Exploring the Phylogenetic Diversity of Phagotrophic Euglenids

by

Gordon Malcolm Lax

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

at
Dalhousie University
Halifax, Nova Scotia
February 2020

© Copyright by Gordon Malcolm Lax, 2020

Dedicated to my parents
Doris and Detlef.

Table of Contents

List of Tables vii

List of Figures viii

Abstract x

List of Abbreviations Used xi

Acknowledgements xiii

Chapter 1 Introduction 1

1.1 Single-cell Methodology 1

1.1.1 Isolation and Lysis of Single Cells 3

1.1.2 Targeted PCR 5

1.1.3 Genomics 6

1.1.4 Transcriptomics 6

1.1.5 Single-cell Methods in Phylogenetics & Biodiversity Research 7

1.2 Euglenids 12

1.2.1 What are Euglenids? 12

1.2.2 The Euglenid Pellicle is Thought to Represent Euglenid Phylogeny 12

1.2.3 Phagotrophs are Surface-associated 13

1.2.4 Symbiontida and Hemimastigophora – Derived Euglenids? . . 14

1.2.5 Few Molecular Sequences are Available for Phagotrophic Euglenids 15

1.2.6 A Molecular Single-cell Approach to Euglenids 16

Chapter 2 Ploeotids Represent Much of the Phylogenetic Diversity of Euglenids **17**
2.1 Abstract . . . 17
2.2 Introduction . . . 18
2.3 Methods . . . 22
2.3.1 Cultures: Establishment and Nucleic Acid Extraction . . . 22
2.3.2 Single Cells: Photodocumentation and DNA Amplification . . . 22
2.3.3 PCR Amplification and Sequencing . . . 23
2.3.4 Alignment and Phylogenetic Analyses . . . 23
2.3.5 Scanning Electron Microscopy . . . 25
2.4 Results . . . 25
2.4.1 Studied Isolates . . . 25
2.4.2 Phylogeny . . . 36
2.4.3 Group I Introns . . . 40
2.4.4 Taxonomic Summary . . . 41
2.5 Discussion . . . 44
2.5.1 Molecular Phylogenetics of Phagotrophic Euglenids . . . 44
2.5.2 *Ploeotia, Serpenomonas* and *Olkasia* . . . 45
2.5.3 Some *Anisonema* Species are 'Ploeotids' . . . 46
2.5.4 Higher-Order Systematics . . . 47
2.5.5 Phylogenetic Structure within Morphospecies . . . 49
2.5.6 Pellicle Strip Architecture . . . 51
2.5.7 Eukaryotrophy in Ploeotids . . . 52
2.5.8 Group I and Non-Canonical Introns . . . 53
2.5.9 Have We Found all Major Clades of Ploeotids? . . . 54
2.5.10 The Need for Multigene Phylogenies . . . 54
2.6 Acknowledgments . . . 55

Chapter 3 Towards a Resolved Tree of Euglenids: A Single-cell and Phylogenomics Approach **56**
3.1 Introduction . . . 56
3.1.1 General Structure of Euglenid Phylogenies . . . 57
3.1.2 Relationships within Spirocuta . . . 58
3.1.3 Relationships at the Base of the Euglenid Tree . . . 58
3.1.4 Symbiontida . . . 59

3.1.5 Increasing Taxon Sampling for SSU rDNA and Multigene Analyses . . . 60
3.2 Methods . . . 60
3.2.1 Single-Cell Isolation & Photodocumentation . . . 60
3.2.2 Single-Cell SSU rDNA Sequencing . . . 61
3.2.3 Single-Cell Transcriptomics . . . 62
3.2.4 Transcriptomes Derived from Mass Cultures . . . 62
3.2.5 SSU rDNA Phylogenetics . . . 63
3.2.6 Assembly of Transcriptomic Data . . . 65
3.2.7 Phylogenomics – Dataset Construction . . . 65
3.2.8 Phylogenomics – Analyses . . . 66
3.3 Results . . . 67
3.3.1 SSU rDNA Phylogenetics . . . 67
3.3.2 Morphology . . . 72
3.3.3 Phylogenomics . . . 80
3.4 Discussion . . . 84
3.4.1 New Insights into Euglenid Diversity at and around 'Genus Level' . . . 84
3.4.2 Large-Scale Euglenid Evolution . . . 91
3.4.3 Methodology . . . 97

Chapter 4 Hemimastigophora is a Novel Supra-kingdom-level Lineage of Eukaryotes 100
4.1 Introduction . . . 101
4.2 Methods . . . 101
4.2.1 Cell Isolation and Transcriptomics . . . 101
4.2.2 Cultivation of *Hemimastix kukwesjijk* . . . 102
4.2.3 Scanning Electron Microscopy of *Hemimastix kukwesjijk* . . . 102
4.2.4 SSU rDNA Analyses . . . 102
4.2.5 Environmental SSU rRNA/rDNA Sequence Comparisons . . . 103
4.2.6 Phylogenomic Dataset Assembly . . . 104
4.2.7 Quality of Hemimastigote Transcriptomes . . . 104
4.2.8 Phylogenomic Analyses . . . 105
4.2.9 Identification of Non-Universal Ancient Genes . . . 107
4.2.10 Identification of *Spironema* c.f. *multiciliatum* . . . 108
4.2.11 Data Availability . . . 108

4.3 Results and Discussion 109
4.3.1 Description of *Hemimastix kukwesjijk* 115
4.4 Acknowledgements 116

Chapter 5 Conclusion **117**
5.1 Current State of Euglenid Research 117
5.1.1 Exploring the Biodiversity of Phagotrophic Euglenids 117
5.1.2 Large-Scale Euglenid Evolution 118
5.1.3 Hemimastigophora 118
5.2 Wider Impact on Euglenid/Eukaryote Research 119
5.2.1 Impact on Our Understanding of Euglenids 119
5.2.2 Remaining Issues in Euglenids 119
5.3 Are Single-Cell Approaches the Future? 120

Appendix A Ploeotids Represent Much of the Phylogenetic Diversity of Euglenids **122**

Appendix B Towards a Resolved Tree of Euglenids: A Single-cell and Phylogenomics Approach **125**

Appendix C Hemimastigophora is a Novel Supra-kingdom-level Lineage of Eukaryotes **139**

Bibliography **148**

List of Tables

2.1 Morphological measurements of ploeotid cultures 26

A.1 Organism codes, GenBank accession codes, and euglenid-biased SSU primer sequences & combinations 122
A.2 Morphological measurements of isolated single cells 122
A.3 Sequence similarities of *Olkasia*, *Serpenomonas*, *Lentomonas*, and *Entosiphon* 122
A.4 Support values for euglenid and ploeotid groups 122
A.5 Sampling sites of cultures and samples 123

B.1 Explanation of sample site codes, with location, coordinates and material sampled 125
B.2 SSU rDNA primer sequences and combinations 125
B.3 Sequencing information, assembly strategies, BUSCO-scores and transrate statistics for transcriptomes and SAGs 125
B.4 Taxon sampling for multigene-analyses and sources for external assemblies 125
B.5 Isolated euglenid cells, with identified morphotype, associated clades and groups, measurements and locomotion information 126
B.6 Support values for important bipartitions in SSU rDNA and multigene phylogenies 126
B.7 Videos of representatives of phagotrophic euglenid groups/clades 126

C.1 Full listing of environmental sequences attributable to Hemimastigophora 139
C.2 Taxa used in phylogenomic analyses 139
C.3 Genes of potential deep evolutionary significance in eukaryotes 139

List of Figures

1.1 Diagrams of two single cell molecular techniques used in protistology . . 5

2.1 Light micrographs of ploeotid taxa (cultures and single cells) 27
2.2 Scanning electron microscopy images of *Ploeotia vitrea* and *P. oblonga* . . 29
2.3 Scanning electron microscopy images of *Serpenomonas costata* 30
2.4 Light micrographs of *Olkasia polycarbonata* 31
2.5 Scanning electron microscopy images of *O. polycarbonata* 32
2.6 Light micrographs of ploeotid single cell isolates and cultures 34
2.7 Maximum Likelihood phylogeny of the SSU-rRNA gene of euglenids . . . 37
2.8 Summary view of euglenid SSU-rRNA phylogeny including *Entosiphon* . . 39
2.9 Introns in *Serpenomonas*, *Ploeotia* and *Keelungia* sp. 41

