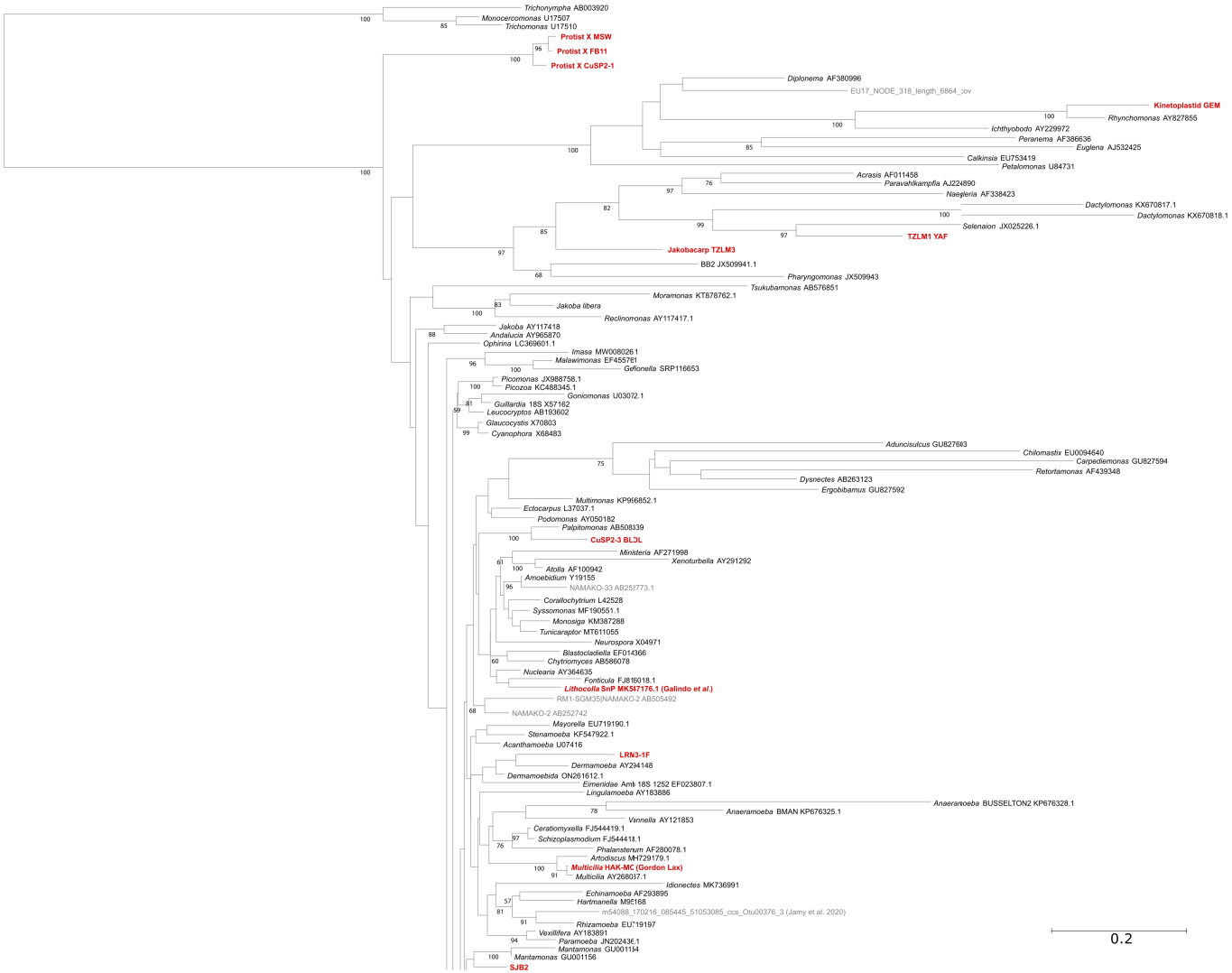
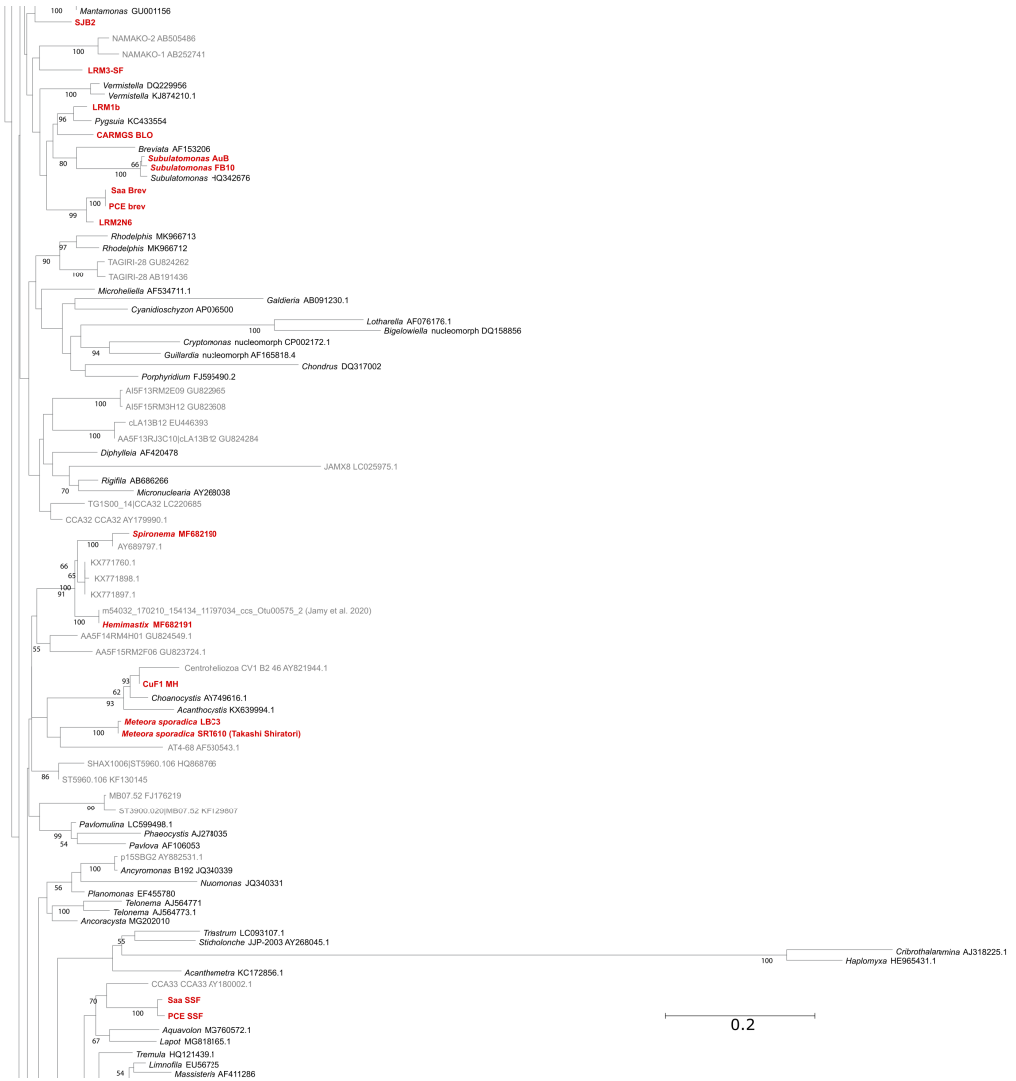
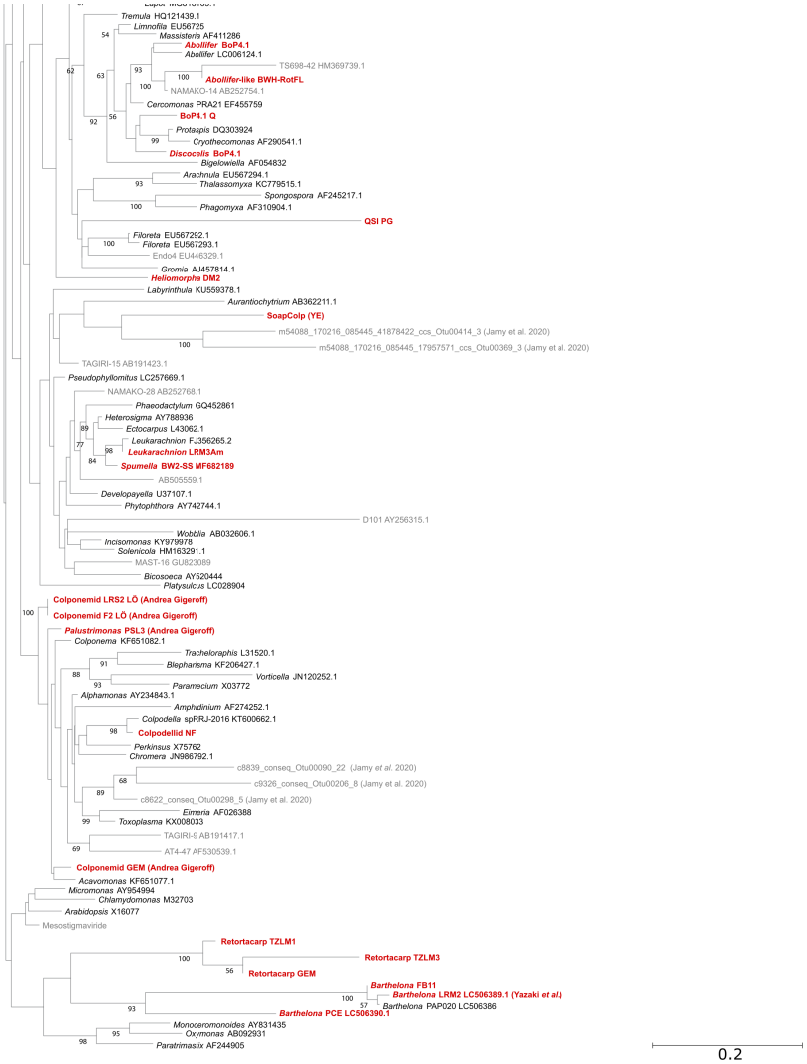


Figure 2.2: Position of new sequences in a global eukaryotic SSU rDNA phylogeny.
(Continued on the next page.)



(Continued on the next page.)



New additions highlighted in red. SSU rDNA sequences obtained by collaborators indicated in brackets. (Continued from previous pages.)

2.3.1 Results by organism

2.3.1.1 Alveolates

Colponemids LÖ, GEM Colp, and *Palustrimonas*

(Gigeroff, Eglit, Simpson; in review)

A number of raptorial eukaryotrophic flagellates with colponemid-like morphology were isolated from hypersaline enrivonments, one of which was also alkaline. Isolates F2-LÖ and LRS2-LÖ are morphologically similar: gliding flagellates with a long trailing flagellum and a flattened cell shape with a depression loosely resembling a spoon, and were isolated from salt pools. Feed on percolomonads and pharyngomonads in culture; poorly fed cells have a prominent protrusion lined with extrusomes that appears to be involved in prey capture. The prey cell is intially contacted by the anterior and then phagocytosed in the groove (fig. 2.3H,J).

GEM-colp was isolated from an alkaline lake microbial mat (pH ~10, 50ppt), and cultivated on a co-occuring kinetoplastid. GEM-Colp is a swimming raptorial eukaryotroph with two unequal flagella emerging subapically, the posterior running through a groove (fig. 2.3C). The groove, in turn, is lined by granules that are likely extrusomes, and has a tooth-like projection at its anterior end just below and to the right of where the posterior flagellum emerges. *Palustrimonas yorkeensis* isolate PSL3-Pal (fig. 2.3A-B) was isolated from a salt pond and cultivated on a co-occuring (much larger) pharyngomonad as prey. Baton-shaped cell with two unequal flagella and a subtle ventral groove. This is the first stable culture of *Palustrimonas* since its discovery and with its recent molecular characterisation being from a now extinct crude culture (Ruinen 1938; Park and Simpson 2015).

SSU rDNA phylogenies place PSL3-Pal with the existing *P. yorkeensis* sequence, while LÖ and GEM-colp are also deep-branching colponemid alveolates, although the branching order remains unresolved (fig. 2.2). No environmental sequence matches were found for any of these SSU rDNA isolates.

Figure 2.2: (Continued from previous page) Maximum likelihood phylogeny calculated from a 250-taxon and 1176-site SSU rDNA alignment under the GTR+Γ+I model in RAxML-NG, aiming to represent the overall phylogenetic diversity of eukaryotes. Support values from 200 non-parametric bootstrap replicates. Taxa from isolates discussed in Chapter 2 are highlighted in red; where the final sequence was provided by a collaborator is indicated in brackets, as are sequences not indexed in NCBI GenBank, from a published long read dataset. Environmental sequences indicated in grey.

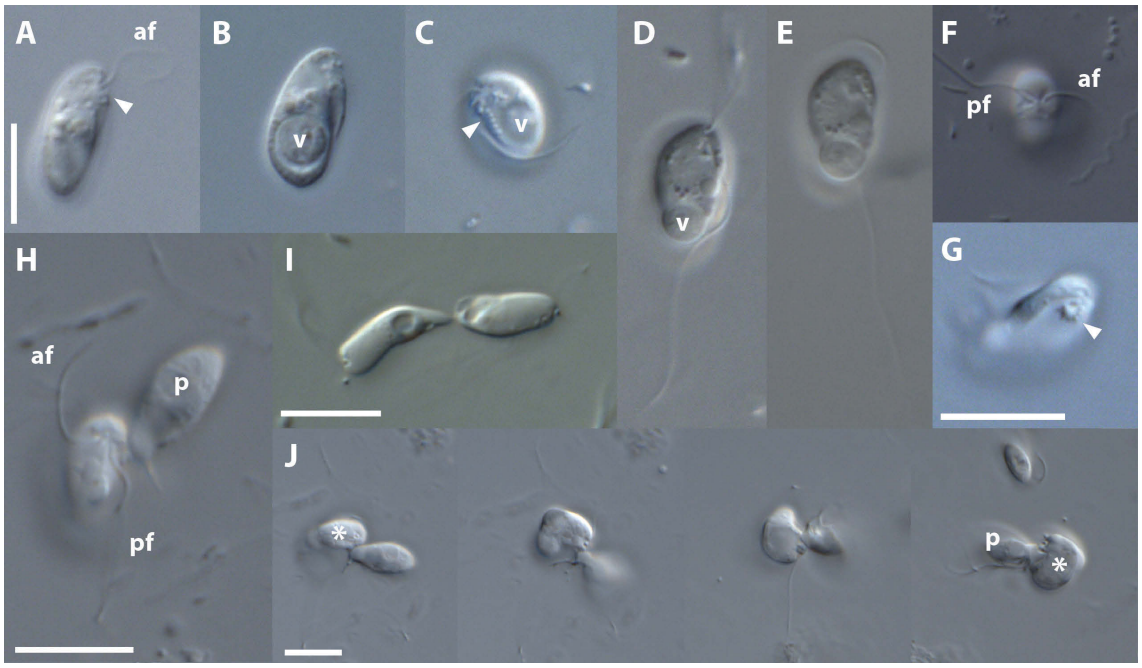


Figure 2.3: Light micrographs of colponemids isolated in this study: *Palustrimonas* sp.,GEM-Colp, and LÖ. A-B) *Palustrimonas* sp. The anterior flagellum (af) inserts sub-apically above a group of granules likely to be extrusomes (arrowhead). A prominent food vacuole is often visible at the posterior of the cell (v). C) Optical section through the flagellar insertion site of GEM-Colp, with the row of granules (extrusomes; arrowhead) visible beneath the posterior flagellum. A food vacuole can be seen in the posterior of the cell (v). D-G) LRS2 LÖ. The cell typically glides passively on its posterior flagellum. A food vacuole (v) can be seen distending the extreme posterior of the cell (D,E). F) View from the top showing the insertion of posterior (pf) and anterior (af) flagella. G) A protrusion with a cluster of granules (extrusomes; arrowhead) can be seen in not well-fed cells. H-J) F2 LÖ. F) Early stage in feeding of LÖ on a prey percolomonad (p), showing the attachment of the prey to the cell's right (also the site of presumed extrusomes in the other colponemid isolates), relative to the insertion of anterior (af) and posterior (pf) flagella. I) A pair of cells in late division. J) Time series of the flagellate (asterisk) in (H) in the early process of the ingestion of its prey (p). Scale bars: 10 m.

Colpodellid "NF"

Extremely narrow (2 x 10-12um) gently twisted flagellate with two flagella roughly the length of the body with a subapical insertion, isolated from rehydrated dried salt crust from Lake Manyara, Tanzania (TZLM1) and cultivated with a co-occuring pharyngomonad as prey. Feeds on the much larger pharyngomonad via myzocytosis and then encysts and divides into four flagellates (fig. 2.4). Branches among colpodellids in SSU rDNA phylogenies (fig. 2.2), which is consistent with mode of feeding and subsequent digestive cysts that divide into multiple progeny. SSU rDNA phylogenies place it among environmental lineages in the *Colpodella angusta* clade with strong support (fig. 2.2). The survey of public environmental sequence data has detected a number of matches from marine anoxic and hypersaline sediments and water, as well as sewage treatment biofilm and cave lampenflora biofilm (fig. 2.23).

Soap "Colp"

Raptorial eukaryotrophic biflagellate superficially resembling a colponemid: a subtle groove underlies the posterior flagellum, while a smaller groove is associated with the anterior (fig. 2.5). Isolated from a haloalkaline sediment sample (50ppt salinity) from Soap Lake, WA, but grows as a stable culture in marine F/2 medium with *Isochrysis* sp. as prey. In culture, a high portion of cells can often be seen mid-division. SSU rDNA phylogenies place SoapColp within Stramenopiles with mediocre support, and no support for their position within stramenopiles (fig. 2.2). No significant matches were found among short read environmental data.

2.3.1.2 Metamonads

Barthelonids

(included in: Yazaki E, Kume K, Shiratori T, Eglit Y, Tanifuji G, Harada R, Simpson AGB, Ishida K.-i, Hashimoto T, Inagaki Y (2020). "Barthelonids Represent a Deep-Branching Metamonad Clade with Mitochondrion-Related Organelles Predicted to Generate No ATP". *PTRSB* 287.1934:20201538)

Barthelona species, first described in Bernard et al. 2000, are flagellates with a conspicuously long flagellum and a distinct J-shaped cytoskeletal element travelling from the flagellar insertion site around the posterior end of the cell (fig. 2.6). Feed on bacteria by ingesting prey through a posterior opening (fig. 2.6K). The position of barthelonids within Metamonada as

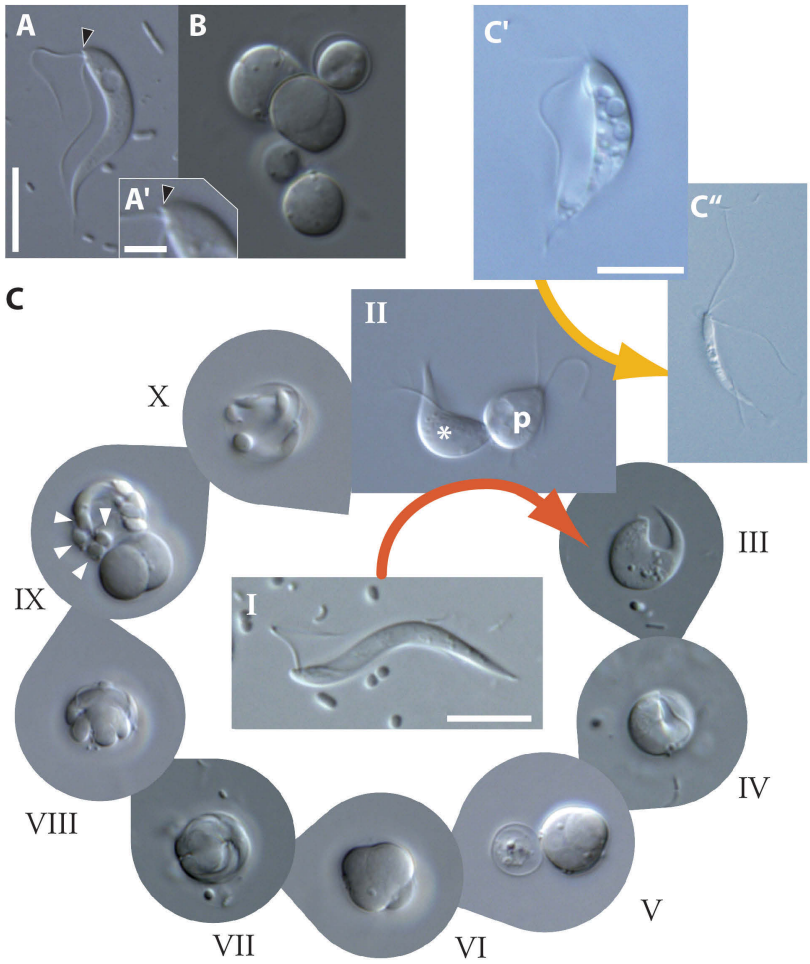


Figure 2.4: Light micrographs and putative life cycle of colpodellid TZLM1-NF assembled from multiple individuals. A-B) Flagellate (A) and a cluster of cysts (B) of the colpodellid NF. An apical protrusion can be seen (black arrowhead) and is consistent with a colpodellid apical complex. C) Life cycle of the colpodellid. Pharyngomonad isolated from TZLM1 as prey before (C) and after (C) predation by the colpodellid. I) General view of the flagellate, showing flagellar insertion at the anterior end and the extreme narrow morphology. II) NF (asterisk) in the process of feeding on its prey (p), which occurs at the anterior end of the predator and is consistent with colpodellid feeding. III-IV) Freshly fed cells likely initiating the encystment process. V) Digestive cyst in an earlier stage. VI) Four-celled stage of the digestive cyst. VII-VIII) New flagellates beginning to form. IX) Four flagellates are visible in cross section of the mature cyst (arrowheads), which is adjascent to a digestive cyst in an earlier stage (below). X) Mature digestive cyst with active flagellates swimming inside. Scale bars: 10 m, except A) 5 m.

sister to fornicates was further confirmed by phylogenomic analysis incorporating a different *Barthelona* isolate (Yazaki et al. 2020). Two additional strains, CuF11 and LRS2, were cultivated from hypersaline environments but not yet sequenced (fig. 2.6H-M), and are maintained in anaerobic media at 150ppt salinity. Despite the frequent appearance of barthelonids in anoxic

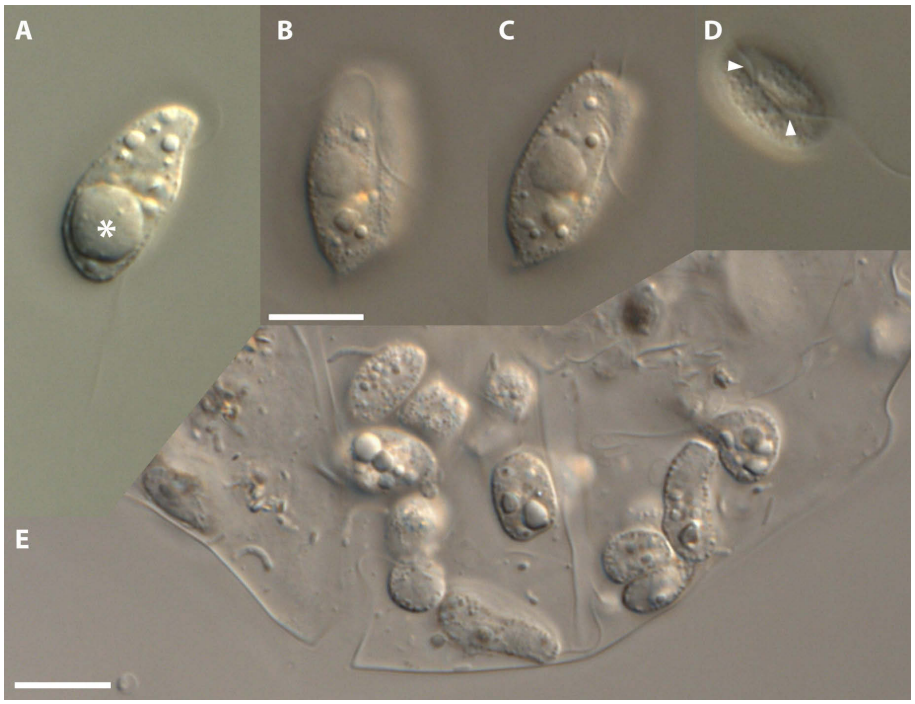


Figure 2.5: DIC light micrographs of SoapColp flagellates. A) Side view of a flagellate with posterior and anterior flagella visible. Large vacuole with prey (asterisk). B-D) View of the surface covered in fine granules, as well as anterior and posterior ventral groves, indicated by arrowheads in D. E) Flagellates inside a carcass of a rotifer in an enrichment of the sample, presumably having fed upon it. Scale bars = 10 m.

enrichments, only two sequences were detected among short read environmental sequence surveys, both from marine anoxic water and hydrothermal vent sediments and at very low read abundances (fig. 2.23).

Retortacarps

Cells have a rounded anterior end and a pointed posterior, ending in a spike up to half the cell length. There is a significant ventral groove exending from the flagellar insertion site at the far anterior of the cell, towards the base of the spike. A cytopharynx opens to the posterior end of this groove, and extends back up along the dorsal side of the cell. The right margin of the groove is broad and extremely thin, and wraps in a gentle helix around the cell towards the cytopharyngeal opening. Flagella insert subapically with a prominent ridge between them. The anterior flagellum is approximately cell-length, while the posterior flagellum is two cell lengths. A broad vane, readily visible by light microscopy, runs along the posterior flagellum for one cell length, facing outward. The flagellates nod as they swim. The nucleus is in the anterior of the

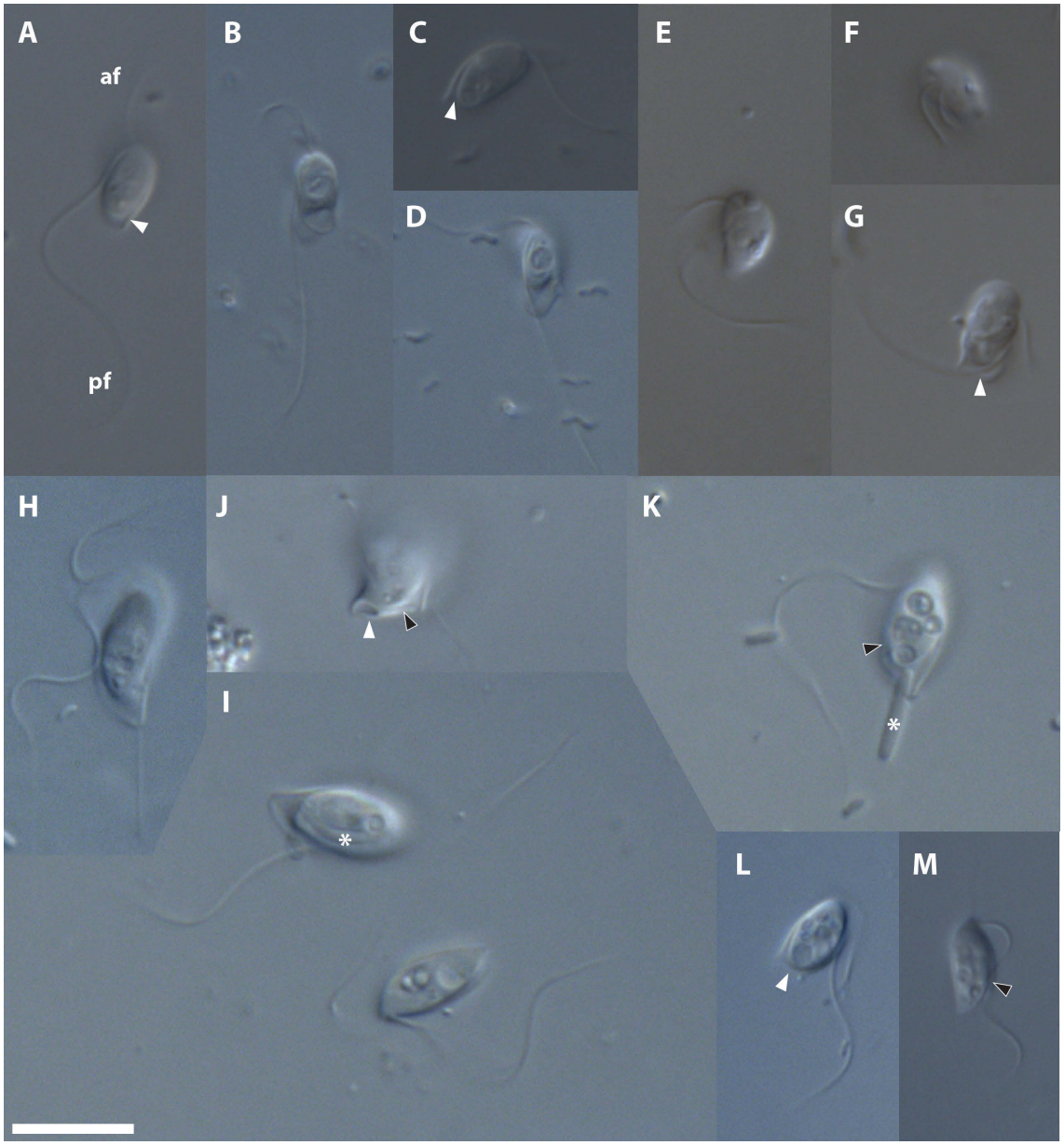


Figure 2.6: Light micrographs of barthelonids isolated in this study. A) LRM2 barth. Anterior (af) and posterior (pf) flagella insert at the anterior of the cell. The distinctive cytoskeletal element of barthelonids can be seen at the posterior (arrowhead). B-D) FB11 barth. Similar morphology to that of LRM2 barth, with the posterior structure and the dorsal cytopharynx extending upward (arrowhead). E-G) PCE barth. An exceptionally small barthelonid that nevertheless retains the key features of the group, namely the posterior structure (G; arrowhead). The flagellar insertion can be seen in (F). H-K) Halophilic barthelonid Cu11 barth. General view of the cell (H), and a pair of cells with an ingested elongated bacterium visible in one (asterisk). J) Optical section through the characteristic barthelonid posterior structure (white arrowhead) and the ventral groove leading to it (black arrowhead). K) Early feeding process, where bacteria (asterisk) are ingested through the posterior opening and up the dorsal cytopharynx opposite of the ventral groove (black arrowhead). L-M) LRS2 barth, a smaller halophilic barthelonid, with the posterior structure (white arrowhead) and ventral groove (black arrowhead) still visible. Scalebar: 10 m.

	Barthelona FB11	GEM RC	Soap18 RC	Soap20A RC	TZLM1 RC	TZLM3 RCL
TZLM3 RCL	78.49	89.23	88.94	89.79	87.72	-
TZLM1 RC	80.9	97.45	97.85	99.22	-	87.72
Soap20A RC	81.12	96.86	98.05	-	99.22	89.79
Soap18 RC	79.7	95.88	-	98.05	97.85	88.94
GEM RC	80.77	-	95.88	96.86	97.45	89.23
Barthelona FB11	-	80.77	79.7	81.12	80.9	78.49

Table 2.5: Percent identity between SSU rDNA of Retortacarps. *Barthelona* sp. isolate FB11 serves as an 'outgroup'.

cell, to the right of the flagellar insertion. There are numerous bacteria in vacuoles, presumably prey. The food vacuoles sometimes distort the shape of a well-fed cell. The flagellates vary considerably in size within one culture isolate. Cysts were observed (see fig. 2.7I,K).

Retortacarps were observed in 7 different samples over the course of this study, all from alkaline lakes; of these, 5 SSU rDNA sequences were obtained (see Table 2.5). Of note is TZLM1-RC, isolated from a hypersaline environment at 160ppt salinity and pH ~10. No corresponding environmental sequence data were found.

2.3.1.3 Obazoa

Breviates

A number of breviates were cultivated from enrichments of samples from anoxic environments (fig. 2.8). Isolate FB10N is a "classical" breviate (like *Breviata* or *Subulatomonas*; referred to here as '*Breviata-Subulatomonas* type') with a long "neck" leading up to the flagellar insertion site, and poor swimming behaviour (contrast with "*Pygsuia*-type" breviates below) (fig. 2.8A-B). Glides with a trailing posterior flagellum. In culture with accompanying prokaryotes grown in organically enriched conditions (3%LB in sterile seawater). SSU rDNA sequencing reveals it to be a lineage of *Subulatomonas*.

LRM1b (fig. 2.8E) is a large (often 15-20 μm in length) amoeboflagellate with two flagella emerging apically from the cell body, one trailing behind and often invisible as the cell glides. Produces lateral filopodia characteristic of breviates. The organism resembles *Pygsuia biforma* but is considerably larger. It swims in a smoother manner than '*Breviata-Subulatomonas* type' breviates by transforming into a narrow cigar-shaped cell with the long anterior flagellum pointing forwards. Thrives in anoxic conditions but is also fairly tolerant of more oxygenated media

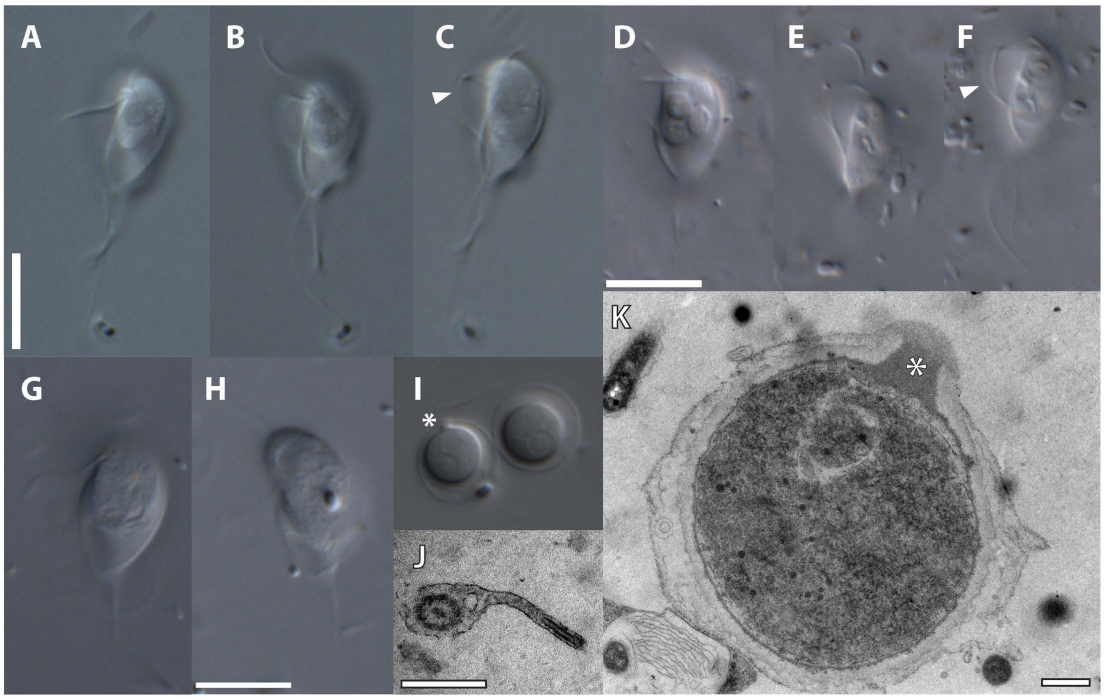


Figure 2.7: DIC light and TEM micrographs of Retortacarps TZLM1-RC, TZLM3-RCL, and GEM-RC showing overall morphology and conspicuous features. A-C) Micrographs of the flagellate TZLM1-RC. A-B) General views of the cell showing flagellar insetion and the overall shape of the feeding groove. C) Position of the posterior flagellum relative to the groove; the flagellar vane is visible (arrowhead). D-F) Flagellates of GEM-RC. Note conspicuous flagellar vane (arrowhead). G-I) Morphology of TZLM3-RCL. G-H) General views of the flagellate. I) Two cysts with visible nuclei, and a characteristic plug (asterisk). J) TEM cross-section of the posterior flagellum of TZLM1-RC showing a 0.7 m wide vane with an internal cytoskeletal structure along its distal end. K) TEM cross section through a cyst of TZLM3-RCL, showing 2 distinct wall layers, as well as a distinctive plug (asterisk). Scale bars: A-I) 10 m; J-K) 0.5 m.

(data not shown). Saa brev (fig. 2.8F) and PCE brev (fig 2.8G-H) are considerably smaller amoeboflagellates with two flagella emerging apically from the cell body, with the posterior trailing behind underneath the cell. Lateral filopodia extend from the ventral side. Both isolates have identical morphologies and SSU rDNA sequences. LRM2N6 (fig. 2.8I) is likewise small and morphologically similar, however is molecularly distinct (96.6 % identical; see fig. 2.2). Curiously, one of anaerobic rhizarians discussed below, QSI-PG, would feed on the culture of PCEbrev but not LRM2N6brev. Numerous breviate-associated environmental sequences were found primarily in marine environment datasets, particularly of sediments; there was a high number of OTUs likely associated with *Subulatomonas*, which is consistent with prior environmental data (Katz et al. 2011).

Evidence for halotolerant breviates

Organisms nearly identical by morphology to *Subulatomonas* spp. were seen several times in low oxygen enrichments of anoxic sediments from hypersaline salterns (fig. 2.8J-L). One isolate was maintained in culture for several years in 150ppt salinity medium #5 organically enriched with grains and LB. This isolate was documented by light microscopy (fig. 2.8J-L), but no sequence was obtained. The survey of environmental sequences in this chapter also detected several OTUs likely related to *Subulatomonas* spp. in hypersaline environment datasets (fig. 2.23).

CARMGS BLO

Biflagellate bacterivore with an exceptionally long posterior flagellum, up to about eight cell lengths (fig. 2.8M). Swims awkwardly with flagellum facing backwards. Attaches to substrate with anterior flagellum facing towards it (fig. 2.8N-O), and beats outward facing long posterior flagellum, presumably to generate feeding currents. A posterior channel dissects the ventral side of the cell into two unequal lobes, and a smaller groove continues to the anterior of the flagellar insetion site. Ingestion occurs at the anterior end of the channel near the insertion site. Posterior flagellum extremely "sticky" and covered in debris; there is a possibility the organism uses flagellar adhesion to bacteria in feeding, but this has been difficult to image, given the organism's sensitivity to microscopy light. Isolated from anoxic mangrove sediment, Piscadera Baai, Curaçao.

Despite little morphological similarity, SSU rDNA phylogenies firmly place CARMGS BLO among breviates or sister to them (with poor resolution between the two options). This

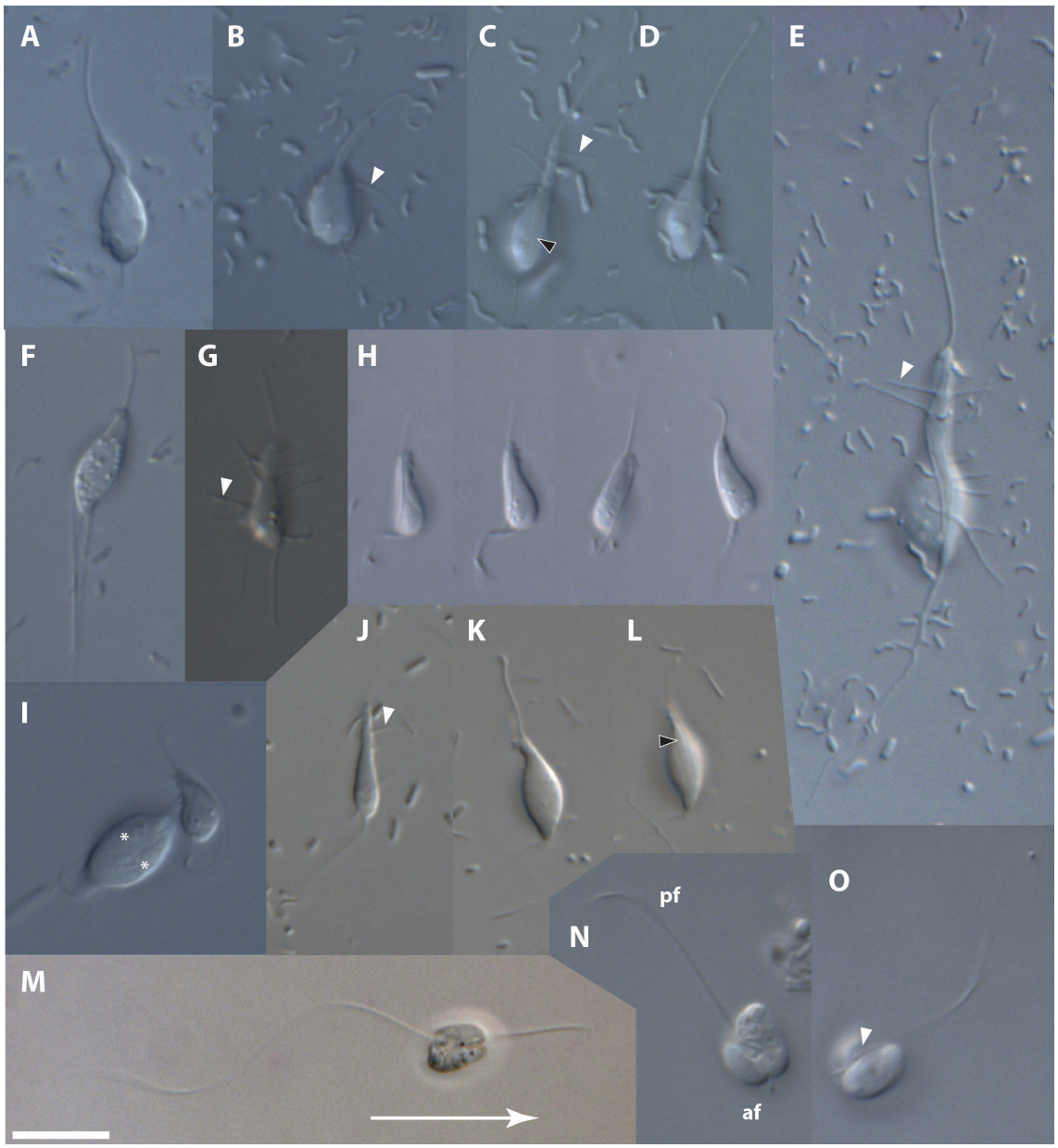


Figure 2.8: Light micrographs of breviates and the related organism BLO. White arrowheads point to lateral filopodia in all cases. A-D) *Subulatomonas* sp. isolates FB10N (A-B) and AuB brev (C-D). A thickened ventral strip can be seen from which lateral filopodia emerge (black arrowhead). Examples of "*Subulatomonas* type" breviates. E) LRM1b, an exceptionally large breviate of the "*Pygsuia* type". Note the trailing posterior flagellum extending almost the full length of the cell body. F-H) small-sized isolates of "*Pygsuia* type" breviates, Saa Brev (F) and PCE brev (G-H). F) and G) and crawling forms whereas a time series of the swimming form can be seen in (H). I) A phylogenetically distinct small "*Pygsuia* type" isolate LRM2N6 brev. A dividing cell (asterisks) can be seen next to a flagellate in growth phase. J-L) Breviates found in a 150ppt HS#5 enrichment from hypersaline CuSP2-1 sample (100ppt). Lateral filopodia (arrowhead) can be seen in (J), and the individuals also have a ventral strip (black arrowhead). K-L are of the same individual. M-O) Flagellate relative of breviates CARMGS BLO. M) Swimming cell, arrow indicates direction of swimming. N-O) Attached individuals, the posterior flagellum (pf) facing away from the substrate while the anterior (af) faces towards it. A channel (arrowhead) between two lobes of the cell can be seen in (O). Scalebar: 10 m.

sequence was confirmed to belong to the isolate by direct single cell PCR amplification partially sequenced by a single Sanger read. Additionally, collaborators found a morphologically similar cell with close sequence identity to CARMBS BLO (Ivan Čepička, unpublished data). Thus, the organism is an atypical breviate that, at least in observed life cycle stages, lacks the characteristic lateral filopodia and gliding motility, instead attaching to surfaces and filter feeding. Anaerobic; preliminary TEM did not reveal canonical mitochondria, thus it may have MROs as do (other) breviates. Numerous short read environmental sequences related to this group were detected in marine anoxic and hydrothermal vent sediments, as well as the water column.

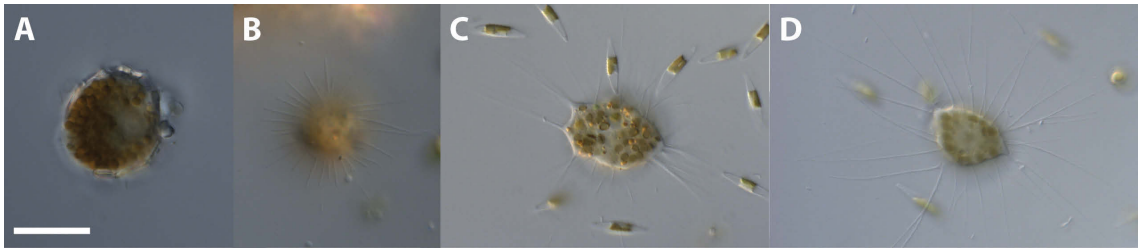


Figure 2.9: Light micrographs of *Lithocolla globosa*. A) Agglutinated form of the amoeba. B) Filopodia mid-retraction. C-D) Views of the non-agglutinated form common in culture. Scalebar: 10 m.

Lithocolla

Sphaerical filose amoeba surrounded by a conspicuous layer of agglutinated debris, maintained in stable culture fed on *Phaeodactylum tricornutum*. In culture of the isolate SnP, flattened forms more closely resembling *Nuclearia* spp. appear, but can be tranformed into sphaerical agglutinated forms by adding sterile chalk dust (Galindo et al. 2019). Isolated from sea floor sediment near Splitnose point, Nova Scotia. The placement of *Lithocolla* within nucleariids was established by molecular phylogenies using this isolate (Galindo et al. 2019). Short read environmental sequences were detected in datasets from marine water and biofilm, as well as hypersaline samples (see fig. 2.23).

MayoP

Intracellular parasite of *Paramoeba* sp. isolate (host identity confirmed by partial SSU rDNA sequence, possibly *P. eilhardi*), with distinctive late infection stages consisting of rosettes of cells, each of which then subdivides into many dozens of non-flagellated spores (fig. 2.10). Following spore dispersal, the parasite leaves behind an empty husk of the amoeba with jug-like openings

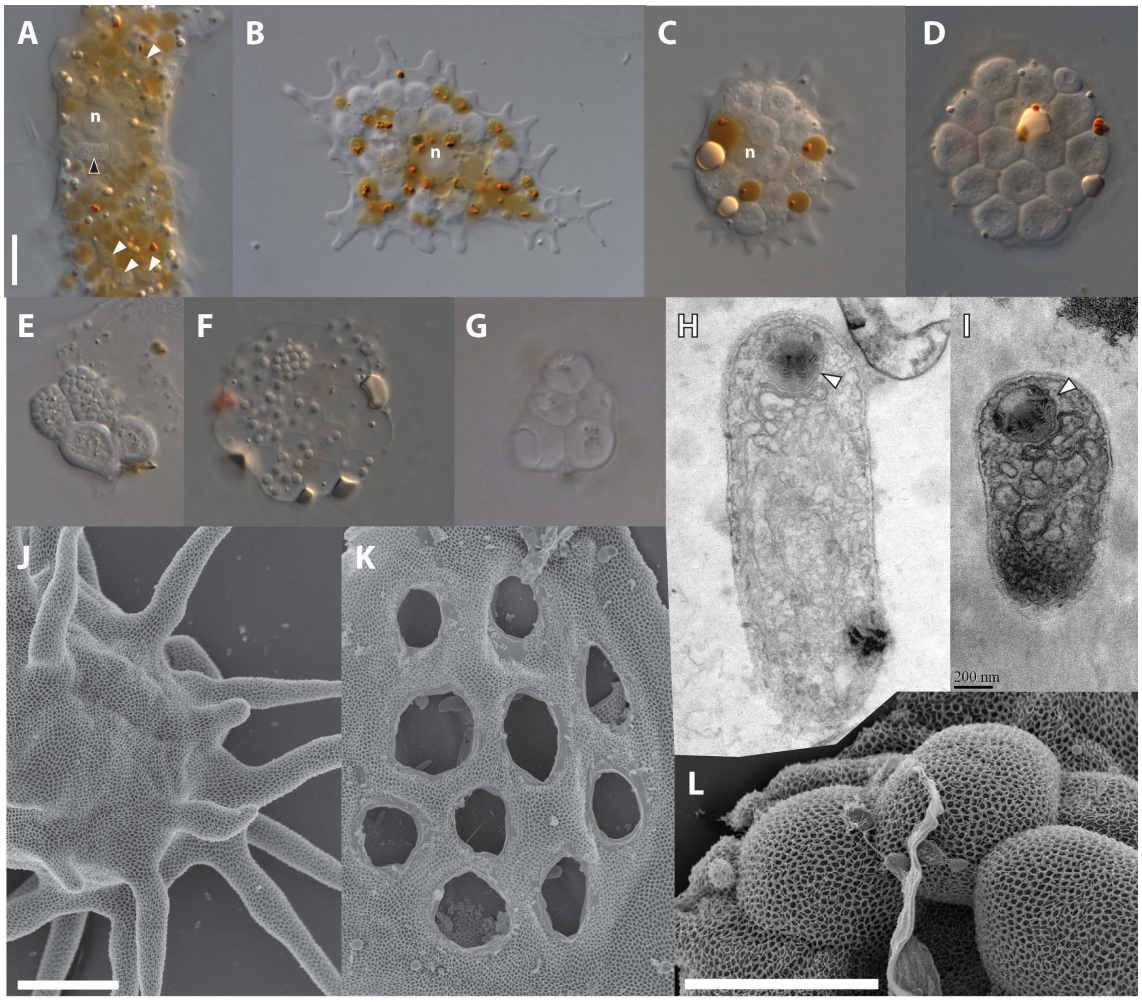


Figure 2.10: Light, TEM and SEM images of unidentified parasite MayoP in *Paramoeba* sp.: appearance of infected cells and detail of spores. A-G) DIC light micrographs of a sequence of infectious stages of the parasite across different individuals. A) Earliest detected infection stage, with numerous small cells inside the amoeba. Brown pigment originates from diatom prey of the amoeba. B) Early-mid infectious stage with larger parasite cells, each containing a clear space in the middle. The host amoeba is still capable of motility. C) Middle infectious stage with the host amoeba stationary and filled with characteristic clusters. Parasite-associated lipid bodies start to form (copper-coloured refractile globules). D) Late-mid infectious stage with the amoeba entirely filled with parasite cells and the host nucleus no longer visible. E) Late stage featuring the beginning of the parasite's cellularisation. F) Nearly empty "rosette" with remnant spores and lipid bodies. G) Characteristic jug-like openings left behind following sporulation. n – host nucleus; white arrowheads – parasite cells; black arrowheads – parasome. H-I) (preliminary, unstained) TEM cross-sections through MayoP spores. Note complex structure at the anterior end (white arrowhead), presumably involved in invasion. J-L) SEM images of the host and parasite. J) Scale-covered surface of an uninfected host *Paramoeba* sp. K) Detail of openings left after infection and sporulation. L) Detail of the host membrane and scales covering the parasite cells, most likely in the mid-late stage. Scale bars: A-G) 10 m; H-I) 200nm; J-K) 10 m; L) 5 m.

(fig. 2.10G,K) see figure legend for further details). Each spore contains a complex structure at the anterior end visible by TEM (fig. 2.10H,I), likely involved in invasion. Isolated from

a subtidal beach sample from Larnaka, Cyprus. Although no sequence data are yet available, MayoP is most likely a cryptomycete, and resembles Rozellids like *Nucleophaga* spp. (Lavier 1935, Corsaro et al. 2014, Corsaro et al. 2016). Unlike *Nucleophaga*, MayoP appears to exclusively infect the host cytoplasm (see fig. 2.10A-C). A cytoplasmic rozellomycete parasite *Morellospora saccamoebae* also has a cluster-like morphology but the spores contain anchoring disks and spiny forms are known (Corsaro et al. 2020). The infectious cell morphology of MayoP under DIC most closely resembles that of *Nucleophaga amoebae* (Gordetskaya et al. 2019).

2.3.1.4 Rhizaria

SSF (*"Cercomonas" granulatus*)

Squirmy, teardrop-shaped cell with a thick pellicle and a conspicuous row of refractile bodies anterior of the flagellar insertion point, which is roughly a third of the cell length from the tip (fig. 2.11A-E). Anterior flagellum extends along the anterior end of the cell. Posterior flagellum likewise extends along the cell surface towards the back of the cell and continues another cell length past it. Posterior end often full of rounded up prey cells in phagocytic vacuoles. Sensitive to oxygen. Roughly triangular nucleus located adjacent to the flagellar insertion site. Size can vary considerably, presumably depending on how well-fed the cell is.

Isolated from subtidal low oxygen sediments from White Rock, BC (Saa) and Cavendish, PEI (PCE), and maintained on breviates Saabrev (Saa) and PCEbrev (PCE brev) as prey, only the latter culture presently extant. Preliminary TEM suggests reduction of mitochondria (fig. 2.11F). SSU rDNA phylogenies place it in Novel Clade 12 among endomyxeans. Originally described as "*Cercomonas granulatus*" (Lee and Patterson 2000), but does not belong with cercomonads in SSU rDNA phylogenies (fig. 2.2) and has a different morphology. No likely environmental sequence matches were found (fig. 2.23).

Pyriform glider

Spindle-shaped cell with one extended "tail" end; two conspicuously long flagella emerge from the opposite, rounded end. Moves in a jerky motion as it glides passively on the long posterior flagellum (fig. 2.11G-I). Cell body filled with conspicuous phagocytic vacuoles. Eukaryotroph, maintained in stable dieukaryotic culture on breviates: TBB1-PG was grown on the large breviate LRM1b as prey (fig. 2.11L; QSI-PG is grown on the much smaller breviate PCEbrev. Isolated from anoxic intertertidal sediments from Boundary Bay (TBB1) and Quadra Island (QSI-PG),

BC, Canada. Preliminary TEM data suggest non-canonical mitochondria (not shown) and extrusomes vial- or flask- shaped (fig. 2.11K), superficially resembling those of heliomorphids *Dimorpha* (=*Heliomorpha*; see below) and *Tetradimorpha* (brugerolleCellCharactersTwo1984a). No environmental sequence matches were found.

Heliomorpha

Eukaryotrophic flagellate with radiating axopodia and two unequal flagella. The axopodia fold back and retract when the flagellate swims, longer flagellum pointing forwards. As the previously attested '*Dimorpha*-like' sequence branching among then-proteomyxids (Cavalier-Smith and EE.- Chao 2003) was almost certainly misidentified, heliomorphids (=dimorphids; *Dimorpha mutans* was moved to *Heliomorpha* due to homonymy, Bass et al. 2009) were among the last of distinct heliozoan groups needing molecular phylogenetic placement. Single cell isolate DM2 was picked by Gordon Lax from a freshwater sediment, and isolate SDM2 was picked from an alkaline lake sample. Sequence data were obtained for the former, and in SSU rDNA phylogenies, heliomorphid DM2 falls within endomyxeans, forming a clade with environmental sequences MPE2-26 (AB695520.1), from aquatic moss pillars in an antarctic freshwater lake, and PR_3E_90 (GQ330591.1) with 72% bootstrap support; this clade, in turn, branches with Novel Clade 10 (Bass et al. 2018) (fig. B.2), sometimes Novel Clade 12 but both with mediocre support. In any case, this *Heliomorpha* isolate branches outside Filosa. A single environmental OTU match was found in short read data, from deep sea sediment, with a very low abundance (fig. 2.23).

Discocelia "HT"

Flat, approximately triangular cell, about 10 μm wide, with two flagella and a rim punctated by granules. Glides smoothly and rapidly, usually in a circular path, on the posterior flagellum as the anterior vibrates laterally (fig. 2.12H-I). Bacterivorous. BoP4.1 HT matches the description of *Discocelia saleuta* (Vørs 1988), then an incertae sedis taxon but later placed among bicosoecids (stramenopiles) based on ultrastructural characteristics (Karpov 2000), then in Apusozoa (Cavalier-Smith 2013), and in Filosa based on unpublished SSU rDNA data (Cavalier-Smith et al. 2014). In this study, the SSU rDNA phylogeny places it within Filosa with strong support. High probability candidate environmental sequence matches were found from marine sediment samples, including some around hydrothermal vents (fig. 2.23).

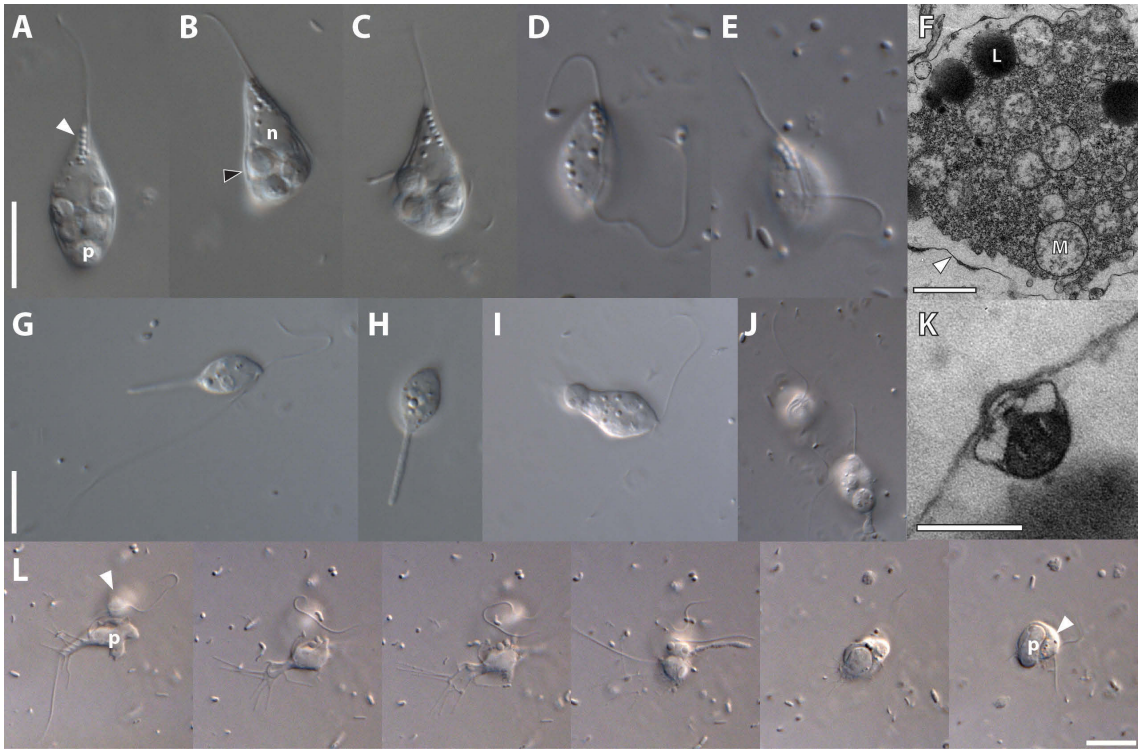


Figure 2.11: Light microscopy and TEM images of endomyxeans Seriously Squishy Flagellate (SSF) and Pyriform Glider (PG). A-C) General morphology of isolate SaaSSF. A) Optical section along the left-right axis of the cell. Distinctive row of granules extends along the portion of the cell body along the attached portion of the anterior flagellum (white arrowhead). Conspicuous phagocytic vacuoles with prey (p) fill up the bulk of the posterior region. B) Section along the dorsoventral axis of the cell showing flagellar insertion and the nucleus immediately underneath it (n). A theca is visible under light microscopy (black arrowhead). C) additional view of the same cell showing the apical end of the cell body directly adjascent to the anterior flagellum. D-E) General morphology and preliminary TEM of PCE-SSF. Anterior granules can be seen as in SaaSSF. The cell morphology is less distinct. F) TEM section through the cytoplasm of PCE-SSF. Osmiophilic bodies, presumably lipid globules (L), and double-membrane-bound organelles, likely MROs (M), can be seen inside. A layer external to the cell membrane (arrowhead) stands out as the cell body shrinks during fixation, likely the structure indicated by the black arrowhead in (B). G-J) General views of isolate QSI-PG. Flagella are long (G), and a rod-like granule-lined protrusion makes up about half of the cell's posterior (G,H). The nucleus is immediately underneath the flagellar attachmennt site (I), and the posterior flagellum extends through a channel in some cases (J). The pair of cells in (J) is recently divided. K) TEM section through an extrusome docked underneath the cell membrane in QSI-PG. L) Time series of predation behaviour of TBB1-PG on the prey breviate LRM1b. The predator (arrowhead) touches the prey (p), after which the prey rapidly disintegrates, and the predator then phagocytoses the breviate. Scale bars: A-E) 10 m; F) 1 m; G-I) 10 m; K) 200nm; L) 10 m.

Abollifer spp.

BoP4.1 RotF

A vaguely shield-shaped cell with a wider anterior end and pointier posterior, a pair of flagella emerging in parallel from inside the opening of an anterior "hood" and an "apron" of granules just underneath it. The "apron" has a pronounced lower boundary visible in DIC (fig. 2.13). Anterior flagellum is shorter and points forward as the cell moves, whereas the posterior flagellum trails behind the cell for an additional half cell body length and can be coiled when the cell travels in a circular motion repeatedly. BoP4.1 RotF is a eukaryotroph, observed to feed on diatoms and is maintained in dieukaryotic culture with a kinetoplastid prey (isolate LAR KIN).

This morphotype was previously misidentified as *Heterochromonas opaca* by Lee and Patterson 2000, and the original *Heterochromonas* is most likely a different taxon entirely (also see comment by Prokina et al. 2021). SSU phylogenies place BoP4.1 RotF with *Abollifer globosa* Shiratori et al. 2014 with full support; though *Abollifer globosa* has a less pronounced "apron", its boundary is still discernable through light microscopy. Differs from *Abollifer prolabens* Vors 1992 (Vørs 1992, Irwin et al. 2019) in that the latter was reported to glide with the cell body perpendular to the substrate. Probably environmental matches occur moderately frequently in examined environmental sequence data, primarily from marine environments ranging from sediment to biofilm to water column (fig. 2.23).

BWH RotFL

Smaller but otherwise superficially similar to BoP4.1 RotF above, albeit with a less pronounced "hood", "apron", and a more plastic morphology overall. Grown in stable dieukaryotic culture with the diatom *Phaeodactylum* as prey, in the dark. Forms a well-supported clade with environmental sequence TS698-42 (HM369739.1) from Boothbay Harbour, Maine (Heywood et al. 2011), and NAMAKO-14 (AB252754.1). In this study's survey of public environmental sequence data, this organism matched the highest number of OTUs, primarily from oxic marine environments and a few OTUs in brackish water.

BoP4.1 Q

A dish- or hat-shaped cell, 12-20 μm in diameter, with a pair of unequal flagella emerging in parallel from a channel beginning behind the nucleus and opening in the centre of the underside of the disk. Glides on a longer trailing flagellum while the shorter flagellum vibrates in a curved

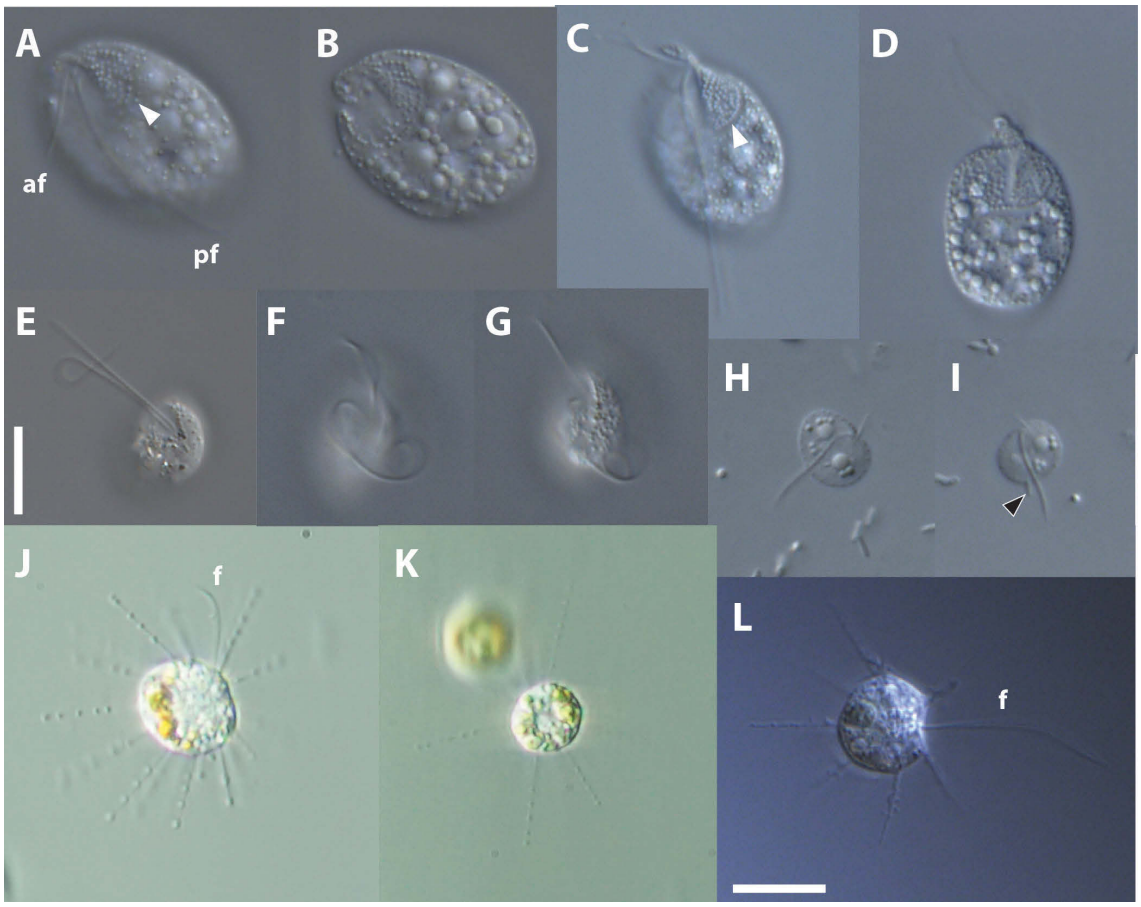


Figure 2.12: Light microscopy of *Heliomorpha* single cells, *Discocelia*, and filosan isolates related to *Abollifer* spp. A-D) *Abollifer* isolate BoP4.1 RotF (previously misidentified as *Heterochromonas opaca*). A-B) Optical sections through a single individual showing anterior (af) and posterior (pf) flagella emerginf from a 'hood', and a clear collar (arrowhead) at the edge of an 'apron' of granules (better visible in B). C) Another individual showing the postion of flagella as well as the clear collar (arrowhead). D) A deeper optical section through another cell showing a canal-like clearing posterior to the flagellar insertion site. E-G) Isolate BWH RotFL, related to *Abollifer*. E) Individual gliding in reverse, with straightened flagella pointing opposite to the direction of motility. F-G) Optical sections through the same individual showing a typical flagellar placement with the end of its posterior flagellum coiled. H-I) *Discocelia* sp. The cell glides on its posterior flagellum as the anterior flagellum vibrates. Small granules can be seen along the rim of the cell. I) An individual cell with a discharged granule (black arrowhead). J-L) *Heliomorpha* single cell isolates picked for SSU rDNA sequencing. DM1 (J) and DM2 (K) were isolated from a freshwater sample by Gordon Lax, while SDM2 was isolated from an alkaline lake. Scale bars: 10 m. Images J and K courtesy of Gordon Lax.

bow to the side underneath the disk. Edge of disk lined with what appear as ridges in DIC (fig. 2.13B), appearing as extrusomes in TEM (fig. 2.13J-L). Eukaryotroph that hunts prey by positioning itself above it and rapidly snapping down to trap the prey like a "lid" (fig. 2.13), then extruding pseudopods from the underside to engulf and phagocytose the prey. Isolated from beach sediment. SSU rDNA phylogenies place it within Filosa, specifically near Cryomonadida.

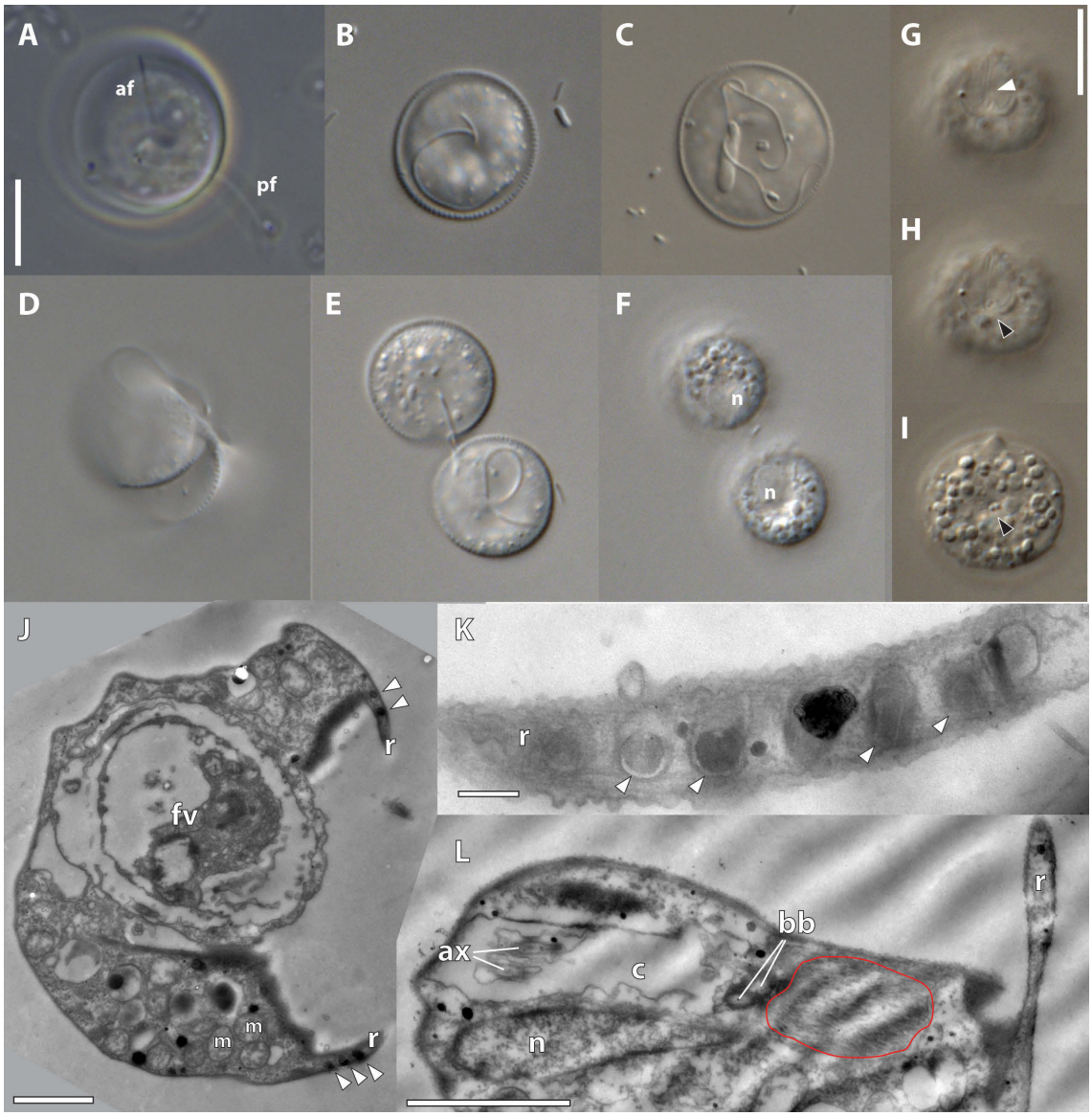


Figure 2.13: Light microscopy and preliminary TEM of filosan flagellate BoP4.1 Q. A) A phase contrast image showing the typical orientation of flagella of a moving cell, the anterior flagellum pointing up in the image. B) A stationary cell clamped down to the surface. Note the flagella contained underneath, and fine ridges along the thick rim of the cell. C) A stationary cell with caught kinetoplastid prey, which was still motile but unable to escape. D) A cell in early division. E-F) Optical sections through a cells at the end of cytokinesis. Nuclei (n) are visible in F. G-I) Optical sections through the top of the cell with a channel (arrowhead) visible through which the flagella pass in parallel to descent down the central channel in the cell (black arrowhead). J-L) Preliminary ultrastructure by TEM. J) Tangential section through a cell, dorsal side to the left and the ventral disk opening to the right, cutting through a portion of the outer rim (r). A prey kinetoplastid can be seen in the food vacuole (fv). Mitochondria (m) have tubular cristae. Extrusomes can be seen along the rim (arrowheads). K) Close-up of the rim (r). Arrowheads point to individual extrusomes in varying longitudinal sections. L) Section through the top of the cell. A pair of flagella initiate from the basal bodies (bb) near the nucleus (n), pass upwards before emerging back down a channel (c) in parallel (see: pair of axonemes (ax)). A fibrous region (circled in red) exists on the other side of the basal bodies from the channel. Scale bars: A-I) 10 m; J) 2 m; K) 200nm; L) 2 m.

This group contains other eukaryotrophs, flagellates and amoebae with a thick theca and a tendency to form filose pseudopods. A single extremely low abundance environmental sequence hit was found in marine hydrothermal vent sediment.

2.3.1.5 Amoebozoa

Multicilia

(Zlatogursky et al., in prep)

Large cell with a variable number of radiating flagella (from half a dozen to over 20); the organism takes turns gliding on different flagella pointing forward in the direction of travel. Predator of a *Vannella* sp., which it consumes by positioning itself above it and extending a pseudopod to collect it. Morphology identical to that of *Multicilia* Cienkowski 1881 (Mikrjukov and Mylnikov 1998, Cienkowski 1881). This morphotype was previously cultivated but the culture has since been lost (Mikrjukov and Mylnikov 1998). A single cell transcriptome was obtained from isolate HAK MC from a mud flat in BC (Canada) by Gordon Lax. In the meantime, a culture established independently from an isolate from another sediment sample, LIS MC. The SSU rRNA sequence derived from the HAK MC transcriptome had 96.5% identity to *Multicilia marina*, and 90.5% to *Artodiscus*.

Organisms with nearly identical morphology to that of *Multicilia* were isolated from two independent hypersaline samples (PSL3 and JT4U, both >100 ppt salinity). The prey is presumed to be small amoebae but di-eukaryotic cultures have not yet been established. Molecular data have not yet been obtained, but the organisms most likely belong to the *Multicilia-Artodiscus* clade. These may be the first observations of a multiciliid in a hypersaline environment to date, as there is no prior mention of hypersaline sightings (Mikrjukov and Mylnikov 1998, Nikolaev et al. 2006, Ntakou et al. 2019). No significant matches were found in short read environmental data.

Trichosphaerium

Isolated from subtidal coastal beach sediment in southern Nova Scotia, amoeboid cells from ~50 μm up to well over a hundred microns long, covered in a pellicle with a layer of conspicuous carbonate crystals 8-10 μm long, with multiple circular openings through which narrow tubular pseudopodia emerge (fig. 2.14). Feed on eukaryotes and bacteria alike, and can be maintained in culture on bacteria alone. Early in the cultivation process, one flask of culture contained

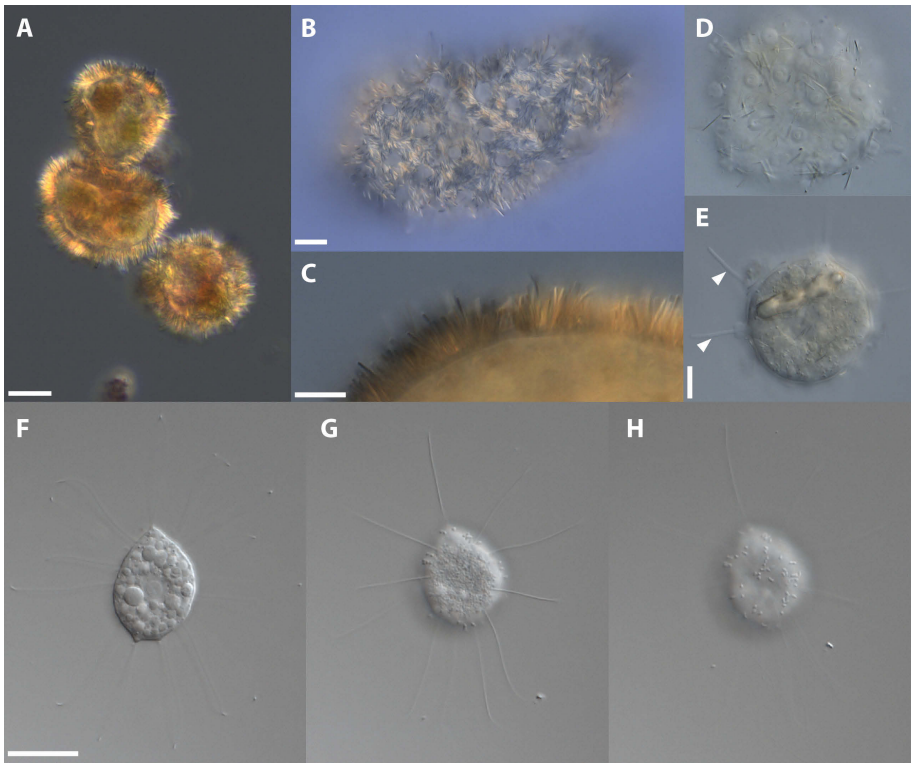


Figure 2.14: DIC light micrographs of amoebozoans *Trichosphaerium* sp., and *Multicilia* isolate HAK-MF. A-E) *Trichosphaerium* sp. A) Low magnification view of several amoebae showing distinctive colour from the calcium carbonate spicules. B) Surface of an amoeba with regular-sized openings among the spicules. C) Optical section through the layer of spicules covering the cell. D-E) Optical sections through a rare de-spiculated individual showing surface apertures from which dactylopodia (E, arrowheads) emerge. F-H) Optical sections through an individual of *Multicilia* isolate HAK MF. Flagella radiatie from all around the cell, and surface granules can be seen in (H). Scale bars: 10 m except A) 50 m; and F-H) 20 m.

only "naked" forms of the amoeba was observed once. Both forms were originally described by (Schaudinn 1899), but the current *Trichosphaerium* in molecular datasets, strain ATCC 40318, was in the less distinctive "naked" form (Tekle et al. 2008), making its identification less certain at the time. Molecular confirmation of identity done by collaborators.

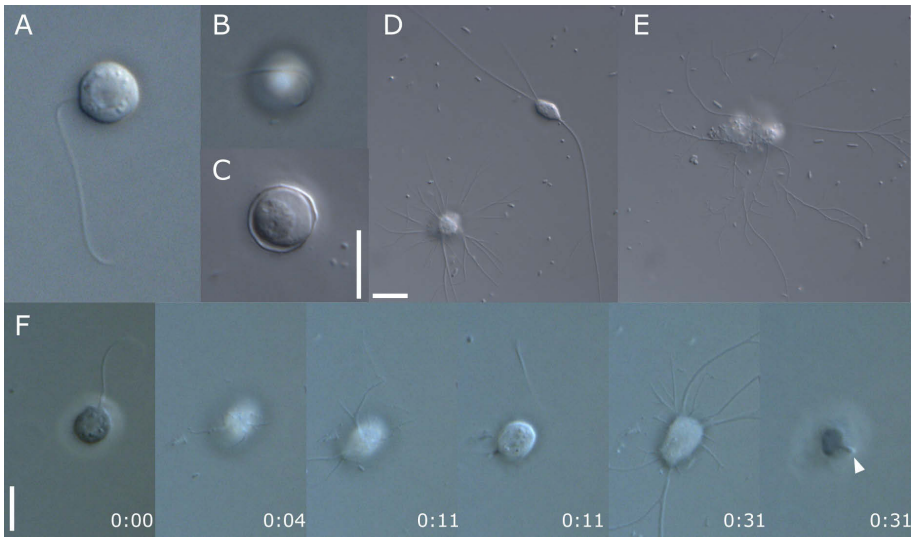


Figure 2.15: DIC light micrographs of amoebozoan with a flagellate stage LRM3-1F A-B) Views of a swarmer cell with the flagellum extended (A) and held in acharacteristic manner wrapping halfway around the cell (B). C) Cyst. D-E) Morphological variation of the amoeboid stage. The arrangement of filopodia can be spider-like (D, bottom) or as very long filaments extending from one or two ends of the cell (D, top), and several amoebae can form clusters (E). F) Time series of a transforming swarmer flagellate, with time indicated in minutes. The total process from attachment to amoeba takes about a half hour. Shortly after transformation, a small protrusion remains where the flagellum was (arrowhead). Scalebar: 10 m.

LRM3-1F

Flat amoeboid cell with a round main body (~10 μm long) with very long branching filopodia radiating from around the cell body, some for considerable distances (fig. 2.15). Has a swarmer stage: a small sphaerical cell (7-8 μm in diameter) with a single emergent flagellum that gently curves backwards around the cell. A swarmer cell was observed to transform fully into the amoeboid form in 30min (fig. 2.15F) Top BLASTn hit within NCBI nt is a *Schizoplasmidium* sp. (accession ID: FJ544418.1, 77% identical); the sequence from LRM3-1F is well-supported within Amoebozoa in SSU rDNA phylogenies but placement within the group remains unresolved. No short-read environmental placements were found, although only the V9 region was present in the current SSU rDNA sequence.

2.3.1.6 Heterolobosea

Dactylomonas/Selenaion "YAF"

Elongate cell with two long, unequal flagella emerging subapically.There is a short groove wrapping around the far anterior end of the cell (fig. 2.16). The posterior flagellum bends backwards around the cell, and has a characteristic small lateral finger-like projection not far from its base. A vesicular nucleus is located a third of a cell length from the anterior end of the cell. Bacterivorous heterotroph, ingestion occurs at the anterior end of the cell (fig. 2.16). Isolated from and maintained in low oxygen haloalkaline (160ppt) conditions enriched with grains and iron citrate. Isolated from salt crust from alkaline Lake Manyara, Tanzania. Morphologically resembles the freshwater heterolobosean flagellate *Dactylomonas* (Hanousková et al. 2019), although branches with its sister lineage *Selenaion* in SSU rDNA phylogeny (fig. 2.16). While flagellates were not recorded in the original description of *Selenaion* (Park et al. 2012), it may have a flagellated stage that resembles that of *Dactylomonas* and YAF. No environmental sequence matches were found.

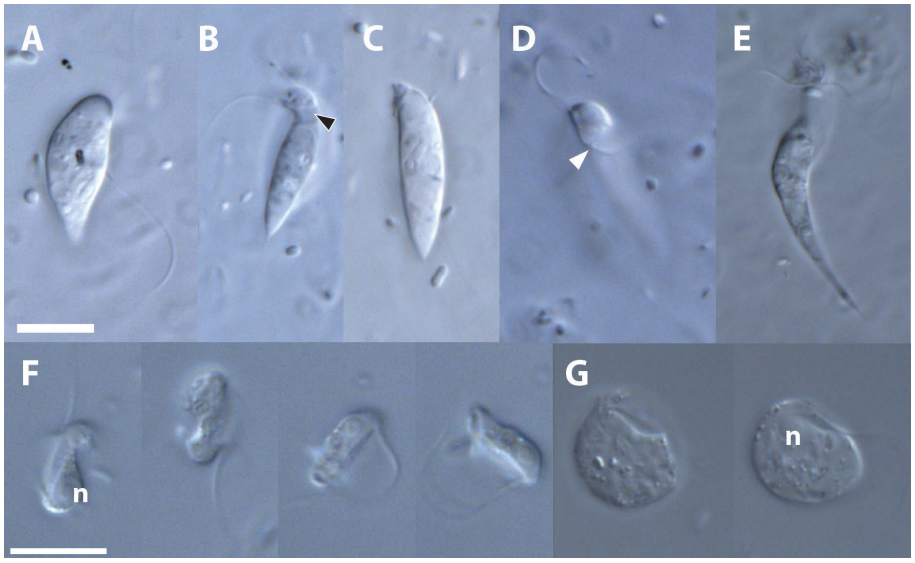


Figure 2.16: Light microscopy of heterolobosean flagellates TZLM1-YAF and TZLM3-JC. A-E) Heterolobosean TZLM1-YAF. General views of the cell. A groove can be seen curving around the anterior end of the cell (black arrowhead). Part way down the posterior flagellum is a small lateral protrusion (white arrowhead). E) Dividing cell with two sets of flagella visible at the anterior end. F-G) Optical sections through two individuals of Heterolobosean TZLM3 JC ("Jakobacarp"). Both flagellates, fed to different degrees, carry an excavate-like groove through which the posterior flagellum beats. A vesicular nucleus is visible (n). Scale bars: 10 m.

Jakobacarp

A biflagellate bacterivore of spindle-to-rounded overall shape that varies considerably within a culture isolate, and with a distinct groove through which the posterior flagellum beats (fig. 2.16F-G). Usually swims with heterodynamically beating flagella. Isolated from a haloalkaline lake (TZLM3) and established in a monoeukaryotic culture with co-isolated bacteria. The SSU rDNA phylogeny places it at the base of Tetramitiidae, and it does share the SSU rDNA Helix 17_1 Tetramitiid synapomorphy (Harding et al. 2013) (see fig. B.3). The overall shape resembles that of a "typical excavate" such as *Jakoba* or CLOs. One potential SSU rDNA V9 environmental sequence was found, from deep sea sediment.

"Ridgy Flagellate" RF

Fairly large cell with conspicuous longitudinal ridges twisting slightly helically along the cell and ending in a spike-like projection, particularly along the dorsal side of the cell. Higher magnification DIC images of flattened cells show a segmented structure immediately underneath each ridge (fig. 2.17) – possibly localised alveolae or bacterial endosymbionts. Ventral side contains a groove resembling that of a pharyngomonad. Four flagella emerge apically, with one pair projecting forwards with a gentle curve and beating in parallel, as the other two point backwards and to opposite sides. Prominent vesicular nucleus located anteriorly close to the emergence of flagella. Cytoplasm full of round vesicles towards the posterior of the cell. In some conditions, flagellates exhibit a peculiar swarming behaviour (fig. 2.17F).

RF contains amoebae that resemble those of pharyngomonads and their close relatives (Harding et al. 2013). These amoebae contain abundant bacteria in the cytoplasm – possibly endosymbionts. Additionally, somewhat irregularly-shaped cysts can be found, smooth and devoid of pore plugs. Isolated from sediment from alkaline Soap Lake, WA, US and grown in 20ppt alkaline medium TZ enriched with grain. No sequence data available to date, but likely a heterolobosean due to life cycle stages, nuclear morphology under light microscopy, arrangement of flagella, and resemblance of the groove to that of pharyngomonads.

2.3.1.7 Residual lineages

Centrohelid "CuFi-MH"

(Zlatogursky, Eglit, Simpson, in prep)

A small heliozoan (fig. 2.18) isolated from rehydrated salt crust from a pool with 120ppt salinity and cultivated on a halotolerant choanoflagellate (isolate 'C3C') in 150ppt HS#5 medium.