3.1 Maximum Likelihood phylogeny of the SSU rDNA gene of euglenids . . . 69
3.2 Light micrographs of peranemid single cells 74
3.3 Light micrographs of anisonemid (-like) and peranemid single cells . . . 76
3.4 Light micrographs of petalomonad and symbiontid single cells 78
3.5 Phylogeny of Discoba inferred from 20 genes 81

4.1 Micrographs of studied hemimastigotes 110
4.2 Environmental sequencing reads assigned to Hemimastigophora 111
4.3 Phylogenetic placement of Hemimastigophora within eukaryotes 112
4.4 Summary of phylogenomic analyses and distribution of select genes . . . 115

A.1 Maximum Likelihood phylogeny of the SSU-rRNA gene of euglenids, including *Entosiphon* 124

B.1 SSU rDNA Maximum Likelihood phylogeny of Euglenozoa ('SSU-LB') . . 127
B.2 SSU rDNA Maximum Likelihood phylogeny of Euglenozoa with long-branching taxa removed ('SSU-noLB') 128

B.3 SSU rDNA Maximum Likelihood phylogeny of Euglenozoa with short sequences removed ('SSU-noShort') 129
B.4 SSU rDNA Maximum Likelihood phylogeny of Euglenozoa with rogue taxa removed ('SSU-noRogues') 130
B.5 SSU rDNA Maximum Likelihood phylogeny of Euglenida ('SSU-Euglenids') 131
B.6 SSU rDNA Maximum Likelihood phylogeny of Discoba ('SSU-Discoba') . . 132
B.7 SSU rDNA Maximum Likelihood phylogeny of Symbiontida ('SSU-Symbiontids') 133
B.8 20-gene phylogeny of Discoba with long-branching taxa removed ('noLB') 134
B.9 20-gene phylogeny of Discoba with rogue taxa removed ('noRogues') . . 135
B.10 20-gene phylogeny of Euglenozoa ('Euglenozoa-only') 136
B.11 20-gene phylogeny of Euglenids ('Euglenida-only') 137
B.12 Graph summarizing support for important bipartitions with progressive removal of fast-evolving sites ('FSR') 138

C.1 DIC images of *Spironema* c.f. *multiciliatum* and *Hemimastix kukwesjijk* . . . 140
C.2 SEM images of *Hemimastix kukwesjijk* 141
C.3 SSU rDNA phylogeny of eukaryotes inferred from 111 taxa 142
C.4 Unrooted phylogeny of eukaryotes inferred from 104 taxa (351 genes) . . 143
C.5 Unrooted phylogeny of eukaryotes from the '58-nLB' dataset 144
C.6 Unrooted phylogeny of eukaryotes from the '58-nDP' dataset 145
C.7 Unrooted phylogeny of eukaryotes from the '61-SR4' dataset 146
C.8 Summary of support for several important bipartitions with sequential removal of fastest evolving sites from the 61 taxa dataset (61-SFSR) . . . 147

Abstract

EUGLENIDS are a widespread, complex group of flagellated, single-celled eukaryotes. The majority of described species are phototrophic forms, yet a large portion of the phylogenetic diversity is composed of phagotrophs, from which phototrophs arose through secondary endosymbiosis. To understand euglenid evolution, it is necessary to understand phylogenetic relationships among phagotrophic euglenids. Yet despite their diversity and evolutionary relevance, phagotrophs are underrepresented in molecular sequence data that are crucial in reconstructing evolutionary relationships. Unfortunately, they are difficult to culture, which complicates molecular data collection for most species, as standard methods of nucleic acid extraction cannot be used. I used a culture-independent single-cell approach to increase sampling of the SSU-rDNA gene for phagotrophs almost five-fold (now 141 sequences). Phylogenetic trees show that ploeotids, an assemblage of rigid phagotrophs, make up much of the basal phylogenetic diversity. Several morphotypes that were previously lumped together into a single genus *Ploeotia* are not monophyletic and belong to multiple separate genera, including *Olkasia* nov. gen. Two species of 'spirocute' *Anisonema* proved to be ploeotids and were transferred to new genera *Hemiolia* and *Liburna*. More previously unsampled phagotrophs belong within Spirocuta (which includes phototrophs and osmostrophs), and phylogenetic analyses revealed several morphologically-defined genera are likely not monophyletic. To better resolve the tree of euglenids, I generated 24 single-cell transcriptomes to provide data for a 20-gene phylogenetic analysis. This divided Spirocuta into phototrophs, a robust 'Anisonemids plus', and a weakly supported 'Peranemids' clade. Ploeotids are paraphyletic, with *Olkasia* robustly inferred as sister to Spirocuta, whereas petalomonads are placed basal to all other euglenids with high support. The multigene analyses suggest that symbiontids are not euglenids, but may be more closely related to diplonemids and kinetoplastids. Hemimastigophora are a group of enigmatic multiflagellated cells that have long evaded molecular sequencing, and were inferred at one point to be related to euglenids based on electron-microscopy data. A single-cell approach generated transcriptomes of group members *Spironema* and *Hemimastix*, enabling them to be included in eukaryote-wide phylogenomic analyses. Remarkably, Hemimastigophora do not fall into any recognised supergroup of eukaryotes, but form their own independent group that branches outside both Diaphoretickes and Amorphea.

List of Abbreviations Used

aa	amino acid
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BS	Bootstrap Support
bp	base pair
BUSCO	Benchmarking Universal Single-Copy Orthologs
CCAP	Culture Collection of Algae and Protozoa
CCMP	Culture Collection of Marine Phytoplankton
cDNA	complementary Deoxyribonucleic Acid
DIC	Differential Interference Contrast microscopy
DNA	Deoxyribonucleic Acid
dNTP	deoxyribose Nucleoside Triphosphates
FACS	Fluorescence-Activated Cell Sorting
FSR	Fast Site Removal
GTR	Generalised Time-Reversible model of DNA evolution
HE	Homing Endonuclease
Hsp90	Heat shock protein 90
HTS	High-Throughput Sequencing
ICZN	International Commission on Zoological Nomenclature
LB	Lysogeny Broth
LBA	Long Branch Attraction Artefacts
LECA	Last Eukaryotic Common Ancestor
LSU	Large Subunit of ribosomal DNA/RNA
LWR	Likelihood-to-Weight Ratio
MALV	Marine ALVeolates
MAST	MArine STramenopiles
MDA	Multiple Displacement Amplification
ML	Maximum Likelihood
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NIES	National Institute for Environmental Studies (culture collection)
PCR	Polymerase Chain Reaction
Pfam	Protein families (protein database)

PMSF	Posterior Mean Site Frequency
pp	posterior probabilities
PSRF	Potential Scale Reduction Factor
RNA	Ribonucleic Acid
RT	Reverse Transcription
SAG	Single Amplified Genome
Sar	Stramenopiles, alveolates, and rhizaria
SEM	Scanning Electron Microscopy
SSU	Small Subunit of ribosomal DNA/RNA
Taq	Thermostable DNA polymerase of *Thermus aquaticus*
TSO	Template Switching Oligo
UFB	Ultra-Fast Bootstrap (also UFboot)
V4	Hypervariable region V4 (in SSU-rDNA)
V9	Hypervariable region V9 (in SSU-rDNA)
WGA	Whole Genome Amplification

Acknowledgements

FIRST and foremost, I have been extremely fortunate to have had a supervisor as wonderful as Alastair Simpson. He is simply one of the kindest and greatest supervisors I could have wished for: Thank you from the bottom of my heart for giving me the opportunity to work on such a challenging yet rewarding project, for your (often 'fiddly' but necessary) advice, teaching me the art of conducting science properly, and your continuing support! I would also like to thank my committee members Claudio Slamovits and Julie LaRoche, and in particular Andrew Roger for teaching me phylogenetics and providing very useful input over the years (and also the wonderful parties). I would like to thank Won Je Lee for help with trickier identification of some phagotrophic euglenids. His Ph.D-thesis was also extensively used for identification. ¶

The Simpson Lab is small, yet has some honestly wonderful and awesome people. Foremost many thanks to Yana for being a good friend, providing lots of encouraging words and making the lab exciting (and providing many a loaf of challah). Previous lab members Robyn Buchwald, Claire Burnard, Sebastian Hess, Anna Busch, and Andrea Gigeroff (and honorary member Raphael McDonald) also provided many laughs and support, and made the lab such a fun place to be in. Many thanks to all protist-friends on the 8th floor in the Tupper Medical Building (in particular Eleni Gentegaki, Sophie and Jan de Vries, Michelle Leger, and Jon Jerlström-Hultqvist), and to the lab-beer crowd (which was crucial for sanity). CGEB (Computational Genomics and Evolutionary Biology) is a wonderful group with many labs and wonderful people, and has been crucial in providing support and interest over the years. Dalhousie would be much worse without it. ¶

My parents Doris and Detlef have always been my top fans and have wholeheartedly supported me ever since I decided to study Biology. Which is why this thesis is dedicated to them—I would not be here without you (in the biological but also spiritual sense): Thank you! Similarly my grandparents Karla and Herwig have always been supportive, even if it meant I was moving to the other side of the world. ¶

Many thanks to all my friends, as they've been always there for me and I would not know what to do without them. Ryke, we've been going through thick and thin together and write almost each single day (take that 6000 km!), and I could not be more happy to have you in my life. Christian, not sure what I would do without our weekly talks and gaming sessions (HEYOOO). I honestly hope we will still be doing this when we're both in our 70s. Lastly, the 'Tuesday Night Special Weather Event'

Trivia Crew has helped to provide reprieve from lab and thesis work. Thank you too all of you, but especially Bob, for providing a view out of a different window of the ivory tower. ¶

My biology teacher in the latter years of high school has been pretty pivotal in my life. Without Eva Göller I would have never even thought about studying biology and then making a career of it (particularly the genetics quizzes involving dragons somehow convinced me this is worth studying). So, thank you for inspiring me—I hope you continue to do the same to more students! ¶

Lastly, my partner Joana has always been there for me no matter what, and always has encouring words of support even through many weeks of my bad mood. Thank you for being loving and awesome: I love you and I am not sure who I would be without you!

Gordon Lax
Halifax, 10 January 2020

Chapter 1

Introduction

1.1 Single-cell Methodology

TRADITIONAL microbiology was founded by establishing and examining cultures of microbes. The cultivation approach enables deep and thorough research of an organism, but is also slow—and starting in the 1980s it became increasingly clear that not all organisms can be easily cultured, at least not without investing considerable resources and time (Olsen et al. 1986, Rappé & Giovannoni 2003). This 'culture bottleneck' became even more apparent with the increasing use of environmental sequencing of the SSU rDNA gene, which revealed a diversity vastly higher than what was known from traditional culturing approaches (Rappé & Giovannoni 2003). While this targeted environmental sequencing effort is superior for capturing diversity to cultivation, information on an individual cell—other than a fragment of the SSU rDNA—is lost. This includes the genome or transcriptome, which can be, for example, used to infer phylogenetic placement when the information in the SSU rDNA is insufficient. Perhaps most importantly though, valuable insights into the biology of the organism can be gained from genomic and transcriptomic data, as well as from microscopy imagery and other data.

Rise of Single-cell Methods

To bridge the gap between current investigative approaches for uncultured and cultured organisms, single-cell molecular methods have increasingly been employed. They can be used to obtain gene, genomic or transcriptomic information (and possibly proteomic data in the near future) from as little as one cell, not requiring any cultures (Kolisko et al. 2014, Woyke et al. 2017). Most of these techniques have ini-

tially been developed in the biomedical field, largely for the investigation of human cancer cells (e.g. Ramsköld et al. 2012, Huang et al. 2015, Prakadan et al. 2017). Researchers investigating prokaryotes (especially bacteria) were in turn quick to adopt some of these methods, as it had become apparent that only an estimated $<1\%$ of bacterial species have been cultured (Marcy et al. 2007b). This culturing bottleneck and subsequent discovery of 'microbial dark matter' was investigated largely by targeted environmental sequencing (Olsen et al. 1986) and metagenome sequencing (Tyson et al. 2004). While advancing knowledge about community structure and gene composition, these approaches, at the time, did not specifically query information that is contained within a single cell. Fluorescent in situ hybridization (FISH) provided a partial advance, by using fluorescent oligonucleotide probes that specifically bind to the ribosomal RNA of cells (DeLong et al. 1989, Amann et al. 1990, Wagner et al. 2003). FISH was pivotal in linking the massive amounts of environmental high-throughput sequence (HTS) data back to their samples (Amann & Fuchs 2008). Even though only a short sequence was known from an imaged single cell, basic morphological characters could be identified, and it provided a way to count cells of interest (Amann & Fuchs 2008). Single-cell sorting and whole genome sequencing approaches were soon developed, enabling researchers to, for example, specifically investigate highly variable metabolic pathways (Marcy et al. 2007b, Rinke et al. 2013). Ultimately, the discovery of novel prokaryote phyla has relied heavily on environmental sequencing of ribosomal RNA using HTS, the interrogation of their community composition and relative abundances on FISH, while single-cell genomics has been crucial in characterizing their biology (Rinke et al. 2013, Hedlund et al. 2014).

Single-cell Approaches in Protists

The diversity of microbial eukaryotes (protists) has so far largely been investigated with culture-based approaches (Keeling & del Campo 2017), environmental sequencing (e.g. Moon-van der Staay et al. 2001, Diez et al. 2001, Stoeck et al. 2010), and purely microscopical work (e.g. Larsen & Patterson 1990). While powerful on their own, these methods often do not complement each other well; for example microscopy and environmental sequencing are difficult to link without the use of FISH. The culturing bottleneck restriction applies to protists as well, especially to heterotrophic cells that feed on other microbes and might well be selective in their choice of food, which can complicate culturing due to unknown prey preferences. Because of their (in general) bigger size when compared to prokaryotes,

PCR-amplification of specific genes—usually the SSU rDNA—from single cells became common practice (e.g. Sebastián & O'Ryan 2001, Chantangsi & Leander 2010, Gómez et al. 2011). Due to apparent ease of use, single cell whole genome sequencing was eventually used on protists (Yoon et al. 2011, Bhattacharya et al. 2012, Roy et al. 2014). While employed successfully several times, genomics is currently being surpassed in popularity by single-cell transcriptomics, mostly owing to the large and complicated nature of many eukaryote genomes which renders assembly tedious (Keeling & del Campo 2017, Kolisko et al. 2014).

There are many possible lines of reasoning behind investigating a single protist cell, rather than a large number of cells, and these depend on the type of question asked. For example, one might be interested in the physiological properties of an organism, and investigating a culture masks the considerable individual variability between cells (Prakadan et al. 2017), which is especially true for transcript expression levels. Nonetheless, the current main use of molecular single-cell methods on protists is to investigate their biodiversity and evolution, which I will be focussing on here. For researchers interested in the biodiversity and evolutionary history of protists, the culturing bottleneck represents a considerable challenge. It is now often more feasible, faster and easier to generate single-cell transcriptomes of an unknown organism, rather than attempting to culture it. In addition, microscopy can be easily integrated into this workflow (e.g. Krabberød et al. 2017a, Kang et al. 2017), which enables a more 'rounded' characterization of taxa than with environmental sequencing alone. Morphological variation among protists can be extreme, so any additional morphological information can aid identification and ultimately help understand an organism (Keeling & del Campo 2017). The obtained sequence data from a cell can then be used in phylogenomic analyses, placing the organism of interest in a phylogenetic context (e.g. Kang et al. 2017).

1.1.1 Isolation and Lysis of Single Cells

Isolating single cells from a sample is the first step in generating molecular data. Isolation needs to be accurate, with only the cell of interest picked. This is relatively straightforward for cells in suspension (such as non-aggregated plankton), but quickly becomes challenging for dense samples from environments like biofilm, soil, and sediments due to large quantities of unwanted particulates. Minute contaminations like spurious prokaryotes or detritus need to be removed as much as possible as they can massively influence downstream applications (Worden et al.

2011, Rinke et al. 2014). In particular, whole-genome amplification (WGA; see below) picks up any DNA-containing contaminant and amplifies it, which can be very problematic in the case of genome sequencing (Huys & Raes 2018). While single-cell transcriptomics is less prone to prokaryotic contamination due to its routinely-used poly(A)-selection, isolation still needs to be clean, since contaminating nucleic acid fragments can stick to cells and will subsequently be co-amplified (Picelli et al. 2014, Kolisko et al. 2014). Conventional PCR of a single cell using specific primers to amplify a gene of interest might be less susceptible to contamination, but this depends highly on the universality of the primers used (e.g. universal eukaryote primers vs. group-biased primers).