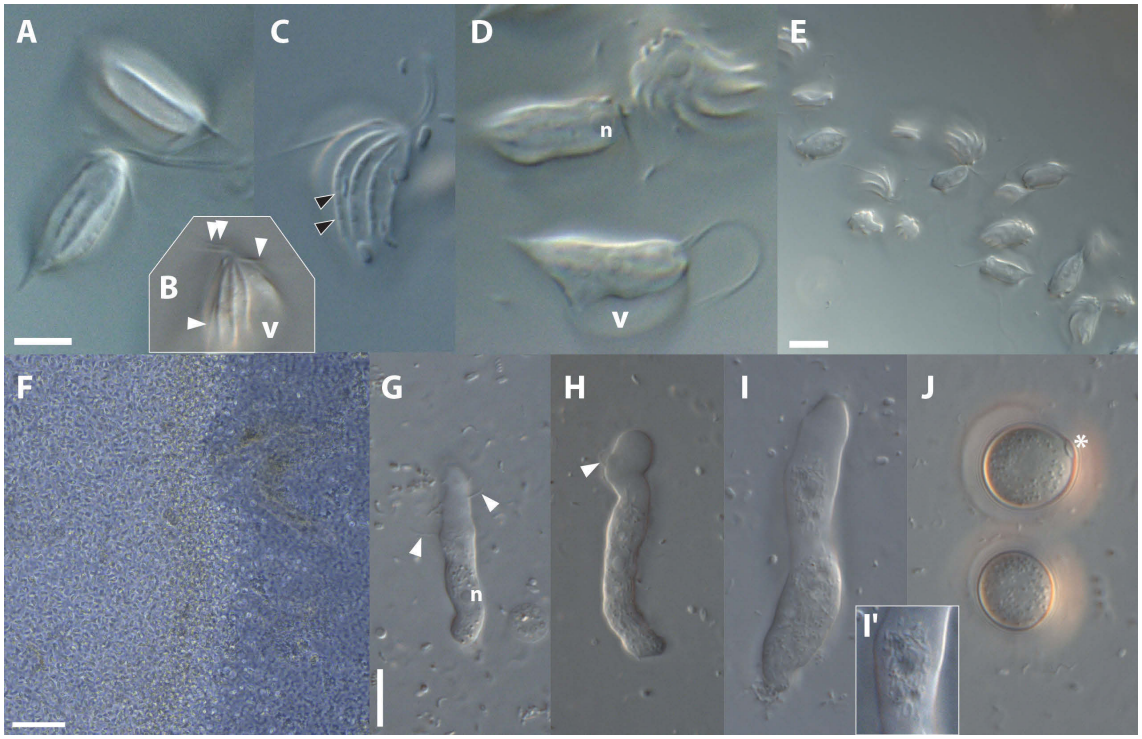


Figure 2.17: Light micrographs of putative heterolobosean "ridgy flagellate" (RF). A-F) Flagellates of RF. A) The surface of the flagellates shows distinctive ridges. B) Four flagella, two in parallel and each of the remaining two opposite of each other, emerge from the apical end of the cell (white arrowheads). The parallel flagella extend behind the dorsal end of the cell, away from the ventral depression (v) C) Septae subdiving the ridges into smaller compartments (black arrowheads). D) A view of three flagellates showing sections through nuclei (n) and a ventral depression (v) (likely homologous to the pharyngomonad groove). E) Edge of the "swarm" that forms on a wet mount slide, with the flagellates arranged in a crowded line. F) Low magnification phase contrast image of the edge of a "swarm", approaching from the left. H-I) RF amoebae. The amoebae are elongate with a hyaline region in the anterior that erupts outward, and thin lateral filopodia (arrowheads) that remain anchored to the surface as the cell moves. A single nucleus is visible (n). The cytoplasm is filled with rod-shaped inclusions, likely bacterial endosymbionts (I'). J) A pair of cysts, containing a plugged opening (asterisk). Scale bars: 10 m except F) 100 m.

Can ingest prey roughly its own size (fig. 2.18B). SSU rDNA phylogenies place CuF1-MH firmly among centrohelids (fig. 2.2), and the organism represents a previously exclusively environmental clade (V. Zlatogursky; data not shown). SEM imaging reveals an absense of scales and an altogether naked surface (fig. 2.18C), which is unusual but not unknown among centrohelids: eg. *Oxnerella* (Cavalier-Smith and EE Chao 2012), which branches in a different part of the centrohelid phylogeny (V. Zlatogursky; data not shown). There are short read environmental data from marine sediment, water, and near hydrothermal vents that meet this study's criteria for high probability matches (fig. 2.23), but those sequences may also represent other closely related undersampled or uncharacterised centrohelid clades.

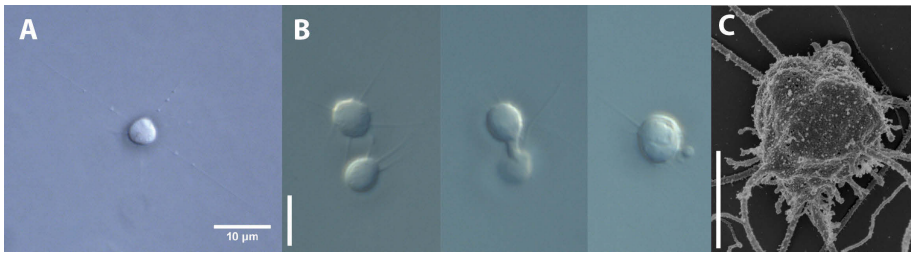


Figure 2.18: Light microscopy and surface features of halophilic centrohelid CuF1-MH. A) General view of the centrohelid. B) Time series of CuF1-MH (top) feeding on a chanoflagellate (bottom). C) SEM image showing the cell surface devoid of spicules. Scale bars: 10 m, except 2 m in C.

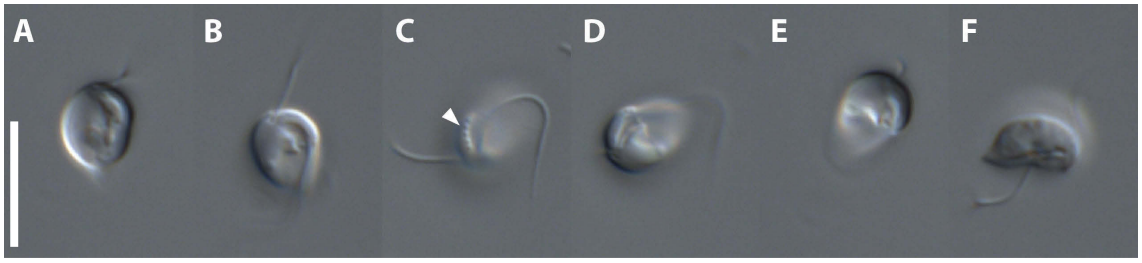


Figure 2.19: DIC light micrographs of ancyromonad CuF1-Anc. A-F) Different DIC optical sections through a single specimen of ancyromonad CuF1-Anc c.f. *Ancyromonas melba*. Note the unusually long anterior flagellum for this group (B,C). The left side of the ventral depression contains a row of granules, presumably extrusomes (C, arrowhead). Scalebar: 10 m.

Ancyromonas melba

Bean-shaped dorso-ventrally flattened cell with a conspicuous ventral depression through which the posterior flagellum passes; anterior flagellum short but longer than in other ancyromonads (fig. 2.19). Prominent battery of granules – presumably extrusomes – lines the anterior left side

of the ventral groove. Strongly resembles the uncultivated species *Ancyromonas melba* (Patterson and Simpson 1996) by morphology. No sequence data yet available. Isolated from rehydrated salt crust from a salt pond (120ppt salinity) in Curaçao, and currently in crude culture in HS#5 medium at 150ppt salinity. Most likely bacterivorous.

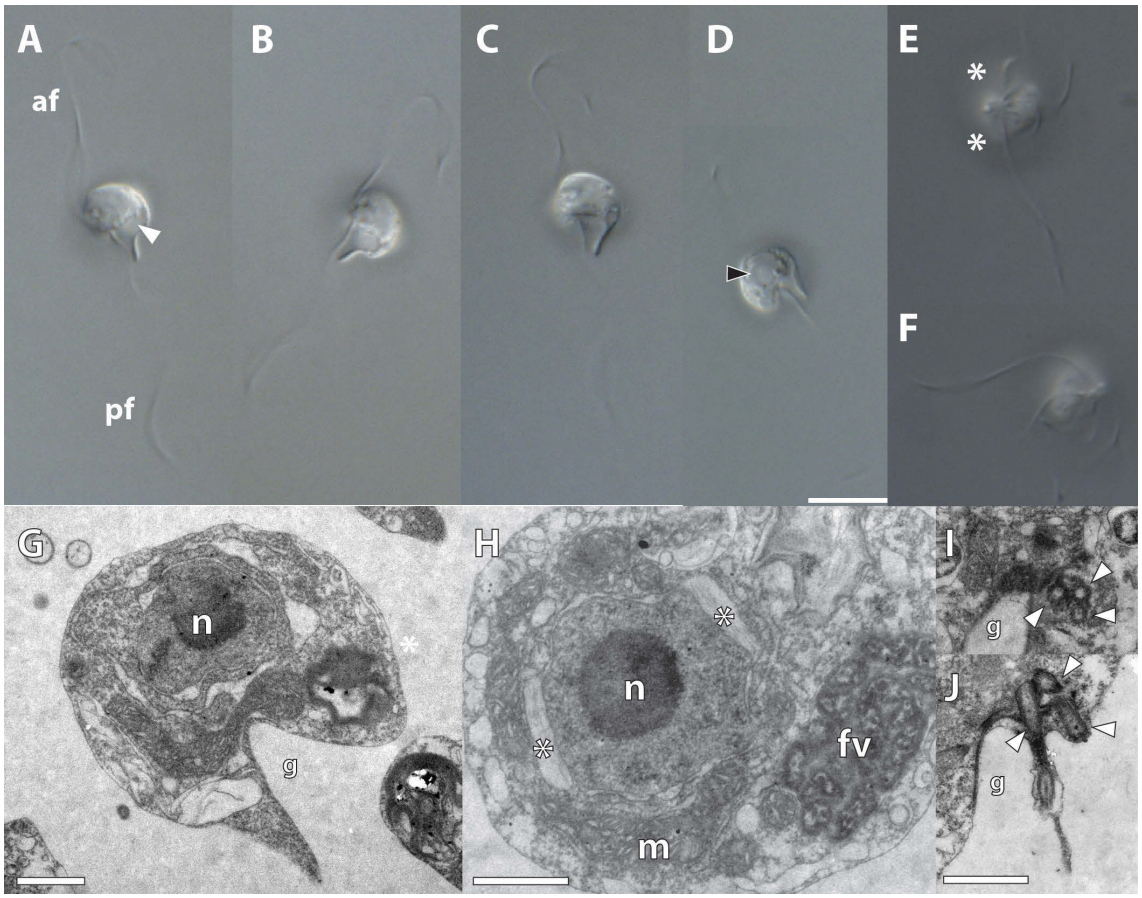


Figure 2.20: Light microscopy and preliminary ultrastructure of palpitomonad CuSP2-3 BLOL. A-C) Optical sections through an individual CuSP2-3 BLOL flagellate showing the long anterior (af) and posterior (pf) flagella, and a wide groove (arrowhead). D) optical section through a cell with the nucleus (n) visible. E-F) Optical sections through a dividing individual, each set of flagella indicated with an asterisk. G-J) Preliminary ultrastructure by TEM. The wide groove can be seen opening to the bottom right (g). H) The nucleus (n) is encircled by mitochondrial lobes (m), and vesicles with fibrillar inclusions (asterisks) where mastigoneme hairs are likely produced can be seen associated with the endoplasmic reticulum. A food vacuole (fv) is to the right. I-J) Sections through two individuals, with a total of three basal bodies (arrowheads). Where appropriate sections where available, all observed cells had more than two basal bodies. Further serial sectioning will likely reveal a fourth. Scale bars: A-F) 10 m; G-J) 1 m.

Palpitomonad "BLOL"

Biflagellate with a long (approximately 4 cell lengths) trailing posterior flagellum extending past a prominent posterior groove. The anterior flagellum is about 3 cell lengths long and held laterally as the cell swims gracefully on a helical path (fig. 2.20). Isolated from sulfidic muds (70ppt salinity) on Curaçao. Initially grown under anaerobic conditions with *Carpediemonas* BICM as prey, which it was observed to ingest at the posterior end of the cell, but currently maintained in aerobic media with only bacteria as prey. Thus, BLOL appears to be both facultatively eukaryotrophic and also capable of tolerating anaerobic conditions.

The SSU rDNA sequence places it with *Palpitomonas bilix* (Yabuki et al. 2010), with 98.7% sequence identity over 1178 nucleotides. The published morphological description for *P. bilix* does not mention a groove, but mentions the cell is delicate under imaging conditions. BLOL is likewise sensitive to light exposure and the groove disintegrates rapidly; it cannot be ruled out that the original *P. bilix* isolate may also sometimes carry a groove. Preliminary ultrastructure data are consistent with those of *Palpitomonas bilix*: vacuolated cytoplasm, flat cristae, and a distinctively concentric endoplasmic reticulum around the nucleus (fig. 2.20G-I). BLOL has a wide band of microtubules associated with the groove, and at least one extra pair of unflagellated basal bodies that were not observed in *P. bilix* (Yabuki et al. 2010).

A slightly smaller isolate with an otherwise identical morphology (BoP3.4 BLOL) was cultivated from another hypersaline environment, but no sequence has been obtained yet. Short read environmental sequences equally identical to both BLOL and *Palpitomonas bilix* (with 98% sequence identity) were detected in marine as well as hypersaline samples (fig. 2.23).

2.3.1.8 *Incertae sedis* and major new lineages

Hemimastigophora

(published in: Lax G*, Eglit Y*, Eme L*, Bertrand EM, Roger AJ, Simpson AGB (2018). "Hemimastigophora Is a Novel Supra-Kingdom-Level Lineage of Eukaryotes". *Nature* 564(7736):410–414)

Eukaryotrophs with two rows of flagella on the sides of the cell and a prominent anterior cap, or 'capitulum', which also serves at the site of prey ingestion. Spironemids taper off into a long "tail" at the posterior end, with which they sometimes attach to surfaces (fig. 2.21A-F). *Hemimastix*, on the other hand, is rounded and with hardly any "tail". Both were picked for single cell transcriptomes, but a culture of *Hemimastix* was in the meantime established with

Spumella sp. as prey. The culture is now available for further research, such as genome sequencing. A smaller *Hemimastix* sp., probably *H. amphikineta* based on size and fewer flagella, was seen in an enrichment inculated with material sampled by Gordon Lax near glacial runoff from Joffre Lakes PP, BC (Canada).

Meteora

(Eglit*, Shiratori* *et al.*, in prep)

Small (2-4um) cell body with long longitudinal extensions on either end upon which it glides, usually with two or more swinging "arms" extending from the cell body proper (fig. 2.21G). Morphologically identical to *Meteora sporadica* Hausmann 2002. Numerous dynamic small granules visible throughout the cell surface, some of which can extend quickly. Aerobes and bacterivores, isolated from intertidal marine sediments. *Meteora* was additionally seen in two Curaçaoan intertidal sediment samples (beach in front of CARMABI, and San Juan Baai; data not shown).

SSU rDNA phylogenies do not place *Meteora* among members of Rhizaria with a similarly dynamic cytoskeleton, nor among any of the former "Heliozoan" groups with radiating axopodial-type extensions, and, in fact, fail to resolve its placement in the eukaryote tree of life. Discussed further in Chapter 3.

Protist X

Anaerobic eukaryotroph with four flagella emerging apically in a cruciform arrangement, each with its own flagellar pocket (fig. 2.21H-K). Its morphology matches that of "Protist X" found in anoxic sediments (Bernard et al. 2000). Position unresolved by SSU rDNA phylogenies (see fig. 2.2). Despite two rounds of environmental sequence searches, each with a greedy approach, no significant matches were found in any dataset. Discussed further in Chapter 4.

SJB2

Bacterivorous flagellate with a conspicuous groove and unequal flagella rapidly skidding on the longer posterior one, overall resembling a "typical Excavate", something between a CLO and a malawimonad by morphology and behaviour (fig. 2.22A-D). Atypical division plane (see fig. 2.22D). Aerobe, isolated from San Juan Baai beach sand. Taxonomic affinity not resolved by SSU rRNA phylogenies. Has surprisingly numerous high probability hits—some previously

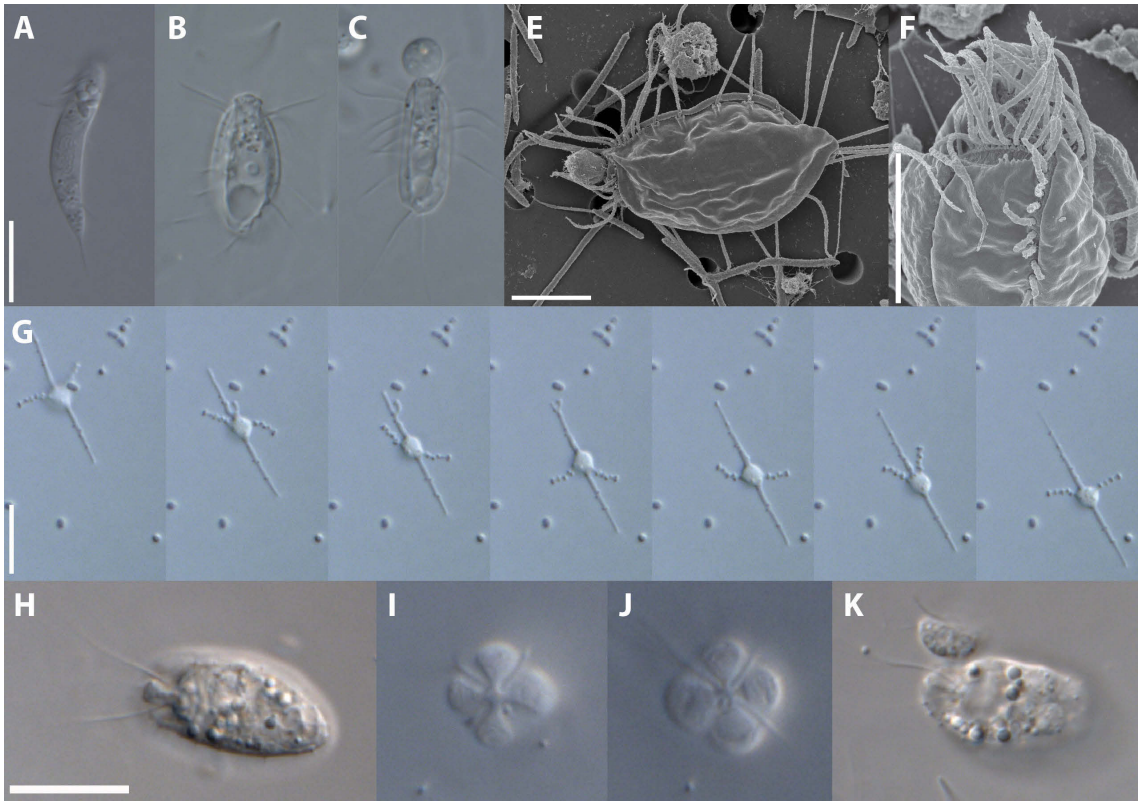


Figure 2.21: Light and TEM micrographs of hemimastigotes, *Meteora*, and Protist X. A-F) Basic characteristics of hemimastigotes *Spironema* cf. *multiciliata* (A) and *Hemimastix kukwesjijk* (B-F). A) *Spironema* cf. *multiciliata* is a lanceolate cell with a posterior "tail" and two rows of flagella at the anterior end. The capitulum is not visible in this image. B) General morphology of *H. kukwesjijk*. From the anterior capitulum (here in lateral section) two rows of flagella continue gently helically down towards the posterior. C) *H. kukwesjijk* feeding on *Spumella sp.* with its anterior capitulum. E) SEM image of the prey ingestion process. F) Early division stage showing a presumably nascent row of short flagella. (For further information on A-E, see Lax et al. (Lax et al. 2018)). G) Time series of a gliding cell of *Meteora sporadica* isolate LBC3 illustrating the distinctive swinging motion of lateral "arm" extensions. (See Chapter 3 for further information) H-K) DIC light micrographs of the anaerobic eukaryotroph Protist X, isolate MSW. H) Longitudinal section through the flagellate including two of its four flagella. I-J) Cross section of the apical insertion site of four flagella and their respective channels. K) Early step in the feeding process showing attachment of the prey *Carpediemonas* cell. (See Chapter 4 for further information). Scale bars: 10 m, except E-F) 10 m.

anotated as green algae—in environmental sequence datasets from marine sediments (including those from the deep sea), hydrothermal vent water, and anoxic marine environments, some with very high abundance for this study (fig. 2.23).

LRM3-SF

Tiny (3-5 μm cell body proper, but up to 8 μm in some larger individuals) amoeba with ultra-fine filopodia (barely visible under near maximum light microscopy resolution with DIC optics) and a flagellated stage that glides upon its posterior flagellum (fig. 2.22E-J). Flexible cell body. Ultra-fine filopodia emerge from the flagellate in some cases as well (fig. 2.22I). A dividing flagellate was found (fig. 2.22I-J). Mature amoebae surrounded by a ring of particulate matter/bactera with ~2 μm of space; sometimes two cells can be found within one such ring and a binucleate individual was found (fig. 2.22F-G), implying division can occur in this stage. Phylogenetic placement resolved by neither a BLASTn search of the nucleotide database (no hits above 86% sequence identity), nor via SSU rDNA phylogeny (fig. 2.2). Obtained from standard enrichment of a 2 year old sample of anoxic silt from an abandoned saltern now at regular marine salinity from near Sant Carles de la Ràpita, Catalonia, Spain. Several short read environmental sequence matches were found in data from hydrothermal vent fluid and deep sea sediments (fig. 2.23).

2.3.2 General environmental sequencing results

A 'probable hit' or 'probably match' is defined as a sequence (operational taxonomic unit (OTU), amplicon sequence variant (ASV), or, occasionally, read) with a RAxML-EPA placement likelihood-weight ratio (lwr) ≥0.9 and its corresponding query sequence ranked above all named (non-environmental) subject sequences in a BLASTn search of the combined NCBI nt and own sequences database.

Of the novel sequences in this chapter, most would have appropriate binding sites for conventional short read ribosomal environmental amplicon sequencing studies. For the V4 region, QSI-PG had a potentially significant mismatch with respect to the TAReuk454FWD1 primer (Stoeck et al. 2010) (fig. B.4; note: no data were available for LRM3-1F, TZLM3-RCL, GEM-RC, Jakobacarp, or CuSP2-3 BLOL in that region). For V9, every sequence except that of LRM3-1F had a perfect match to the "TARA Oceans" 1398F primer (de Vargas et al. 2015a), and no significant mismatches were present for the corresponding 1510R reverse primer (fig. B.4;

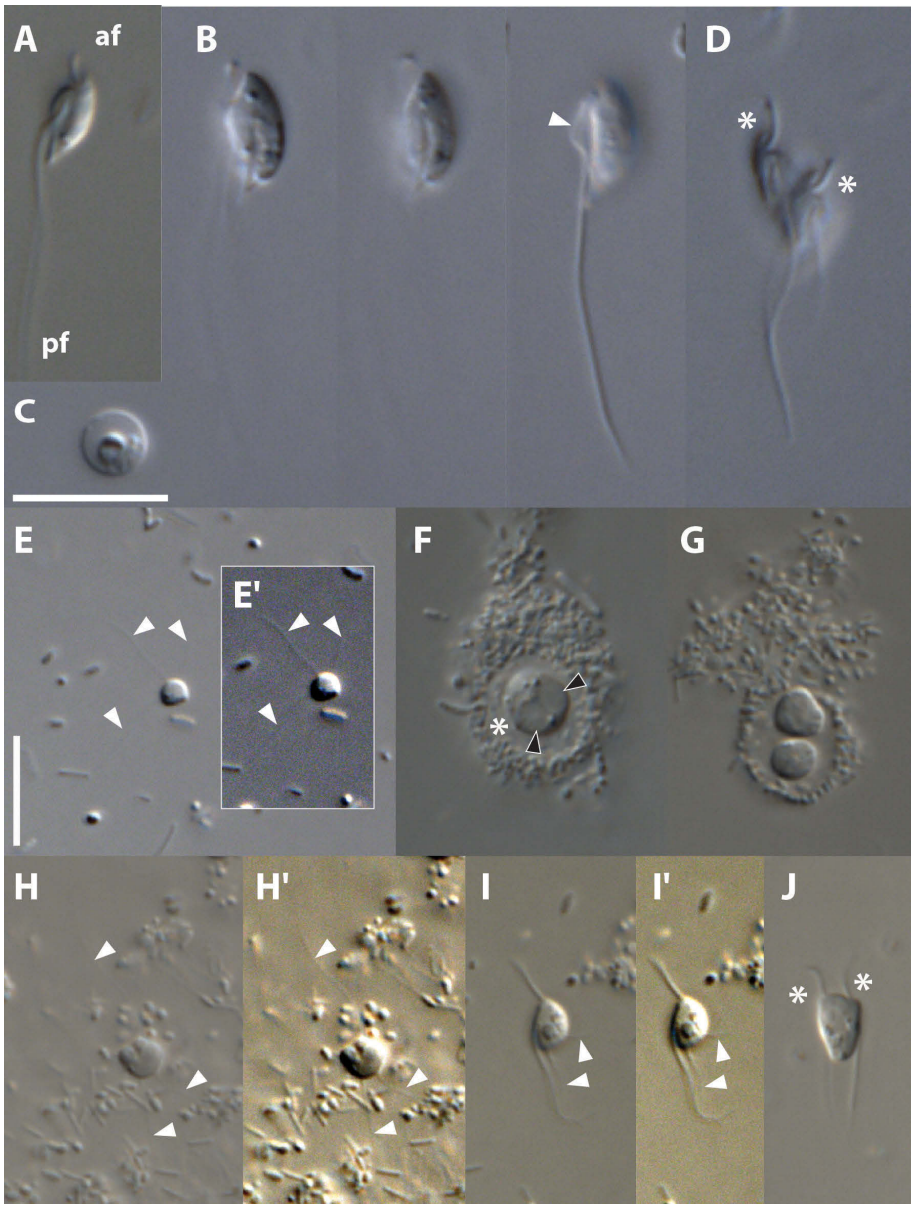


Figure 2.22: DIC light microscopy of *incertae sedis* organisms SJB2 and LRM3-SF. A-D) Excavate-like flagellate SJB2. A-B) General views and optical sections of two individuals, showing the insertion sites of anterior (af) and posterior (pf) flagella, as well as a vane on the posterior flagellum (arrowhead). C) Probable cyst of SJB2. D) A dividing flagellate (asterisks). Note unusual axis of division. (Continued on the next page)

note low quality and generally sparse sequence data for the reverse primer binding site at the 3' end of the gene). Thus, one would expect QSI-PG to evade amplification by this example V4 primer set, while most sequences should be amenable to amplification by the "TARA Oceans" V9 primer set.

This analysis was able to detect probable relatives for 16 out of 27 surveyed groups (20/39 isolates for which sequences were available) in a set of public short read sequence datasets totalling 14M sequences (OTUs, ASVs, and reads) (fig. 2.23a). The examined taxa with the most probable hits were the colpodellid NF, the breviates (especially *Breviata-Subulatomonas* type; perhaps unsurprisingly, as a number of environmental sequences has been detected previously: Katz et al. 2011), cercozoans RotF (*Abollifer* sp.) and RotFL – and, surprisingly, the representative of a novel excavate-like lineage SJB2. The majority of probable hits come from marine samples.

Despite appearing in high abundance in every alkaline sample and enrichment examined so far (n=5), no sequences with significant identity to any of the Retortacarps were found, even in the metagenome (not shown) and amplicon datasets from one of the sampling locations (Table 2.6; Zorz et al. 2019). Similarly, only two probable hits were identified for barthelonids, which appear frequently in anoxic enrichments, including those from hypersaline environments. The high abundance of sequences from cave lamenflora from the VAMPS dataset (Havlena et al. 2021) may be a result of lack of clustering or sequencing errors; without it, the colpodellid NF was still found in a variety of habitat types but with fewer probable hits for each.

Of the high probability hits, 40 matching environmental sequences met the traditional 97% sequence identity cut off for OTU classification, of which 24 had a sequence identity of 99% or above. The organisms with hits of above 97% sequence identity were LRM3-SF, breviates (including BLO), *Lithocolla*, SJB2, *Discocelia* and palpitomonads, with the latter three with

Figure 2.22: (continued from the previous page) E-J) Extremely small amoeba LRM3-SF and its flagellate stage. E) General view of a small individual. Note ultra-fine filopodia (arrowheads, see contrast-enhanced inset E') emerging from the cell body. F-G) Larger amoebae, usually surrounded by a ring of bacteria and/or debris, with a conspicuous 2-3 m clearing around the cell body (asterisk). The individual in (F) contains two nuclei (black arrowheads), and two cells can be found inside some rings (G), implying division in this stage. H) Another amoeba, not surrounded by a ring of debris, with ultra-fine filopodia emerging from the cell body (arrowheads, see contrast-enhanced image H'), I-J) flagellates of LRM3-SF. Two flagella, one anterior and one trailing posterior, emerge from the anterior of the cell. Ultra-fine filopodia can be seen extending from some individuals (arrowheads, see contrast-enhanced image I'). An individual with two sets of flagella (J) and two nuclei (not shown) was observed, indicating cell division is possible in this stage as well. Scale bars: 10 m.

Organism	Number of OTUs found
Barthelonids	2
Discocelis	16
BoP4.1 Q	1
BLO (breviate?)	45
Breviates: subulatomonas	57
Breviates: other	24
CuF1 MicroH	21
Heliomorpha	1
Jakobacarp	1
LRM3-SF	5
Lithocolla	39
Colpodellid NF	166
Palpitomonads	7
RotF (*Abollifer*)	46
RotFL (*Abollifer*-like)	109
SJB2	52

Table 2.6: OTU matches from environmental sequence datasets to novel taxa found in this study.

hits of 99% sequence identity or above. None were exactly identical.

Among the probable hits, a considerable majority are accompanied by low total abundance counts (non-normalised) in their respective datasets. Where per-site abundance information was available, a pattern of generally low counts occasionally interspersed with one or two higher abundance samples was observed for numerous organisms (fig. 2.23b. Note palpitomonads and *Discocelia* especially, as well as RotFL). 1097 of 1318 hits, or 83% (258 / 292 OTUs, or 88%, with the VAMPS data excluded) had corresponding abundance counts less than 3.

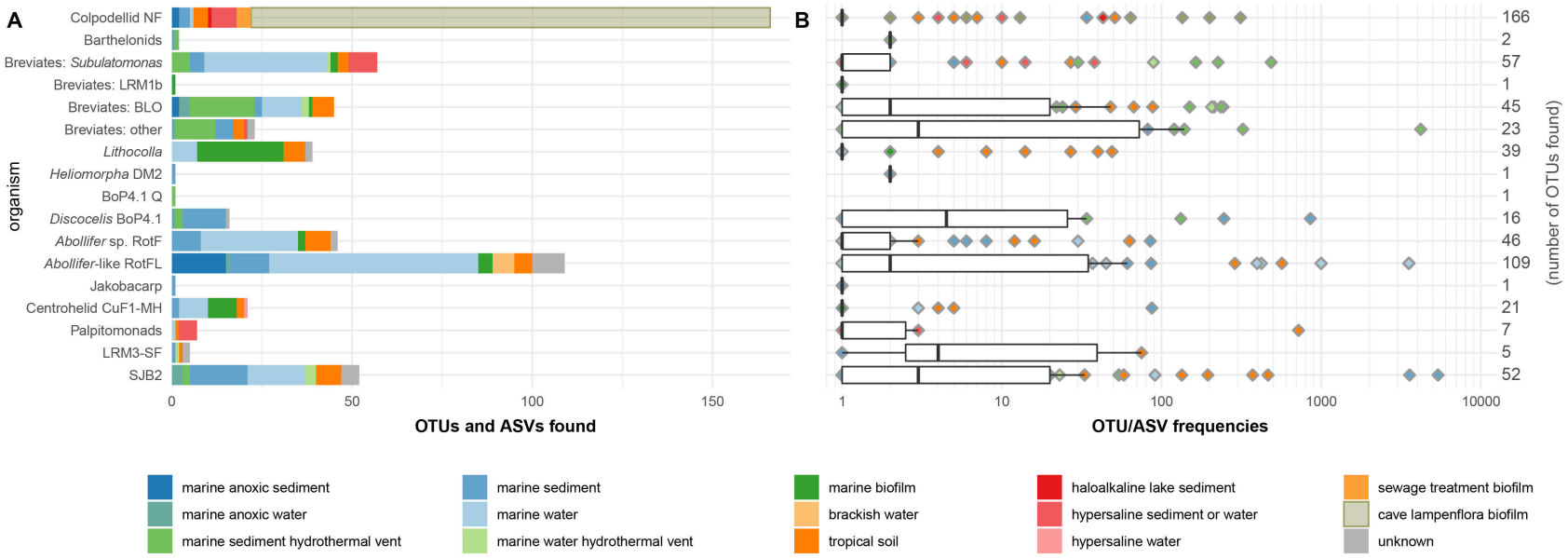


Figure 2.23: OTUs and ASVs found in environmental sequence datasets associated with novel taxa. A) Number of OTUs and ASVs for each taxon, by habitat type. B) Per sample frequencies (counts; based on read coverage) of each detected OTU or ASV, where data were available. Note that different normalisation and clustering approaches behind each dataset make cross-dataset comparisons difficult. The cave lampenflora biofilm dataset appears to not have been clustered at all, thus the OTU diversity is greatly inflated. The tropical soil dataset contains marine data; not all sample code data were available to date.

2.4 Discussion

The few dozen novel isolates in this study primarily come from sediment enrichments of marine, alkaline, and hypersaline samples. The newly identified and cultivated heterotrophic lineages are distributed broadly across the tree of eukaryotes, including Sar (eg. colponemids and SoapColp), metamonads (barthelonids and retortacarps), Obazoa (breviates and *Lithocolla*), Rhizaria (PG, *Discocelia*), Amoebozoa (*Multicilia*), Heterolobosea (Jakobacarp), and the nascent supergroup surrounding Hemimastigophora (*Meteora*, Protist X) (fig. 4.7, Chapter 4).

2.4.1 Evolutionary significance of novel lineages

Of note are some shared morphological features between several of the novel lineages. It has long been noticed that several groups of eukaryotes with morphology dominated by a large ventral groove, called "excavate taxa" (Simpson and Patterson 1999), fall in different parts of the eukaryotic tree. Of the three major "excavate" groups, *Carpediemonas*-like organisms (CLOs) are metamonads, jakobids fall in Discoba, and malawimonads are currently of uncertain placement in the tree, but branch with metamonads (possibly then representing an aerobic sister taxon) or within Amorphea near ancyromonads (Brown et al. 2018). In this study, isolate 'SJB2' has classic "excavate" features down to the dominating groove and a visible flagellar vane; however, SSU rDNA phylogenies do not place it within the previously known "excavate" groups, and its position remains unresolved. Isolate 'Jakobacarp' was another organism with an excavate-like morphology, including the groove. Surprisingly, SSU rDNA phylogenies place it in Discoba, but firmly within Heterolobosea: specifically within Tetramitiida (with whom it shares a derived SSU rRNA feature). Within Heterolobosea, percolomonads are another group with arguably "excavate" features, but with a notably different arrangement of flagella (all in parallel, as opposed to the heterodynamic flagella of classical excavates; Pánek and Čepička 2012). Altogether, it is possible the "excavate" traits in both SJB2 and Jakobacarp may be retained relics of an ancestral form; however, ultrastructural studies of the flagellar apparatus are needed to examine this further.

Colponemids are eukaryotrophic flagellates that also carry a ventral groove, although this time more subtle than in excavates, and accompanied by a smaller groove associated with the anterior flagellum. In addition to the importance of colponemids in understanding the evolution of alveolates (Tikhonenkov et al. 2014, Tikhonenkov et al. 2020b), a number of novel major lineages in the recent decade have been characterised from organisms resembling colponemids by

morphology and behaviour (raptorial eukaryotrophy). Examples include *Rhodelphis* (Gawryluk et al. 2019), *Ancoracysta* (Janouškovec et al. 2017) and *Aquavolon* (Bass et al. 2018). This study continues the trend: despite its resemblance to colponemids, 'SoapColp' turns out to not associate with them in SSU rDNA phylogenies, but instead branches within Stramenopiles, albeit with modest support. The broad yet patchy phylogenetic distribution of "colponemid-like" lineages may potentially reflect a selective retetion of ancestral traits in those lineages – in other words, the common ancestor of Diaphoretickes may have been a raptorial eukaryotroph with two ventral grooves or concavities: a major one underneath the posterior flagellum and a smaller one underneath the anterior. Relics of this arrangement may arguably also be present in the glaucophyte *Cyanophora paradoxa* (Kugrens et al. 1999). This possibility deserves further attention through more detailed examinations of the cell biology and morphology of these taxa, many of which are incompletely characterised.

It is notable how diverse flagellates with grooves are, and that they are broadly distributed across the tree of eukaryotes. The sampling of organisms with 'colponemid-like' and 'excavate-like' morphologies may continue to prove fruitful for the characterisation of novel lineages.

The survey of organsims from anoxic environments revealed several novel distinct anaerobic and anaerotolerant clades; or, in other words, independent adaptations to low oxygen environments. SSF and PG each represent a distinct adaptation to anoxic conditions among Rhizaria, complementing *Brevimastigomonas* (Gawryluk et al. 2016c) and *Mikrocytos* (Burki et al. 2013). Protist X currently represents the first cultivated anaerobic relative of hemimastigotes (although reports of unusual protists that might be hemimastigotes have been made previously: Ivan Čepička, personal communication) The tolerance of low oxygen envirionments by BLOL is also novel to palpitomonads, as the type species is aerobic (Yabuki et al. 2010). Altogether, these established cultures may prove useful for expanding our understanding of adaptation to anoxic environments, perhaps of the earlier stages in particular.

2.4.2 Few novel sequences fall within known environmental clades

Of the newly obtained SSU rDNA sequences in this study, several were placed in previously identified environmental clades. The anaerobic eukaryotrophic rhizarian lineage SSF maps onto Novel Clade 12 (NC12) Bass et al. 2018. Interestingly, the existence of anaerobes among

endomyxeans was previously suspected from short read environmental sequence surveys of sulfidic and anoxic waters (Triadó-Margarit and Casamayor 2015). It is possible some of these sequences are associated with the similarly anaerobic and eukaryotrophic rhizarian lineage PG, but are too divergent to be detected through BLASTn searches or evolutionary placement analyses. The heliomorphid DM2 was placed in the environmental clade containing envirionmental sequences MPE2-26 and PR_3E_90, branching either with NC10 or NC12; NC10 was recently characterised by the discovery of colponemid-like aerobic eukaryotrophs *Aquavolon* and *Lapot* (Bass et al. 2018). BWH-RotFL ended up in the environmental clade containing environmental lineages NAMAKO-14 and NAMAKO-15 (Takishita et al. 2007), sister to *Abollifer* species. This supports an *Abollifer*-related identity for that clade. Lastly, CuF1-MH mapped onto a previously exclusively environmental clade of centrohelids (Vasily Zlatogursky, unpublished data).

On the other hand, most novel lineages identified in this study did not correspond to a previously recognised environmental clade. For example, despite organisms resembling *Barthelona* species thriving in anoxic enrichments of a variety of different samples, no long read environmental SSU rDNA sequences were found in NCBI nt databases, nor did the *Barthelona* spp. SSU rDNA branch with any previously identified environmental lineage (a caveat here is that BLASTn searches for related lineages may not be sensitive enough: with default BLASTn parameters, the PCE barthelonid sequence does not pick up that of barthelonid LRM2 and vice versa). In a similar vein, Retortacarp-like organisms were found in virtually every alkaline lake sample examined so far, often at high abundance in the original sample; still, we could not find any related environmental sequence, even in short read data. Of the novel *major* lineages, while hemimastigotes did have corresponding molecular sequences in ecological studies (Lax et al. 2018), extensive surveys of both long read and short read environmental sequence data found nothing for *Meteora* or Protist X (each seen in and successfully cultivated from multiple different samples). None of the organisms discussed here had significant phylogenetic affinity to any of the environmental clades not assigned to supergroups that are in EukMap (Cédric Berney). Curiously, most of these clades consist of environmental sequences identified via Sanger sequencing of clone libraries in the early 2000s.

2.4.3 Undersampled habitats are productive for finding novel diversity

In this study, two groups of anaerobes showed an unexpected prevalence and diversity in marine anoxic environments: barthelonids and breviates. Both groups were recognised as distinct taxa

relatively recently, within the past couple of decades (Bernard et al. 2000 and Walker et al. 2006), and a similar expansion of diversity and distribution was detected at the same time by other research groups (Yazaki et al. 2020; Ivan Čepička, unpublished data). In an analogous case, the recognition of of *Carpediemonas membranifera* as a distinct lineage in the 1990s (Ekebom et al. 1996, Simpson and Patterson 1999) was then followed by the characterisation of an extensive breadth of distinct lineages in that part of the tree (Kolisko et al. 2010b). This suggests that not only may currently non-speciose lineages represent larger groups of distinct lineages, but that anoxic habitat diversity in particular may be severely undersampled. In contrast to the novel emergence of anaerobes within aerobic groups discussed above, this is an example of expansion of shallow rather than deep taxon sampling.

Most of the new halophiles characterised over the last 15 years have been classified as novel genera (*i.e.* not known from marine or freshwater habitats) (Harding and Simpson 2018), and there is a diversity of short read environmental sequences restricted to high salinity habitats (Filker et al. 2017). The known diversity of halophilic protists is normally dominated by heterolobosean groups (Park and Simpson 2015, Harding and Simpson 2018, Ruinen 1938), some ciliates (Park and Simpson 2015, Post et al. 1983), stramenopiles (Park et al. 2006, Rybarski et al. 2021), chlorophytes like Dunalliela (Post et al. 1983), colpodellids (Post et al. 1983 , Patterson and Simpson 1996), choanoflagellates (Schiwitza et al. 2018), and unidentified centrohelids (Post et al. 1983); in short, non-halophilic protists have invaded hypersaline enrivonments a multitude of times independently. This study continues to expand that list, with the addition of hypersaline representatives of Multicilia, breviates, barthelonids, and a collection of additional colponemid lineages, all of which were found the live in salinities at or over 150ppt. One possible cause behind the degree of novelty discovered here is that the anoxic layers of hypersaline habitats have been particularly underexplored. One partial exception is the anaerobic ciliate *Trimyema* (Park and Simpson 2015). Molecular adaptations to halophily have been explored in very few of these halophilic protists (Harding et al. 2016). The transition between marine and freshwater environments within clades was recently found to be not as rare as originally believed (Jamy et al. 2022), thus one would expect finding further halophilic lineages within otherwise marine or freshwater clades.

Alkaline environments present another chemical challenge for single-celled eukaryotes. The majority of microbial studies of these environments to date have been on prokaryotes (Sorokin et al. 2015, felfoldiMicrobialCommunitiesSoda2020; but Pirlot et al. 2005, Yasindi and

Taylor 2016); in fact, the diversity of eukaryotes is expected to be low compared to that in freshwater lake systems (Lanzén et al. 2013, Schagerl and Burian 2016). Overwhelmingly, the new organisms found in this study represented new species and, in several cases, a distinct novel lineage (*e.g.* Retortacarps, SoapColp). Further sampling of alkaline environments should continue to yield yet uncharted novel euakryotic diversity. Next to nothing is known about the adaptation of microbial eukaryotes to extreme alkaline environments, and the established cultures could be suitable systems for its study—for example, Retortacarps, as all lineages to date have been detected exclusively in alkaline samples, and several grow to high density in culture.

2.4.4 Eukaryotrophs are an especially underexplored source of new diversity

Besides a handful of algae, the majority of protists currently characterised from hypersaline environments are bacterivores (Harding & Simpson 2018); furthermore, the majority of studied anaerobic microbial eukaryotes are either bacterivores or animal parasites (see Chapter 4 Discussion). Here, I documented and isolated several new eukaryotrophic lineages, many from anaerobic, hypersaline, and hyperalkiline environments. The largest collection were the colponemid isolates, representing novel deep lineages of alveolates. SoapColp (Weston et al. in prep) may represent a novel major lineage of stramenopiles, while endomyxeans PG and SSF each represent a new Rhizarian anaerobic eukaryotroph. The most phylogenetically distinct groups found in this study, the Hemimastigotes (Lax et al. 2018) and Protist X (see Chapter 4), are also both eukaryotrophs. Thus, it appears that in most types of environments, searching for eukaryotrophs in particular may prove fruitful in the hunt for novel diversity (see also: Tikhonenkov 2020). This may be due a combination of past requirements for purity of cultures for molecular work, as well as the possibility that eukaryotrophic lineages in particular may evade detection in environmental sequencing analyses due to high trophic level (and thus usually low relative abundance; Bachy et al. 2022).

2.4.5 Environmental sequence analysis

My analysis of environmental sequences in this chapter requires several caveats. For the sake of simplicity, in this discussion, I will use OTUs to refer to ASVs as well (as an additional caveat, some data in the VAMPS environmental sequence dataset appear to not have been clustered, *e.g.* the cave lampenflora dataset). Next, molecular sequence classification is dependent upon the completeness of the reference database at any given time: a more closely related representative

may be found to be phenotypically divergent from the previous closest relative and change the assumptions made about the environmental lineage. This is especially true among the more sparsely characterised areas of the tree of eukaryotes. And lastly, the input environmental data were inconsistently clustered and not normalised; thus, abudance data should be interpreted qualitatively rather than quantitatively. Absense of normalisation should have minimal impact on interpreting low abudances (frequencies of 1 or 2), but whether a high per sample abundance represents a true 'bloom' should not be inferred from these data. As an aside, some taxa may be missing from some public datasets due to quality filters removing extremely low abundance OTUs (eg. with a frequency of 2 or lower).

The taxa in this study are overwhelmingly rare in surveyed environmental datasets. Where plausible OTU matches were detected at all, the majority had low total abundance counts in their respective datasets. Some had a pattern of low counts across individual samples, with the occasional spike in abundance in one or two. A similar pattern of generally low abundances with a high relative abundance in one dataset was observed for hemimastigotes (Lax et al. 2018). This pattern could have two contributing explanations. First, these low read abundance taxa may, in fact, be present at least at modest levels: either due to much of the sequencing depth being soaked up by a small number of exceptionally high abundance taxa, and/or as a result of biases in amplicon sequencing methods. Additionally, some environment types have been sampled substantially more thoroughly than others. Alternatively, the organisms could be not abundant in nature, sparsely distributed, or persist at low abundances outside of short-lived blooms.

That the majority of reads usually belong to a small portion of the OTUs is a widely observed result in microbial ecology (de Vargas et al. 2015a, Forster et al. 2016, Mahé et al. 2017). Some of this may be a consequence of technical biases. For example, some of these technical problems reviewed in V. Wintzingerode et al. 1997 are: 1. primer affinity bias and incompatibility (In the course of this study, some cultivated organisms provided difficult to sequence with standard eukaryotic primers), mixed eukaryotic DNA in ecological samples would only exacerbate the problem further; 2. biases in PCR amplification (eg. due to %GC content); chemical contaminants in the sample can further impact PCR efficiency; 3. introns and other secondary structure complications in SSU rDNA and rRNA; some taxa may have DNA-associated molecules that can inhibit PCR, e.g. by causing loops; 4.polymerase error may appear as low abundance sequence variants that are difficult to distinguish from true low abundance variants.

At low abundances, factors affecting quantitative analyses can lead to a failure to detect a portion of the biodiversity altogether. Thus, additional complications—such as due to variations in ribosomal copy number between organisms (V. Wintzingerode et al. 1997; Lavrinienko et al. 2021)—can nevertheless impact qualitative studies.

Technical issues aside, some of the underdetection of taxa in molecular environmental surveys can be a result of taxonomic groups being specific to certain environments that are rarely sampled. A prime case of this is the benthos, which, partly due to being a more difficult type of sample to process, is considerably less explored than its planktonic counterparts (Forster et al. 2016, Vaulot et al. 2022b). Taxa may also be difficult to detect due to the dominance of a small number of extremely abundant OTUs (de Vargas et al. 2015a, Suter et al. 2021); that some novel taxa appear prevalent in 'extreme' environments (Stoeck et al. 2003, Pasulka et al. 2019) may be a consequence of a lower relative abundance of 'weedy' taxa that normally dominate but do poorly under those unusual environmental conditions. Sampling 'extreme' or otherwise unusual environments may allow detection of more taxa with broad ecological distributions, in addition to the specialists restricted to those 'extreme' environments.

Another critical factor in molecular ecology is the timing of sampling. Microbial communities fluctuate over time, and this is true both of prokaryotic (Robicheau et al. 2022, Parada and Fuhrman 2017, Yeh and Fuhrman 2022) and eukaryotic microbes (Piredda et al. 2017, Robicheau et al. 2022). Sparse temporal sampling can lead one to overlook organisms that have a particular seasonality, or that may appear only in very specific conditions (biotic or abiotic). An obvious example is algal blooms in the ocean, where predators, parasites, and opportunistic feeders may accompany a rapid crash in the algal population. This can be picked up through continuous monitoring, eg. via FlowCytobot (Peacock et al. 2014) or frequent environmental sequencing sampling (Ollison et al. 2022). Sequences of organisms likely related to hemimastigophora were at a low abundance in surveyed environmnetal sequence data except in one sample (Lax et al. 2018); we speculate that the timing of this particular sampling event may have coincided with a bloom of its prey.

Some microbial eukaryotes may in fact be present at extremely low abundances, or be altogether dormant, for the vast majority of time. There have been arguments for a "microbial seed bank"—a dormant population lingering, for example, in sediments, awaiting opportunity to arise as new colonisers (Finlay 2002, Scheckenbach et al. 2010). This may explain some of the observed differences between diversity characterised through cultivation-based approaches

and that in environmental sequence data, which has been observed for prokaryotes (Ventosa et al. 2014)—where some frequently isolated taxa were consistently low abdunce OTUs in corresponding environmental data from the same samples—as well as protists (Risse-Buhl et al. 2013). Thus, the cultivation-based approaches to exploring protist diversity in this study may have been accessing the "microbial seed bank", *i.e.* some of the isolated organsisms may represent normally low abundance taxa that flourished when presented with an underpopulated or unestablished environment (the enrichment medium).

2.4.6 Further directions

Although this study was cultivation-based, comparisons with publicly available environmental sequence data highlighted some limitations in current datasets. Short read, 454- and Illumina-based, sequencing has largely limited amplicon-based studies in microbial eukaryotes to the V4 and V9 regions of the gene. This leads to the omission of unusually long versions of these regions (*e.g.* V4 in euglenids; Lax and Simpson 2020) and, more crucially, insufficient signal for placing sequences without known close relatives in databases, nor in reference phylogenies. Furthermore, these short reads are sometimes misidentified by BLAST-based pipelines and environmental sequence placement analyses alike. Longer environmental amplicons would relax the sampling requirements and significantly improve the phylogenetic resolution of novel environmental sequences.

A significant innovation gaining attention at present is the use of long read sequencing technologies for close-to-full-length SSU rDNA amplicon—or the full rDNA operon—sequencing. Initially plagued by relatively high error rates, both Oxford Nanopore and PacBio technologies have become more reliable as the manufacturers and the research community have become better acquainted with the new types of sequence data. Research in long read error correction is ongoing (eg. Zhang et al. 2020, Dohm et al. 2020). A recent study using PacBio sequencing rDNA in eukaryotes improved on this further by applying clustering methods and using both direct addition to the ribosomal reference tree as well as environmental sequence placements to assign and verify taxonic classification (Jamy et al. 2020a). Furthermore, long read sequencing can include portions of the LSU rDNA. While primer biases may continue to be a major problem (that they were for clone libraries (Stoeck et al. 2006), and also for direct PCR on cultured organisms in this chapter), long read sequencing may enable finding "mostly universal" primers in regions that were previously unsuitable for short read amplicon sequencing. Alternatively, a mix of primer sets may prove effective for covering a broader spectrum of taxonomic diversity,

especially for eukaryotes.

This chapter shows the value of a diverse approach to sampling from novel environments. Here, I utilised a relatively small repertoire of both sample types and the methods used to interrogate them. Amongst many possibilities for new directions, here I'd like to highlight two in particular. First, as lake chemistry is more complex and diverse than simply salinity and alkalinity, it may be worth investigating further different types of chemically unusual lakes, *e.g.* the magnesium rich Mahoney Lake, BC (Canada). Second, it would be interesting to explore further down the redoxcline for eukaryotes that feed upon or otherwise require bacteria that use alternative terminal electron acceptors, such as iron or sulfate. The eukaryotes themselves need not utilise these compounds in the same way. While it is clear that emplying different approaches broadens the exploration of novel protistan diversity, it is also true that if every protistology lab applies the same methods, there would still be plenty of novel diversity to be found. The golden age of microbial exploration is still only beginning.

2.5 Acknowledgements

I would like to acknowledge Noèlia Carrasco for providing access to the Punta la Banya (Catalonia) sampling location; Alexandra Worden (GEOMAR) and CIFAR for funding support and Mark Vermeij (CARMABI) for the Curaçao fieldwork. Jackie Zorz, Z. Nathan Taylor, and Ashley R. Smith supplied soda lake samples. Bruce Curtis provided immense help with bioinformatics, particularly BLAST tools. Sabine Filker, Alexandra Schoenle, Jackie Zorz, and Elizabeth Suter provided additional OTU tables and sample information for their datasets. Sebastian Hess was key in streamlining the TEM fixation and resin infiltration protocol, and Gordon Lax supplied cells for *Heliomorpha* and *Multicilia* isolate HAK. This work was funded by NSERC.

Chapter 3

Rediscovery, cultivation and phylogenomic analysis of *Meteora sporadica* reveals a new eukaryote 'supergroup'

3.1 Introduction

'Kingdom-level' branches are being added to the tree of eukaryotes at a rate approaching one per year, with no signs of slow-down (Burki et al. 2020; Gawryluk et al. 2019; Lax et al. 2018; Tikhonenkov et al. 2020a). Some are completely new discoveries, while others are morphologically unusual protists that were previously described but lacked molecular data. For example, Hemimastigophora are predatory (eukaryotrophic) protists with two rows of flagella that were known since the 19th century, but proved to represent a new deep-branching eukaryote lineage when phylogenomic analyses were conducted recently (Lax et al. 2018). *Meteora sporadica* Hausmann et al. (2002) has an unique cell morphology including a long axis of anterior and posterior projections, and, typically, a pair of lateral 'arms' (fig. 3.1a). The cell glides along substrates with the 'arms' swinging regularly back and forth. Unfortunately, the original description of *Meteora* was by light microscopy only and from a short-term enrichment of deep sea sediment; thus, *Meteora* represents both a novel cell type and eukaryote of unknown phylogenetic placement. Here, we cultured *Meteora sporadica* and obtained transcriptomic data from this organism. Transmission electron microscopy shows that anterior-posterior projections are

supported by microtubules originating from a cluster of perinuclear microtuble organising centres (MTOCs). Likewise, the 'arms' are supported by microtubules, and neither have a flagellar axoneme-like structure. Remarkably, phylogenomic analyses of 254 phylogenetic marker proteins robustly support a close relationship with Hemimastigophora. Our study suggests that *Meteora* and Hemimastigophora together represent a morphologically diverse eukaryotic 'supergroup', and thus are important for resolving the tree of eukaryote life and early eukaryote evolution.

3.2 Results and discussion

3.2.1 Morphology

The cell body of *Meteora sporadica* isolate LBC3 is typically about 4-5 μm in diameter, with 1 or more lateral 'arms' (usually 2) and glides along the surface via its long axis. The arms normally swing regularly back and forth, but gliding persists when the 'arms' are static or absent, indicating that motility does not depend on arm movement. Detached floating cells bend and squirm but appear to lack directed motility. There are numerous small granules along the 'arms', as well as the long axis (fig. 3.1D, arrow). These move back and forth along those extensions, as well as between them (fig. 3.1D). Occasionally, some granules rapidly expand into a short protrusion (Fig. 3.1D, white barbed arrow).

Meteora feeds on bacteria by contacting the prey with a granule (fig. 3.1C), typically on one of the 'arms'. The bacterium becomes attached, is gradually moved to the base of the arm, then phagocytosed once at the cell body proper. Each bacterium in a digestive vacuole fully covered by serial TEM (of isolate SRT610; see methods) contained a single complex structure (fig. 3.2), presumably the discharged extrusome. Most microbial eukaryotes with extrusomes use them for capturing eukaryotic prey or for defense (Archibald et al. 2017; Hausmann et al. 2003); it is notable that here extrusomes are used to capture bacterial prey, which is much rarer (Mikrjukov 1995).

TEM sections also reveal a population of distinctive hash-mark-shaped microtubule organising centres (MTOCs), underneath the nucleus. These MTOCs nucleate a bundle of microtubules in both anterior and posterior directions, forming the long axis. In cross section, the microtubules form a loose bundle, without any semblance of a flagellar axonemal structure. Transverse microtubules emerge from the same MTOCs and focus into the base of each 'arm'.

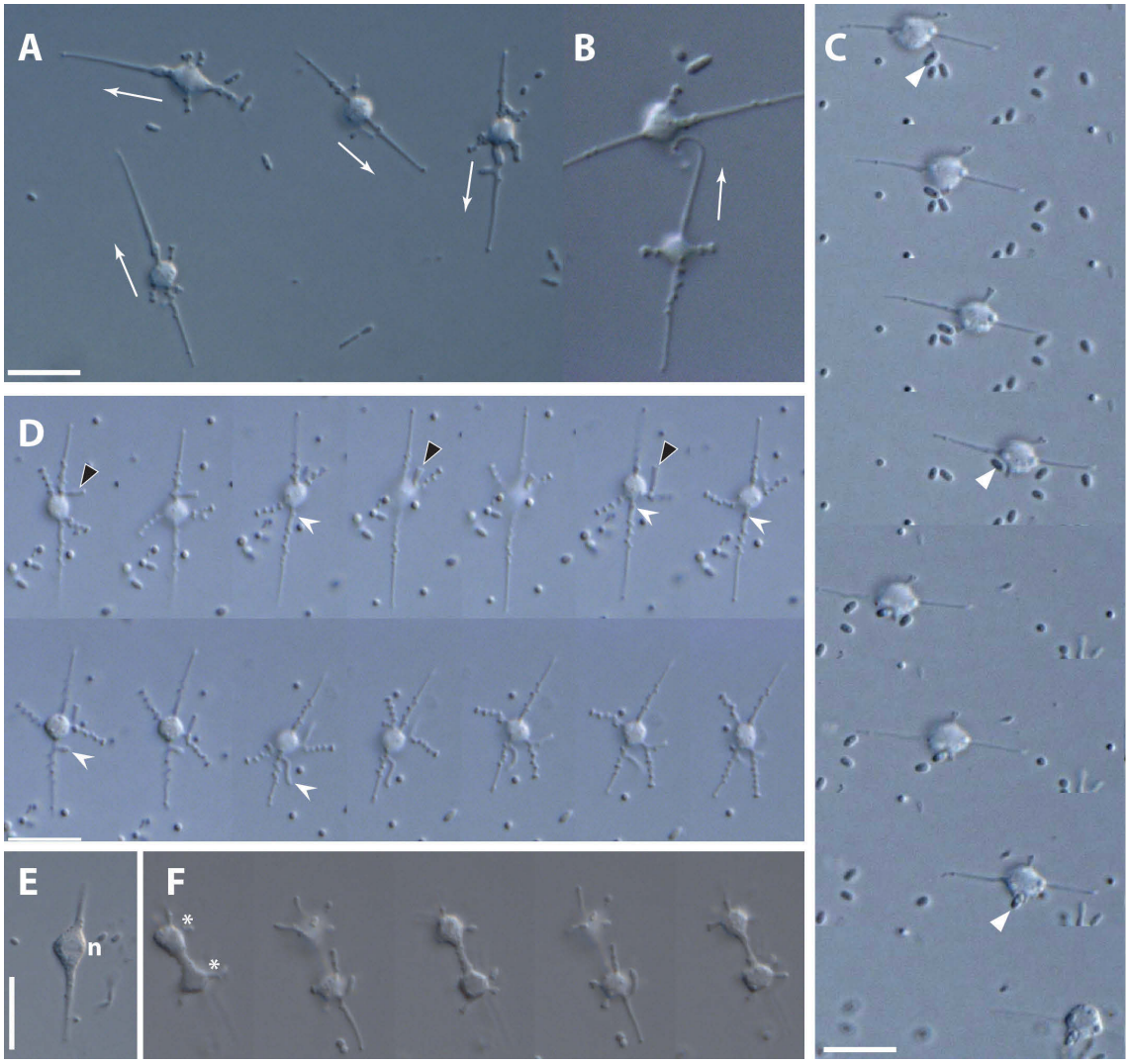


Figure 3.1: Light micrographs of *Meteora sporadica* isolate LBC3.A) General view of four cells of *Meteora sporadica* isolate LBC3, showing their direction of movement (arrows), and variety in lateral "arm" morphology. B) Collision between two cells (arrow shows movement) showing the bending of the long axis. C) Feeding on a bacterium. A granule on an "arm" plays a role in contacting and attaching the bacterium (arrowhead), which is then moved towards the cell body proper and phagocytosed there. D) Behaviour of cytoskeletal elements and surface granules. Protrusions can jump between the long axis and the "arms" across the surface of the cell body proper (black arrowhead). Areas of the axes associated with surface granules can protrude outwards, sometimes rapidly (white barbed arrow), and later fuse with the long axis (not shown). E) Cell in early division, "arms" retracted, as the cell body proper, containing the nucleus (n) in mitosis, gradually moves up and down along the long axis. F) Later stage of another dividing cell. Cells separate along the long axis and gradually begin to reconstitute "arms" starting from this stage. Scalebars: 10 m (all to same scale)

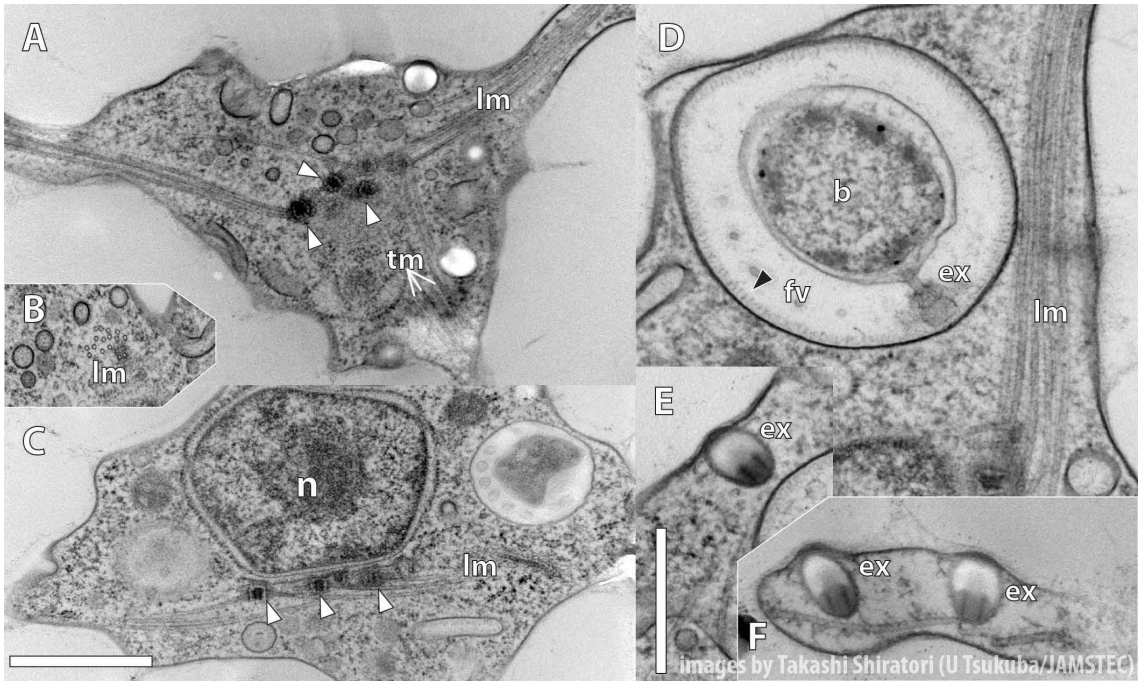


Figure 3.2: Ultrastructure of *Meteora sporadica* isolate SRT610.A) Cross section through a cluster of microtubule organising centres (MTOCs; white arrowheads) associated with the bundle of longitudinal microtubules (lm), and from which transverse microtubules (tm) extend at a right angle. B) Cross section through a bundle of longitudinal microtubules (lm). C) Longitudinal section through a cell, with the ventral side on the bottom of the image, showing the MTOCs (white arrowheads) situated underneath the nucleus (n). D) Food vacuole (fv) containing a prey bacterium (b) with a discharged extrusome (ex) still attached. The membrane inside the food vacuole is covered in a fibrillar material (black arrowhead). E-F) An undischarged extrusome (ex) docked to the cell membrane on the cell body proper (E) and in one of the lateral "arms" (F). Samples prepared and images by Takashi Shiratori (Univeristy of Tsukuba; JAMSTEC), used with kind permission. Scalebars: 500nm

The cells reproduce by dividing across the long axis: the cell stops in place while the cell body proper moves up and down for several minutes until visible cytokinesis begins; the daughter cells pinch off and then re-establish the missing ends of the long axis (fig. 3.1E-F).

Meteora is a bacterivorous surface glider; most other organisms that glide across surfaces, eating bacteria, are flagellates, and they usually glide on one of their flagella. They are highly abundant and found widely distributed across the tree of eukaryotes: examples include phagotrophic euglenids (Lax and Simpson 2020), glissomonads (Howe et al. 2009), mantamonads (Glücksman et al. 2011), and apusomonads (Torruella et al. 2017). The ecology and behaviour of *Meteora* closely resembles that of bacterivorous gliding flagellates; however, it does not have flagella, nor their obvious derivatives. Thus, it defies sorting into one of the general categories of eukaryotic organisms.

3.2.2 rDNA phylogenies and environmental sequence analysis

The SSU rDNA sequences for Meteora *isolates* LBC3 and SRT610 were 97% identical. An SSU rDNA gene phylogeny of 192 taxa broadly representing eukaryote diversity (fig. 3.3) failed to resolve the phylogenetic position of *Meteora* with any support, nor place it in any major group of eukaryotes. A similarly broad dataset of concatenated SSU+LSU rDNA likewise failed to resolve its position (fig. 3.4).

A phylogenetic placement analysis of environmental sequences from various publicly available datasets (see Chapter 2) using RAxML-EPA identified almost no candidate relatives of *Meteora*. A marine environmental sequence (asv_053_06994, Biomarks) was identified as a candidate *Meteora* relative with a modest likelihood-weight ratio of 0.7, but in turn was 97% identical to uncultured marine hydrothermal vent sediment clone AT4-68 (AF530543.1); this latter sequence sometimes resolves as sister to *Meteora* but without support (20% bootstrap support). Another environmental sequence, from a Neotropical soils metatranscriptome, was identified with a likelihood-weight ratio of only 0.56 and 86% sequence identity to *Meteora* LBC3 – thus, its identity remains inconclusive. No other environmental sequence hits were found. Thus, *Meteora* appears to represent a new and distinct phylogenetic entity in the current molecular tree of eukaryotes.

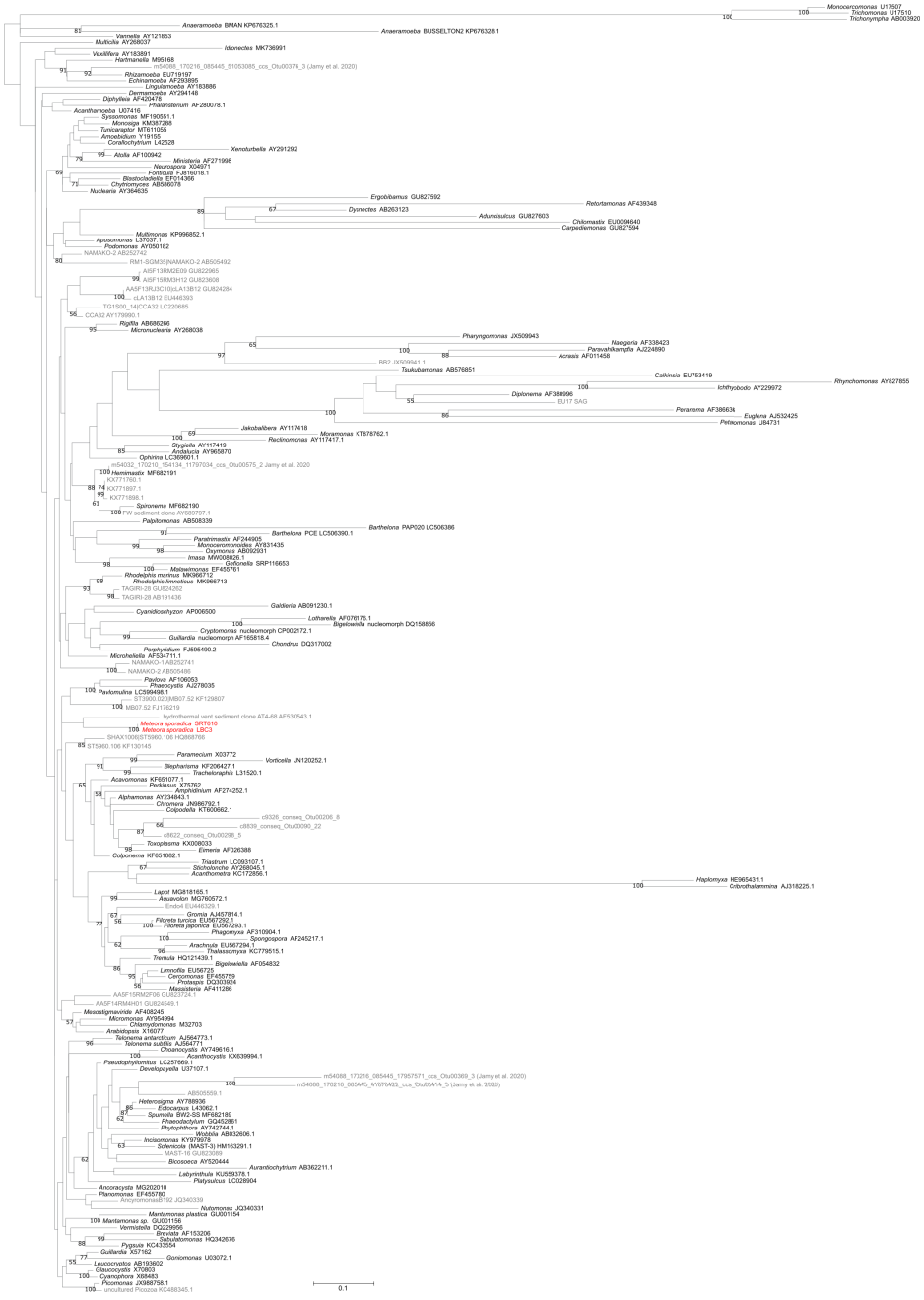


Figure 3.3: Position of *Meteora sporadica* in an SSU rDNA phylogeny. (Continued on the following page.)

3.2.3 Phylogenomics

To better assess *Meteora's* placement in the eukaryote phylogeny, we generated a transcriptome from isolate LBC3, and analysed this together with transcriptome data generated independently from a second isolate, SRT610 (Takashi Shiratori and Ken-Ichiro Ishida; University of Tsukuba). We assembled a 254-gene dataset, representing a broad sampling of eukaryotic diversity through 108 taxa, reduced to 66 for computationally-intensive analyses. The phylogenomic dataset was well represented in the sequenced transcriptomes (236/254 genes for both), which additionally had relatively high BUSCO scores (211 and 231 complete/255 for LBC3 and SRT610, respectively). Phylogenies inferred for both 108- and 66-taxon datasets, broadly agree with other eukaryote-wide phylogenetic studies (Burki et al. 2020; Gawryluk et al. 2019; Lax et al. 2018; Strassert et al. 2019), for example recovering Sar, Obazoa, Amorphea (i.e. Obazoa+Amoebozoa) and Discoba with full support (figs. 3.8 and 3.5). We did not, however, recover Telonemia as the sister group to Sar, *i.e.* the TSAR group (Strassert et al. 2019; but see: Yazaki et al. 2022). Remarkably, *Meteora* did not fall into any of the established supergroups, but instead formed a maximally-supported clade with Hemimastigophora, a phylogenetically isolated taxon recently proposed to represent a new eukaryote supergroup (Lax et al. 2018). The orphan protist lineage *Ancoracysta* branches as sister to this *Meteora*+Hemimastigophora clade, though with weaker support (85% PMSF bootstrap support; 97% UFBOOT support; posterior probability 0.99).

To assess the robustness of the *Meteora*+Hemimastigophora clade, we examined multiple variations on the 66-taxon dataset that test for potential sources of phylogenetic error. Analysis of a dataset that excludes three branches identified as long-branching outliers ("nLB") still returned maximal support for the *Meteora*+Hemimastigophora clade (fig. B.6). Recoding the amino acid data into a reduced alphabet of 4 classes based on (i) the pre-defined "SR4" categorisation (Susko and Roger 2007), or (ii) custom classes optimised to minimise across-taxon compositional bias in this dataset `minmax-chisquared`)(Susko 2021), both robustly supported *Meteora*+Hemimastigophora (SR4 and `minmax-chisquared`: 100% and 99% UFBOOT support, respectively; figs. 3.6 and B.9). By contrast, these same analyses did not recover *Ancoracysta*+*Meteora*+Hemimastigophora and placed *Ancoracysta* elsewhere entirely, as sister to

Figure 3.3: (continued from previous page) SSU rDNA phylogeny representing eukaryote-wide diversity for use as the reference tree for environmental sequence placement analyses. Alignment contains 1187 sites across 173 taxa. Tree inferred under the GTR+Γ+I model with 1000 non-parametric bootstrap replicates. *Meteora* sequences highlighted in red.

Figure 3.4: Concatenated SSU-LSU rDNA phylogeny. SSU-LSU rDNA phylogeny inferred from 3051 sites in final concatenated alignment, across 137 taxa, under the GTR+Γ+I model with 1000 non-parametric bootstrap replicates. *Meteora* sequences highlighted in red.

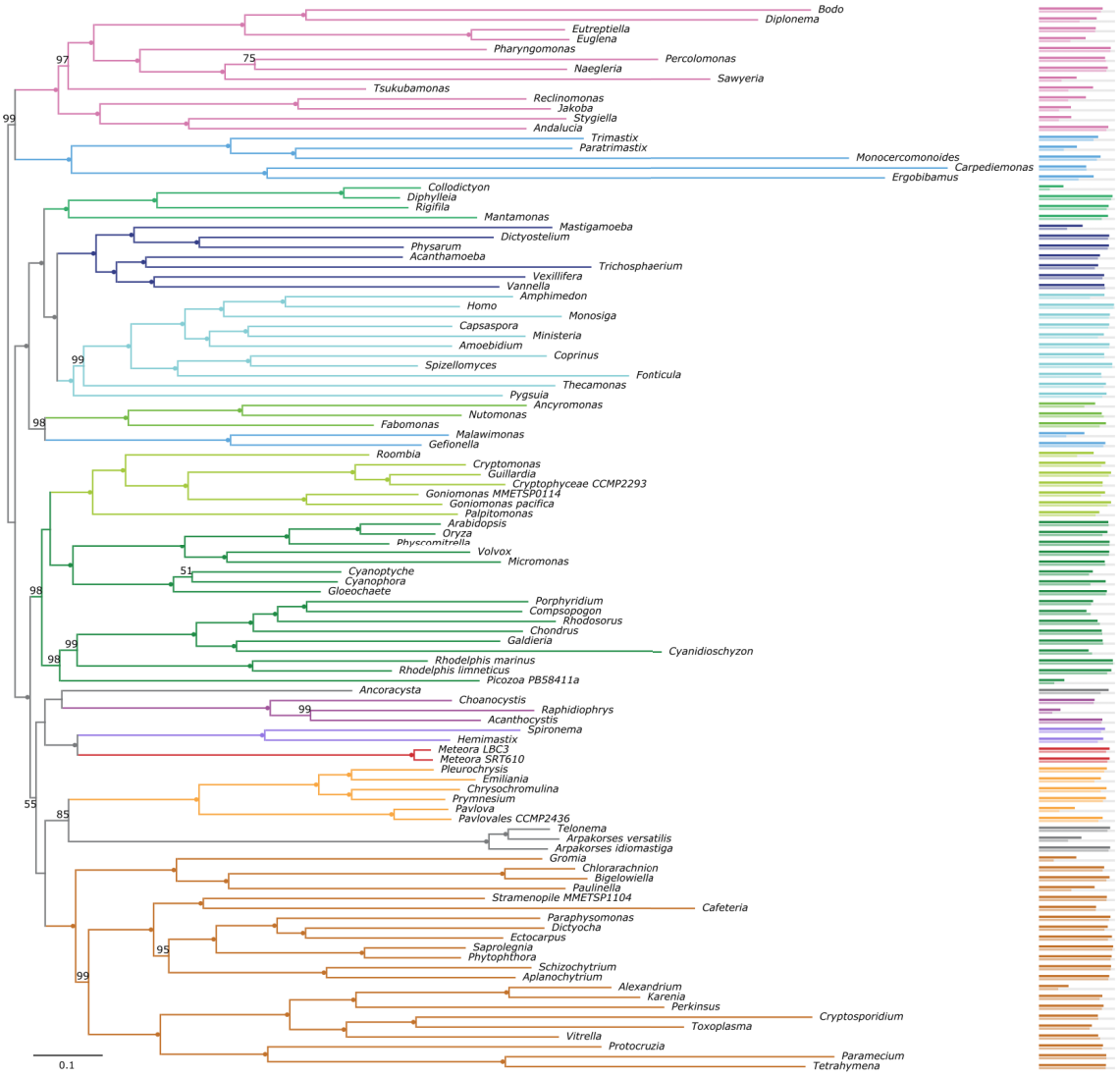


Figure 3.5: 108-taxon phylogeny inferred under the LG+C20+F+Γ model, showing the placement of *Meteora* among eukaryotes. 108 taxon phylogeny inferred from 70471 sites across a concatenated 254-gene alignment under the LG+C20+F+Γ model, representing major eukaryotic groups. Node support values represent % UFBOOT support from 1000 replicates. Filled circles indicate full support. Bars on the right represent coverage across the alignment as percent genes (top) and percent sites (bottom).

'Diaphoretickes' (SR4, 92% UFBOOT) or to haptophytes (`minmax-chisquared`, 97% UFBOOT). Incidentally, the 66-taxon dataset without *Ancoracysta* (noAnco) recovered the *Meteora*+ Hemimastigophora relationship with full support (fig. B.7).

Removal of the fastest evolving sites (FSR) in 10% increments (fig. 3.7a) showed *Meteora*+Hemimastigophora as maximally supported until 30% sites remaining, where support drops to ~80% UFBOOT; the widely-accepted Discoba taxon and the CRuMs clade behave similarly. Conversely, support for *Ancoracysta*+*Meteora*+Hemimastigophora is lower throughout (generally <95% UFBOOT although 98% at 50% sites remaining) and drops precipitously at 30% sites remaining.

Random subsampling of 50% of the genes in 5 jackknife replicates maintained robust support for *Meteora*+Hemimastigophora but not for *Ancoracysta*+*Meteora*+Hemimastigophora (fig. 3.7b). The gene concordance factor (gCF, Minh et al. 2020a) ; Minh et al. 2020a) for *Meteora*+Hemimastigophora (8.65%) was similar, or higher, than that of several accepted supergroups like Sar (9%), CRuMs (4.35%) and Amorphea (1.2%). By contrast, *Ancoracysta*+*Meteora*+ Hemimastigophora was recovered in <1% of the single gene trees (gCF 0.87%, fig. 3.6a). We infer that the phylogenetic signal for the *Meteora*+Hemimastigophora relationship is broadly distributed across genes.

Overall, the *Meteora*+Hemimastigophora association remained robust through tests for biases from subsets of genes and sites, and, notably, those for compositional bias. On the other hand, *Ancoracysta*+*Meteora*+Hemimastigophora was poorly supported in these tests, especially compositional bias. The position of *Ancoracysta* remains unresolved, as in prior studies (Gawryluk et al. 2019; Janouškovec et al. 2017).

In summary, phylogenomic analyses convincingly show *Meteora* as a sister group to Hemimastigophora. This seems remarkable based on their morphology and basic life history. As shown here, *Meteora* cells are completely aflagellate bacterivores, whereas hemimastigotes are multi-flagellated cells that prey on microbial eukaryotes (W Foissner et al. 1988b; Lax et al. 2018). Both exhibit symmetry, which is relatively unusual among eukaryotes; however, Hemimastigotes have diagonal symmetry and are essentially a constant shape, whereas *Meteora* appears to be predominantly bilaterally symmetrical and is highly plastic, breaking and re-establishing symmetry in the arrangement of arms. While these groups seem to have little in common, established eukaryotic supergroups like Sar and CRuMs also encompass a bewildering variety of morphologies and lifestyles. Sar encompasses fungal-like, flagellated, and amoeboid forms, and even large

macroscopic algae (Grattepanche et al. 2018). Although there are fewer than 10 described species of CRuMs, they range from small filose amoebae, bacterivorous nanoflagellates to larger eukaryotrophic flagellates (Brown et al. 2018).

This finding resonates with the high rate of discovery of novel eukaryotes, and indeed entire new phylum- and supergroup-level lineages, over recent years (Burki et al. 2020; Gawryluk et al. 2019; Janouškovec et al. 2017; Lax et al. 2018; Strassert et al. 2019). As with all other recently discovered major lineages (Burki et al. 2020), *Meteora* is a free-living heterotrophic protist, underlining the importance of pursuing this category for cataloguing deeper eukaryote diversity. It is also an illustration that the first known representative of a major clade (i.e. hemimastigotes in *Meteora*+Hemimastigophora) need not reflect the morphology or biology of the rest of the group. In particular, environmental lineages, i.e. groups known only from molecular data, may not necessarily be similar in basic biology to their closest morphologically-characterised relatives. Both the remarkable cellular architecture and unexpected phylogenetic placement of *Meteora sporadica* suggest that the bewildering diversity of microbial eukaryotes is far from fully understood, and will continue to surprise us.

3.3 Methods

3.3.1 Isolation and cultivation

Meteora sporadica LBC3 was isolated from an intertidal sediment sample from Cuba (21 °35' 24.99", -77 °5' 35.33"), kindly provided by Claire Burnard. The sample was enriched in seawater + LB at room temperature (21 °C). An individual cell on a flake of biofilm was picked with a drawn-out glass micropipette and placed in 0.1%LB in autoclaved natural seawater medium.It was subsequently maintained in tissue culture flasks with unidentified co-cultured bacteria at 16 °C and transferred every 2 weeks.

3.3.2 Light microscopy

Aliquots of culture are incubated on a sealed slide preparation overnight and imaged on Zeiss AxioVert 200M with an AxioCam ICc5 camera (Carl Zeiss AG). Downstream image processing and analysis was done in FIJI (Rasband 1997; Schneider et al. 2012).

3.3.3 Transmission electron microscopy

Transmission electron micrographs (TEMs) of serially sectioned material from *Meteora sporadica* isolate SRT610 (which is morphologically indistinguishable from LBC3), were provided by Takashi Shiratori and Ken-ichiro Ishida (University of Tsukuba).

3.3.4 DNA extraction, PCR, SSU rDNA phylogenetics

DNA was extracted from cultured isolate LBC3 using DNeasy Blood & Tissue kit (Qiagen). The small subunit ribosomal DNA (SSU rDNA) was obtained by semi-nested PCR, with initial amplification using forward primer EukA (5′-AACCTGGTTGATCCTGCCAGT-3′) and reverse primer 1498R (5′-CACCTACGGAAACCTTGTTA-3′) at 63 °C annealing temperature for 35 cycles, followed by secondary amplification with forward primer 82F (5′-GAAACTGCGAATGGCTC-3′) and reverse primer 1498R at 63 °C for 25 cycles. The final sequence was determined by Sanger sequencing (Génome Québec), with some PCR product being gel-extracted prior to sequencing (QIAquick Gel Extraction kit; Qiagen). Sanger reads were trimmed and assembled in Geneious v. 6.1.8 (Biomatters Ltd.).

The SSU rDNA sequence from LBC3, plus that from isolate SRT610 (provided by Takashi Shiratori and Ken Ishida, University of Tsukuba), were added to a global eukaryotic SSU rDNA alignment (derived from the reference SSU rDNA dataset for the environmental analysis in Lax et al. 2018) via profile alignment in SeaView (Edgar 2004; Gouy et al. 2010a). The alignment was further augmented for taxon sampling with additional environmental sequences from NCBI and Jamy et al. (2020b), corrected manually, then masked via gblocks (Castresana 2000) followed by manual correction to yield a 1187 site alignment across 173 taxa. This was subject to phylogenetic analyses in RAxML (Stamatakis 2014) (raxmlHPC-PTHREADS-SSE3 v. 8.2.6) under the GTR+Γ+I model with 50 starting trees and 1000 non-parametric bootstraps.

3.3.5 SSU+LSU rDNA phylogenies

Source alignments for SSU and LSU rDNA from Jamy et al. (2020b) were expanded for broader taxon selection using publicly available data in NCBI nt, or extracted from published transcriptome and genome assemblies using barrnap (seemanBacterialRibosomalRNA2018) v. 0.9. *Meteora* LSU rRNA sequences were extracted from the transcriptomes (see below) using barrnap and concatenated with the SSU rDNA mentioned above. Site selection was performed on each alignment using g-blocks (Castresana 2000) in SeaView (Gouy et al. 2010a) followed by manual

curation, then the SSU and LSU rDNA alignments were concatenated for a total of 3051 sites. The phylogeny was inferred via RAxML (Stamatakis 2014) (raxmlHPC-PTHREADS-SSE3 v. 8.2.6) under the GTR+Γ+I model with 50 starting trees and 1000 non-parametric bootstraps.

3.3.6 Environmental sequence analysis

We searched 14 153,628 publicly available V4 and V9 sequences from TARA Oceans (V9) (de Vargas et al. 2015a), VAMPS (V9) (Huse et al. 2014) , MetaPR2 ("Biomarks") (V4) (Pawlowski et al. 2012; Vaulot et al. 2022b), deep sea sediments (V9) (Schoenle et al. 2021), Malaspina (V4) (Obiol et al. 2020), neotropical (Mahé et al. 2017) and temperate (Geisen et al. 2015) soil metatranscriptomes (V4), and the Cariaco basin oxic-anoxic gradient (V4)(Suter et al. 2021) for sequences very similar to *Meteora*. The V4 and V9 regions of the *Meteora* LBC3 SSU rDNA were extracted and used to query the respective databases using BLASTn, with a sequence identity threshold of 80%. The collected sequences were then aligned using PaPaRa (S Berger and Stamatakis 2012) against a 1187 site, 173 taxon reference SSU rDNA alignment derived from Lax et al. 2018, manually curated through MUSCLE (Edgar 2004) profile alignments in SeaView (Gouy et al. 2010a), augmented for taxon sampling with additional environmental sequences from NCBI and Jamy et al. 2020 (Jamy et al. 2020b) (fig. 3.3). Phylogenetic placements were inferred via RAxML-EPA (SA Berger et al. 2011). Output was analysed in R using ggtree (Yu et al. 2017) and filtered with a likelihood-weight ratio threshold of 0.5.

3.3.7 Transcriptome sequencing and assembly

For LBC3, RNA was extracted from culture grown in 0.1%LB in sterile seawater on 20 Petri plates (15 cm diameter), scraped and spun 30 min at 2500 *g* and 16 °C, followed by adding 15 mL TRIzol (ThermoFisher) to 5 mL of resuspended pellet. Then, 3 mL of chloroform was added and phase separation obtained by centrifugation for 30 min at 4500 *g* at 4 °C. The aqueous phase was removed and further treated as per manufacturer instructions. The RNA was further purified with a phenol:chloroform extraction and treated with DNase. Quantity was assessed by Qubit (ThermoFisher). The sequencing library was prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490; New England Biolabs), and sequenced on Illumina MiSeq with 2 x 250 bp reads (V2 kit), indexed with Illumina adaptors i703 and i503 (multiplexed with an undescribed metamonad with adaptors i704 and i504).

Read quality was inspected using FastQC (Andrews 2010), adaptors clipped and reads trimmed with Trimmomatic (Bolger et al. 2014) v.0.30 (`LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15`

`CROP:160 MINLEN:36`) and assembly performed in Trinity (Haas et al. 2013) v.2.0.2. To remove most cross-contamination from multiplexed samples, we used a custom script (M. Kolisko, Institute of Parasitology Biology Centre, Czech Academy of Sciences, České Budějovice) and then reassembled in Trinity.

The independently sequenced and assembled transcriptome of isolate SRT610 was provided by Takashi Shiratori and Ken-ichiro Ishida (University of Tsukuba).

Transcriptome completeness for both isolates was assessed by BUSCO (Simão et al. 2015) v3.0.2 using eukaryote_odb10 dataset. This yielded 211/255 complete BUSCOs (24 fragmented) for LBC3 and 231/255 complete BUSCOs (13 fragmented) for SRT610. Of the 254 phylogenomic marker genes, 236 genes and 88.5% and 90.2% of sites were present in the final alignment for LBC3 and SRT610, respectively. Both transcriptomes were derived from monoeukaryotic cultures, sequenced independently in different locations, and the resulting marker genes were similar between the transcriptomes. Throughout the single gene trees, in almost all cases where candidate marker gene sequences were present for both isolates, they were sister lineages to each other with high support.