The simplest method to isolate cells is with a finely-drawn out glass pipette, either by hand, or by using a micromanipulator. This approach has been used extensively for single-cell PCR (e.g. Chantangsi & Leander 2010, Lax & Simpson 2013), genomes (Yoon et al. 2011), and recently, for expressed sequence tags and transcriptomes (Kang et al. 2017, Balzano et al. 2015). This method is relatively low-throughput, but is inexpensive and can be easily combined with photodocumentation: Cells can be photographed before or after initial isolation, providing high-quality imagery or even video that can provide additional information (Chantangsi & Leander 2010, Lax & Simpson 2013, Lahr et al. 2011, Bennett & Triemer 2012). Some researchers have even combined light microscopy, multiple-locus PCR, and scanning electron microscopy (SEM) on single dinoflagellate cells (Takano & Horiguchi 2005). Another low-cost method is serial dilution—here samples are diluted to such an extent that only single cells are contained within final subsamples, although this has only been applied to bacterial genome sequencing (Zhang et al. 2006). In recent years automated cell-sorting methods have gained traction, including fluorescence-activated cell sorting (FACS), and microfluidic devices (Prakadan et al. 2017), though the latter has not been extensively employed yet in microbial eukaryote research.

A critical part of post-isolation treatment is efficient lysis to enable access to the nucleic acids (or proteins) of interest. Furthermore, mRNA in particular is rapidly degraded by cellular RNAses once the cell is dying—this results in a sequencing product of lesser quality (Jackowiak et al. 2011). Various methods of lysis employed on protist cells include freeze-thaw cycles (e.g. Lax & Simpson 2013), chemical lysis (e.g. Krabberød et al. 2017a), and the use of enzymes such as cellulases (mainly for cells with cell walls, like dinoflagellates) and proteases (e.g. Marín et al. 2001, Lynn & Pinheiro 2009).

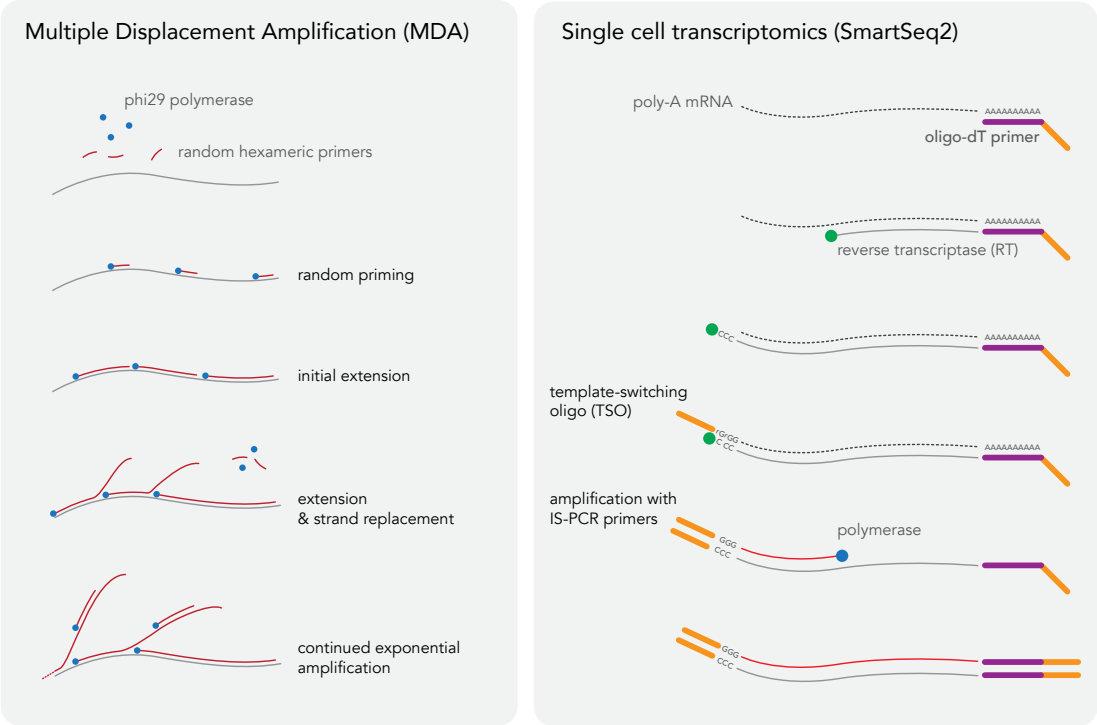

Figure 1.1: Diagrams of two single cell molecular techniques used in protist research. The left panel shows multiple displacement amplification (MDA), which is used to amplify DNA within a single cell. The right panel depicts SmartSeq2, a method to generate cDNA from all mRNA in a single cell, which in turn is amplified via PCR. Adapted from Spits et al. (2006) and Picelli et al. (2014), respectively.

1.1.2 Targeted PCR

The most widely used method to generate molecular sequence data from single protist cells is targeted PCR, most often to amplify their SSU rDNA (e.g. Sebastián & O'Ryan 2001, Lynn & Pinheiro 2009, Chantangsi & Leander 2010, Lax & Simpson 2013). While tested extensively on a large portion of protist diversity, single-cell PCR relies heavily on the properties of the primers used and is low-throughput, usually only allowing amplification of a single gene at a time. One possible extension of single-cell PCR is to first use whole genome amplification (usually MDA, see below) to amplify all contained genomic material within a single cell, and subsequent PCR on the target gene of interest. This enables a multitude of PCR-reactions to be run for each single cell, allowing for Sanger-sequencing of multiple, rather than only single genes (e.g. Krabberød et al. 2011, Bennett & Triemer 2012).

1.1.3 Genomics

Several techniques exist to generate genomic sequences from single cells. The most widespread method at present is multiple displacement amplification (MDA), an approach to whole genome amplification (Lasken & McLean 2014). Here, linear genomic DNA is amplified in an isothermal reaction with a phi29 polymerase and random hexameric primers, resulting in multi-branched networks of amplified DNA (Dean et al. 2001, Fig. 1.1). This creates products with high, but uneven genome coverage, with some loci being overrepresented (Huang et al. 2015, Gawad et al. 2016).

A recurring issue in whole genome amplification techniques is the impact of method-specific errors and biases. For example, uneven coverage of the genome is inherent to MDA—this is problematic since many genome assemblers assume even coverage, ultimately resulting in sub-par, partial assemblies. The use of nanoscale volumes in microfluidic devices (Marcy et al. 2007a, Clingenpeel et al. 2015), combining multiple single amplified genomes (SAGs) from the same species during assembly, and use of assemblers that are specialised in dealing with uneven genome coverage (e.g. SPAdes), can all mitigate these biases to some extent (Lasken & McLean 2014).

The nature of eukaryote genomes, with large genome sizes, repeats, long stretches of non-coding sequences, spliceosomal introns, and sequence composition variation represents another hurdle that has largely remained unresolved, even by using high-quality genomes derived from cultures, rather than SAGs (Keeling & del Campo 2017). Largely for these reasons, single-cell transcriptomics (see below) is increasingly being employed more broadly in protists.

1.1.4 Transcriptomics

Curiously, the first single-cell transcriptomes were generated just 2 years after the first application of RNA-seq on cells derived from cultures (Tang et al. 2009). Most currently available methods for single-cell transcriptomics use three basic steps, with some variations: 1) reverse-transcription of mRNA into cDNA, 2) amplification of resulting cDNA (routinely done using limited-cycle PCR), 3) cleanup of amplified cDNA and library preparation for sequencing on HTS-systems. While a handful of different specific methods for single-cell RNA-seq exist (see Kolodziejczyk et al. 2015, for overview), only one has been used widely with protists: SmartSeq2 (Fig. 1.1). A commercial SmartSeq2 kit is available (Takara Bio SMARTer), but a much

cheaper 'homebrew' version (Ramsköld et al. 2012, Picelli et al. 2013) seems to be more widely employed now (e.g. Kang et al. 2017, Tice et al. 2016, Irwin et al. 2019, Gawryluk et al. 2019). Here, reverse transcription is initiated with poly(T)-priming to capture the mRNA, with second-strand synthesis occurring via the use of a template-switching oligo (TSO) at the 5' end of the mRNA. The transcribed cDNA is then amplified via PCR, yielding enough material to construct a sequencing library for HTS on the Illumina platform (Picelli et al. 2014).

1.1.5 Single-cell Methods in Phylogenetics & Biodiversity Research

Targeted PCR: Diversity & Phylogenetics

As in prokaryotes, applications of single-cell methods in protists have greatly advanced the field of biodiversity by increasing the rate of discovery and placement of important taxa. Single-cell PCR was used in protists for the first time in 2001 (Sebastián & O'Ryan 2001, Marín et al. 2001), providing the means to analyse the 'unculturable'. Since then, the method has found its way into research into most major protist groups, and is now a standard method (Lynn & Pinheiro 2009). Single-cell PCR is especially useful when attempts at culturing a taxon fail, when cells are rarely found in a sample and are thus valuable, or when the target taxon must be co-cultured with other eukaryotes. The SSU rDNA is the most widely used phylogenetic marker in protists, and is overwhelmingly the target for single-cell PCR approaches (e.g. Chantangsi & Leander 2010, Lynn & Pinheiro 2009). And while powerful, each PCR relies heavily on well-matching primers to this locus, which can be problematic (see below). This approach is also usually limited to a single locus per PCR-reaction, but in rare cases has been expanded to several loci (e.g. Takano & Horiguchi 2005).

Methods like environmental sequencing allow for higher throughput and provide a good overview of the diversity (especially abundant taxa), but only sequence relatively short fragments of the SSU rDNA gene (e.g. V4 or V9 regions; Stoeck et al. 2010). Single-cell PCR on the other hand can generate full length SSU rDNA sequences, and has been used successfully to investigate clades that have been described from these environmental sequencing studies. For example Gómez et al. (2011) isolated colonies of a stramenopile epiphyte of diatoms from marine samples, and sequenced their SSU rDNA with single-colony PCR: This taxon actually belongs to MAST-3, a previously identified, environmental-sequence-only clade of marine stramenopiles (Massana et al. 2004). SAGs have also been used to phylogenetically place cytometry-sorted cells among MAST-clades. After sorting and SSU-PCR, cells

that were placed among MAST-clades in SSU rDNA phylogenies had their single-cell genomes sequenced (Roy et al. 2014, Mangot et al. 2017).

Recovery of SSU rDNA Other Than Targeted PCR

While suffering from certain issues in protists (see 'Phylogenomics' below), single-cell genomics and subsequent high-throughput sequencing can be used to generate data from the rDNA operon, including the SSU and LSU. It might sometimes not be possible to obtain SSU rDNA data from novel or divergent organisms by relying on primers, as so called universal eukaryotic primers are seldom universal in practice (Hadziavdic et al. 2014). Rather than relying on primers biased towards a certain taxon (e.g. Amoebozoa: Pawlowski & Burki 2009; and Euglenids: Lax & Simpson 2013), SAGs can provide a primer-independent workaround (e.g. Strassert et al. 2018, Gawryluk et al. 2016). Additionally, WGA enables the recovery of more than a single gene, which can be used in multigene phylogenetics (including phylogenomics). This approach was used to study syndinians (also known as Marine Alveolates: 'MALVs'), a very large assemblage related to typical dinoflagellates that has mainly been characterised by environmental sequence data, with just a few strains in culture (Strassert et al. 2018). While the sequenced SAGs were very incomplete (1–12% BUSCO completion), researchers were able to use the obtained SSU- and LSU-rDNAs in a phylogenetic analysis supporting the paraphyly of MALVs. The WGA-approach was also used to explore the diversity of diplonemids (Gawryluk et al. 2016), a group that has recently been found to be massively abundant in the sunlit pelagic ocean (by environmental sequencing; de Vargas et al. 2015, Flegontova et al. 2016). Sequencing of 10 single-cell genomes yielded highly incomplete assemblies (maximum 9% BUSCO completion), but near full-length SSU rDNA sequences could be extracted and a phylogenetic tree estimated (Gawryluk et al. 2016). Similarly, the SSU rDNA gene can usually be extracted from single-cell transcriptome assemblies (Irwin et al. 2019). For example, single-cell transcriptome or genome data that was initially generated for phylogenomic studies of Amoebozoa (Kang et al. 2017), was subsequently used for more detailed studies of relationships between closely related taxa based on SSU rDNA phylogenetics (*Echinosteliopsis* and *Echinostelium* in Fiore-Donno et al. 2018).

Phylogenomics: Single-cell Genomics

Both single-cell transcriptomics and genomics are powerful tools in helping to increase the rate of discovery and placement of novel protist lineages. The first phylogenetic trees of eukaryotes were estimated using the SSU rDNA gene (Woese et al. 1990), which in several taxonomic groups is long-branching and divergent and, in combination with inadequate phylogenetic methods, led to these taxa being wrongly placed at the base of the eukaryote tree (Burki 2014). With the sequencing of large numbers of genomes and transcriptomes becoming feasible with the introduction of 454 pyrosequencing and later Illumina HTS-systems, many marker genes, rather than just a single or small number of genes (SSU rDNA, or protein-coding genes like tubulins) could be sequenced cheaply. This allowed for phylogenetic analyses of datasets of dozens-to-hundreds of genes: Multigene phylogenetics or Phylogenomics (e.g. Hampl et al. 2009, Zhao et al. 2012, Brown et al. 2013). In combination with the development of increasingly complex phylogenetic models, this enabled estimation of more accurate phylogenetic trees (e.g. Wang et al. 2017, Brown et al. 2018).

Initial amplification of DNA from a single cell with WGA is straight forward, and has been used to generate phylogenomic data of uncultured protists in several groups, with picozoans being a prominent example (Yoon et al. 2011). Here the single-cell genome sequencing data enabled estimation of a phylogeny of 7 protein-coding genes, confirming Picozoa as an orphan taxon without any clear placement within the tree of eukaryotes (Yoon et al. 2011, Brown et al. 2013). Single-cell genomics has been used successfully to expand sampling within the fungal tree of life (Ahrendt et al. 2018, Davis et al. 2019). Like in many protist groups, fungal diversity consists of mostly uncultured taxa, which has been hampering phylogenomic studies to establish relationships and exploration of early traits (Ahrendt et al. 2018). SAGs generated from eight unsampled uncultured species enabled reconstruction of a 192-gene wide phylogeny and inference of ancient lifestyles across fungal groups (Ahrendt et al. 2018, Davis et al. 2019).

Though several studies reported large variations in genome completeness among their SAGs, they did recover more complete genomes by combining multiple SAGs from one species into a single assembly (Ahrendt et al. 2018, Davis et al. 2019, Yoon et al. 2011, Mangot et al. 2017). While it improves gene recovery by mitigating some of the inherent biases of MDA (mostly uneven genome coverage; Worden et al. 2011, Gawad et al. 2016), care must be taken to only bin assemblies of identical, or very closely related taxa from the same population together (Mangot et al. 2017). Contamination of amplified material is also widespread, and stringent contamination

filtering is required before final analysis (Worden et al. 2011). As such, treating SAGs as metagenomes, rather than originating from a single taxon, can be helpful (Krabberød et al. 2017a, Davis et al. 2019).

Phylogenomics: Single-cell Transcriptomics

While more recently developed, single-cell transcriptomics has been adopted very quickly, and is increasingly becoming a standard method to generate phylogenomic-grade data from uncultured protists (e.g. Feng et al. 2020, Kang et al. 2017, Krabberød et al. 2017a). The main advantage over genomics is that assembly is much more straightforward, since only mRNA is sequenced, omitting the large stretches of repeats that make assembly of SAGs difficult (Bankevich et al. 2012). The method also leads to better gene coverage when compared to single-cell genomes, especially for highly expressed genes (Kolisko et al. 2014). A higher recovery of genes is especially desired in phylogenomic applications, whereas highly expressed genes are potentially interesting for ecological questions (Liu et al. 2017).