3.3.8 Phylogenomic dataset assembly

The 351 gene phylogenomic dataset from Lax et al. 2018 (based originally on Brown et al. 2018), was expanded by adding the two *Meteora* isolates, plus recently sequenced taxa including *Ancoracysta* (Janouškovec et al. 2017), telonemids (Strassert et al. 2019), and the three *Rhodelphis* transcriptomes (Gawryluk et al. 2019) via a custom pipeline (Brown et al. 2018). Telonemids, *Rhodelphis*, and *Meteora* were added (and Hemimastigophora re-added) using a custom script that enables multiple candidate genes per transcriptome to be selected and added at once, up to 4 in this case. After addition, each gene was re-aligned with MAFFT-linsi (Katoh and Standley 2013), trimmed with BMGE (Criscuolo and Gribaldo 2010a) (`-h 0.5 -g 0.2, -m BLOSUM30`), and phylogenies inferred under the LG4X+Γ model (Le et al. 2012) in IQ-Tree v1.5.5 (Nguyen et al. 2015), then manually inspected for paralogues, contaminants, lateral gene transfers, and signs of deep paralogies within the base dataset. Sequences marked for deletion were removed using a custom script. Where deep paralogies were detected affecting the whole gene tree, we discarded the gene from the dataset, resulting in a final phylogenomic dataset of 254 genes. The single gene alignments were filtered using PREQUAL (Whelan et al. 2018) with `-filterthresh 0.95` (0.28% masked), then trimmed with BMGE (`-h 0.5 -g 0.2, -m BLOSUM30`) and concatenated, for a final alignment of 70471 sites. The taxa were subsampled to produce a 108-taxon

dataset aiming to represent pan-eukaryote diversity, and a 66-taxon dataset for computationally intensive analyses. In both cases, phylogenetically redundant taxa were removed, with retention of higher coverage taxa where possible.

3.3.9 Phylogenomic analyses

An initial phylogeny was inferred from the concatenated 254 gene 108 taxon dataset in IQ-TREE v1.5.5 (Nguyen et al. 2015) using the LG+C20+F+Γ model, with support assessed via UFBOOT bootstrap approximation (1000 replicates) in IQ-TREE (Minh et al. 2013). Next, a phylogeny was inferred from the subsampled 66-taxon dataset under the LG+C60+F+Γ model, then used as a guide tree for the 60 custom profile site heterogeneous mixture model LG+MAM60+Γ (Susko et al. 2018) (hereafter referred to as "MAM60") using the program mammal (Susko 2022), with support values generated via UFBOOT bootstrap approximation in IQ-TREE (Minh et al. 2013). MAM60 was preferred over C60 by AIC (7324035 - 7302542 = 21493) and BIC (7325959 - 7314353 = 11606). A site-heterogeneous mixture model approximation method, PMSF (Wang et al. 2018), was used to generate 200 non-parametric bootstrap trees using the MAM60 tree as the guide tree. A Bayesian phylogeny was inferred using the CAT+GTR model in Phylobayes (Lartillot and Philippe 2004) v. 1.8 via 4 chains, with 1.1 x 10^4 cycles and a burnin of 500. Posterior probabilities were inferred from the 3 converged chains (chains 2-4), with the remaining chain (chain 1) stabilising on a topology with a different placement of haptophytes and *Telonema* (Figs. B.10 and B.11).

The Hemimastigophora+*Meteora* relationship was interrogated further via downstream analyses based on the 66-taxon dataset. A step-wise removal of fastest evolving sites (Fast Site Removal – FSR) was done in 10% increments using Phylofisher v.0.1.20 (Tice et al. 2021) and corresponding phylogenies inferred under MAM60 with UFBOOT support. Support values for relationships of interest were summarised via a custom script. A 'no long-branching taxa' (nLB) alignment was produced by determining the outlier long branches via a custom script (L. Eme; CNRS at Université Paris-Sud, France), in this case *Tetrahymena*, *Diplonema*, and *Bodo*. This dataset, along with one with *Ancoracysta* removed (noAnco) was used to infer a phylogeny under the MAM60 model.

To test whether the Hemimastigophora+*Meteora* relationship was the result of a few outlier genes, two analyses were conducted: gene jack-knifing and gene concordance factor (gCF (Minh et al. 2020a)) calculation. 5 gene-jack-knifing replicate alignments of 50% of the genes

(following recommendations for adequate statistical power in (Brown et al. 2018) were generated using `random_sample_iteration.py` utility in Phylofisher (Tice et al. 2021), and corresponding phylogenies inferred under MAM60 in IQ-TREE with statistical support from 1000 UFBOOT replicates. Single gene trees were estimated under MAM60 in IQ-TREE v.1.5.5 for each of the 254 individual gene alignments and gCF calculated in IQ-TREE v2.0 (Minh et al. 2020b).

To test for biases arising from sequence composition, two recoding approaches were used. The Susko and Roger set of 4 amino acid classes (SR4, Susko and Roger 2007) was used to reduce the amino acid alphabet. Additionally, a set of 4 amino acid classes that minimises compositional differences between sequences was determined via minmax-chisq (Susko 2021). In both cases, these schemes were used to recode the amino acid alignment as well as the 60 category MAMMaL model definition via custom scripts, and then a phylogeny was inferred under the GTR+[4binCustomModel]+R6 model in IQ-TREE 2.0, with support values inferred from 1000 UFBOOT replicates. Trees were formatted using the Ete3 toolkit (Huerta-Cepas et al. 2016).

3.4 Acknowledgements

I would like to thank Claire Burnard for supplying the original sample. André Comeau (IMR, Dalhousie U.) provided invaluable help with the Illumina MiSeq run. I would also like to thank Takashi Shiratori (U. Tsukuba/JAMSTEC) and Ken-ichiro Ishida for kindly sharing data and agreeing to collaborate. This work was funded by NSERC.

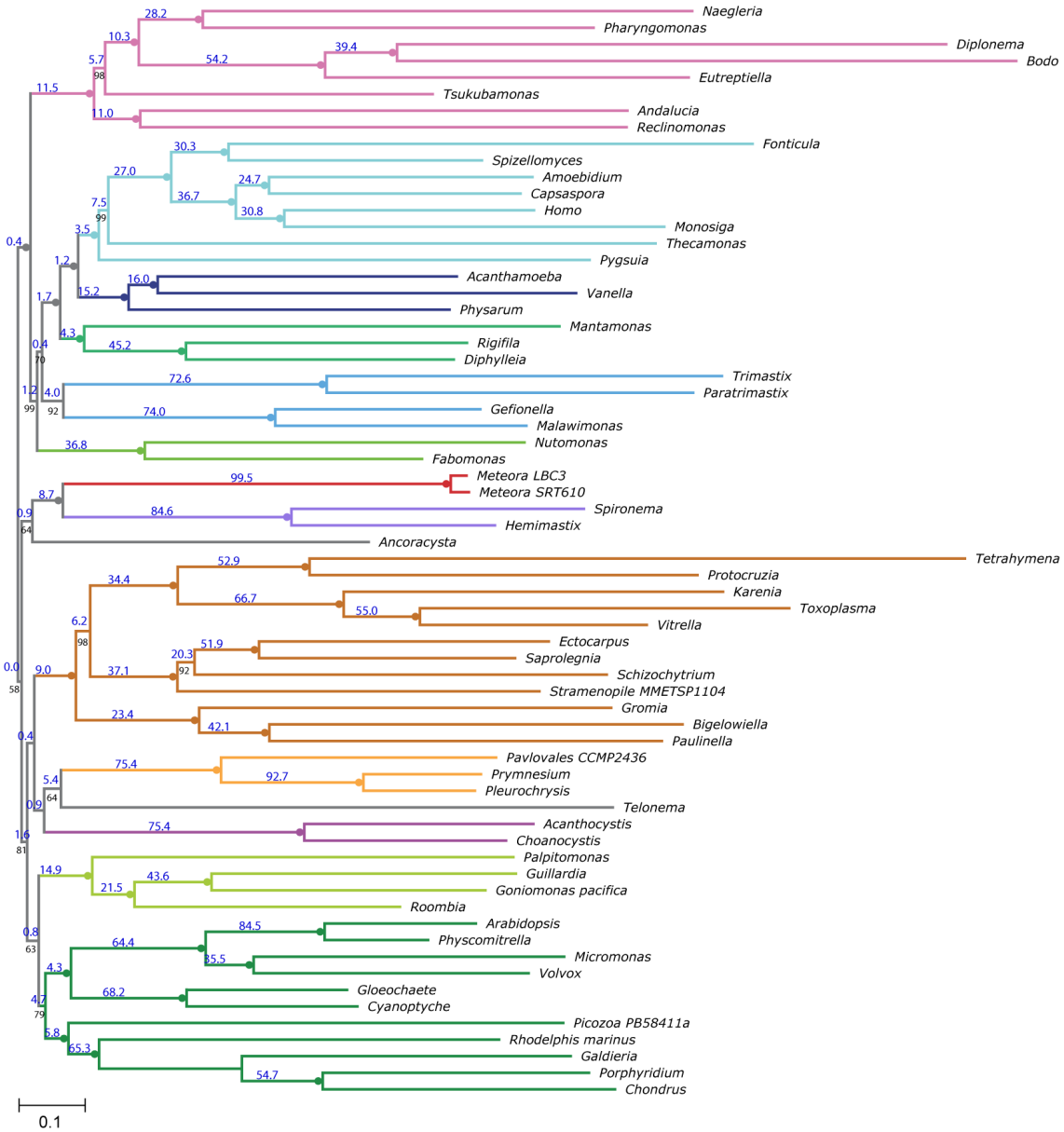


Figure 3.6: Gene concordance factor (gCF) analysis. a) 66 taxon topology with gene concordance factor values indicated on branches as percentages in blue.

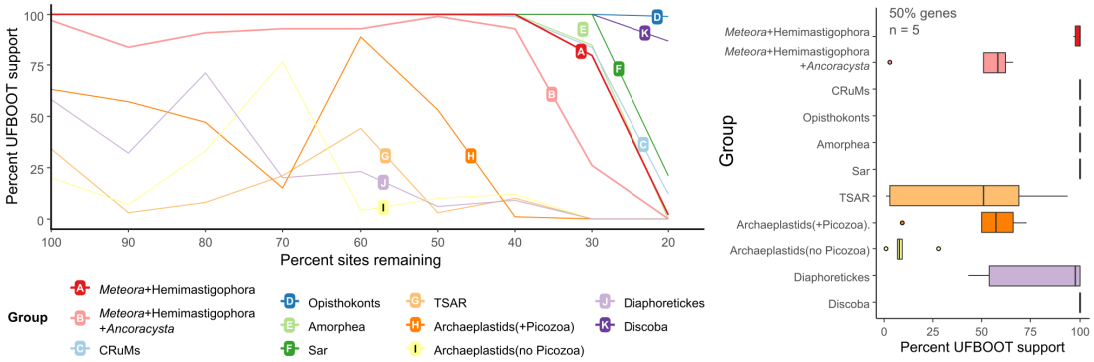


Figure 3.7: FSR and gene jackknifing analyses. a) Fast-site removal (FSR) profile of selected groupings with step-wise removal in 10% increments. Plot traces UFBOOT support (1000 replicates) under the LG+MAM60+Γ model. b) Support for selected groupings following 50% gene jackknifing (i.e., 50% of genes randomly removed) across 5 replicates.

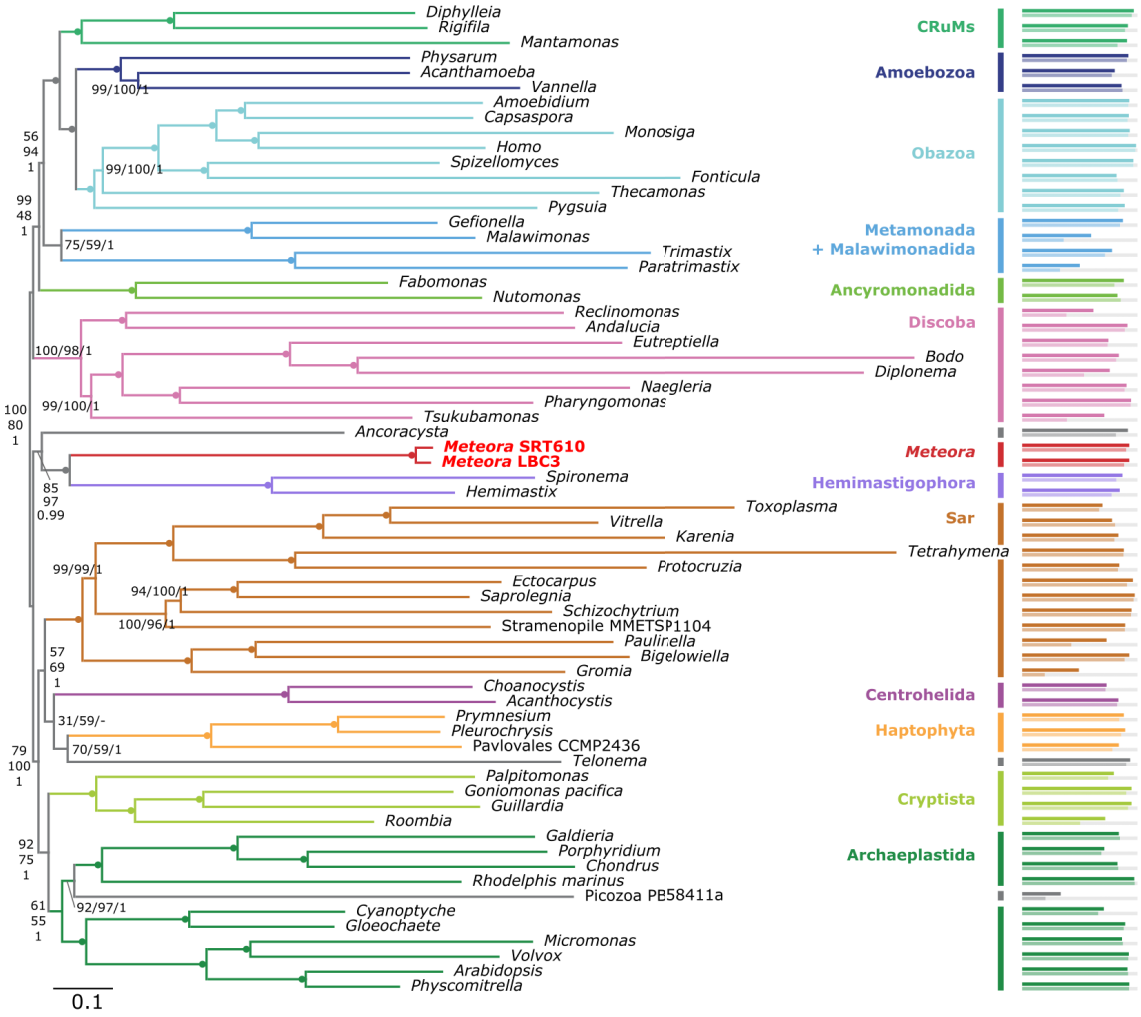


Figure 3.8: Phylogenetic placement of *Meteora* among eukaryotes under the MAM60 model. Maximum likelihood phylogeny inferred from 70471 sites across 254 genes over 66 taxa under the LG+MAM60+Γ model. Support values on branches show posterior mean site frequency bootstrap support (PMSF; 200 true replicates), UFBOOT support (1000 replicates), and Bayesian posterior probabilities (PP) under the CAT+GTR model, in that order, left to right or top to bottom. Filled circles indicate full support (100%, 100%, 1). Bars on the right indicate % coverage by gene (above) and by site (below).

Chapter 4

The anaerobic eukaryotroph 'Protist X' is a deep-branching relative of Hemimastigophora and *Meteora*

4.1 Introduction

Over the past 15 years, considerable advances have been made in the understanding of the eukaryotic tree of life, with much of eukaryotic diversity coalescing into "supergroups" well-supported by molecular phylogenetics (Burki et al. 2020). A handful of difficult to place "orphan" lineages still remains – e.g. Ancyromonads (Brown et al. 2018), – a problem exacerbated by what is likely incomplete taxon sampling across the eukaryote tree. Furthermore, lineages new to molecular sequencing are both completing and complicating our view of the tree of life. The discovery of *Rhodelphis* (Gawryluk et al. 2019) and enrichment of Picozoa taxon and gene coverage (Schön et al. 2021) have yielded moderate statistical support for Archaeplastid monophyly in multigene molecular phylogenies; this association was previously not recovered or statistically unsupported (eg. Janouškovec et al. 2017; but see Lax et al. 2018). On the other hand, the position of the newly discovered protist *Ancoracysta* (Janouškovec et al. 2017) remains unclear. Still, other lineages represent potential novel major groups containing a diverse collection of organisms – recently, collodictyonids, rigifilids, and mantamonads were shown to be

a well-supported clade (Brown et al. 2018) despite their varied morphology and lifestyles. Another lineage previously lacking molecular data, Hemimastigophora, was identified as a distinct branch outside any previously identified supergroup (Lax et al. 2018). It was later joined by an organism with very little resemblance to them – the remarkable gliding aflagellate *Meteora sporadica* (see Chapter 3).

An unusual heterotrophic flagellate was seen twice in marine anoxic sediment samples by Bernard et al. (2000), and mentioned – but not formally described – as 'Protist X'. This flagellate had an ovoid, slightly elongate cell body (12 μm and 19 μm in length) exhibiting rotational symmetry, with a cruciform arrangement of four equal flagella, each inserting subapically and emerging from a depression. This distinctive morphology did not resemble that of any previously established group of flagellates, anaerobic or otherwise, known at the time of the study. Nothing described since resembles 'Protist X' either, nor have associated molecular data ever been published for this organism.

Here, we characterise several isolates of flagellates corresponding closely to the description of 'Protist X'. The flagellates were found in marine and hypersaline anoxic sediment samples, and observed to feed on other eukaryotic flagellates in the sample, using the cell surface as the site of attachment and ingestion. The cells are sensitive to oxygen exposure, suggesting they are anaerobes. We have obtained ribosomal small subunit for several isolates, transcriptome data for 2, and have successfully established a permanent culture of one with a eukaryotic anaerobic flagellate as prey. We also characterise the ultrastructure of the cultivated strain by electron microscopy. Remarkably, phylogenomic analyses reveal that 'Protist X' forms a highly supported clade with Hemimastigophora and *Meteora*.

4.2 Methods

4.2.1 Cultivation and light microscopy

Several days after collection, a silty sediment sample with a bacterial mat from an intertidal lagoon (35ppt salinity) on McNabs Island at the mouth of Halifax harbour, Nova Scotia (44°36'20.5" N, 63°31'19.5" W) yielded a bloom of 'Protist X' under organic-rich (and anaerobic) conditions already present in the sample. Cells were picked by micropipette and washed in filter-sterilised medium from the same source. Three healthy cells were added to an established culture of the anaerobic jakobid *Stygiella incarcerata* strain MB1 (Simpson et al. 2008) as prey, growing in 3%

LB in autoclaved natural seawater filled to 12mL in a 15mL conical centrifuge tube; and, once established, transferred weekly. Later, the prey were switched to the faster growing CLO *Carpediemonas membranifera* strain BICM (Kolisko et al. 2010a), and also transferred weekly.

Additional cells were imaged and picked for sequencing as follows: isolate CuSP2-1 from Salina Sint Michiel, Curaçao, sediment (12°09'23.6"N 68°59'14.4"W) was temporarily maintained in culture on *Ergobibamus cyprinoides* strain CL (Park et al. 2010) as prey, from which cells were imaged, and three were picked and pooled for single cell transcriptome sequencing. Secondly, a single cell from anoxic intertidal sediments in False Bay, San Juan Island, WA, US (48°29'10.2"N, 123°04'32.7"W) (FB11), was picked for SSU rDNA sequencing via whole genome amplification by multiple displacement amplification (MDA, see below).

Light microscopy was done using DIC and phase contrast optics on a Zeiss Axiovert 200M with an AxioCam ICc5 camera (Carl Zeiss AG).

4.2.2 SSU rDNA sequencing

To separate the predator from the prey, a multiple displacement amplification (MDA) kit (General Electric Health) was used to amplify the genomic DNA of single or pooled cells of MSW and FB11, picked by micropipette as described above. Then material was amplified by PCR using unviersal eukaryotic primers (EukA and EukB) as per Table 2.2.1. (Chapter 2), and Sanger sequenced at Génome Québec.

4.2.3 rDNA phylogenies

An alignment of 196 SSU rDNA sequences broadly sampling the available tree of eukaryotes was profile aligned using MUSCLE (Edgar 2004) and manually curated in SeaView (Gouy et al. 2010b), then masked via gblocks (Castresana 2000) further edited manually for a total of 1212 sites. RAxML-NG (Stamatakis 2014) was used to infer the maximum likelihood phylogeny under the GTR+Γ+I model with 50 starting trees, and support values determined from 200 non-parametric bootstraps. LSU rRNA sequences for isolates MSW and CuSP2-1 were both obtained from the respective transcriptome data suing barrnap (seemanBacterialRibosomalRNA2018), and added to the concatenated SSU-LSU alignment from Chapter 3. Phylogenetic analysis performed as described in chapter 3.

4.2.4 RNA extraction and transcriptome sequencing

For strain MSW, TRIzol was used to extract bulk RNA from whole culture at peak predator:prey ratio, following the standard TRIzol protocol. An mRNA library for sequencing was

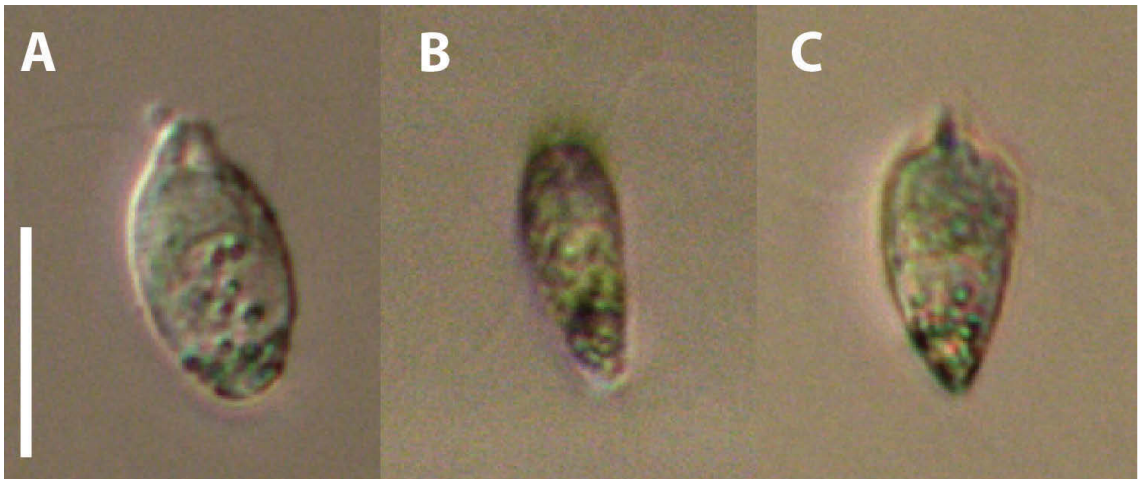


Figure 4.1: Three individual cells picked for single cell transcriptome of the isolate CuSP2-1. A-C) DIC light microscopy of the three individual cells of the isolate CuSP2-1 harvested for the single cell transcriptome. Cells are approximately 15 m long.

then prepared using the Illumina TruSeq kit, and sequenced on a MiSeq along with an undescribed heterolobosean (Yang et al. 2017). The reads were trimmed by Trimmomatic and assembled in Trinity.

For isolate CuSP2-1 ("CuCandC4"), 3 cells (fig. 4.1) were harvested and washed in sterilised source medium by micromanipulation with a glass pipetter, then preserved in lysis buffer and frozen overnight. A modified version of the Smart-seq2 protocol (Picelli et al. 2014) was followed to prepare input for the sequencing library. The sequencing library was prepared with a Nextera XT kit and sequenced on a V2 kit on an Illumina MiSeq. Reads were trimmed by Trimmomatic and assembled using RNAspades (Bushmanova et al. 2019).

4.2.5 Phylogenomic analyses

A modified version of a phylogenomic pipeline used in Brown *et al.* 2018 was employed, with modifications described in Chapter 3. Prey and other contaminating sequences were removed from 'Protist X' transcriptomes by screening single gene phylogenies generated under theLG+C20+Γ+F model. A 110 taxon concatenated dataset was built from BMGE-trimmed with default parameters (Criscuolo and Gribaldo 2010b) alignments of 254 phylogenomic marker genes for a total of 70471 sites (67145 distinct), based on the analysis in chapter 3, with isolates MSW and CuCandC4 covering 89% and 55% of genes (84%, 31% sites), respectively. A preliminary phylogeny was inferred from the full 110 taxon dataset under the LG+C20+F+Γ model with support values from 1000 replicates of UFBoot2 (Hoang et al. 2018) in IQTree2 (version 2.2.0, Nguyen et al. 2015). For computationally intensive analyses, a subset of 68 taxa was extracted from this

dataset, from which a phylogeny was then inferred under the LG+MAM60+Γ model (hereafter referred to as "MAM60") in IQTree2, with support values inferred via UFBoot2 optimised by nearest neighbour interchanges (`-bnni`). The resulting best tree topology was then used as a guide tree for PMSF (Wang et al. 2018) in IQTree (version 1.5.5) for 200 non-parametric bootstraps. Trees were formatted for publication using the Ete3 toolkit (Huerta-Cepas et al. 2016).

4.2.6 Electron microscopy

Cells from 5-7 day old cultures of MSW with *Carpediemonas* isolate BICM as prey were grown in multiple 15 mL conical tubes, filled to 5 mL, and harvested by gently pipetting the last 0.5-0.7 mL from the bottom of the tube into a new 15mL tube, let to settle overnight, then the bottom 1 mL taken and mixed with cells harvested from other tubes into a 1.5 mL tube. Residual cells were confirmed to appear healthy in that state even a couple hours after the rest were fixed.

Generally, 500 µL of cells was added directly to 500 µL of fixation solution (see Table A.2) in 2 mL tube and immediately gently mixed and placed on ice for 1 hour. Cells were spun 4min at 9K rpm and the pellet rinsed once with RO water, then resuspended in 200 µL RO water and 100 µL dehydrated acetone added to it for the first dehydration step at 33% acetone. Each dehydration step was 8 minutes and direct addition of dehydrated acetone was used to increase its concentration, for 50%, 75%, and 90% steps. Material was again spun 4min at 9K rpm and supernatant replaced with 100% acetone for 7 minutes, then with 50% acetone 50% EPON (SpiPON) 812 stored horizontally overnight. Cells were spun, aliquotted into 200 µL tubes and left overnight for acetone to fume off, then topped off with fresh resin, pelleted 5 min at 8k rpm in a horizontal centrifuge, then cured for two days at 65°C. In the meantime, 10 µL of resin-infiltrated cells were transferred to a slide to make a permanent mount.

Resin blocks were sectioned with a diamond knife on the Leica UC6 Ultramicrotome in approximately 65 nm increments, which were placed on pioloform coated slot grids. Imaging was done on a Tecnai-12 transmission electron microscope with a FEI Gatan Camera.

For SEM, cells added to poly-L lysine-coated 12mm round coverslips in multiwell plates, and fixed in 25% gluteraldehyde and osmium vapour. The samples were washed in seawater and then subject to an ethanol dehydration series (30-50-70-80-90-95-3 x 100%), then dried in a critical point dryer with CO_2 (Leica EM CPD300) and gold-palladium coated with a Leica EM ACE200 sputter-coater. Samples were imaged on the Hitachi S4700 scanning electron microscope at the DalTech SEM facility.

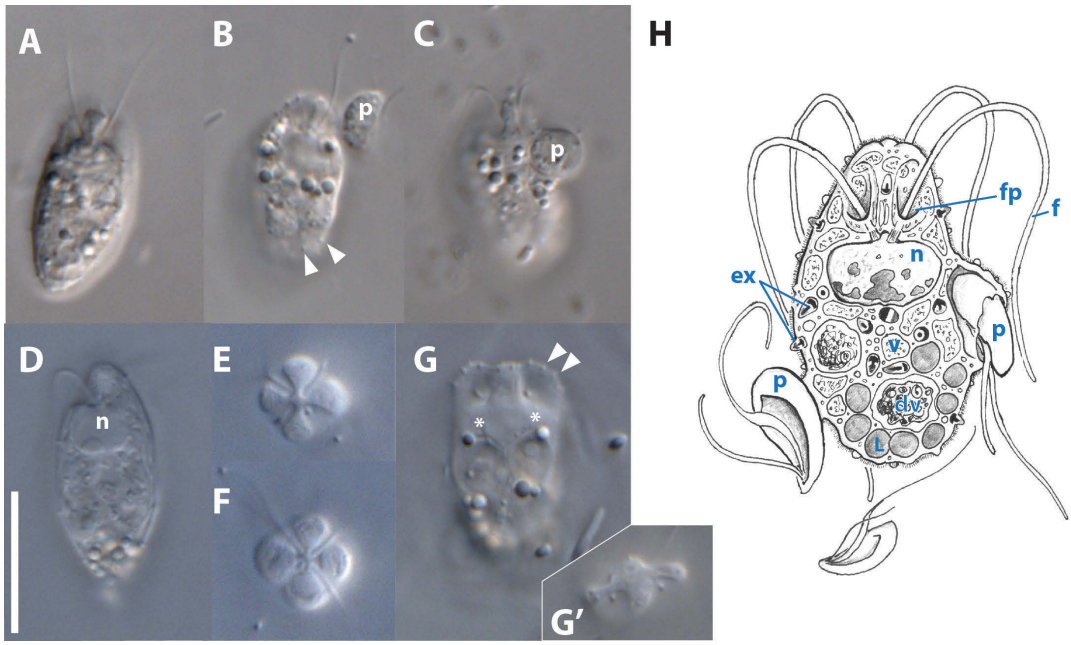


Figure 4.2: Light microscopy and a general diagram of Protist X. A-C) 'Protist X' isolate MSW. A) General view of the cell showing two of the flagella and associated pockets. B-C) View of a cell feeding on a prey *Carpediemonas* cell (p), and small particles on the surface, presumably where extrusomes dock (arrowheads). D-G') 'Protist X' isolate CuSP2-1. D) Longitudinal section through a cell showing two flagellar pockets and the nucleus (n) immediately underneath. E-F) Cross section through the apical flagellar pockets showing their cruciform arrangement. G) Cell in early division with two nuclei and the flagellar pockets growing apart (asterisks). Extrusomes (arrowheads) are particularly visible at the apical end of the cell. G') Grazing top view of the separating flagellar pockets from the cell in G. H) Diagram representing the general structure of the cell, drawn upon data from light and electron microscopy. Shows four flagella (f) and their pockets (fp), and the nucleus (n), vesiculated vacuoles (v), lipid globules (L) and digestive vacuoles (dv) inside the cell. Extrusomes (ex) are distributed underneath the surface, and up to two prey (p) cells can be ingested at once. Scalebar: 10 m

4.3 Results

4.3.1 General morphology and behaviour

Cells are spindle-shaped or pear-shaped (fig. 4.2A, D), typically about 14 – 18 μm long and 7 – 9 μm wide at the widest point, with four equal flagella one cell length long, each emerging from a deep anterior flagellar pocket, in a cruciform arrangement (fig. 4.2E-F). Cells swim in a jerky (stepwise) fashion, with ciliary beats by the flagella and diagonally opposite sets of flagella generally moving in synchrony. A single conspicuous nucleus is located just below the flagellar insertion sites, with condensed material towards the posteior (fig. 4.2D). Multiple refractile globules are located predominantly in the posterior end of the cell (fig. 4.2A – D), and the cytoplasm is overall highly vacuolated. Regular-sized small bumps, presumably docking sites of extrusomes, are distributed across the entire cell body surface (fig. 4.2B), including the anterior

end around flagellar pockets (fig. 4.2G). Cells are sensitive to oxygen exposure on a slide.

'Protist X' captures flagellate prey by surface contact anywhere on the cell surface, presumably through extrusomes. This is followed by gradual engulfment (fig. 4.2B; fig. 4.3). Sometimes more than one prey can be ingested at once (fig. 4.3B). Cell division occurs longitudinally (fig. 4.2G, G')

4.3.2 Ultrastructure

SEM reveals the four flagellar pockets with a single corresponding flagellum each at the apical end of the cell (fig. 4.4A, B). Bumps resembling the surface particles visible under DIC light microscopy are likely the extrusomes (fig. 4.4A). The cell surface, as well as the flagella, is covered in fibrillar material (fig. 4.4A-C).

Under transmission microscopy, the cytoplasm appears highly vacuolated (fig. 4.4D-F), with a large nucleus at the anterior third of the cell (fig. 4.4D,E) containing condensed material around the periphery and as a body towards the posterior end of the nucleus (fig. 4.4E). A population of large electron-dense globules is found particularly concentrated at the posterior end of the cell (fig. 4.4D). A subset of the vacuoles apparently contain internal vesicular material (fig. 4.4F,G), sometimes with coated vesicles visible facing outward from the surface (fig. 4.4G). Extrusomes are found throughout the cell in different stages of development (fig. 4.4E,F, H-J). Presumably mature extrusomes are found docked underneath the cell surface, and are conical, 600 nm long, with a rounded base consiting of dense-staining material, and a dense-staining central core extending halfway up the structure (fig. 4.4H-J). The more rounded immature extrusomes are associated with the endomembrane system, and consist of three layers: light-staining material between a dense staining and medium-staining layers (fig. 4.4F). The cell surface, including that of flagella, is covered in fine hair-like structures (fig. 4.4G, K). No double-membrane bound organelles were found.

4.3.3 rDNA phylogenies and environmental sequence survey

The SSU rDNA sequences of isolates MSW, FB11, and CuSP2-1 are longer than usual, at 2.1kb; MSW and FB11 are 95% identical to each other, and each is 89% indetical to CuSP2-1. A eukaryote-wide SSU rDNA phylogeny inferred under the GTR+Γ+I model does not reveal a well-supported placement of 'Protist X' anywhere in the tree (fig. 4.5). Notably, 'Protist X' does not go within

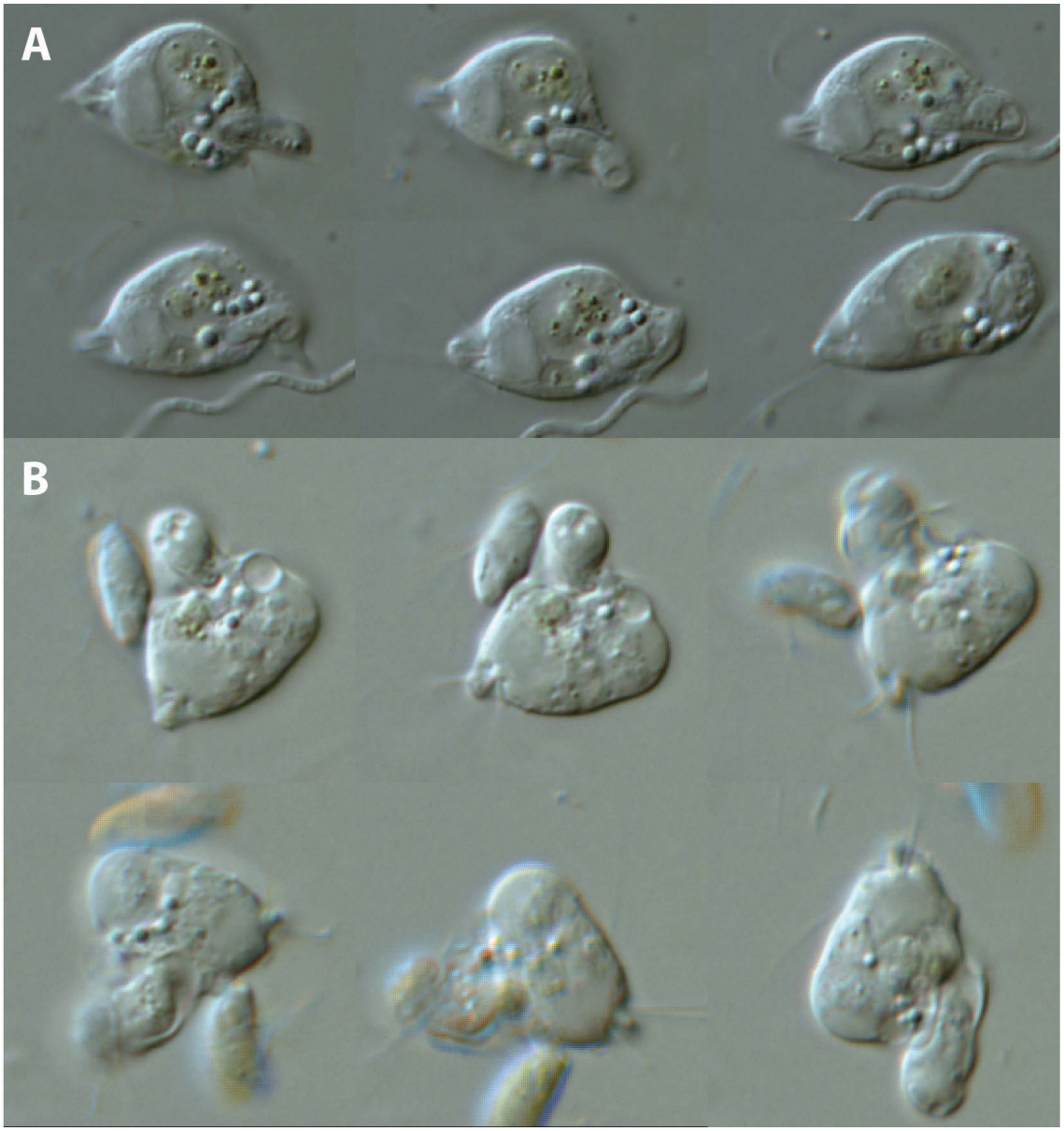


Figure 4.3: Image sequences of 'Protist X' feeding on flagellates. A-B) 'Protist X' cells from the same location as isolate FB11 but 2 years prior. A) Gradual ingestion of a flagellate from the posterior end of the cell. B) Another individual attempting to ingest a second cell as it phagocytoses the first. Cells were about 15 m long.

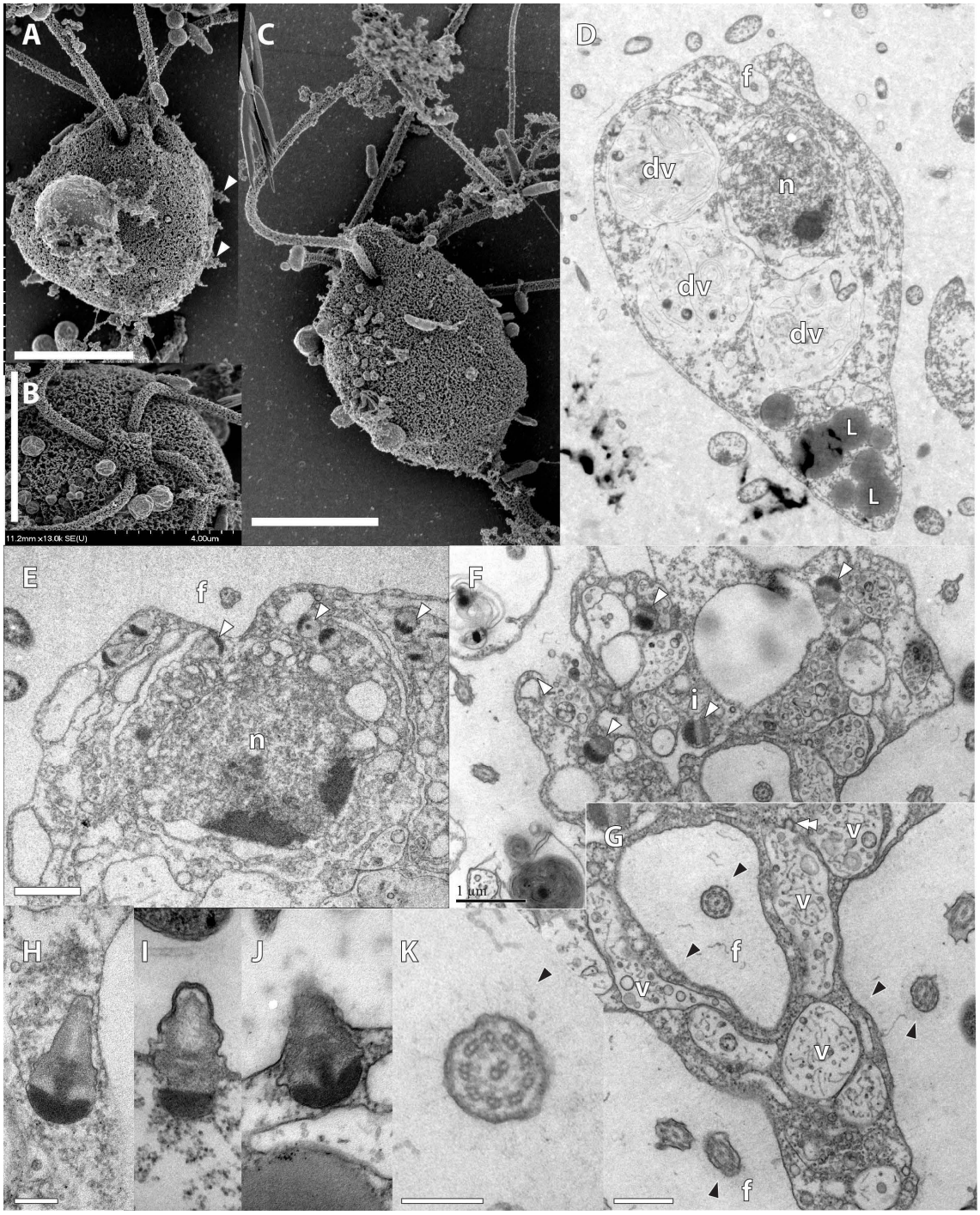


Figure 4.4: Ultrastructure of 'Protist X' isolate MSW by SEM and TEM. A-C) SEM images showing the surface and general structure of Protist X. Extrusomes (perhaps partially discharged) can be seen protruding from the surface (arrowheads). B) View of the apical end of the cell with the four flagella and their pockets. The entire cell surface including that of the flagella is covered in a fibrillar material (also visible in TEM images). D) Longitudinal section through a 'Protist X' cell. At the anterior end, a grazing section through a flagellum and its pocket (f) is position anterior to the nucleus (n). Three digestive vacuoles (dv) can be seen. Osmiophilic organelles, presumably lipid globules, are especially concentrated towards the posterior end of the cell. (Continues on the next page.)

any known anaerobic group with meaningful support. The placement of 'Protist X' (represented by isolates MSW and FB11) is likewise not resolved by the concatenated SSU+LSU phylogeny inferred under the same model (fig. 4.6).

As reported in Chapter 2, no sequences with significant affinity to 'Protist X' taxa were found in surveys of publicly available short read environmental sequence data, despite greedy search strategies.

4.3.4 Phylogenomics

A 68-taxon 254-gene phylogeny was inferred under the site heterogenous MAM60 mixture model (see: methods). Widely accepted major supergroups like Obazoa, Sar, Metamonada, and Discoba (Burki et al. 2020) were recovered with full support, as was a clade encompassing Hemimastigophora and Meteora (fig. 4.7, chapter 3). Surprisingly, Protist X does not branch with any previously described group of anaerobes characterised by molecular phylogenetics, nor, in fact, any established supergroup. Instead, it forms a robust, fully supported clade with Hemimastigophora and *Meteora* (100% PMSF bootstrap support). We refer to this as the "MHX" clade. 'Protist X' and Hemimastigophora form a clade to the exclusion of *Meteora* with 85% PMSF bootstrap support. Interestingly, PMSF support went up slightly for an association of *Ancoracysta* with Hemimastigophora, *Meteora*, and Protist X, compared to the dataset without 'Protist X' (Chapter 3).