Increasingly more so than single-cell genomics, single-cell transcriptomics protocols have enabled researchers to investigate parts of the eukaryote tree of life that were difficult to address before. In these parts of the tree this usually means hard-to-culture taxa and/or species that are predatory (i.e. that feed on other eukaryotes). For example, Hehenberger et al. (2017) examined the previously undescribed predators *Syssomonas* and *Pigoraptor*. With the use of a 255-gene phylogenetic analysis they placed both taxa in Opisthokonta, with *Syssomonas* forming the novel group Pluriformea with *Corallochytrium*, and *Pigoraptor* branching within Filasterea. Both taxa were in culture, but large scale RNA-extraction might have been unfeasible due to the eukaryotic prey present in the cultures. Strictly speaking, the researchers did not examine single cells, but rather pools of up to 20 manually isolated cells (in the case of *Pigoraptor*), or 7000 cell-sorted *Syssomonas* cells to generate their data (Hehenberger et al. 2017). Similarly, resolving the phylogenetic relationships within Rhizaria has been hampered by the difficulty of culturing many of these taxa. Krabberød et al. (2017a) used single-cell transcriptomics to improve taxon sampling in Retaria. The 255-gene analysis strongly confirmed the existence of a Retaria clade that includes Radiolaria and Foraminifera, and also the previously undersampled Taxopodida (Krabberød et al. 2017a). These transcriptomes also provided the ability to further investigate the genetic bases of morphological innovations and differences in this morphologically diverse group of eukaryotes.

The main advantage of single-cell transcriptomics over single-cell genomics is that it is easier to assemble the data, while retaining the option of producing high-quality photodocumentation and providing data that can—in terms of number of recovered genes (BUSCO and phylogenomic dataset coverage)—be on par with transcriptomes derived from mass cultures (Kang et al. 2017). This combination enables comprehensive investigations of deep relationships of large protist groups—Kang et al. (2017) and Tice et al. (2016) used a combination of 'bulk' and single-cell transcriptomes to enable multigene phylogenomics of Amoebozoa, collecting data for previously unsampled taxa (61 newly sampled taxa out of 86 included). Multigene phylogenetics with 325 genes revealed a deep split in Amoebozoa, between Tevosa and Discosea (Kang et al. 2017). The availability of high-quality imagery and transcriptome data enables testing of hypotheses regarding the ancient state of the vastly diverse life cycle characters in this group. A similar study was recently done on Arcellinids, a group of testate Amoebozoa with an extensive fossil record (Lahr et al. 2019). Information from 13 single-cell transcriptomes was used in phylogenomic analyses, and aided in reconstructing ancestral morphological states in this group (Lahr et al. 2019).

One known issue with the current iteration of SmartSeq2 is its size limitation: Currently only single cells bigger than 15–20 μm can be effectively used (Kolisko et al. 2014, Liu et al. 2017). This is simply due to the fact that there is less mRNA available in smaller cells, resulting in lower-quality cDNA, and ultimately in less efficient gene recovery and high stochastic variation of transcripts between cells (Liu et al. 2017). A workaround for this is pooling of several individual cells of the same taxon/morphotype into the same tube and generating a transcriptome from this material (e.g. Hehenberger et al. 2017). Similar to SAG-coassemblies (see above), Liu et al suggest generating true single-cell transcriptomes of the organism of interest, and combining their reads at the assembly stage, negating the stochastic variation of transcript recovery to some extent (Liu et al. 2017).

1.2 Euglenids

1.2.1 What are Euglenids?

Euglenids are a complex and diverse group of single-celled flagellates. They are a major subgroup of the taxon Euglenozoa and thus related to Diplonemea and Kinetoplastea. They are well-known mostly because of their charismatic phototrophic species (e.g. *Euglena, Phacus, Trachelomonas*), yet also contain osmotrophic forms (e.g. *Rhabdomonas, Astasia, Menoidium*), and a large diversity of lesser-known phagotrophic species. The latter are particularly interesting, since they are thought to have given rise to both the phototrophs and osmotrophs (Jackson et al. 2018, Maruyama et al. 2011, Turmel et al. 2008). Yet despite their evolutionary relevance, they are not well researched outside of morphological diversity studies. Of more than 1500 described euglenid species, roughly 600 are phagotrophs (Lax & Simpson 2013, Leander et al. 2017), though we know little about both the evolutionary relationships among phagotrophs themselves, and their relationships with phototrophs and other Euglenozoa.

Euglenids can be found in virtually any aquatic environment (Leander et al. 2017), and in some terrestrial systems (Cavalier-Smith et al. 2016). While many phototrophs are found in the freshwater and marine water column, the majority of phagotrophs are surface-associated and glide on a substrate (Leander et al. 2017). Light microscopy-based diversity studies of certain benthic marine and freshwater environments have found a large diversity of euglenids, often in large numbers and representing a major part of the heterotrophic flagellate community (Boenigk & Arndt 2002, Lee & Patterson 2002). This suggests that they could occupy key roles in benthic ecosystems as major consumers of bacteria and/or other protists.

1.2.2 The Euglenid Pellicle is Thought to Represent Euglenid Phylogeny

A submembraneous structure called the euglenid pellicle is unique to euglenids. The pellicle is composed of microtubule-supported articulated proteinaceous strips and varies in terms of number but also structure of pellicle strips among euglenid morphotypes (Leander et al. 2017). Some phagotrophs have 4–12 pellicle strips, whereas some phototrophs have more than 100. Many cells with a large number of strips (>20) can undergo ‘euglenoid motion’, or ‘metaboly’ by sliding their pellicle strips against each other. Interestingly, most phototrophic and many osmotrophic

euglenids have more than 20 pellicle strips and are flexible (or capable of metaboly), whereas phagotrophic euglenids with $\leq$12 pellicle strips are rigid (i.e. not capable of metaboly). This ultimately led to a hypothesis that the number, form, and organisation of pellicle strips reflects euglenid phylogeny: 'primitive' few-strip euglenids gave rise to euglenids with more pellicle strips, which ultimately led to cells with more than 100 strips (Leander et al. 2007). This is thought to have occurred through so-called 'strip-doubling' events, which has been observed in several euglenids during cell division (Yubuki & Leander 2012). Despite the attractiveness of such a model of euglenid evolution, it has been virtually untested with molecular phylogenetic tools (Lax & Simpson 2013).

A commonly used classification separates phagotrophs into bacterivorous and eukaryovorous taxa, with contrasting pellicle forms: rigid cells with 4–12 pellicle strips are thought to consume bacteria, whereas cells with $>$20 pellicle strips are capable of eukaryovory (ingesting other single-celled eukaryotes). While apparently useful, this distinction has obvious limitations, as some few-strip euglenids have been shown to ingest yeast and single-celled algae (Linton & Triemer 1999, Lee & Patterson 2000). As with the organization of the pellicle, the separation into bacterivorous vs. eukaryovorous species has been barely explored with molecular phylogenetic approaches.

1.2.3 Phagotrophs are Surface-associated

Locomotion among euglenids is also particularly diverse: Most phototrophs swim using one or two flagella, whereas phagotrophic euglenids mostly glide on substrate using unknown motor systems localised to the flagella (Leander et al. 2017). Further, phagotrophs glide employing a variety of different modes that involve different flagella. The main distinction is between anterior and posterior gliders (Cavalier-Smith 2016): Some anterior gliders have a single emergent flagellum (e.g. *Petalomonas, Urceolus*) that is directed anteriorly, with most of the flagellum in contact with the substrate, whereas anterior gliders with two emergent flagella (e.g. *Notosolenus, Sphenomonas, Peranema*) trail their often-short posterior flagellum, while the anterior flagellum is inferred to power the gliding (Saito et al. 2003). Posterior gliders (e.g. *Ploeotia, Anisonema, Dinema*) often freely beat their anterior flagellum in front of the cell, while the posterior flagellum, which is typically longer, is used to move the cell along the substrate. The emergence and evolutionary relevance of these patterns of flagellar movement and locomotion have not been examined in depth—for example

Petalomonas and *Urceolus* are thought to be not closely related, yet both are anterior gliders with a similar pattern of movement. Finally some (e.g. *Neometanema*) 'skid' along the surface powered by beating of the anterior flagellum rather than a gliding process, with their posterior flagellum passively dragging along the surface (Lax & Simpson 2013, Lee & Simpson 2014b, Larsen & Patterson 1990).