4.4 Discussion

A handful of other flagellates also carry a cross-like arrangement of four anteriorly situated flagella. Of non-algal groups, most notable are *Collodictyon* (Orr et al. 2018), *Tetradimorpha*

Figure 4.4: (Continued from previous page) E) Close up of the anterior end of a longitudinal section. A population of extrusomes (white arrowheads) is visible near the surface, anterior to the nucleus. The cytoplasm is highly vacuolated. F) Cross section across the far anterior end of the cell. Extrusomes presumably in early stages of maturity can be seen inside the cytoplasm (white arrowheads; especially i). G) Close-up of the far anterior section showing three of the four flagella (f), one in its pocket. The fibrillar coating of the flagella and surface of the cell is discernible (black arrowheads). Vesiculated vacuoles (v) pack the cytoplasm in this region, one with an apparent coated vesicle budding outwards from its surface (double arrowhead). H-J) Higher magnification views of the mature extrusomes. One of the cells can be seen docked at the surface of the cell and protruding outwards (I). K) Close-up of a cross section through a flagellum, showing the axoneme and the fibrillar coating around the flagellum (black arrowhead). Scalebars: A-C) 5 m; D-F) 1 m; G) 500 nm; H-K) 200 nm

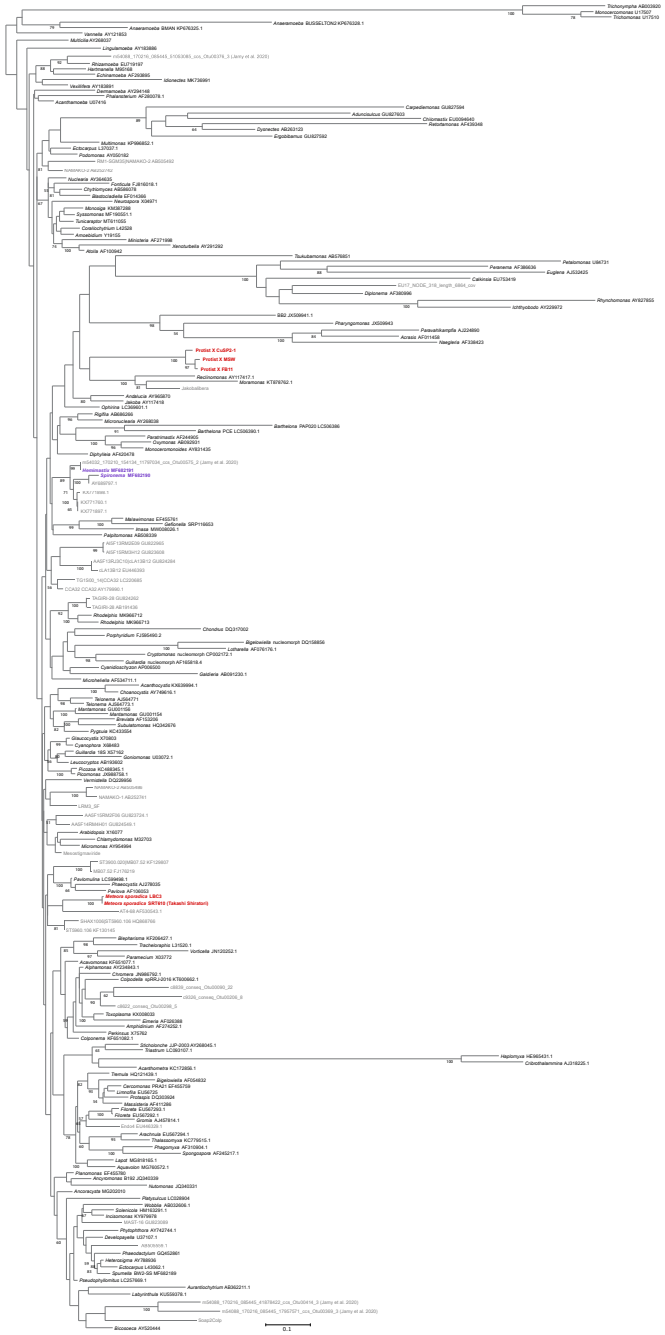


Figure 4.5: 'Protist X' in a global eukaryote SSU rDNA phylogeny. Inferred under the GTR+Γ+I model in RAxML-NG. See text for further details.

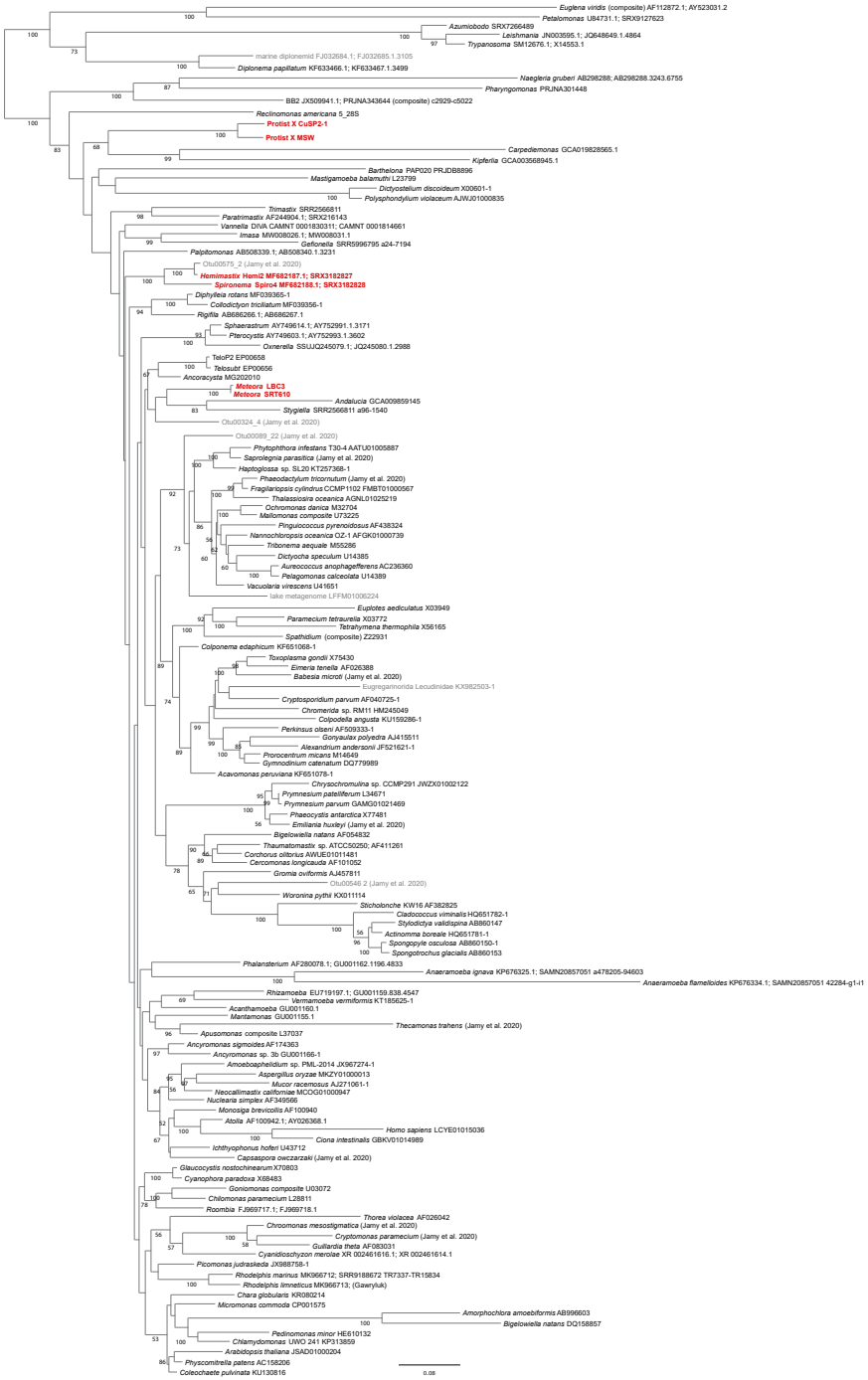


Figure 4.6: Concatenated eukaryote-wide SSU-LSU phylogeny inferred under the GTR+I+Γ model showing the position of Protist X. (Continued on the next page.)

(Brugerolle and Mignot 1983), and rarely-seen cruciform tetraflagellates of *Anaeramoeba* (Táborský et al. 2017). 'Protist X' lacks the axopodial extensions and prominent long basal bodies of *Tetradimorpha* (visible even by light microscopy); nor does it does not swim like *Collodictyon*. Assuming *Tetradimorpha* is closely related to *Dimorpha* (i.e. a rhizarian, see Chapter 2), 'Protist X' did not branch with any of these groups in neither rDNA nor protein phylogenies. Despite similar resemblance to the flagellates of Anaeramoeba—a metamonad (Stairs et al. 2021) and an anaerobe—'Protist X' did not branch near Anaeramoeba nor metamonads in general. Moreover, 'Protist X' does not share the extra-thick (up to 2 μm) flagella of *Anaeramoeba* flagellates (Táborský et al. 2017, fig 10).

'Protist X' is morphologically and behaviourally different from the other characterised members of the MHX clade. Currently described Hemimastigophora (W Foissner et al. 1988a, I Foissner and W Foissner 1993) are eukaryotrophs with two rows of several to a dozen or more flagella. They are predominantly freshwater and soil aerobes, though possible marine members were identified in analyses of environmental sequence databases (Lax et al. 2018), and a probable hemimastigophoran was recorded from marine anoxic sediments as "Protist Y" (Bernard et al. 2000). *Meteora* on the other hand, is a gliding aerobic bacterivore without flagella, but with a unique "rowing" morphology, and is found in marine sediments (Hausmann et al. 2002; chapter 3). 'Protist X' may represent the first characterised anaerobic lineage of this emerging diverse clade.

There is abundant precedent for a vast range of morphological and behavioural diversity within a group that lacks obvious synapomorphies, but is robustly defined by molecular phylogenetics. Rhizaria are a striking example of such a group, encompassing bacterivorous flagellates, parasites (haplosporidia, paramyxids, phytomyxids), eukaryotrophs (*Viridiraptor*, *Aurigamonas*), anaerobes (*Brevimastigomonas*; also chapter 2), and giant amoebae with reticulated pseudopods (radiolaria, foraminifera) (Grattepanche et al. 2018)—and even fungal-like osmotrophs (Feng et al. 2021)— and are well supported even in SSU rDNA phylogenies (Nikolaev et al. 2004). More recently, CRuMs was proposed based on phylogenomic analyses, and contains the bacterivorous gliding flagellate *Mantamonas*, eukaryotrophic flagellates like *Collodictyon*, and the tiny amoeboid Rigifilids (Brown et al. 2018).

Free-living anaerobic eukaryotes have evolved numerous times, in almost every supergroup

Figure 4.6: (Continued from previous page) See Materials and methods text for further details.

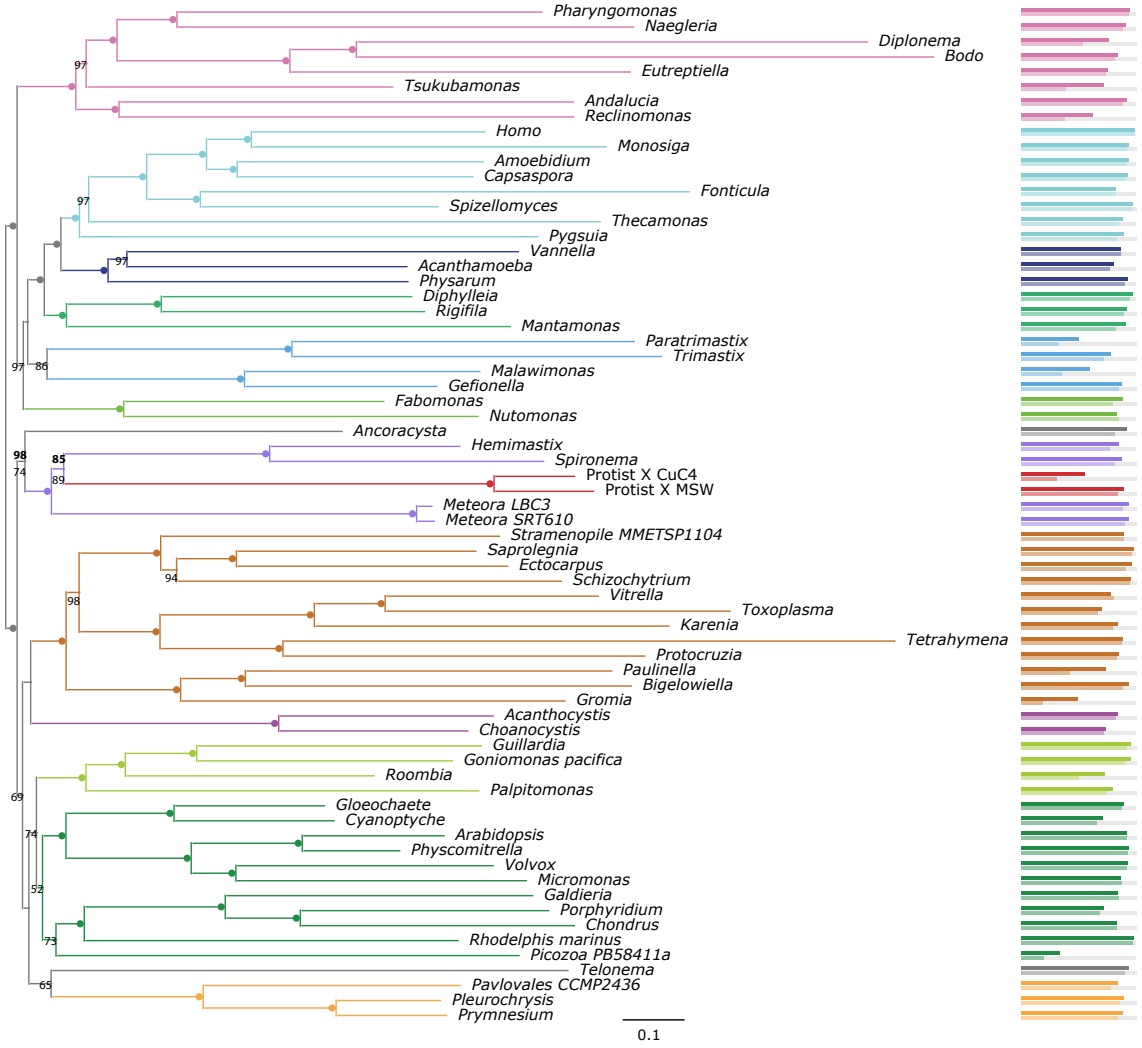


Figure 4.7: Position of 'Protist X' in a phylogenomic tree. 254-gene phylogeny inferred under the LG+MAM60+F+Γ model in IQTree. See Materials and methods text for further details. Support values from 1000 UFBOOT replicates, except among *Ancoracysta*, *Meteora*, Hemimastigophora, and Protist X, where the top value in bold is from 200 PMSF replicates (tree not shown).

(fig. 1.2) (Leger et al. 2019). In addition to expanding previously established anaerobic groups (eg. *Anaeramoeba* Stairs et al. 2021), researchers continue to discover new lineages descended from within aerobic clades, such as: breviates within Obazoa (Brown et al. 2013), *Brevimastigomonas* within Filosa (Gawryluk et al. 2019), *Rictus* (Yubuki et al. 2010), and *Cantina* (Yubuki et al. 2015b, Noguchi et al. 2015) both within Stramenopiles. These new lineages have greatly improved our understanding of the diversity of metabolic and genetic adaptations to low oxygen environments, as well as their evolutionary histories (Gawryluk et al. 2016c). Thus, it is perhaps not surprising that a phylogenetically deep and presumably highly diverse group would contain previously uncharacterised anaerobic lineages.

Anaerobic environments have been considered to host little predation activity, especially eukaryotrophy (eg. Fenchel 2012). Ciliates have been proposed as an exception (Fenchel 2012), and there are large anaerobic amoebae known to be detritivores and may be capable of ingesting eukaryotic prey. However, what is sparse are mentions of eukaryotrophs of a similar size to their prey in anaerobic environments. These eukaryotrophs are widely attested in other types of environments (eg. colpodellids and colponemids, Janouškovec et al. 2015). 'Protist X' represents one such anaerobic lineage, as it is in the same size category as its flagellate prey.

Anaerobic eukaryotes were previously predicted to almost exclusively prey on bacteria, with the exception of ciliates and large amoebae mentioned above; partly this was predicted based on the greater biomass of prokaryotes relative to eukaryotes observed in those environments (Fenchel 2012). Indeed, it is difficult to detect anaerobic eukaryotroph sequences in public environmental short read datasets (see Chapter 2). It was further argued that phagotrophy was unlikely to arise prior to global oxygenation (Fenchel 2012). The characterisation of Protist X along with several other anaerobic eukaryotrophs (see Chapter 2) calls this into question, and suggests anaerobic eukaryotrophy merits further investigation for ecological and evolutionary reasons alike.

4.4.1 Future directions

Whether 'Protist X' truly forms a robust clade with *Meteora* and Hemimastigophora remains to be tested. First of all, a series of downstream analyses similar to those done for *Meteora* (Chapter 3) would examine whether the association is robust within the given dataset. It would also be useful to test the association using alternative datasets sampling a different set of marker genes. Lastly, one should look for morphological synapomorphies in ultrastructure data, focusing in particular on comparing elements of the flagellar root apparatus between 'Protist X' and

Hemimastigophora.

'Protist X' represents an independent adaptation to anaerobiosis. The transcriptome data reported here are available for investigating the presence of anaerobiosis-related genes. Furthemore, isolate MSW is still in viable culture, and can thus be subjected to genome sequencing and further ultrastructural studies, such as examining whether there are shared features between its flagellar apparatus and those of hemimastigophora. Considering further molecular and electron microscopy data together, it would be interesting to determine whether 'Protist X' has a mitochondrion-related organelle (MRO). Lastly, the characterisation of this novel anaerobic eukaryotroph suggests they are an underexplored category that may continue to yield fruitful novel taxon discovery.

4.5 Acknowledgements

I would like to thank Tommy Harding for help with sequencing library preparation and Marlena Dlutek for help with sequencing of isolate MSW. Gordon Lax supervised single cell transcriptome generation and prepared the sequencing library for isolate CuSP2-1, as well as performing the SEM fixes of isolate MSW. I'd also like to thank Alex Worden and CIFAR for Curaçao sampling funding, and Mark Vermeij at CARMABI. Sebastian Hess provided helpful advice for how to improve our TEM fixation and resin infiltration protocol. Ping Li and Patricia Scallion supervise and run the TEM and SEM imagine core facilities, respectively. This work was funded by NSERC.

Chapter 5

Conclusion

Here, the search for novel protist taxa, primarily through cultivation and combined with morphological characerisation and molecular phylogenetics, has substantially enhanced our knowledge of eukaryotic diversity. In this section, I will discuss the role of culturing in contemporary protist discovery. Then I will outline a non-exhaustive set of future research avenues made accessible by the cultures established in this thesis.

Since the emergence of environmental molecular sequencing methods, there has been considerable emphasis on the concept of the "uncultured [or unculturable] majority" (Staley and Konopka 1985, Dawson and Fritz-Laylin 2009, Lewis et al. 2021, Marcy et al. 2007, Woyke et al. 2017), often stated to comprise 99% of microbes (e.g. Staley and Konopka 1985, Marcy et al. 2007). Additionally, there was an unspoken assumption that the set of organisms accessible to cultivation would fall fully within the set discoverable by cultivation-independent molecular approaches. In other words, given deep and extensive environmental sequencing, we would not discover any new significant taxa through cultivation. By contrast, most of the major novel lineages found in chapter 2 did not map onto any preexisting environmental clades (eg. *Meteora*, Protist X, QSI-PG, or the Retortacarps), and about half had no detectable 'candidate related OTUs' in environmental data. While a very large number of environmental sequences indeed remain without a representative that has been cultured—or even morphologically characterised—as it stands, cultivation-based approaches are providing access to a different pool of biodiversity than environmental sequencing.

Establishing polyeukaryotic cultures expands access to a pool of diversity that may be more

difficult to maintain in the lab in monoeukaryotic systems. By definition, obligate eukaryotrophs require other eukaryotes. Previously, it had been particularly difficult to examine polyeukaryotic systems using molecular sequencing approaches. Excitingly, recent advances in sequencing technology as well as bioinformatics have opened up groups that were previously considered intractable for molecular biology (and molecular cell biology), such as colponemids and colpodellids, Hemimastigophora, eukaryotrophic filosans, heterotrophic dinoflagellates. The pattern continues in this thesis through numerous examples, such "Protist X" or "Q" or "Pyriform Glider".

A similar pattern in apparent in bacterial and archaeal research (Lewis et al. 2020). For example, cultivation by gradually eliminating some 'excess' microbial diversity instead of starting out from a single isolate has enabled the capture of previously uncultured bacteria known only from sequence data (Luzan and Chistoserdov 2017). The remarkable cable bacteria (Geerlings et al. 2019) and a recently discovered centimeter-long prokaryotic giant Ca. *Thiomargarita magnifica* (Volland et al. 2022) were characterised through cultivation-based approaches, and the latter has not been found in metagenomic data. The previous expectations for monomicrobial bacterial cultures completely exclude entire ecological types of microbes, for example bacteria that are obligate predators of other bacteria. The recently discovered bacterivorous planctomycete Ca. *Uab* (Shiratori et al. 2019) may yield insights to previously underappreciated capabilities of bacterial membranes, and perhaps point to scenarios of early endomembrane evolution in eukaryotes through analogy; however, due to bacterial nomenclatural rules[1], an obligate predator like *Uab* cannot even be formally described, leading to an entire ecological niche of now culturable but still undescribable prokaryotes.

An emerging addition to pure environmental sequencing and pure cultivation methods are single cell molecular approaches, including those accompanied by morphological data. Single cell transcriptomics can yield phylogenomic marker gene data of sufficient quality to place a novel deep lineage of eukaryotes in the tree of life (Lax et al. 2018). This approach has also enabled harvesting complex phylogenomic data from many taxa within a specific group, for a fraction of the cost involved in culturing and bulk sequencing each of these organisms; the overhaul of phagotrophic euglenid phylogenetics (Lax and Simpson 2020 Lax et al. 2019) would be

[1]Governed by rule 31b of ICNB; previous editions disallowed naming based on mixed cultures (Lapage et al. 1992) but the latest edition appears to permit it provided the description is unambiguous (Parker et al. 2019).

incredibly expensive, in both cost and effort, without single cell methods, and perhaps impossible in places. Single cell genomics is rapidly improving as well and allows inferring ecological and physiological information not available in a ribosomal DNA fragment (e.g. Labarre et al. 2020). Single cell approaches can help tailor the cultivation techniques most likely to work for an organism of interest, *e.g.* by obtaining hints of trophic mode from the single cell genomic or transcriptomic data, or identifying potential contaminating prey sequences. Thus, the combination of traditional cultivation with single cell molecular approaches will further expand the eukaryotic diversity that is available and accessible to study. Needless to say, a certain set of microbial diversity will resist domestication outright, and, in addition, cultivation-independent approaches will continue to be invaluable for microbial ecology studies.

5.1 Further directions

The several dozen cultures established in this thesis provide access to many further avenues of study. The cultured representative isolates of the possible novel supergroup containing *Meteora*, hemimastigotes, and Protist X is rich in opportunity for further studies. First of all, two distinct but related cell motility behaviours of *Meteora sporadica* in particular attract attention. Gliding in protists is generally associated with flagella (Saito et al. 2003), or with an acto-myosin motility system like in diatoms (Poulsen et al. 1999) or some apicomplexans (Soldati-Favre 2008). Since *Meteora* is missing flagellar proteins (see Chapter 3), and is not directly related to any of the groups with well-studied acto-myosin motility systems, it would be fascinating to investigate the molecular mechanisms of its gliding behaviour. Moreover, the peculiar swinging motility of the lateral "arm" extensions has no immediate obvious analogue in any other eukaryotic cell type. Following the cytoskeletal systems involved may well reveal a novel molecular 'type' of cellular movement.

Protist X represents a novel, independently-derived anaerobic lineage. Double-membrane-bound organelles indicative of mitochondrial origin have not yet been found in TEM data (see Chapter 4). Investigating the anaerobic metabolism genes would at the very least reveal an independent evolutionary pathway to anaerobiosis. In addition to searching for the MRO, more

thorough ultrastructural studies could explore potential synapomorphies for the Hemimastigote-Protist X association in the flagellar root apparatus.

Besides 'Protist X', the other eukaryotes isolated from anaerobic environments in this thesis will yield interesting insights to the study of both the diversity of anaerobic metabolism as well as its evolutionary development. This will be a continuation of an existing trend, helping uncover an apparently complex set of evolutionary stories with horizontal as well as vertical inheritance (Stairs et al. 2018, Leger et al. 2019). Expanding the characterised diversity within previously known anaerobic clades (*i.e.*, shallow taxon sampling), such as through the additions of an atypical breviate relative BLO or of barthelonids and Retortacarps to the metamonads can improve the models of evolution of anaerobiosis in those groups. Newly characterised anaerobes emerging from within aerobic clades are of particular interest for seeking parallels in early adaptation (*e.g.* Gawryluk et al. 2016c, Stairs et al. 2014), thus QSI-PG and SSF each could make an excellent system in which to explore those questions. Lastly, perhaps the claims that eukaryotrophy might be energetically expensive for inhabitants of permanently anoxic environments (Fenchel 2012) can be investigated through quantitative feeding experiments (like Massana and Pedrós-Alió 1994) and growth curves compared to those for aerobic eukaryotrophs.

The newly cultivated isolates in this study can be examined for traces of retained ancestral characters within and between existing supergroups. The atypical breviate relative BLO raises questions about early Obazoa and the evolution of opisthokonts: the flagellar apparatus would be of interest there as BLO, unlike typical breviates, is a flagellate *sensu stricto* (i.e. lacks lateral filopodia altogether and does not glide—at least in the life stages seen so far). On a bigger scale, the broad distribution of colponemid-like cells across lineages of Diaphoretickes raises a tantalising question: are there shared features—for example, in the flagellar apparatus—that have been retained from a (possibly eukaryotrophic) last common ancestor of the group (or groups, depending on the position of the root of the eukaryote tree)? Or does the general morphology of a eukaryotroph with a large posterior ventral groove or depression and a smaller anterior one simply arise often enough through convergence?

Studies on eukaryote adaptation to high salt environments (Harding et al. 2016) can be further augmented by parallel investigations of adaptation to extreme alkalinity, especially in polyextremophiles. The 'retortacarp' TZLM1 RC appears to be adapted to an environment that should be hostile to eukaryotes on three counts: it is anoxic, hypersaline, and hyperalkaline. Additionally, the degree of these conditions varies considerably by season in TZLM1

RC's source habitat. Retortacarps were found in every alkaline sample examined in this study, and thus far no hint of them—neither morphological nor molecular—has been found in non-alkaline samples. The existing Retortacarp cultures are monoeukaryotic and grow to high densities; altogether, this highlights these organisms as potential excellent candidates for research on adaptation to extreme environments in microbial eukaryotes.

Cultivation remains an important toolkit for the exploration of protist diversity. Identification by morphology discernible under light microscopy is effective for detecting important lineages. These traditional "low throughput" methods are incredibly powerful in combination with current molecular technologies.

Appendix A

Additional tables

Table A.1: Sampling locations. (continues on the next page)

Sample code	Name	Sample type	Sample notes	Latitude	Longitude	salinity (ppt)	Sample credits	Date (approximate)
AuB	Bermagui, Aus	mangrove sediment	might be slightly brackish	36°25'31.7"S	150°03'15.6"E	marine	Alastair Simpson	January, 2016
BoP4.1	Bay of Pigs, Cuba	intertidal mudflat		22°11'52.0"N	81°08'19.7"W	marine	Claire Burnard	Feb, 2016
BW2	Bluff Wildnerness, Timberlea, NS	soil		44°39'39.7"N	63°46'02.9"W	freshwater		17 Apr, 2016
BWH	Bras d'Or, Baddeck, Cape Breton, NS	sediment		46°06'00.4"N	60°44'54.3"W	brackish	Kiran More	July, 2016
CARMGS	CARMABI, Piscadera baai mangrove sediment	mangrove sediment	anoxic	12°08'14.5"N	68°58'03.2"W	marine	Camille Poirier; funding: CIfAR (Alex Worden)	April, 2017
CuF1	Fort Beekenburg, Curaçao	hypersaline crust and sediment	salt pond salt crust	12°04'23.5"N	68°51'41.5"W	120	funding: CIfAR (Alex Worden); CARMABI and Mark Vermeij	April, 2017
CuSP2-1	Salina Sint Michiel, Curaçao	hypersaline anoxic sediment	salt pond sediment	12°09'25.2"N	68°59'16.2"W	100	funding: CIfAR (Alex Worden); CARMABI and Mark Vermeij	April, 2017
CuSP2-3	Salina Sint Michiel, Curaçao	hypersaline anoxic sediment	salt pond anoxic sulfidic sediment	12°09'18.4"N	68°59'13.1"W	80	funding: CIfAR (Alex Worden); CARMABI and Mark Vermeij	April, 2017
F2-2U	Fort Beekenburg, Curaçao	hypersaline sediment	salt pond sediment	12° 04' 23" N	68° 51' 41" W	120	Alastair Simpson	April, 2016
FB10N	False Bay, San Juan Is, WA	intertidal mud flat	anoxic sediment	48°29'07.9"N	123°04'31.1"W	marine		2015-06-03
FB11	False Bay, San Juan Is, WA	intertidal mud flat	anoxic sediment	48°29'10.2"N	123°04'32.7"W	marine		2015-06-03
GEM	Goodenough Lake Mat, BC	alkaline, microbial mat		51° 21' 54" N	121° 47' 41" W	50	Jackie Zorz	May, 2017
KSS	Kejimkujik Seaside Adjunct	high intertidal sediment	beach, pink pool	43°50'03.5"N	64°50'10.8"W	marine		September, 2017
LAR	Larnaka, Cyprus	subtidal sediment	beach	34°54'34.6"N	33°38'17.2"E	marine (38)		June, 2018
LBC3	Playa la Boca, Cuba	anoxic fine sand		21°36'07.9"N	77°05'14.9"W	marine	Claire Burnard	Feb, 2016
LIS	Luke Island Sediment, NS	ocean floor sediment		44°34'32.5"N	63°56'19.7"W	marine	Karen Filbee-Dexter	2016
LRM1	La Rapita, Catalonia, Spain; mud	subtidal sediment	beach/anoxic	40°35'33.2"N	0°42'37.0"E	marine	sampling access+permissions: Noèlia Carrasco	2015-02-27
LRM2	La Rapita, Catalonia, Spain; mud	mud, abandoned saltern	anoxic	40°38'25.6"N	0°44'26.1"E	marine	sampling access+permissions: Noèlia Carrasco	2015-02-27
LRM3	La Rapita, Catalonia, Spain; mud	mud, abandoned saltern	anoxic	40°38'22.9"N	0°44'26.2"E	marine	sampling access+permissions: Noèlia Carrasco	2015-02-26
LRS2	La Rapita, Catalonia, Spain; salt	saltern sediment		40° 35' 14" N	0° 41' 24" W		sampling access+permissions: Noèlia Carrasco	2015-02-27
MSW	McNabs Island, NS	intertidal mud flat/ bacterial mat (pink)	anoxic, mixed with decaying seaweed	44°36'20.4"N	63°31'18.3"W	marine		August, 2014
PCE	PEI Cavendish Eelgrass	subtidal sediment	anoxic	46°30'00.0"N	63°25'11.3"W	marine		July, 2016
PSL3	Playa Santa Lucia, Cuba		hypersaline	21° 34' 17" N	77° 03' 21" W		Claire Burnard	Feb, 2016
QSI	Small Inlet, Quadra Island, BC	marine sediment	rich in shellfish	50°15'43.1"N	125°16'17.5"W	marine	Gordon Lax	2016
Saa	Centennial beach or White Rock, BC	mud at low tide	top cm	49°00'N	123°01'W	marine	Greg Gavelis	late May, 2015

Table A.1: Sampling locations. (continued from the previous page)

Sample code	Name	Sample type	Sample notes	Latitude	Longitude	salinity (ppt)	Sample credits	Date (approximate)
SJB2	Playa Hulu, San Juan Baai, Curaçao	high energy exposed beach sand	beach	34°15'38.2"N	136°46'44.8"E	marine	funding: CIfAR (Alex Worden); CARMABI and Mark Vermeij	April, 2017
SnP	Splitnose Point, NS	ocean floor sediment		44°28'37.2"N	63°32'27.6"W	marine	John O'Brien	2016
Soap18	Soap Lake, WA	sediment	alkaline, anoxic	47°23'28.2"N	119°29'11.9"W	30	Ashley Smith	Jul, 2018
Soap20	Soap Lake, WA	sediment	alkaline, anoxic	47°23'28.2"N	119°29'11.9"W	20	Ashley Smith	Dec, 2020
TBB1	Tsawwassen (Boundary Bay)	intertidal mud flat	anoxic	49°00'27.1"N	123°01'52.7"W	marine	Claudio Slamovits	end of May, 2015
TZLM1	Lake Manyara, TZ	alkaline salt crust rehydrated in lake water	hypersaline alkaline anoxic	3°37'01.7"S	35°44'22.8"E	160	Z. Nathan Taylor	September, 2015
TZLM3	Lake Manyara, TZ	alkaline lake water, near hot spring	saline alkaline anoxic	3°37'01.7"S	35°44'22.8"E	5	Z. Nathan Taylor	September, 2015

Table A.2: TEM (Transmission Electron Microscopy) fixation methods (Chapter 2)

Organism	Fix ID	Fix Date	Collection method	Fix Type	Fixatives	Medium
ProtistX	C01	15.10.14	centrifugation	sim	2.5GA, 1OS	same medium
ProtistX	C05	24.03.15	centrifugation (2500g 10°C 15min)	sim	2.5%GA 1%Os	filter-sterilised native medium
ProtistX	C10	23.08.15	centrifugation	sim	1%Os, 5%GA	same medium
ProtistX	C11	24.08.15	centrifugation at lower speed	sim	1%Os, 5%GA	same medium
PCESSF	PS1	10.02.19		sim		same medium
QSIPG	QP1	08.05.19	transfer from bottom of tube	sim, 18h	3.5%GA, 1%Os, 20mM HEPES	
BoP4.1 Q	Q1	12.09.18		sim		same medium
SP2.3BLOL	BLOL1	08.11.17		seq	2.5%GA, 1%Os	SW
SP2.3BLOL	BLOL mar	11.12.17			0.5%OS, 1.25%GA	SW
CARMGS BLO	BLO2	04.10.17		seq	2.5%GA, 1%OS	SW
CARMGS BLO	BLO6	08.05.19	transfer from bottom of tube	sim, 18h	4%GA, 1%Os, 20mM HEPES	SW
TZLM1 RC	RC3	30.05.19	spun at max speed (bacteria removed at lower speed)	seq, 1h	5%GA in TZ50; 1%Os	
TZLM3 RCL	RCL3	04.10.17		seq	2.5%GA, 1%OS	
LAR MayoP	M01/MA	12.09.18				
LAR MayoP	M02	12.09.18				
LAR MayoP	M03	06.03.19			1.25% GA 1%Os	

Fix type: (sim)ultaneous, (seq)uential, near-(sim)ultaneous. Fixatives: (OS)mium tetroxide, (G)lutar(A)ldehyde, HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, TZ50 (50ppt alkaline medium "TZ").

Table A.3: SEM (Scanning Electron Microscopy) fixation methods (Chapter 2)

organism	fix	date	collection method	fix conditions	dehydration
LBC3Met	GL_Met1	07.02.18	cells settle for 30min-1h	1h GA, OS vapour; 3x wash in dH2O	EtOH series 7min steps, CPD 12 exchanges
LBC3Met	GL_Met2	07.02.18	cells settle for 30min-1h	30min GA then 30min OsO4 direct; 3x wash filt SW	EtOH series 7min steps, CPD 12 exchanges
LBC3Met	GL_Met3	21.02.18		1 drop 25% GA, Os vapour, RT; 3x wash in dH2O	EtOH series, CPD 12 exchanges
LBC3Met	GL_Met4	21.02.18		30min 1 drop 25% GA, THEN Os vapour 30min, RT; ...	EtOH series, CPD 12 exchanges
LBC3Met	GL_Met5	01.03.18	grown on slide?	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges
LBC3Met	GL_Met6	01.03.18	grown on slide?	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges
RotF	GL_RotF1	01.03.18	grown on slide?	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges
RotF	GL_RotF2	01.03.18	grown on slide?	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges
MSW3C	GL_MSW1	07.12.17	200-300µL onto polyLcoverslip	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges
MSW3C	GL_MSW2	07.12.17	200-300µL onto polyLcoverslip	1h 1 drop 25% GA, Os vapour, RT	EtOH series, CPD 12 exchanges

Appendix B

Additional phylogenetic trees

Figure B.1: Position of new sequences in a global eukaryotic SSU rDNA phylogeny. (Continued on the next page.)

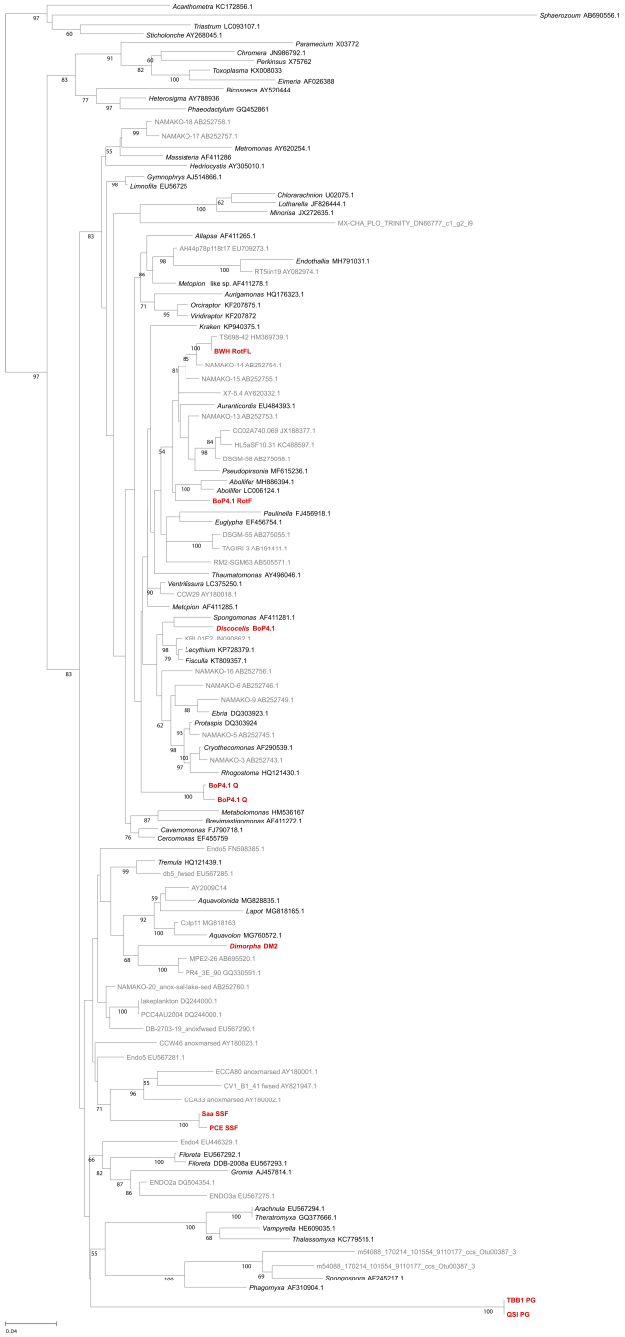


Figure B.2: Placement of novel taxa in a Rhizaria-focused SSU rDNA phylogeny. Maximum likelihood phylogeny of a 1192 site alignment inferred in RAxML-NG under the GTR+Γ+I model with support values from 200 non-parametric boostraps. See online supplement for bigger version of the tree.

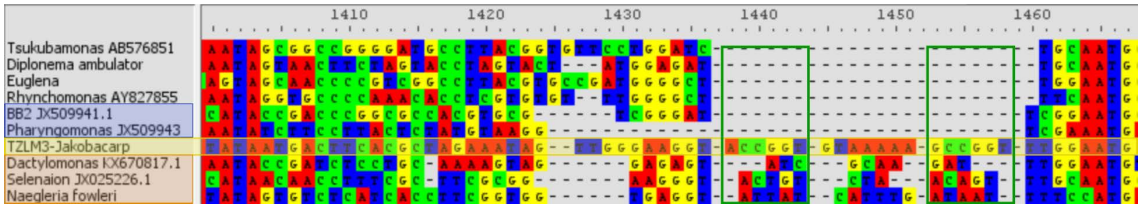


Figure B.3: Tetramitid synapomorphy SSU rDNA Helix 17_1 in 'Jakobacarp'. Tetramitid taxa highlighted in orange, other heterolobosea in blue, and 'Jakobacarp' sequence in yellow. Green boxes indicate palindromic stem sequences. Aligned and annotated in AliView (For further information on Helix 17_1, see Harding et al. 2013)

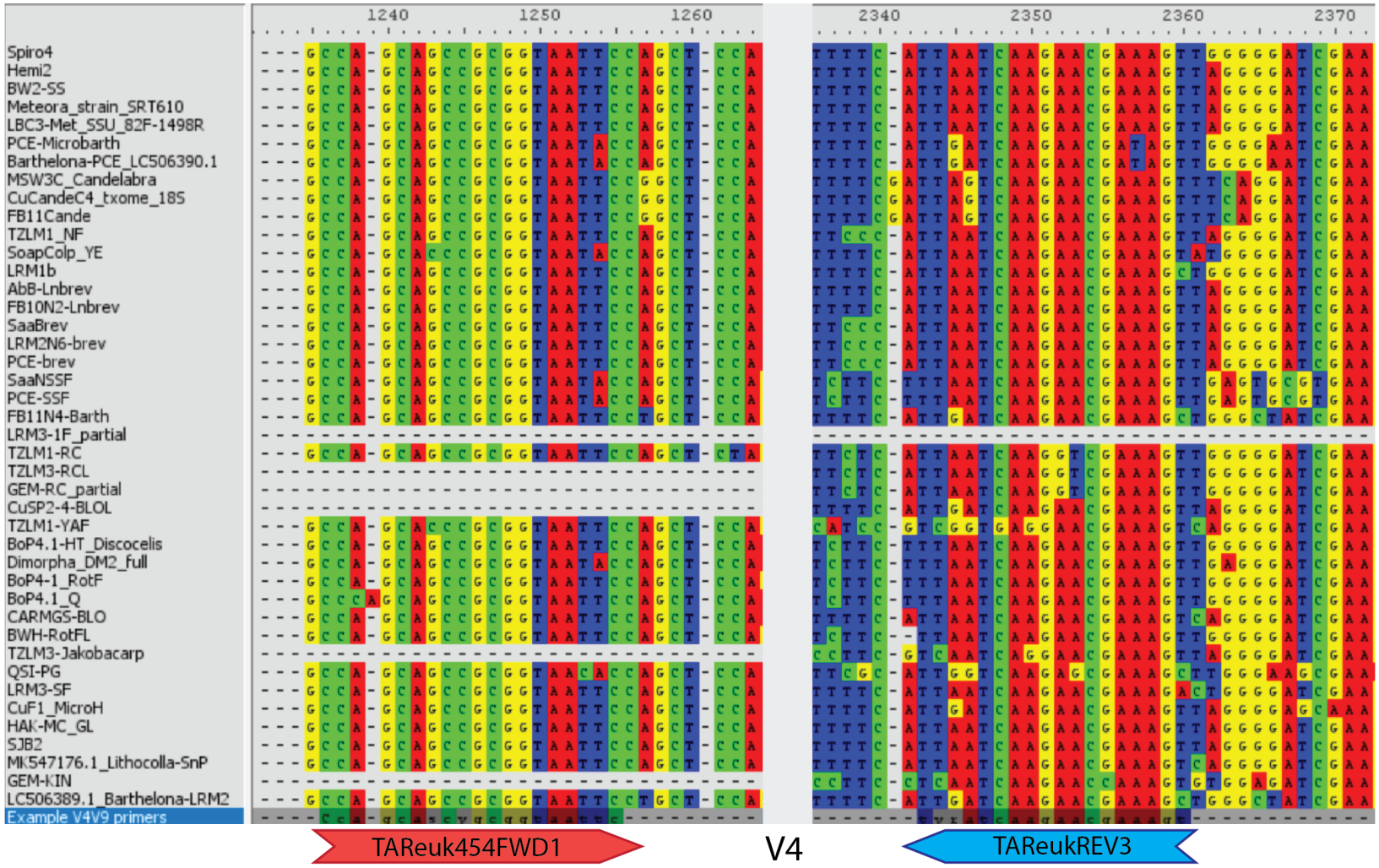


Figure B.4: SSU rDNA V4 primer binding region with example primer set. Example primers TAReuk454FWD1 and TAReukREV3 (Stoeck et al. 2010) with 5' ends of binding sites indicated by the direction of the arrow.

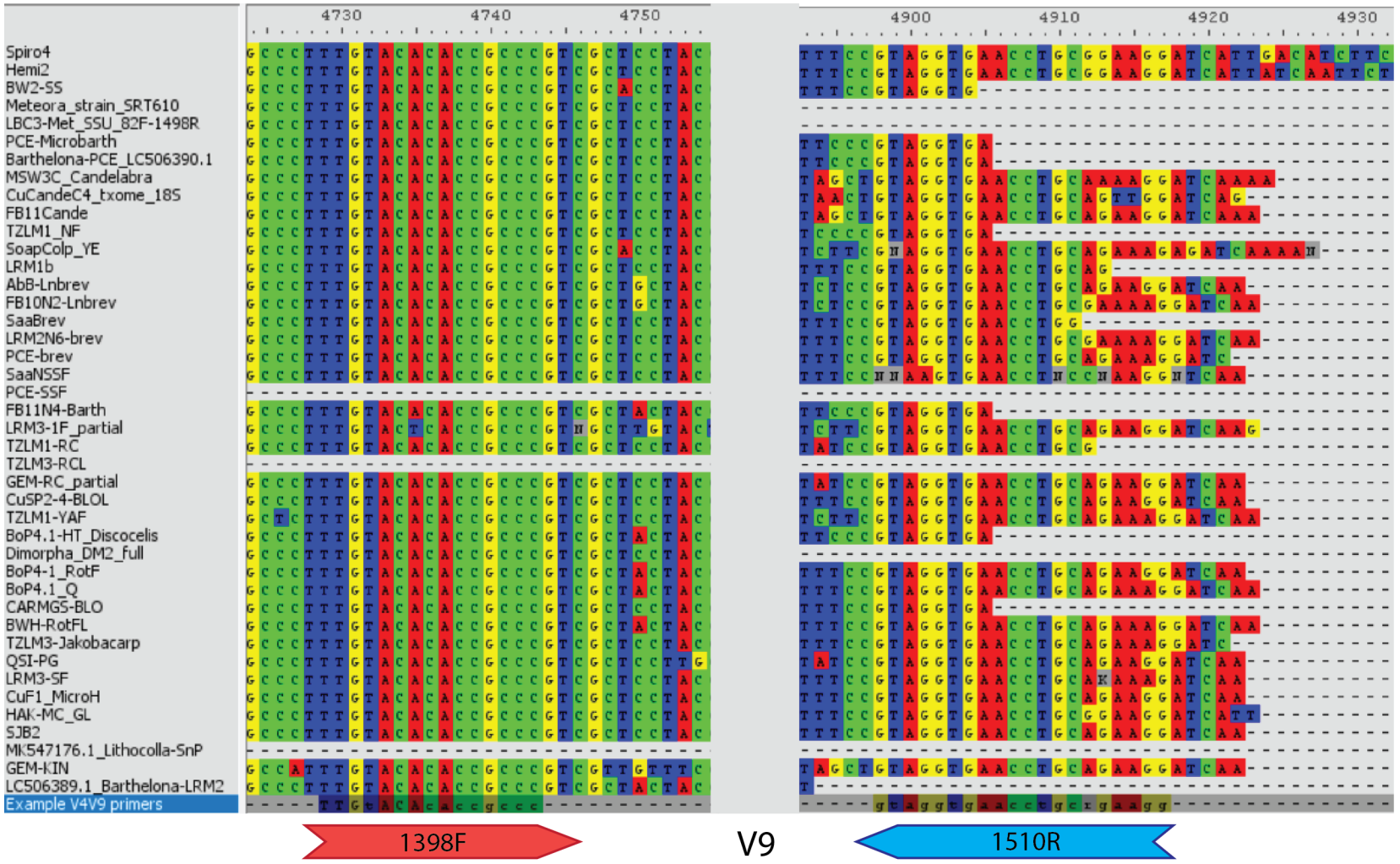


Figure B.5: SSU rDNA V9 primer binding region with example primer set. Example primers 1398F and 1510R (de Vargas et al. 2015a) with 5' ends of binding sites indicated by the direction of the arrow.

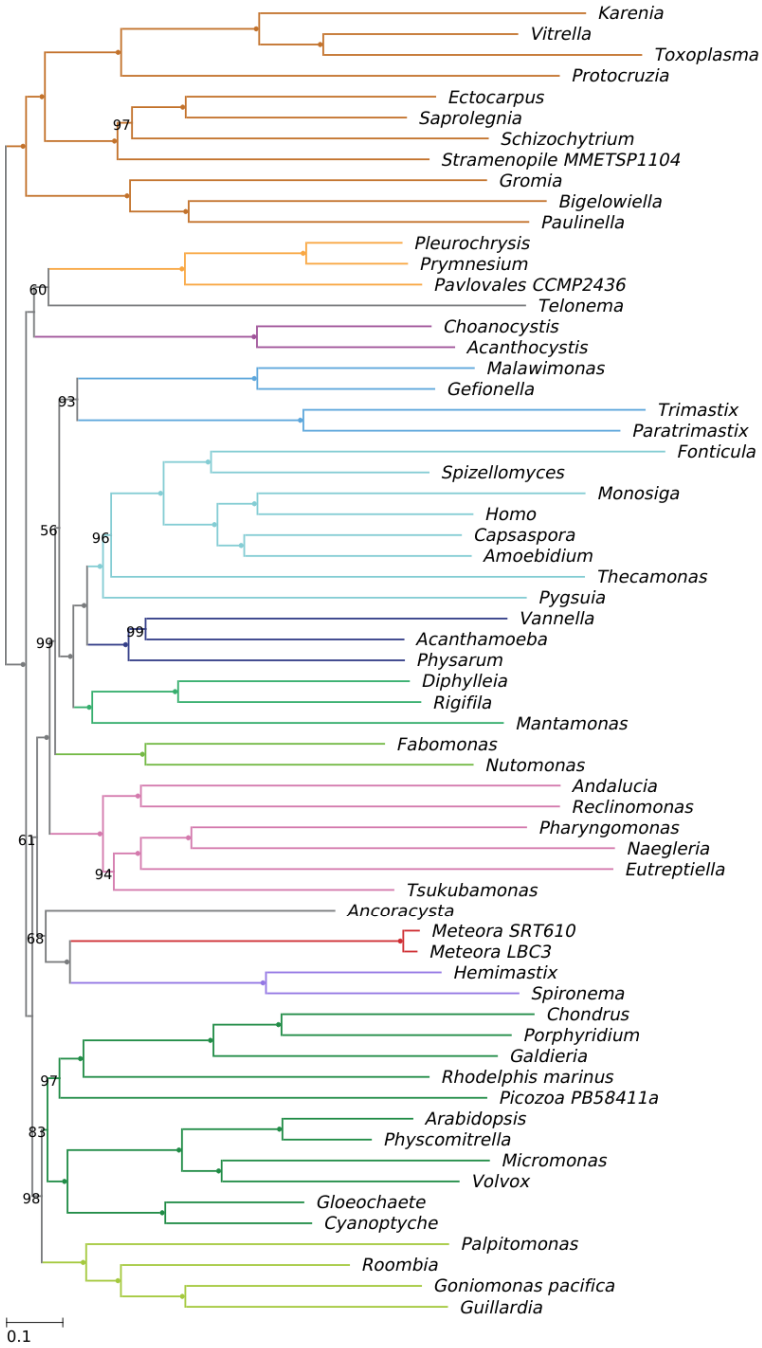


Figure B.6: Figure S4. No long-branching taxa (nLB) phylogeny, core 66 taxon dataset with three taxa removed, inferred from a concatenated 254-gene alignment under the LG+MAM60+Γ model, support values from 1000 UFBOOT replicates. Filled circles indicate full support.

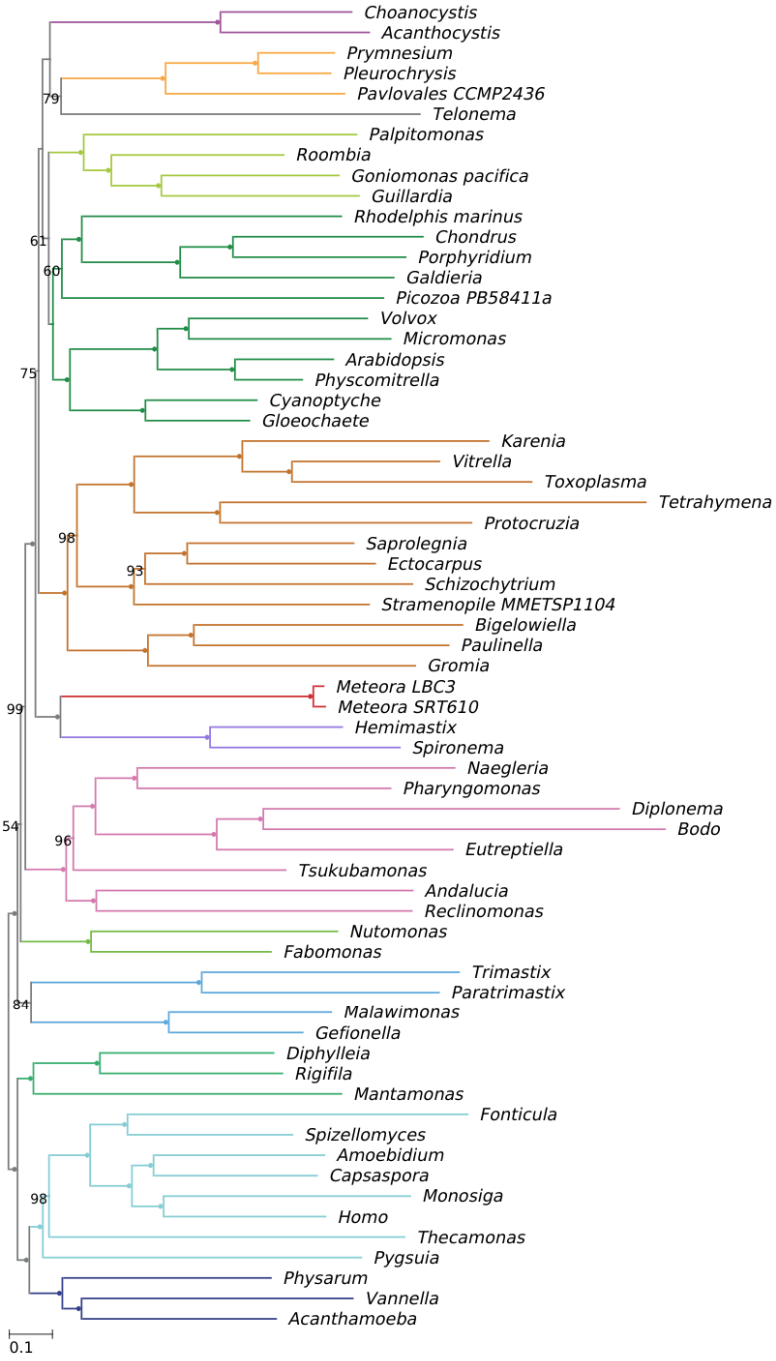


Figure B.7: MAM60 analysis with *Ancoracysta* removed (nAnco). No *Ancoracysta* (nAnco) phylogeny, core 66 taxon dataset with *Ancoracysta* removed, inferred from a concatenated 254-gene alignment under the LG+MAM60+Γ model, support values from 1000 UFBOOT replicates. Filled circles indicate full support.

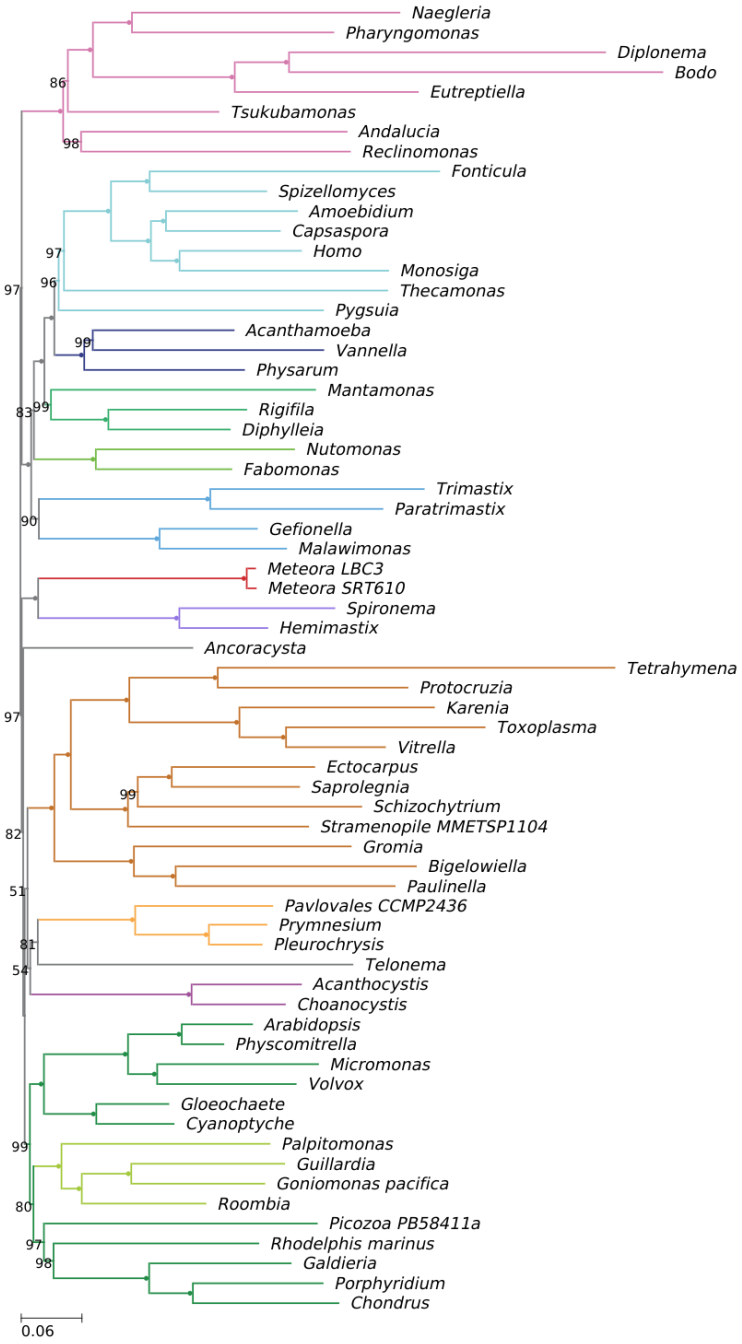


Figure B.8: SR4 Recoding analysis. Phylogeny inferred from SR4-recoded 254-gene alignment derived from core 66 taxon dataset, under the LG+MAM60+Γ model, support values from 1000 UFBOOT replicates. Filled circles indicate full support.

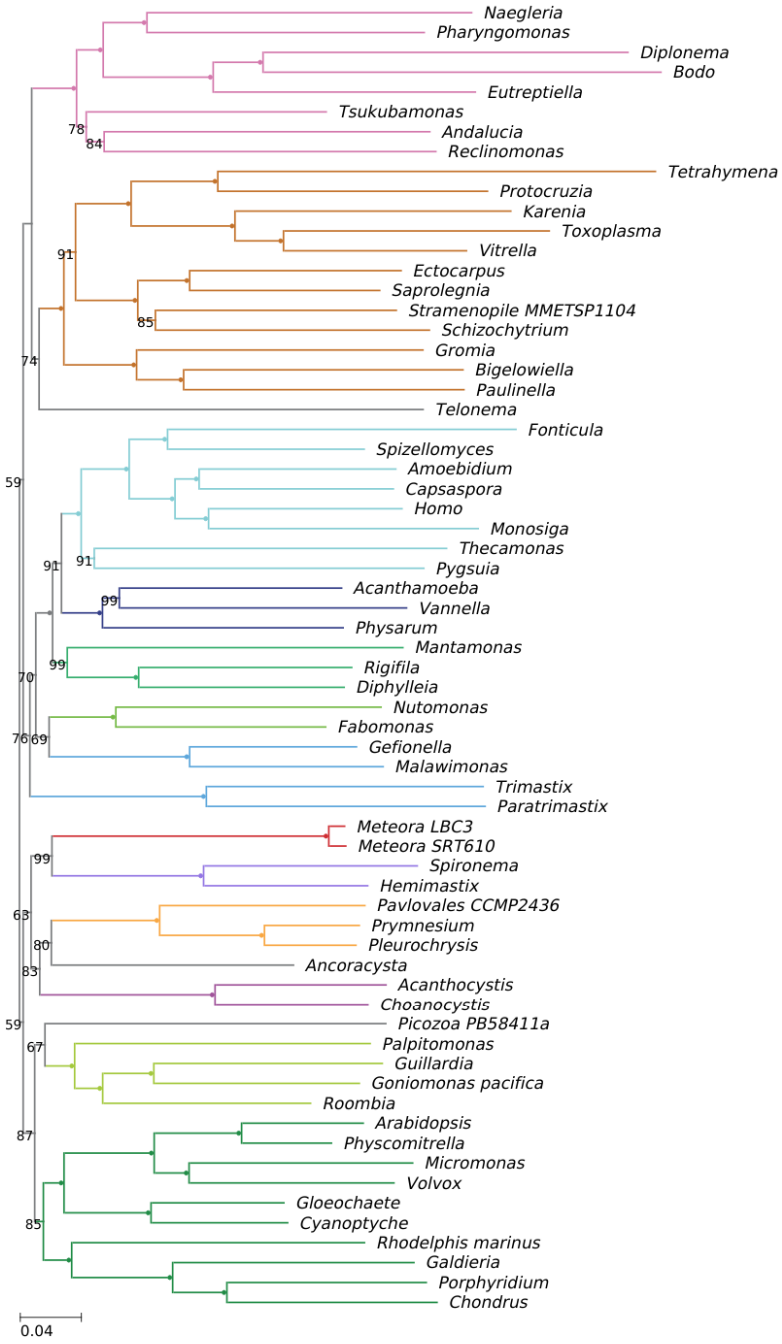


Figure B.9: `minmax-chisquared` recoding analysis. Phylogeny inferred from MinMax-Chisq-recoded 254-gene alignment derived from core 66 taxon dataset, under the LG+MAM60+Γ model, support values from 1000 UFBOOT replicates. Filled circles indicate full support.

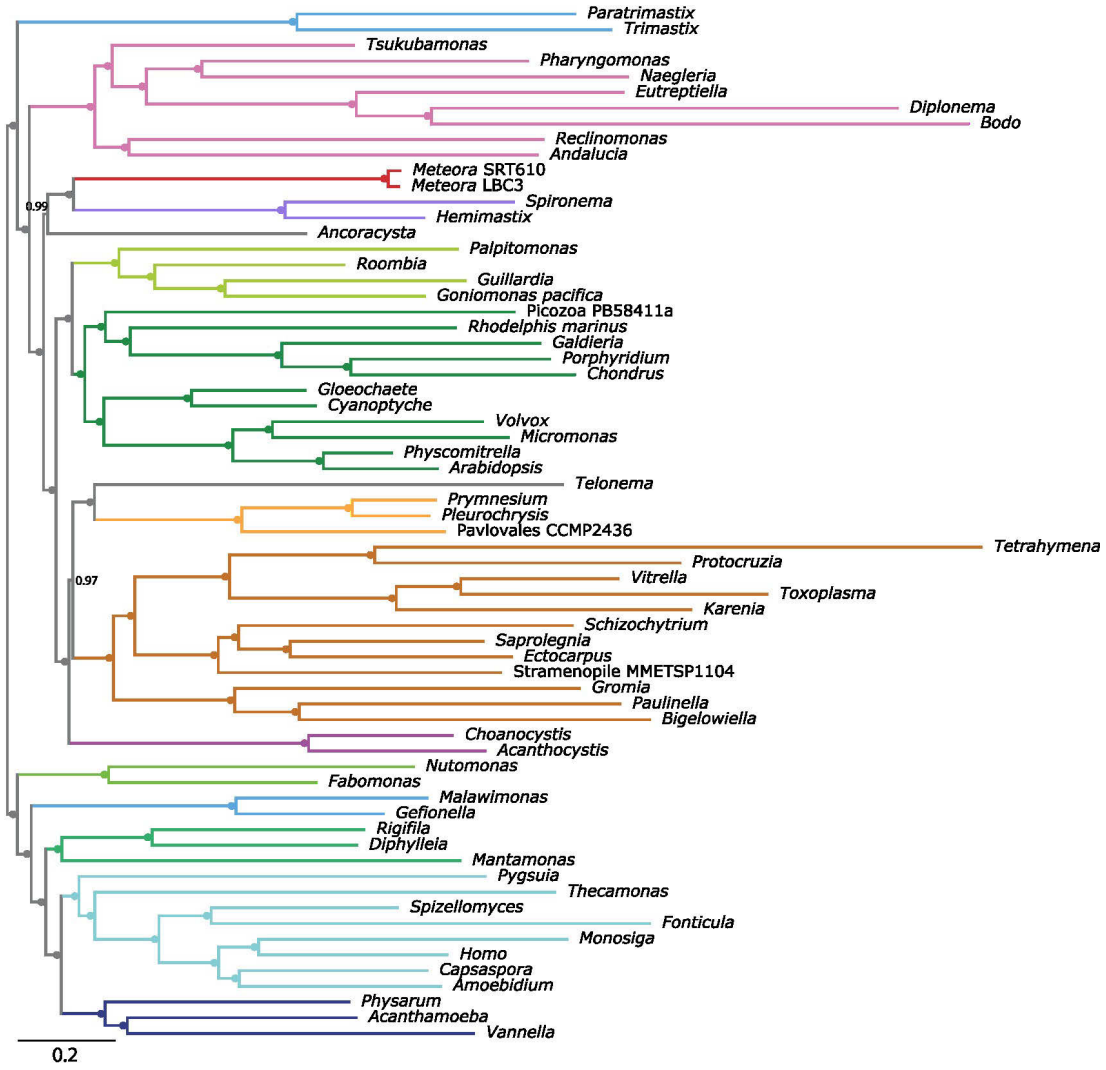


Figure B.10: *Meteora* PhyloBayes CAT+GTR, chains 2-4 (Chapter 3). 66-taxon PhyloBayes CAT+GTR consensus phylogeny of chains 2-4, following 1.1 x 104 cycles with a burnin of 500. Support values show posterior probabilities. Filled circles indicate full support.

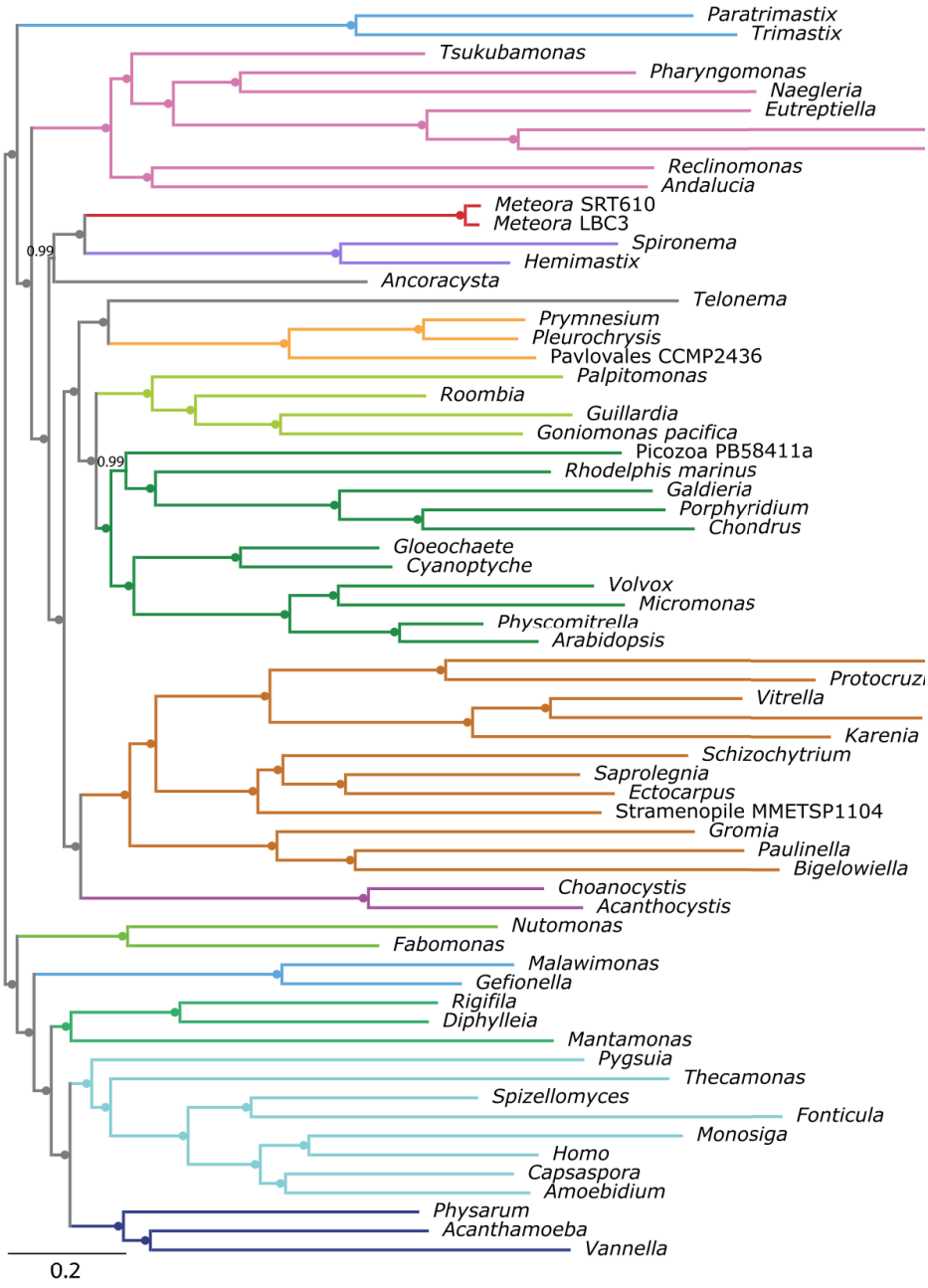


Figure B.11: *Meteora* PhyloBayes CAT+GTR, chain 1. Chapter 3) 66-taxon PhyloBayes CAT+GTR phylogeny of chain 1, following 1.1 x 104 cycles with a burnin of 500. Support values show posterior probabilities. Filled circles indicate full support.

Appendix C

Online Supplement

Table C.1: Environmental sequence matches for organisms isolated in Chapter 2. Tab-delimited text file. The prefix before the first vertical bar in `Target sequence id` is the code for the database the sequence originated from (see Table 2.3). `Pass reciprocal BLAST?` indicates whether the sequence matched the original query (SSU rDNAs obtained for the Chapter 2 isolates) as the top non-environmental hit via BLASTn search of the NCBI nucleotide database combined with Chapter 2 SSU rDNAs sequences.

Electronic supplement: Deposited on DalSpace

Bibliography

Adl SM, Habura A, Eglit Y (2014). "Amplification Primers of SSU rDNA for Soil Protists". *Soil Biol. Biochem.* 69, pp. 328–342.

Adoutte A, Balavoine G, Lartillot N, de Rosa R (1999). "Animal Evolution: The End of the Intermediate Taxa?" *Trends in Genetics* 15.3 (3), pp. 104–108.

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997). "Evidence for a Clade of Nematodes, Arthropods and Other Moulting Animals". *Nature* 387.6632 (6632), pp. 489–493.

Akıl C, Robinson RC (2018). "Genomes of Asgard Archaea Encode Profilins That Regulate Actin". *Nature* 562.7727 (7727), pp. 439–443.

Akıl C, Tran LT, Orhant-Prioux M, Baskaran Y, Manser E, Blanchoin L, Robinson RC (2020). "Insights into the Evolution of Regulated Actin Dynamics via Characterization of Primitive Gelsolin/Cofilin Proteins from Asgard Archaea". *PNAS* 117.33 (33), pp. 19904–19913.

Andrews S (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].*

Archibald JM, Simpson AG, Slamovits CH, eds. (2017). *Handbook of the Protists*. Cham: Springer International Publishing.

Bachy C, Hehenberger E, Ling Y.-C, Needham DM, Strauss J, Wilken S, Worden AZ (2022). "Marine Protists: A Hitchhiker's Guide to Their Role in the Marine Microbiome". *The Marine Microbiome*. Ed. by LJ Stal, MS Cretoiu. The Microbiomes of Humans, Animals, Plants, and the Environment. Cham: Springer International Publishing, pp. 159–241.

Balavoine G, Telford MJ (1995). "Identification of Planarian Homeobox Sequences Indicates the Antiquity of Most Hox/Homeotic Gene Subclasses". *PNAS* 92.16 (16), pp. 7227–7231.

Baldauf SL (2003). "The Deep Roots of Eukaryotes". *Science* 300.5626 (5626), pp. 1703–1706.

Bass D, Chao EE.-, Nikolaev S, Yabuki A, Ishida K.-i, Berney C, Pakzad U, Wylezich C, Cavalier-Smith T (2009). "Phylogeny of Novel Naked Filose and Reticulose Cercozoa: Granofilosea Cl. n. and Proteomyxidea Revised". *Protist* 160.1, pp. 75–109.

Bass D, Tikhonenkov DV, Foster R, Dyal P, Janouškovec J, Keeling PJ, Gardner M, Neuhauser S, Hartikainen H, Mylnikov AP, Berney C (2018). "Rhizarian 'Novel Clade 10' Revealed as Abundant and Diverse Planktonic and Terrestrial Flagellates, Including Aquavolon n. Gen." *J Eukaryot Microbiol* 65.6, pp. 828–842.

Benmeradi N, Payre B, Goodman SL (2015). "Easier and Safer Biological Staining: High Contrast UranyLess Staining of TEM Grids Using mPrep/g Capsules". *Microsc Microanal* 21.S3, pp. 721–722.

Berger S, Stamatakis A (2012). *PaPaRa 2 . 0 : A Vectorized Algorithm for Probabilistic Phylogeny-Aware Alignment Extension*. URL: https://cme.h-its.org/exelixis/pubs/Exelixis-RRDR-2012-5.pdf (visited on 06/22/2022).

Berger SA, Krompass D, Stamatakis A (2011). "Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood". *Systematic Biology* 60.3, pp. 291–302.

Berger SA, Stamatakis A (2011). "Aligning Short Reads to Reference Alignments and Trees". *Bioinformatics* 27.15, pp. 2068–2075.

Bernard C, Simpson AGB, Patterson DJ (2000). "Some Free-Living Flagellates (Protista) from Anoxic Habitats". *Ophelia* 52.2 (2), pp. 113–142.

Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M (2004). "Evolutionary Genomics of Nuclear Receptors: From Twenty-Five Ancestral Genes to Derived Endocrine Systems". *Mol Biol Evol* 21.10 (10), pp. 1923–1937.

Bolger AM, Lohse M, Usadel B (2014). "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinforma. Oxf. Engl.* 30.15 (15), pp. 2114–20.

Boros E, Kolpakova M (2018). "A Review of the Defining Chemical Properties of Soda Lakes and Pans: An Assessment on a Large Geographic Scale of Eurasian Inland Saline Surface Waters". *PLOS ONE* 13.8, e0202205.

Bråte J, Krabberød AK, Dolven JK, Ose RF, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K (2012). "Radiolaria Associated with Large Diversity of Marine Alveolates". *Protist* 163.5, pp. 767–777.

Brown MW, Heiss AA, Kamikawa R, Inagaki Y, Yabuki A, Tice AK, Shiratori T, Ishida K.-I, Hashimoto T, Simpson AGB, Roger AJ (2018). "Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group". *Genome Biology and Evolution* 10.2, pp. 427–433.

Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, Roger AJ (2013). "Phylogenomics Demonstrates That Breviate Flagellates Are Related to Opisthokonts and Apusomonads." *Proc. Biol. Sci.* 280.1769 (1769), p. 20131755.

Brugerolle G, Mignot J.-P (1983). "Caractéristiques Ultrastructurales de l'Hélioflagelle *Tetradimorpha* (Hsiung) et leur Intéret pour l'Etude Phyletique des Héliozoaires". *J. Protozool.* 30.3, pp. 473–480.

Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ (2013). "Phylogenomics of the Intracellular Parasite Mikrocytos Mackini Reveals Evidence for a Mitosome in Rhizaria". *Current Biology* 23.16 (16), pp. 1541–1547.

Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ (2016). "Untangling the Early Diversification of Eukaryotes: A Phylogenomic Study of the Evolutionary Origins of Centrohelida, Haptophyta and Cryptista". *Proc. R. Soc. B Biol. Sci.* 283.1823 (1823), pp. 20152802–20152802.

Burki F, Roger AJ, Brown MW, Simpson AGB (2020). "The New Tree of Eukaryotes". *Trends in Ecology & Evolution* 35.1 (1), pp. 43–55.

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019). "rnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data". *GigaScience* 8.9, giz100.

Cahill MA (2017). "The Evolutionary Appearance of Signaling Motifs in PGRMC1". *Biosci. Trends* 11.2 (2), pp. 179–192.

Cannon JT, Vellutini BC, Smith J, Ronquist F, Jondelius U, Hejnol A (2016). "Xenacoelomorpha Is the Sister Group to Nephrozoa". *Nature* 530.7588 (7588), pp. 89–93.

Castresana J (2000). "Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis". *Molecular Biology and Evolution* 17.4, pp. 540–552.

Cavalier-Smith T (2013). "Early Evolution of Eukaryote Feeding Modes, Cell Structural Diversity, and Classification of the Protozoan Phyla Loukozoa, Sulcozoa, and Choanozoa". *European Journal of Protistology* 49.2, pp. 115–178.

Cavalier-Smith T, Chao EE, Snell EA, Berney C, Fiore-Donno AM, Lewis R (2014). "Multigene Eukaryote Phylogeny Reveals the Likely Protozoan Ancestors of Opisthokonts (Animals, Fungi, Choanozoans) and Amoebozoa." *Mol. Phylogenet. Evol.* 81, pp. 71–85.

Cavalier-Smith T, Chao EE (2012). "Oxnerella Micra Sp. n. (Oxnerellidae Fam. n.), a Tiny Naked Centrohelid, and the Diversity and Evolution of Heliozoa". *Protist* 163.4, pp. 574–601.

Cavalier-Smith T, Chao EE.- (2003). "Phylogeny and Classification of Phylum Cercozoa (Protozoa)". *Protist* 154.3, pp. 341–358.

Cienkowski L (1881). "Отчет о беломорской экскурсии 1880 г. [An account on the White Sea excursion in 1880]". *Proc. St.-Petersb. Imp. Soc. Nat.* 12, pp. 130–171.

Corfield AP, Berry M (2015). "Glycan Variation and Evolution in the Eukaryotes". *Trends in Biochemical Sciences*. Special Issue: The Magic of the Sugar Code 40.7 (7), pp. 351–359.

Corsaro D, Michel R, Walochnik J, Venditti D, Müller K.-D, Hauröder B, Wylezich C (2016). "Molecular Identification of Nucleophaga Terricolae Sp. Nov. (Rozellomycota), and New Insights on the Origin of the Microsporidia". *Parasitol Res* 115.8, pp. 3003–3011.

Corsaro D, Walochnik J, Venditti D, Hauröder B, Michel R (2020). "Solving an Old Enigma: Morellospora Saccamoebae Gen. Nov., Sp. Nov. (Rozellomycota), a Sphaerita-like Parasite of Free-Living Amoebae". *Parasitol Res* 119.3, pp. 925–934.

Corsaro D, Walochnik J, Venditti D, Müller K.-D, Hauröder B, Michel R (2014). "Rediscovery of Nucleophaga Amoebae, a Novel Member of the Rozellomycota". *Parasitol Res* 113.12, pp. 4491–4498.

Criscuolo A, Gribaldo S (2010a). "BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments." *BMC Evol. Biol.* 10.1 (1), pp. 210–210.

Criscuolo A, Gribaldo S (2010b). "BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments." *BMC Evol. Biol.* 10.1 (1), p. 210.

Crisp MD, Cook LG (2005). "Do Early Branching Lineages Signify Ancestral Traits?" *Trends in Ecology & Evolution* 20.3 (3), pp. 122–128.

Dawson SC, Fritz-Laylin LK (2009). "Sequencing Free-Living Protists: The Case for Metagenomics". *Environ Microbiol* 11.7, pp. 1627–1631.

De Vargas C et al. (2015a). "Ocean Plankton. Eukaryotic Plankton Diversity in the Sunlit Ocean." *Science* 348.6237 (6237), pp. 1261605–1261605.

Dohm JC, Peters P, Stralis-Pavese N, Himmelbauer H (2020). "Benchmarking of Long-Read Correction Methods". *NAR Genomics and Bioinformatics* 2.2, lqaa037.

Edgar RC (2004). "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput". *Nucleic Acids Res* 32.5, pp. 1792–1797.

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, Suarez P, Varela R, Epstein S (2011). "Protistan Microbial Observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger Insights into Species Richness". *ISME J* 5.8 (8), pp. 1344–1356.

Ekebom J, Patterson DJ, Vørs N (1996). "Heterotrophic Flagellates from Coral Reef Sediments (Great Barrier Reef, Australia)". *Archiv für Protistenkunde* 146.3, pp. 251–272.

Embley TM (2006). "Multiple Secondary Origins of the Anaerobic Lifestyle in Eukaryotes". *Philosophical Transactions of the Royal Society B: Biological Sciences* 361.1470 (1470), pp. 1055–1067.

Epstein S, López-García P (2008). ""Missing" Protists: A Molecular Prospective". *Biodivers Conserv* 17.2, pp. 261–276.

Fenchel T (2012). "Anaerobic Eukaryotes". *Anoxia: Evidence for Eukaryote Survival and Paleontological Strategies*. Ed. by AV Altenbach, JM Bernhard, J Seckbach. Cellular Origin, Life in Extreme Habitats and Astrobiology. Dordrecht: Springer Netherlands, pp. 3–16.

Feng J.-J, He C.-Y, Jiang S.-H, Zhang T, Yu L.-Y (2021). "Saccharomycomorpha Psychra n. g., n. Sp., a Novel Member of Glissmonadida (Cercozoa) Isolated from Arctic and Antarctica". *J. Eukaryot. Microbiol.* 68.3, e12840.

Filker S, Forster D, Weinisch L, Mora-Ruiz M, González B, Farías ME, Rosselló-Móra R, Stoeck T (2017). "Transition Boundaries for Protistan Species Turnover in Hypersaline Waters of Different Biogeographic Regions". *Environ. Microbiol.* 19.8, pp. 3186–3200.

Filker S, Gimmler A, Dunthorn M, Mahé F, Stoeck T (2015). "Deep Sequencing Uncovers Protistan Plankton Diversity in the Portuguese Ria Formosa Solar Saltern Ponds". *Extremophiles* 19.2, pp. 283–295.

Finlay BJ (2002). "Global Dispersal of Free-Living Microbial Eukaryote Species". *Science* 296.5570, pp. 1061–1063.

Foissner I, Foissner W (1993). "Revision of the Family Spironemidae Doflein (Protista, Hemimastigophora), With Description of Two New Species, Spironema Terricola N. Sp. and Stereonema Geiseri N, G., N. Sp." *J. Eukaryot. Microbiol.* 40.4 (4), pp. 422–438.

Foissner W, Blatterer H, Foissner I (1988a). "The Hemimastigophora (Hemimastix Amphikineta Nov. Gen., Nov. Spec.), a New Protistan Phylum from Gondwanian Soils." *Eur. J. Protistol.* 23.4 (4), pp. 361–83.

Foissner W, Blatterer H, Foissner I (1988b). "The Hemimastigophora (Hemimastix Amphikineta Nov. Gen., Nov. Spec.), a New Protistan Phylum from Gondwanian Soils". *European Journal of Protistology* 23.4, pp. 361–383.

Forster D et al. (2016). "Benthic Protists: The under-Charted Majority". *FEMS Microbiology Ecology* 92.8, fiw120.

Friedman R, Hughes AL (2001). "Pattern and Timing of Gene Duplication in Animal Genomes". *Genome Res.* 11.11 (11), pp. 1842–1847.

Friedmann HC (2004). "From Butyribacterium to E. Coli : An Essay on Unity in Biochemistry". *Perspect. Biol. Med.* 47.1 (1), pp. 47–66.

Galindo LJ, Torruella G, Moreira D, Eglit Y, Simpson AGB, Völcker E, Clauß S, López-García P (2019). "Combined Cultivation and Single-Cell Approaches to the Phylogenomics of Nucleariid Amoebae, Close Relatives of Fungi". *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1786 (1786), p. 20190094.

Garey JR, Schmidt-Rhaesa A (1998). "The Essential Role of "Minor" Phyla in Molecular Studies of Animal Evolution". *Integr Comp Biol* 38.6 (6), pp. 907–917.

Gawryluk RMR, del Campo J, Okamoto N, Strassert JFH, Lukeš J, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2016a). "Morphological Identification and Single-Cell Genomics of Marine Diplonemids". *Current Biology* 26.22 (22), pp. 3053–3059.

— (2016b). "Morphological Identification and Single-Cell Genomics of Marine Diplonemids". *Current Biology* 26.22, pp. 3053–3059.

Gawryluk RMR, Kamikawa R, Stairs CW, Silberman JD, Brown MW, Roger AJ (2016c). "The Earliest Stages of Mitochondrial Adaptation to Low Oxygen Revealed in a Novel Rhizarian". *Current Biology* 26.20 (20), pp. 2729–2738.

Gawryluk RMR, Tikhonenkov DV, Hehenberger E, Husnik F, Mylnikov AP, Keeling PJ (2019). "Non-Photosynthetic Predators Are Sister to Red Algae". *Nature* 572.7768 (7768), pp. 240–243.

Geerlings NMJ, Zetsche E.-M, Hidalgo-Martinez S, Middelburg JJ, Meysman FJR (2019). "Mineral Formation Induced by Cable Bacteria Performing Long-Distance Electron Transport in Marine Sediments". *Biogeosciences* 16.3, pp. 811–829.

Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Urich T (2015). "Metatranscriptomic Census of Active Protists in Soils". *ISME J* 9.10 (10), pp. 2178–2190.

Glücksman E, Snell EA, Berney C, Chao EE, Bass D, Cavalier-Smith T (2011). "The Novel Marine Gliding Zooflagellate Genus Mantamonas (Mantamonadida Ord. n.: Apusozoa)." *Protist* 162.2 (2), pp. 207–21.

Gordetskaya O, Mesentsev Y, Kamyshatskaya O, Michel R, Walochnik J, Smirnov A, Nassonova E (2019). "Real-Time Observations on the Development of Intranuclear Parasite Nucleophaga Amoebae (Rozellomycota) in the Culture of Thecamoeba Quadrilineata". *Protistology* 4.

Gouy M, Guindon S, Gascuel O (2010a). "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Mol. Biol. Evol.* 27.2 (2), pp. 221–4.

— (2010b). "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Mol. Biol. Evol.* 27.2 (2), pp. 221–4.

Grattepanche J.-D, Walker LM, Ott BM, Paim Pinto DL, Delwiche CF, Lane CE, Katz LA (2018). "Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data". *BioEssays* 40.4, p. 1700198.

Haas BJ et al. (2013). "De Novo Transcript Sequence Reconstruction from RNA-seq Using the Trinity Platform for Reference Generation and Analysis." *Nat. Protoc.* 8.8 (8), pp. 1494–512.

Hanousková P, Táborský P, Čepička I (2019). "Dactylomonas Gen. Nov., a Novel Lineage of Heterolobosean Flagellates with Unique Ultrastructure, Closely Related to the Amoeba Selenaion Koniopes Park, De Jonckheere & Simpson, 2012". *J. Eukaryot. Microbiol.* 66.1, pp. 120–139.

Harding T, Brown MW, Plotnikov A, Selivanova E, Park JS, Gunderson JH, Baumgartner M, Silberman JD, Roger AJ, Simpson AGB (2013). "Amoeba Stages in the Deepest Branching Heteroloboseans, Including Pharyngomonas: Evolutionary and Systematic Implications". *Protist* 164.2, pp. 272–286.

Harding T, Brown MW, Simpson AG, Roger AJ (2016). "Osmoadaptative Strategy and Its Molecular Signature in Obligately Halophilic Heterotrophic Protists". *Genome Biology and Evolution* 8.7, pp. 2241–2258.

Harding T, Simpson AGB (2018). "Recent Advances in Halophilic Protozoa Research". *J. Eukaryot. Microbiol.* 65.4, pp. 556–570.

Hausmann K, Hülsmann N, Radek R (2003). "Protistology". *Protistology*.

Hausmann K, Weitere M, Wolf M, Arndt H (2002). "Meteora Sporadica Gen. Nov. et Sp. Nov. (Protista Incertae Sedis) – an Extraordinary Free-Living Protist from the Mediterranean Deep Sea". *European Journal of Protistology* 38.2 (2), pp. 171–177.

Havlena Z, Kieft TL, Veni G, Horrocks RD, Jones DS (2021). "Lighting Effects on the Development and Diversity of Photosynthetic Biofilm Communities in Carlsbad Cavern, New Mexico". *Appl Environ Microbiol* 87.6, e02695–20.

Heath TA, Hedtke S, Hillis D (2008). "Taxon Sampling and Accuracy of Phylogenetic Analyses". *Journal of Systematics and Evolution* 46.

Hehenberger E, Eitel M, Fortunato SAV, Miller DJ, Keeling PJ, Cahill MA (2020). "Early Eukaryotic Origins and Metazoan Elaboration of MAPR Family Proteins". *Molecular Phylogenetics and Evolution* 148, p. 106814.