1.2.4 Symbiontida and Hemimastigophora – Derived Euglenids?

Over the years, several other groups have been thought to be derived euglenids, or closely related to them (Simpson 1997). Symbiontids are one of these: They have been found in anaerobic environments and the first description of a member of this group (*Calkinsia aureus*) was as a euglenid (Lackey 1960). In recent decades though, this placement has been viewed critically, and they have rather been more considered more generally as euglenozoans, based on features like the presence of paraxonemal rods, tubular extrusomes, and a unique ultrastructural organization of the feeding apparatus (Simpson et al. 1997, Yubuki et al. 2013). Their name is derived from the fact that they are covered by rows of one or more kinds of episymbiotic bacteria (Yubuki et al. 2009, Monteil et al. 2019). A plethora of short-read SSU rDNA sequencing data are available from this group, yet full-length SSU rDNA sequences combined with high-quality imagery are available for just two species (*Calkinsia aureus* and *Bihosphites bacati*). This single-gene data places symbiontids firmly within Euglenozoa, yet unfortunately is insufficient to determine their exact position, in particular whether they are derived euglenids, or a distinct branch within Euglenozoa (Yubuki & Leander 2018, Cavalier-Smith 2016, Breglia et al. 2010, Yubuki et al. 2009). No culture of a symbiontid has been established, limiting options to generate additional molecular data.

The Hemimastigophora are an enigmatic group of multi-flagellated, free-living predators that have been known since the late nineteenth century, yet had escaped molecular sequencing efforts (Klebs 1893, Foissner et al. 1988). They had never been brought into culture, yet morphological features collected by ultrastructural study (Foissner et al. 1988, Foissner & Foissner 1993) and light microscopy lead researchers to place them in several supergroups widely spread across the tree of eukaryotes (Foissner et al. 1988, Cavalier-Smith 1998, 2000, Cavalier-Smith et al. 2008). In particular, the first modern study of Hemimastigophora suggested that they might be closely related to euglenids (Foissner et al. 1988). This was largely based on the presence of two pellicular plates in *Hemimastix*, which were inferred

to be homologous with euglenid pellicle strips. Yet without any sequence data, the placement of this group among other eukaryotes remained unresolved, including whether they are at all related to euglenids.

1.2.5 Few Molecular Sequences are Available for Phagotrophic Euglenids

Many of the issues associated with euglenid taxonomy and understanding their evolutionary history exist because molecular phylogenies of this group are poorly sampled and poorly resolved. Phototrophic euglenids, and to a lesser extent primary osmotrophs, are well-sampled (e.g. Bicudo & Menezes 2016, Karnkowska et al. 2015, Linton et al. 2010, Marin et al. 2003), whereas phagotrophs suffer from poor sampling of any molecular data (Lax & Simpson 2013, Cavalier-Smith et al. 2016, Paerschke et al. 2017, Schoenle et al. 2019). Most of the molecular data that are available comes from the SSU rDNA gene, yet in euglenids (as in certain other groups like Amoebozoa), this gene is often highly divergent, making phylogenetic estimation difficult (Busse et al. 2003, Łukomska-Kowalczyk et al. 2016). The divergent nature of this gene in euglenids can also make acquiring sequences difficult, as 'universal' eukaryotic PCR primers often do not work, and euglenid-biased primers have to be used (Busse et al. 2003, Lax & Simpson 2013). Furthermore few environmental sequencing surveys pick up any phagotrophic euglenids, most likely for the same reason (Kolisko et al in prep.).

Few permanent cultures of phagotrophic euglenids have existed over the years, and in many cases in the period leading up to the molecular era, isolated wild cells were studied (e.g. Leander & Farmer 2001), or only a semi-permanent culture maintained for some months (e.g. Schnepf et al. 2002). While there are multiple reasons for this, two chief ones are the historical focus on phototrophic taxa, and the difficulty of establishing and maintaining a culture of a protist that might have a narrow selection of prey. Obviously, this lack of cultures has also contributed to the poor sampling of molecular sequences.

Solely morphology-based taxonomy is still the norm in phagotrophic euglenids. Worryingly though, many of the morphological characters used might be highly variable and thus over-generalized: For example a feeding apparatus that is visible with light microscopy is one of the features traditionally distinguishing *Ploeotia* from *Anisonema* (Larsen & Patterson 1990), yet some *Anisonema* cells with a feeding apparatus have been found (e.g. Lee 2012, Al-Qassab et al. 2002). In many cases the

taxonomic decisions based on morphological characters might actually reflect phylogenetic relationships, but this remains mostly untested with molecular phylogenetic tools due to insufficient taxon sampling.

1.2.6 A Molecular Single-cell Approach to Euglenids

In summary, evolutionary relationships among euglenids are largely unresolved, especially deep relationships towards the base of the euglenid tree. The main reason for this is poor taxon sampling among phagotrophic euglenids, even for the SSU rDNA gene, and few cultures exist from which to obtain the missing data. To circumvent the need for culturing, and thus to quickly increase sampling for molecular data, I employed a single-cell centric approach throughout my thesis. This methodology also enabled the capture of high-quality microscopy imagery, which is crucial in linking morphological and molecular data. Chapter 2 focusses on the phylogenetic diversity of 'ploeotids', a diverse paraphyletic assemblage of rigid cells close to the base of the euglenid tree, using a morphological and molecular approach to generate novel SSU rDNA sequencing data. Chapter 3 enhances the molecular sampling of the diversity of Spirocuta by sequencing novel SSU rDNA data. It also reports the first multigene analysis that includes phagotrophic euglenids, aiming to resolve some of the deep relationships within the euglenid evolutionary tree, including the position of Symbiontids. This latter work employed single-cell transcriptomics to generate the bulk of the analysed data. Chapter 4 reports the first molecular data for Hemimastigophora, principally based on transcriptomes derived using single-cell methods from two species, and resolves their phylogenetic position through a eukaryote-wide phylogenomic analysis.

Chapter 2

Ploeotids Represent Much of the Phylogenetic Diversity of Euglenids

A version of this chapter has been published in:
Lax G, Lee WJ, Eglit Y, Simpson A (2019).
Ploeotids represent much of the phylogenetic diversity of euglenids.
Protist **170**:233–257.
doi:10.1016/j.protis.2019.03.001

Authors' contributions: GL and AS conceived project idea, GL isolated, identified, imaged and generated molecular data from single cells. YE and WJL isolated cultures. GL conducted phylogenetic analyses, and wrote manuscript with input from AS and other authors.

¶

2.1 Abstract

PLOEOTIDS are an assemblage of rigid phagotrophic euglenids that have 10–12 pellicular strips and glide on their posterior flagellum. Molecular phylogenies

place them as a poorly resolved, likely paraphyletic assemblage outside the Spirocuta clade of flexible euglenids, which includes the well-known phototrophs and primary osmotrophs. Here, we report SSU rRNA gene sequences from 38 ploeotids, using both single-cell and culture-based methods. Several contain group I or non-canonical introns. Our phylogenetic analyses place ploeotids in 8 distinct clades: *Olkasia* n. gen., *Hemiolia* n. gen., *Liburna* n. gen., *Lentomonas*, *Decastava*, *Keelungia*, Ploeotiidae, and *Entosiphon*. *Ploeotia vitrea*, the type of *Ploeotia*, is closely related to *P. oblonga* and *Serpenomonas costata*, but not to *Lentomonas*. *Ploeotia* c.f. *vitrea* sensu Lax and Simpson 2013 is not related to *P. vitrea* and has a different pellicle strip architecture (as imaged by scanning electron microscopy): it instead represents a novel genus and species, *Olkasia polycarbonata*. We also describe new genera, *Hemiolia* and *Liburna*, for the morphospecies *Anisonema trepidum* and *A. glaciale*. A recent system proposing 13 suprafamilial taxa that include ploeotids is not supported by our phylogenies. The exact relationships between ploeotid groups remain unresolved and multigene phylogenetics or phylogenomics are needed to address this uncertainty.

2.2 Introduction

Euglenids (euglenoids) are a diverse group of flagellates distinguished by the euglenid pellicle—a structure underneath the cell membrane that is composed of 4 to >100 abutting proteinaceous strips underlain by microtubules (Leander et al. 2001b). Cells with ≥ 16 strips are usually flexible, and often capable of dramatic cell-shape changes called 'euglenoid motion' or 'metaboly', while those with 12 or fewer strips are rigid (Leander et al. 2007). Euglenids include organisms with phototrophic, osmotrophic and phagotrophic nutritional modes (Leander et al. 2017). Phagotrophy is the ancestral mode, and some kind of flexible phagotrophic euglenid was the likely host in a secondary endosymbiosis with a member of Pyramimonadales to give rise to phototrophic euglenids (Turmel et al. 2008). This occurred around 600 MYA, according to recent estimates (Jackson et al. 2018).