Hehenberger E, Tikhonenkov DV, Kolisko M, del Campo J, Esaulov AS, Mylnikov AP, Keeling PJ (2017). "Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals". *Current Biology* 27.13 (13), 2043–2050.e6.

Hess S, Eme L, Roger AJ, Simpson AGB (2019). "A Natural Toroidal Microswimmer with a Rotary Eukaryotic Flagellum". *Nat Microbiol* 4.10 (10), pp. 1620–1626.

Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R (2011). "Capturing Diversity of Marine Heterotrophic Protists: One Cell at a Time". *ISME J* 5.4 (4), pp. 674–684.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018). "UFBoot2: Improving the Ultrafast Bootstrap Approximation". *Mol Biol Evol* 35.2, pp. 518–522.

Holton TA, Pisani D (2010). "Deep Genomic-Scale Analyses of the Metazoa Reject Coelomata: Evidence from Single- and Multigene Families Analyzed Under a Supertree and Supermatrix Paradigm". *Genome Biol Evol* 2, pp. 310–324.

Howe AT, Bass D, Vickerman K, Chao EE, Cavalier-Smith T (2009). "Phylogeny, Taxonomy, and Astounding Genetic Diversity of Glissomonadida Ord. Nov., the Dominant Gliding Zooflagellates in Soil (Protozoa: Cercozoa)." *Protist* 160.2 (2), pp. 159–89.

Hu SK, Herrera EL, Smith AR, Pachiadaki MG, Edgcomb VP, Sylva SP, Chan EW, Seewald JS, German CR, Huber JA (2021). "Protistan Grazing Impacts Microbial Communities and Carbon Cycling at Deep-Sea Hydrothermal Vents". *Proc. Natl. Acad. Sci.* 118.29, e2102674118.

Huerta-Cepas J, Serra F, Bork P (2016). "ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data". *Molecular Biology and Evolution* 33.6, pp. 1635–1638.

Hughes AL, Friedman R (2004a). "Differential Loss of Ancestral Gene Families as a Source of Genomic Divergence in Animals". *Proc. R. Soc. Lond. B* 271.suppl_3 (suppl_3).

— (2004b). "Differential Loss of Ancestral Gene Families as a Source of Genomic Divergence in Animals". *Proc. R. Soc. Lond. B* 271.suppl_3 (suppl_3).

Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM, Sogin ML (2014). "VAMPS: A Website for Visualization and Analysis of Microbial Population Structures". *BMC Bioinformatics* 15.1, p. 41.

Irwin NAT, Tikhonenkov DV, Hehenberger E, Mylnikov AP, Burki F, Keeling PJ (2019). "Phylogenomics Supports the Monophyly of the Cercozoa". *Molecular Phylogenetics and Evolution* 130, pp. 416–423.

Ishikawa H (1977). "Evolution of Ribosomal RNA". *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 58.1, pp. 1–7.

Jamy M, Biwer C, Vaulot D, Obiol A, Jing H, Peura S, Massana R, Burki F (2022). "Global Patterns and Rates of Habitat Transitions across the Eukaryotic Tree of Life". *Nat Ecol Evol*, pp. 1–13.

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2020a). "Long-Read Metabarcoding of the Eukaryotic rDNA Operon to Phylogenetically and Taxonomically Resolve Environmental Diversity". *Mol. Ecol. Resour.* 20.2, pp. 429–443.

— (2020b). "Long-Read Metabarcoding of the Eukaryotic rDNA Operon to Phylogenetically and Taxonomically Resolve Environmental Diversity". *Mol. Ecol. Resour.* 20.2 (2), pp. 429–443.

Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, Keeling PJ (2015). "Factors Mediating Plastid Dependency and the Origins of Parasitism in Apicomplexans and Their Close Relatives." *Proc. Natl. Acad. Sci. U. S. A.* 112.33 (33), pp. 10200–7.

Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Rohwer FL, Mylnikov AP, Keeling PJ (2017). "A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome Reduction". *Current Biology* 27.23 (23), 3717–3724.e5.

Jones BE, Grant WD, Duckworth AW, Owenson GG (1998). "Microbial Diversity of Soda Lakes". *Extremophiles* 2.3, pp. 191–200.

Karakashev D, Galabova D, Simeonov I (2003). "A Simple and Rapid Test for Differentiation of Aerobic from Anaerobic Bacteria". *World Journal of Microbiology and Biotechnology* 19.3, pp. 233–238.

Karnkowska A et al. (2016). "A Eukaryote without a Mitochondrial Organelle". *Current Biology* 26.10 (10), pp. 1274–1284.

Karpov SA (2000). "Ultrastructure of the Aloricate Bicosoecid Pseudobodo Tremulans, with Revision of the Order Bicosoecida". *Protistology* 1.3, pp. 101–109.

Katoh K, Standley DM (2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Mol. Biol. Evol.* 30.4 (4), pp. 772–80.

Katz LA, Grant J, Parfrey LW, Gant A, O'Kelly CJ, Anderson OR, Molestina RE, Nerad T (2011). "Subulatomonas Tetraspora Nov. Gen. Nov. Sp. Is a Member of a Previously Unrecognized Major Clade of Eukaryotes." *Protist* 162.5 (5), pp. 762–73.

Kaur B, Valach M, Peña-Diaz P, Moreira S, Keeling PJ, Burger G, Lukeš J, Faktorová D (2018). "Transformation of Diplonema Papillatum, the Type Species of the Highly Diverse and Abundant Marine Microeukaryotes Diplonemida (Euglenozoa)". *Environ. Microbiol.* 20.3, pp. 1030–1040.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012). "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinforma. Oxf. Engl.* 28.12 (12), pp. 1647–9.

Kim DH, Ausubel FM (2005). "Evolutionary Perspectives on Innate Immunity from the Study of Caenorhabditis Elegans". *Current Opinion in Immunology* 17.1 (1), pp. 4–10.

Kolisko M, Boscaro V, Burki F, Lynn DH, Keeling PJ (2014). "Single-Cell Transcriptomics for Microbial Eukaryotes". *Current Biology* 24.22 (22), R1081–R1082.

Kolisko M, Silberman JD, Cepicka I, Yubuki N, Takishita K, Yabuki A, Leander BS, Inouye I, Inagaki Y, Roger AJ, Simpson AGB (2010a). "A Wide Diversity of Previously Undetected Free-Living Relatives of Diplomonads Isolated from Marine/Saline Habitats". *Environ. Microbiol.* 12.10, pp. 2700–2710.

Kolisko M, Silberman JD, Čepička I, Yubuki N, Takishita K, Yabuki A, Leander BS, Inouye I, Inagaki Y, Roger AJ, Simpson AGB (2010b). "A Wide Diversity of Previously Undetected Free-Living Relatives of Diplomonads Isolated from Marine/Saline Habitats". *Environ. Microbiol.* 12.10 (10), pp. 2700–2710.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019). "RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference". *Bioinformatics* 35.21, pp. 4453–4455.

Krabberød AK, Orr RJS, Bråte J, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K (2017). "Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria". *Mol Biol Evol* 34.7 (7), pp. 1557–1573.

Kugrens P, Clay BL, Meyer CJ, Lee RE (1999). "Ultrastructure and Description of Cyanophora Biloba, Sp. Nov., with Additional Observations on C. Paradoxa (Glaucophyta)". *J. Phycol.* 35.4, pp. 844–854.

Labarre A, Obiol A, Wilken S, Forn I, Massana R (2020). "Expression of Genes Involved in Phagocytosis in Uncultured Heterotrophic Flagellates". *Limnol. Oceanogr.* 65.S1, S149–S160.

Lane N (2015). "The Unseen World: Reflections on Leeuwenhoek (1677) 'Concerning Little Animals'". *Philos Trans R Soc Lond B Biol Sci* 370.1666, p. 20140344.

Lanzén A, Simachew A, Gessesse A, Chmolowska D, Jonassen I, Øvreås L (2013). "Surprising Prokaryotic and Eukaryotic Diversity, Community Structure and Biogeography of Ethiopian Soda Lakes". *PLOS ONE* 8.8, e72577.

Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA (1992). *Rules of Nomenclature with Recommendations*. ASM Press.

Lara E, Moreira D, Vereshchaka A, López-García P (2009). "Pan-Oceanic Distribution of New Highly Diverse Clades of Deep-Sea Diplonemids". *Environ. Microbiol.* 11.1, pp. 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* 21.6, pp. 1095–1109.

Laumer CE, Bekkouche N, Kerbl A, Goetz F, Neves RC, Sørensen MV, Kristensen RM, Hejnol A, Dunn CW, Giribet G, Worsaae K (2015). "Spiralian Phylogeny Informs the Evolution of Microscopic Lineages". *Current Biology* 25.15 (15), pp. 2000–2006.

Lavier G (1935). "Sur une *Nucleophaga* parasite d' *Entamœba ranarum*". *Ann. Parasitol. Hum. Comp.* 13.4, pp. 351–361.

Lavrinienko A, Jernfors T, Koskimäki JJ, Pirttilä AM, Watts PC (2021). "Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities?" *Trends in Microbiology*. Special Issue: Microbial Genetics: Stress Management 29.1, pp. 19–27.

Lax G, Eglit Y, Eme L, Bertrand EM, Roger AJ, Simpson AGB (2018). "Hemimastigophora Is a Novel Supra-Kingdom-Level Lineage of Eukaryotes". *Nature* 564.7736 (7736), pp. 410–414.

Lax G, Lee WJ, Eglit Y, Simpson AGB (2019). "Ploeotids Represent Much of the Phylogenetic Diversity of Euglenids". *Protist* 170.2 (2), pp. 233–257.

Lax G, Simpson AGB (2020). "The Molecular Diversity of Phagotrophic Euglenids Examined Using Single-cell Methods". *Protist*, p. 125757.

Le SQ, Dang CC, Gascuel O (2012). "Modeling Protein Evolution with Several Amino Acid Replacement Matrices Depending on Site Rates". *Molecular Biology and Evolution* 29.10, pp. 2921–2936.

Lee WJ, Patterson DJ (2000). "Heterotrophic Flagellates (Protista) from Marine Sediments of Botany Bay, Australia". *J. Nat. Hist.* 34.4, pp. 483–562.

Lees-Miller JP, Cobban A, Katsonis P, Bacolla A, Tsutakawa SE, Hammel M, Meek K, Anderson DW, Lichtarge O, Tainer JA, Lees-Miller SP (2020). "Uncovering DNA-PKcs Ancient Phylogeny, Unique Sequence Motifs and Insights for Human Disease". *Progress in Biophysics and Molecular Biology*.

Leger MM, Kolísko M, Stairs CW, Simpson AGB (2019). "Mitochondrion-Related Organelles in Free-Living Protists". *Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes*. Ed. by J Tachezy. Microbiology Monographs. Cham: Springer International Publishing, pp. 287–308.

Leger MM et al. (2017). "Organelles That Illuminate the Origins of Trichomonas Hydrogenosomes and Giardia Mitosomes". *Nat. Ecol. Evol.* 1.4 (4), pp. 1–7.

Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG (2020). "Innovations to Culturing the Uncultured Microbial Majority". *Nat. Rev. Microbiol.*, pp. 1–16.

— (2021). "Innovations to Culturing the Uncultured Microbial Majority". *Nat Rev Microbiol* 19.4 (4), pp. 225–240.

López-García P, Philippe H, Gail F, Moreira D (2003). "Autochthonous Eukaryotic Diversity in Hydrothermal Sediment and Experimental Microcolonizers at the Mid-Atlantic Ridge". *Proc. Natl. Acad. Sci.* 100.2, pp. 697–702.

López-García P, Vereshchaka A, Moreira D (2007). "Eukaryotic Diversity Associated with Carbonates and Fluid–Seawater Interface in Lost City Hydrothermal Field". *Environ. Microbiol.* 9.2, pp. 546–554.

Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, Dai J, Koonin EV, Li G, Chu H, Li M (2020). "Coevolution of Eukaryote-like Vps4 and ESCRT-III Subunits in the Asgard Archaea". *mBio* 11.3 (3).

Luzan T, Chistoserdov A (2017). "Characterization of Enrichment Cultures Involved in the Carbon, Sulfur and Iron Biogeochemical Cycles in French Guiana Mobile Muds". *Geomicrobiol. J.* 34.1, pp. 11–26.

Mahé F et al. (2017). "Parasites Dominate Hyperdiverse Soil Protist Communities in Neotropical Rainforests". *Nat Ecol Evol* 1.4 (4), pp. 1–8.

Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR (2007). "Dissecting Biological "Dark Matter" with Single-Cell Genetic Analysis of Rare and Uncultivated TM7 Microbes from the Human Mouth". *Proc. Natl. Acad. Sci.* 104.29, pp. 11889–11894.

Massana R, Pedrós-Alió C (1994). "Role of Anaerobic Ciliates in Planktonic Food Webs: Abundance, Feeding, and Impact on Bacteria in the Field". *Appl. Environ. Microbiol.* 60.4, pp. 1325–1334.

Mesbah NM, Hedrick DB, Peacock AD, Rohde M, Wiegel J (2007). "Natranaerobius Thermophilus Gen. Nov., Sp. Nov., a Halophilic, Alkalithermophilic Bacterium from Soda Lakes of the Wadi An Natrun, Egypt, and Proposal of Natranaerobiaceae Fam. Nov. and Natranaerobiales Ord. Nov." *Int. J. Syst. Evol. Microbiol.* 57.Pt 11 (Pt 11), pp. 2507–12.

Mikrjukov KA (1995). "Structure, Function, and Formation of Extrusive Organelles ? Microtoxicysts in the rhizopodPenardia Cometa". *Protoplasma* 188.3-4 (3-4), pp. 186–191.

Mikrjukov KA, Mylnikov AP (1998). "The Fine Structure of a Carnivorous Multiflagellar Protist Multicilia Marina Cienkowski, 1881 (Flagellata Incertae Sedis)". *Eur. J. Protistol.* 34.4 (4), pp. 391–401.

Minh BQ, Hahn MW, Lanfear R (2020a). "New Methods to Calculate Concordance Factors for Phylogenomic Datasets". *Molecular Biology and Evolution* 37.9, pp. 2727–2733.

Minh BQ, Nguyen MAT, von Haeseler A (2013). "Ultrafast Approximation for Phylogenetic Bootstrap." *Mol. Biol. Evol.* 30.5 (5), pp. 1188–95.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020b). "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". *Molecular Biology and Evolution* 37.5, pp. 1530–1534.

Monteil CL et al. (2019). "Ectosymbiotic Bacteria at the Origin of Magnetoreception in a Marine Protist". *Nat. Microbiol.* 4.7 (7), pp. 1088–1095.

Moore RB, Oborník M, Janouskovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, Slapeta J, Hoegh-Guldberg O, Logsdon JM, Carter DA (2008). "A Photosynthetic Alveolate Closely Related to Apicomplexan Parasites." *Nature* 451.7181 (7181), pp. 959–63.

Nguyen L.-T, Schmidt HA, von Haeseler A, Minh BQ (2015). "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Mol. Biol. Evol.* 32.1 (1), pp. 268–74.

Nicholson AC, Malik S.-B, Logsdon JM, Van Meir EG (2005). "Functional Evolution of ADAMTS Genes: Evidence from Analyses of Phylogeny and Gene Organization". *BMC Evol Biol* 5.1 (1), p. 11.

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J (2004). "The Twilight of Heliozoa and Rise of Rhizaria, an Emerging Supergroup of Amoeboid Eukaryotes". *PNAS* 101.21 (21), pp. 8066–8071.

Nikolaev SI, Berney C, Petrov NB, Mylnikov AP, Fahrni JF, Pawlowski J (2006). "Phylogenetic Position of Multicilia Marina and the Evolution of Amoebozoa." *Int. J. Syst. Evol. Microbiol.* 56.Pt 6 (Pt 6), pp. 1449–58.

Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura K, Takishita K (2015). "Metabolic Capacity of Mitochondrion-related Organelles in the Free-living Anaerobic Stramenopile Cantina Marsupialis". *Protist* 166.5, pp. 534–550.

Ntakou E, Siemensma F, Bonkowski M, Dumack K (2019). "The Dancing Star: Reinvestigation of Artodiscus Saltans (Variosea, Amoebozoa) Penard 1890". *Protist* 170.4, pp. 349–357.

Obiol A, Giner CR, Sánchez P, Duarte CM, Acinas SG, Massana R (2020). "A Metagenomic Assessment of Microbial Eukaryotic Diversity in the Global Ocean". *Mol. Ecol. Resour.* 20.3 (3), pp. 718–731.

Ohama T, Kumazaki T, Hori H, Osawa S (1984). "Evolution of Multicellular Animals as Deduced from SS rRNA Sequences: A Possible Early Emergence of the Mesozoa". *Nucleic Acids Res* 12.12 (12), pp. 5101–5108.

Ollison GA, Hu SK, Hopper JV, Stewart BP, Smith J, Beatty JL, Rink LK, Caron DA (2022). "Daily Dynamics of Contrasting Spring Algal Blooms in Santa Monica Bay (Central Southern California Bight)". *Environ. Microbiol.* n/a.n/a.

Orr RJS, Zhao S, Klaveness D, Yabuki A, Ikeda K, Watanabe MM, Shalchian-Tabrizi K (2018). "Enigmatic Diphyllatea Eukaryotes: Culturing and Targeted PacBio RS Amplicon Sequencing Reveals a Higher Order Taxonomic Diversity and Global Distribution". *BMC Evolutionary Biology* 18.1, p. 115.

Orsi W, Song YC, Hallam S, Edgcomb V (2012). "Effect of Oxygen Minimum Zone Formation on Communities of Marine Protists". *ISME J* 6.8 (8), pp. 1586–1601.

Pánek T, Čepička I (2012). "Diversity of Heterolobosea". *Genetic Diversity in Microorganisms*. IntechOpen, p. 26.

Parada AE, Fuhrman JA (2017). "Marine Archaeal Dynamics and Interactions with the Microbial Community over 5 Years from Surface to Seafloor". *ISME J* 11.11 (11), pp. 2510–2525.

Park JS, Cho BC, Simpson AGB (2006). "Halocafeteria Seosinensis Gen. et Sp. Nov. (Bicosoecida), a Halophilic Bacterivorous Nanoflagellate Isolated from a Solar Saltern." *Extrem. Life Extreme Cond.* 10.6 (6), pp. 493–504.

Park JS (2012). "Effects of Different Ion Compositions on Growth of Obligately Halophilic Protozoan Halocafeteria Seosinensis." *Extrem. Life Extreme Cond.* 16.1 (1), pp. 161–4.

Park JS, De Jonckheere JF, Simpson AGB (2012). "Characterization of Selenaion Koniopes n. Gen., n. Sp., an Amoeba That Represents a New Major Lineage within Heterolobosea, Isolated from the Wieliczka Salt Mine". *J. Eukaryot. Microbiol.* 59.6, pp. 601–613.

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* 57.6, pp. 520–528.

Park JS, Simpson AGB (2015). "Diversity of Heterotrophic Protists from Extremely Hypersaline Habitats". *Protist* 166.4 (4), pp. 422–437.

Parker CT, Tidall BJ, Garrity GM (2019). "International Code of Nomenclature of Prokaryotes: Prokaryotic Code (2008 Revision)". *Int. J. Syst. Evol. Microbiol.* 69 (1A), S1–S111.

Pasulka A, Hu SK, Countway PD, Coyne KJ, Cary SC, Heidelberg KB, Caron DA (2019). "SSU-rRNA Gene Sequencing Survey of Benthic Microbial Eukaryotes from Guaymas Basin Hydrothermal Vent". *J. Eukaryot. Microbiol.* 66.4, pp. 637–653.

Patterson DJ, Simpson AGB (1996). "Heterotrophic Flagellates from Coastal Marine and Hypersaline Sediments in Western Australia". *European Journal of Protistology* 32.4, pp. 423–448.

Pawlowski J et al. (2012). "CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms". *PLOS Biology* 10.11, e1001419.

Peacock EE, Olson RJ, Sosik HM (2014). "Parasitic Infection of the Diatom Guinardia Delicatula, a Recurrent and Ecologically Important Phenomenon on the New England Shelf". *Mar. Ecol. Prog. Ser.* 503, pp. 1–10.

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014). "Full-Length RNA-seq from Single Cells Using Smart-seq2". *Nat Protoc* 9.1 (1), pp. 171–181.

Pilátová J, Tashyreva D, Týč J, Vancová M, Hussain Bokhari SN, Skoupý R, Klementová M, Küpper H, Mojzeš P, Lukeš J (2022). *Massive Accumulation of Strontium and Barium in Diplonemid Protists*. preprint. Microbiology.

Piredda R, Tomasino MP, D'Erchia AM, Manzari C, Pesole G, Montresor M, Kooistra WHCF, Sarno D, Zingone A (2017). "Diversity and Temporal Patterns of Planktonic Protist Assemblages at a Mediterranean Long Term Ecological Research Site". *FEMS Microbiol Ecol* 93.1, fiw200.

Pirlot S, Vanderheyden J, Descy J.-P, Servais P (2005). "Abundance and Biomass of Heterotrophic Microorganisms in Lake Tanganyika". *Freshw. Biol.* 50.7, pp. 1219–1232.

Plazzi F, Ferrucci RR, Passamonti M (2010). "Phylogenetic Representativeness: A New Method for Evaluating Taxon Sampling in Evolutionary Studies". *BMC Bioinformatics* 11.1 (1), p. 209.

Post FJ, Borowitzka LJ, Borowitzka MA, Mackay B, Moulton T (1983). "The Protozoa of a Western Australian Hypersaline Lagoon". *Hydrobiologia* 105.1, pp. 95–113.

Poulsen NC, Spector I, Spurck TP, Schultz TF, Wetherbee R (1999). "Diatom Gliding Is the Result of an Actin-Myosin Motility System". *Cell Motil Cytoskeleton* 44.1, pp. 23–33.

Prokina KI, Keeling PJ, Tikhonenkov DV (2021). "Heterotrophic Flagellates and Centrohelid Heliozoans from Marine Waters of Curacao, the Netherlands Antilles". *European Journal of Protistology* 77, p. 125758.

Prokopchuk G, Korytář T, Juricová V, Majstorović J, Horák A, Šimek K, Lukeš J (2022). "Trophic Flexibility of Marine Diplonemids - Switching from Osmotrophy to Bacterivory". *ISME J* 16.5 (5), pp. 1409–1419.

Rasband WS (1997). *ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA*.

Richter DJ, Fozouni P, Eisen MB, King N (2018). "Gene Family Innovation, Conservation and Loss on the Animal Stem Lineage". *eLife* 7. Ed. by MJ Telford, e34226.

Risse-Buhl U, Herrmann M, Lange P, Akob DM, Pizani N, Schönborn W, Totsche KU, Küsel K (2013). "Phagotrophic Protist Diversity in the Groundwater of a Karstified Aquifer – Morphological and Molecular Analysis". *J. Eukaryot. Microbiol.* 60.5, pp. 467–479.

Robicheau BM, Tolman J, Bertrand EM, LaRoche J (2022). "Highly-Resolved Interannual Phytoplankton Community Dynamics of the Coastal Northwest Atlantic". *ISME COMMUN.* 2.1 (1), pp. 1–12.

Roger AJ, Muñoz-Gómez SA, Kamikawa R (2017). "The Origin and Diversification of Mitochondria". *Current Biology* 27.21 (21), R1177–R1192.

Rosenberg MS, Kumar S (2001). "Incomplete Taxon Sampling Is Not a Problem for Phylogenetic Inference". *PNAS* 98.19 (19), pp. 10751–10756.

Ruinen J (1938). "Notizen Über Salzflagellaten. II. Über Die Verbreitung Der Salzflagellaten". *Arch Protistenkd* 90, pp. 210–258.

Rybarski AE, Nitsche F, Soo Park J, Filz P, Schmidt P, Kondo R, GB Simpson A, Arndt H (2021). "Revision of the Phylogeny of Placididea (Stramenopiles): Molecular and Morphological Diversity of Novel Placidid Protists from Extreme Aquatic Environments". *European Journal of Protistology* 81, p. 125809.

Saito A, Suetomo Y, Arikawa M, Omura G, Khan SMMK, Kakuta S, Suzaki E, Kataoka K, Suzaki T (2003). "Gliding Movement in Peranema Trichophorum Is Powered by Flagellar Surface Motility". *Cell Motil Cytoskeleton* 55.4, pp. 244–253.

Schagerl M, Burian A (2016). "The Ecology of African Soda Lakes: Driven by Variable and Extreme Conditions". *Soda Lakes of East Africa*. Ed. by M Schagerl. Cham: Springer International Publishing, pp. 295–320.

Schagerl M, Renaut RW (2016). "Dipping into the Soda Lakes of East Africa". *Soda Lakes of East Africa*. Ed. by M Schagerl. Cham: Springer International Publishing, pp. 3–24.

Schaudinn F (1899). "Untersuchungen Über Den Generationswechsel von Trichosphaerium Sieboldi Schn", p. 116.

Scheckenbach F, Hausmann K, Wylezich C, Weitere M, Arndt H (2010). "Large-Scale Patterns in Biodiversity of Microbial Eukaryotes from the Abyssal Sea Floor". *Proc. Natl. Acad. Sci.* 107.1, pp. 115–120.

Schiwitza S, Arndt H, Nitsche F (2018). "Four New Choanoflagellate Species from Extreme Saline Environments: Indication for Isolation-Driven Speciation Exemplified by Highly Adapted Craspedida from Salt Flats in the Atacama Desert (Northern Chile)". *European Journal of Protistology* 66, pp. 86–96.

Schneider CA, Rasband WS, Eliceiri KW (2012). "NIH Image to ImageJ: 25 Years of Image Analysis". *Nat Methods* 9.7 (7), pp. 671–675.

Schoenle A, Hohlfeld M, Hermanns K, Mahé F, de Vargas C, Nitsche F, Arndt H (2021). "High and Specific Diversity of Protists in the Deep-Sea Basins Dominated by Diplonemids, Kinetoplastids, Ciliates and Foraminiferans". *Commun Biol* 4.1 (1), pp. 1–10.

Schön ME, Zlatogursky VV, Singh RP, Poirier C, Wilken S, Mathur V, Strassert JFH, Pinhassi J, Worden AZ, Keeling PJ, Ettema TJG, Wideman JG, Burki F (2021). "Single Cell Genomics Reveals Plastid-Lacking Picozoa Are Close Relatives of Red Algae". *Nat Commun* 12.1 (1), p. 6651.

Shiratori T, Suzuki S, Kakizawa Y, Ishida K.-i (2019). "Phagocytosis-like Cell Engulfment by a Planctomycete Bacterium". *Nat. Commun.* 10.1 (1), p. 5529.

Shiratori T, Yokoyama A, Ishida K.-I (2014). "Phylogeny, Ultrastructure, and Flagellar Apparatus of a New Marimonad Flagellate Abollifer Globosa Sp. Nov. (Imbricatea, Cercozoa)". *Protist* 165.6, pp. 808–824.

Sidow A, Thomas WK (1994). "A Molecular Evolutionary Framework for Eukaryotic Model Organisms". *Current Biology* 4.7 (7), pp. 596–603.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015). "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs". *Bioinformatics* 31.19, pp. 3210–3212.

Simpson AGB, Perley TA, Lara E (2008). "Lateral Transfer of the Gene for a Widely Used Marker, Alpha-Tubulin, Indicated by a Multi-Protein Study of the Phylogenetic Position of Andalucia (Excavata)." *Mol. Phylogenet. Evol.* 47.1 (1), pp. 366–77.

Simpson AGB, Eglit Y (2016). "Protist Diversification". *Encyclopedia of Evolutionary Biology*.

Simpson AGB, Patterson DJ (1999). "The Ultrastructure of Carpediemonas Membranifera (Eukaryota) with Reference to the "Excavate Hypothesis"". *European Journal of Protistology* 35.4, pp. 353–370.

Soldati-Favre D (2008). "Molecular Dissection of Host Cell Invasion by the Apicomplexans: The Glideosome". *Parasite* 15.3 (3), pp. 197–205.

Sorokin DY, Banciu HL, Muyzer G (2015). "Functional Microbiology of Soda Lakes". *Current Opinion in Microbiology*. Environmental Microbiology • Extremophiles 25, pp. 88–96.

Sorokin DY, Foti M, Pinkart HC, Muyzer G (2007). "Sulfur-Oxidizing Bacteria in Soap Lake (Washington State), a Meromictic, Haloalkaline Lake with an Unprecedented High Sulfide Content". *Appl Environ Microbiol* 73.2, pp. 451–455.

Spang A, Eme L, Saw JH, Caceres EF, Zaremba-Niedzwiedzka K, Lombard J, Guy L, Ettema TJG (2018). "Asgard Archaea Are the Closest Prokaryotic Relatives of Eukaryotes". *PLOS Genetics* 14.3 (3), e1007080.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG (2015). "Complex Archaea That Bridge the Gap between Prokaryotes and Eukaryotes". *Nature* 521.7551 (7551), pp. 173–179.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M, Roger AJ (2014). "A SUF Fe-S Cluster Biogenesis System in the Mitochondrion-Related Organelles of the Anaerobic Protist Pygsuia." *Curr. Biol. CB* 24.11 (11), pp. 1176–86.

Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, Fawcett JP, Roger AJ (2018). "Microbial Eukaryotes Have Adapted to Hypoxia by Horizontal Acquisitions of a Gene Involved in Rhodoquinone Biosynthesis". *eLife* 7. Ed. by J McCutcheon, e34292.

Stairs CW, Táborský P, Salomaki ED, Kolisko M, Pánek T, Eme L, Hradilová M, Vlček Č, Jerlström-Hultqvist J, Roger AJ, Čepička I (2021). "Anaeramoebae Are a Divergent Lineage of Eukaryotes That Shed Light on the Transition from Anaerobic Mitochondria to Hydrogenosomes". *Current Biology* 31.24, 5605–5612.e5.

Staley JT, Konopka A (1985). "Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats". *Annu Rev Microbiol* 39, pp. 321–346.

Stamatakis A (2014). "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinforma. Oxf. Engl.* 30.9 (9), pp. 1312–3.

Stock A, Breiner H.-W, Pachiadaki M, Edgcomb V, Filker S, La Cono V, Yakimov MM, Stoeck T (2012). "Microbial Eukaryote Life in the New Hypersaline Deep-Sea Basin Thetis". *Extremophiles* 16.1, pp. 21–34.

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H.-W, Richards TA (2010). "Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water". *Mol. Ecol.* 19.s1, pp. 21–31.

Stoeck T, Hayward B, Taylor GT, Varela R, Epstein SS (2006). "A Multiple PCR-primer Approach to Access the Microeukaryotic Diversity in Environmental Samples". *Protist* 157.1, pp. 31–43.

Stoeck T, Taylor GT, Epstein SS (2003). "Novel Eukaryotes from the Permanently Anoxic Cariaco Basin (Caribbean Sea)". *Appl. Environ. Microbiol.* 69.9, pp. 5656–5663.

Strassert JFH, Jamy M, Mylnikov AP, Tikhonenkov DV, Burki F (2019). "New Phylogenomic Analysis of the Enigmatic Phylum Telonemia Further Resolves the Eukaryote Tree of Life". *Molecular Biology and Evolution* 36.4, pp. 757–765.

Strassert JFH, Karnkowska A, Hehenberger E, del Campo J, Kolisko M, Okamoto N, Burki F, Janouškovec J, Poirier C, Leonard G, Hallam SJ, Richards TA, Worden AZ, Santoro AE, Keeling PJ (2018). "Single Cell Genomics of Uncultured Marine Alveolates Shows Paraphyly of Basal Dinoflagellates". *ISME J.* 12.1 (1), pp. 304–308.

Susko E (2022). *MAMMaL: (M)Ultinomial (A)Pproximate (M)Ixture (Ma)Ximum (L)Ikelihood Accelerated Estimation of Frequency Classes in Site-heterogeneous Profile Mixture Models Version 1.1.3*.

— (2021). *Minmax-Chisq: Obtaining a Reduced Amino Acid Alphabet with More Homogeneous Composition, Version 1.1*.

Susko E, Lincker L, Roger AJ (2018). "Accelerated Estimation of Frequency Classes in Site-Heterogeneous Profile Mixture Models". *Molecular Biology and Evolution* 35.5, pp. 1266–1283.

Susko E, Roger AJ (2007). "On Reduced Amino Acid Alphabets for Phylogenetic Inference". *Molecular Biology and Evolution* 24.9, pp. 2139–2150.

Suter EA, Pachiadaki M, Taylor GT, Edgcomb VP (2021). "Eukaryotic Parasites Are Integral to a Productive Microbial Food Web in Oxygen-Depleted Waters". *Front Microbiol* 12, p. 764605.

Táborský P, Pánek T, Čepička I (2017). "Anaeramoebidae Fam. Nov., a Novel Lineage of Anaerobic Amoebae and Amoeboflagellates of Uncertain Phylogenetic Position". *Protist* 168.5, pp. 495–526.

Takishita K, Tsuchiya M, Kawato M, Oguri K, Kitazato H, Maruyama T (2007). "Genetic Diversity of Microbial Eukaryotes in Anoxic Sediment of the Saline Meromictic Lake Namakoike (Japan): On the Detection of Anaerobic or Anoxic-tolerant Lineages of Eukaryotes". *Protist* 158.1, pp. 51–64.

Tashyreva D, Prokopchuk G, Votýpka J, Yabuki A, Horák A, Lukeš J (2018a). "Life Cycle, Ultrastructure, and Phylogeny of New Diplonemids and Their Endosymbiotic Bacteria". *mBio* 9.2, e02447–17.

Tashyreva D, Prokopchuk G, Yabuki A, Kaur B, Faktorová D, Votýpka J, Kusaka C, Fujikura K, Shiratori T, Ishida K.-I, Horák A, Lukeš J (2018b). "Phylogeny and Morphology of New Diplonemids from Japan". *Protist* 169.2, pp. 158–179.

Tekle YI, Grant J, Anderson OR, Nerad TA, Cole JC, Patterson DJ, Katz LA (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* 47.1, pp. 339–352.

Telford MJ, Budd GE, Philippe H (2015). "Phylogenomic Insights into Animal Evolution". *Current Biology* 25.19 (19), R876–R887.

Thomas DN, Dieckmann GS (2009). *Sea Ice*. John Wiley & Sons. 641 pp.

Tice AK, Žihala D, Pánek T, Jones RE, Salomaki ED, Nenarokov S, Burki F, Eliáš M, Eme L, Roger AJ, Rokas A, Shen X.-X, Strassert JFH, Kolísko M, Brown MW (2021). "PhyloFisher: A Phylogenomic Package for Resolving Eukaryotic Relationships". *PLOS Biology* 19.8, e3001365.

Tikhonenkov DV (2020). "Predatory Flagellates – the New Recently Discovered Deep Branches of the Eukaryotic Tree and Their Evolutionary and Ecological Significance". *Protistology* 14.1.

Tikhonenkov DV, Janouškovec J, Mylnikov AP, Mikhailov KV, Simdyanov TG, Aleoshin VV, Keeling PJ (2014). "Description of Colponema Vietnamica Sp.n. and Acavomonas Peruviana n. Gen. n. Sp., Two New Alveolate Phyla (Colponemidia Nom. Nov. and Acavomonidia Nom.Nov.) and Their Contributions to Reconstructing the Ancestral State of Alveolates and Eukaryotes". *PLoS ONE* 9.4 (4), e95467–e95467.

Tikhonenkov DV, Mikhailov KV, Hehenberger E, Karpov SA, Prokina KI, Esaulov AS, Belyakova OI, Mazei YA, Mylnikov AP, Aleoshin VV, Keeling PJ (2020a). "New Lineage of Microbial Predators Adds Complexity to Reconstructing the Evolutionary Origin of Animals". *Current Biology*.

Tikhonenkov DV, Strassert JFH, Janouškovec J, Mylnikov AP, Aleoshin VV, Burki F, Keeling PJ (2020b). "Predatory Colponemids Are the Sister Group to All Other Alveolates". *Molecular Phylogenetics and Evolution* 149, p. 106839.

Torruella G, Moreira D, López-García P (2017). "Phylogenetic and Ecological Diversity of Apusomonads, a Lineage of Deep-Branching Eukaryotes". *Environ Microbiol Rep* 9.2, pp. 113–119.

Treude T (2012). "Biogeochemical Reactions in Marine Sediments Underlying Anoxic Water Bodies". *Anoxia: Evidence for Eukaryote Survival and Paleontological Strategies*. Ed. by AV Altenbach, JM Bernhard, J Seckbach. Cellular Origin, Life in Extreme Habitats and Astrobiology. Dordrecht: Springer Netherlands, pp. 17–38.

Triadó-Margarit X, Casamayor EO (2015). "High Protists Diversity in the Plankton of Sulfurous Lakes and Lagoons Examined by 18s rRNA Gene Sequence Analyses". *Environ. Microbiol. Rep.* 7.6, pp. 908–917.

V. Wintzingerode F, Göbel UB, Stackebrandt E (1997). "Determination of Microbial Diversity in Environmental Samples: Pitfalls of PCR-based rRNA Analysis". *FEMS Microbiology Reviews* 21.3, pp. 213–229.

Valentine JW, Collins AG (2000). "The Significance of Moulting in Ecdysozoan Evolution". *Evol. Dev.* 2.3 (3), pp. 152–156.

Vargas C de et al. (2015b). "Ocean Plankton. Eukaryotic Plankton Diversity in the Sunlit Ocean." *Science* 348.6237 (6237), p. 1261605.

Vaulot D, Hung Sim CW, Ong D, Teo B, Biwer C, Jamy M, dos Santos AL (2022a). *metaPR² : A Database of Eukaryotic 18S rRNA Metabarcodes with an Emphasis on Protists*. preprint. Ecology.

— (2022b). "metaPR²: A Database of Eukaryotic 18S rRNA Metabarcodes with an Emphasis on Protists". *bioRxiv*, p. 2022.02.04.479133.

Ventosa A, Fernández AB, León MJ, Sánchez-Porro C, Rodriguez-Valera F (2014). "The Santa Pola Saltern as a Model for Studying the Microbiota of Hypersaline Environments". *Extremophiles* 18.5, pp. 811–824.

Villanueva L, von Meijenfeldt FAB, Westbye AB, Yadav S, Hopmans EC, Dutilh BE, Damsté JSS (2020). "Bridging the Membrane Lipid Divide: Bacteria of the FCB Group Superphylum Have the Potential to Synthesize Archaeal Ether Lipids". *ISME J.*, pp. 1–15.

Volland J.-M et al. (2022). "A Centimeter-Long Bacterium with DNA Contained in Metabolically Active, Membrane-Bound Organelles". *Science* 376.6600, pp. 1453–1458.

Vørs N (1988). "Discocelis Saleuta Gen. Nov. et Sp. Nov. (Protista Incertae Sedis): — A New Heterotrophic Marine Flagellate". *European Journal of Protistology* 23.4, pp. 297–308.

— (1992). "Heterotrophic Amoebae, Flagellates and Heliozoa from the Tvärminne Area, Gulf of Finland, in 1988–1990". *Ophelia* 36.1, pp. 1–109.

Wägele, Misof (2001). "On Quality of Evidence in Phylogeny Reconstruction: A Reply to Zrzavý's Defence of the 'Ecdysozoa' Hypothesis". *J. Zool. Syst. Evol. Res.* 39.3 (3), pp. 165–176.

Walker G, Dacks JB, Embley TM (2006). "Ultrastructural Description of Breviata Anathema, N. Gen., N. Sp., the Organism Previously Studied as "Mastigamoeba Invertens"". *J. Eukaryot. Microbiol.* 53.2 (2), pp. 65–78.

Wang H.-C, Minh BQ, Susko E, Roger AJ (2018). "Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation". *Syst Biol* 67.2, pp. 216–235.

West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF (2018). "Genome-Reconstruction for Eukaryotes from Complex Natural Microbial Communities". *Genome Res.* 28.4 (4), pp. 569–580.

Whelan S, Irisarri I, Burki F (2018). "PREQUAL: Detecting Non-Homologous Characters in Sets of Unaligned Homologous Sequences". *Bioinformatics* 34.22, pp. 3929–3930.

Wideman JG, Lax G, Leonard G, Milner DS, Rodríguez-Martínez R, Simpson AGB, Richards TA (2019). "A Single-Cell Genome Reveals Diplonemid-like Ancestry of Kinetoplastid Mitochondrial Gene Structure". *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1786 (1786), p. 20190100.

Wolf YI, Rogozin IB, Koonin EV (2004). "Coelomata and Not Ecdysozoa: Evidence From Genome-Wide Phylogenetic Analysis". *Genome Res.* 14.1 (1), pp. 29–36.

Woyke T, Doud DFR, Schulz F (2017). "The Trajectory of Microbial Single-Cell Sequencing". *Nat Methods* 14.11 (11), pp. 1045–1054.

Yabuki A, Inagaki Y, Ishida K.-i (2010). "Palpitomonas Bilix Gen. et Sp. Nov.: A Novel Deep-branching Heterotroph Possibly Related to Archaeplastida or Hacrobia". *Protist* 161.4, pp. 523–538.

Yang J, Harding T, Kamikawa R, Simpson AGB, Roger AJ (2017). "Mitochondrial Genome Evolution and a Novel RNA Editing System in Deep-Branching Heteroloboseids". *Genome Biol Evol* 9.5 (5), pp. 1161–1174.

Yasindi AW, Taylor WD (2006). "The Trophic Position of Planktonic Ciliate Populations in the Food Webs of Some East African Lakes". *African Journal of Aquatic Science* 31.1, pp. 53–62.

— (2016). "The Protozoa of Soda Lakes in East Africa". *Soda Lakes of East Africa*. Ed. by M Schagerl. Cham: Springer International Publishing, pp. 179–204.

Yazaki E, Kume K, Shiratori T, Eglit Y, Tanifuji G, Harada R, Simpson AGB, Ishida K.-i, Hashimoto T, Inagaki Y (2020). "Barthelonids Represent a Deep-Branching Metamonad Clade with Mitochondrion-Related Organelles Predicted to Generate No ATP". *Proceedings of the Royal Society B: Biological Sciences* 287.1934 (1934), p. 20201538.

Yazaki E, Yabuki A, Imaizumi A, Kume K, Hashimoto T, Inagaki Y (2022). "The Closest Lineage of Archaeplastida Is Revealed by Phylogenomics Analyses That Include Microheliella Maris". *Open Biol.* 12.4, p. 210376.

Yeh Y.-C, Fuhrman JA (2022). "Contrasting Diversity Patterns of Prokaryotes and Protists over Time and Depth at the San-Pedro Ocean Time Series". *ISME COMMUN.* 2.1 (1), pp. 1–12.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT.-Y (2017). "Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data". *Methods Ecol. Evol.* 8.1, pp. 28–36.

Yubuki N, Leander BS, Silberman JD (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* 161.2, pp. 264–278.

Yubuki N, Pánek T, Yabuki A, Čepička I, Takishita K, Inagaki Y, Leander BS (2015a). "Morphological Identities of Two Different Marine Stramenopile Environmental Sequence Clades: Bicosoeca Kenaiensis (Hilliard, 1971) and Cantina Marsupialis (Larsen and Patterson, 1990) Gen. Nov., Comb. Nov." *J. Eukaryot. Microbiol.* 62.4 (4), pp. 532–542.

— (2015b). "Morphological Identities of Two Different Marine Stramenopile Environmental Sequence Clades: Bicosoeca Kenaiensis (Hilliard, 1971) and Cantina Marsupialis (Larsen and Patterson, 1990) Gen. Nov., Comb. Nov." *J. Eukaryot. Microbiol.* 62.4, pp. 532–542.

Zaremba-Niedzwiedzka K et al. (2017). "Asgard Archaea Illuminate the Origin of Eukaryotic Cellular Complexity". *Nature* 541.7637 (7637), pp. 353–358.

Zhang H, Jain C, Aluru S (2020). "A Comprehensive Evaluation of Long Read Error Correction Methods". *BMC Genomics* 21.6, p. 889.

Zorz JK, Sharp C, Kleiner M, Gordon PMK, Pon RT, Dong X, Strous M (2019). "A Shared Core Microbiome in Soda Lakes Separated by Large Distances". *Nat Commun* 10.1 (1), p. 4230.

Zuendorf A, Bunge J, Behnke A, Barger KJ.-A, Stoeck T (2006). "Diversity Estimates of Microeukaryotes below the Chemocline of the Anoxic Mariager Fjord, Denmark". *FEMS Microbiology Ecology* 58.3, pp. 476–491.