Most phagotrophic euglenids are gliding cells that associate with surfaces. They are found in almost all aquatic habitats (Larsen & Patterson 1990, Patterson & Simpson 1996, Schroeckh et al. 2003, Lee 2012), and in some cases represent the major group of heterotrophic flagellates by biomass (Dietrich & Arndt 2004, Boenigk & Arndt 2002, Lee & Patterson 2002). Phagotrophic euglenids can be crudely divided into three categories: Flexible forms with ≥ 18 strips (e.g. *Peranema, Urceolus, Neometanema, Anisonema*); petalomonads, which are rigid cells, usually with 4–10

strips that glide on their anterior flagellum (e.g. *Petalomonas, Notosolenus, Sphenomonas*); and ploeotids, rigid cells with 10 or 12 strips, which glide on the posterior flagellum (e.g. *Ploeotia, Entosiphon, Keelungia*). The anaerobic symbiontids are sometimes considered phagotrophic euglenids that lack a complete pellicle, but there has not yet been a definitive phylogenetic placement of this group (Adl et al. 2019, Yubuki et al. 2013, Lax & Simpson 2013, Cavalier-Smith 2016).

The phototrophs and osmotrophs are relatively well characterised in terms of phylogenetic relationships and biodiversity. By contrast, our current understanding of phagotrophic euglenid phylogeny and taxonomy is fragmentary, with the major source of phylogenetic knowledge being species-poor SSU rDNA phylogenies. Numerous analyses have shown that flexible phagotrophic euglenids are closely related to phototrophic and osmotrophic euglenids. This clade of euglenids, characterised by having many helically-arranged strips, was usually referred to as 'H' or 'HP' and recently formalised as 'Spirocuta' or 'Helicales' (Busse et al. 2003, Lee & Simpson 2014b, Cavalier-Smith et al. 2016, Paerschke et al. 2017). Petalomonads and symbiontids are also both apparently monophyletic (Lax & Simpson 2013, Cavalier-Smith et al. 2016). Ploeotids, on the other hand, are invariably not recovered as a clade, instead forming a paraphyletic group that seemingly gave rise to Spirocuta, and possibly petalomonads and/or symbiontids as well (Lax & Simpson 2013, Chan et al. 2015, Cavalier-Smith 2016, Paerschke et al. 2017). This identifies ploeotids as a key assemblage for understanding the early evolutionary history of euglenids.

Ploeotids are 7–60 μm long, often 15–25 μm. They are frequently characterised as bacterivorous (Leander et al. 2001a, Leander 2004, Leander et al. 2007), though there are some reports of ingested eukaryote cells (e.g. Linton & Triemer 1999, Lax & Simpson 2013). Cells glide on the thickened, sometimes very long, posterior flagellum. Many cells also exhibit a 'jerking back' motion with their posterior flagellum, similar to the spirocute *Anisonema* (e.g. Al-Qassab et al. 2002). The anterior flagellum is directed forward and either sweeps from side to side or is just held in front, and likely is used to detect prey. Ploeotids (see below for taxonomy) are mostly found in marine environments and almost always have 10 pellicle strips, though the arrangement and fine structure of strips differs considerably. For example, *Ploeotia vitrea* has 10 similar strips that meet at sharp keels, whereas *Serpenomonas/Ploeotia costata* has 5 broad strips alternating with 5 very narrow strips that lie within grooves. *Lentomonas* has 7 humped strips that feature prominently on the dorsal side, while the three ventral strips are flat. *Keelungia* and *Decastava* both have 10 similar relatively flat strips around the cell body (Chan et al. 2013, Cavalier-Smith et al. 2016).

Entosiphon is unusual in often or always having 12 pellicular strips that mostly alternate in form (Triemer & Fritz 1987, Larsen & Patterson 1990).

The genus *Ploeotia* itself was introduced in 1841 when Dujardin described *Ploeotia vitrea* (Dujardin 1841), but was then mostly forgotten until the mid 1980s. Over the next two decades the genus was 'redescribed' by Farmer & Triemer (1988) using transmission and scanning electron microscopy data, and ~20 additional species were described via light microscopy (e.g. Larsen & Patterson 1990, Patterson & Simpson 1996, Al-Qassab et al. 2002). Over the same time, two additional genera were introduced based on light- and electron microscopy data. *Serpenomonas* (type species *Serpenomonas costata*) was established as a genus without considering *Ploeotia* (Triemer 1986) and was shortly after merged with *Ploeotia* (Farmer & Triemer 1988). *Lentomonas* was discriminated from *Ploeotia* on the basis of ultrastructural data (Farmer & Triemer 1994), which came from an organism indistinguishable from *Ploeotia corrugata*, described slightly earlier (Larsen & Patterson 1990). Consequently, *Lentomonas* was often also treated as a junior synonym of *Ploeotia* (Ekebom et al. 1995, Patterson & Simpson 1996), leading to a 'lumping' situation where all ploeotids other than *Entosiphon* were assigned to a single genus. Prior to 2013 the only molecular data available for ploeotids was from *Ploeotia/Serpenomonas costata*, and from *Entosiphon* strains with extremely rapidly evolving SSU-rRNA genes (von der Heyden et al. 2005, Paerschke et al. 2017). Recent phylogenies that include a couple of additional species have shown the ploeotids to be divergent one from other (Chan et al. 2013, Lax & Simpson 2013, Cavalier-Smith et al. 2016). This has led to taxonomic 'splitting' at the genus level, including the description of two new genera, *Keelungia* and *Decastava* (Chan et al. 2013, Cavalier-Smith et al. 2016), and moves to again recognise *Serpenomonas* and *Lentomonas* as distinct from *Ploeotia* (Cavalier-Smith 2016). Cavalier-Smith (2016) has further proposed a highly detailed assignment of ploeotid genera into families, orders, subclasses, classes and superclasses. Nonetheless, prior to our study SSU-rRNA genes were available for only four nominal ploeotid species outside of *Entosiphon*: *Ploeotia/Serpenomonas costata* (Busse & Preisfeld 2003b, Chan et al. 2015), *Keelungia pulex* (Chan et al. 2013), *Decastava edaphica* (Cavalier-Smith et al. 2016, Paerschke et al. 2017); the sequences reported under the names *Decastava edaphica* and *Ploeotia edaphica* are derived from the same culture and are 99.3% identical) and a cell identified as *Ploeotia* c.f. *vitrea* (Lax & Simpson 2013), plus HSP90 sequences from just three species (Breglia & Leander 2007, Cavalier-Smith et al. 2016). Crucially, the type species of *Ploeotia, P. vitrea,* has never been examined using molecular methods. It was recently

assumed that its placement was to follow *Ploeotia* c.f. *vitrea* (Cavalier-Smith et al. 2016), notwithstanding an explicit caveat that this would require confirmation (Lax & Simpson 2013). In fact, this assumption was used to justify placement of *P. vitrea* and *S./P. costata* in separate classes within the taxonomy discussed above (Cavalier-Smith 2016, Cavalier-Smith et al. 2016). There has not been any molecular data for *Lentomonas*.

Considering the genus *Anisonema* adds an additional complication. *Anisonema* currently includes the type species, *Anisonema acinus*, and ~20 other nominal species. Many of these are similar to ploeotids in the arrangement of their flagella. As with several other phagotrophic euglenid genera, it is unclear how many of these morphospecies actually belong to *Anisonema*, since the genus assignments are partly based on questionable morphological characters, such as the visibility of the feeding apparatus with light microscopy (Larsen & Patterson 1990). The only molecular data available up to now has been SSU rRNA gene sequences from several *A. acinus* (-like) populations/cells, which prove to belong to Spirocuta (Busse et al. 2003, Lax & Simpson 2013). Unfortunately, the absence of sequence data from other morphotypes has made it impossible to test whether all nominal *Anisonema* species belong to Spirocuta. Notably, the morphotypes *A. trepidum* and *A. glaciale* have not been reported to show flexibility, and both differ noticeably from *A. acinus* in their locomotion and flagellar movement patterns (Larsen 1987, Larsen & Patterson 1990).

This study aims to examine the broad-scale molecular diversity of ploeotids, to better understand their phylogenetic affinities and clarify their systematics. We established 10 cultures representing the morphospecies *Ploeotia vitrea*, *Ploeotia / Serpenomonas costata*, *Ploeotia oblonga*, and *Keelungia* sp. We also derived SSU-rRNA gene sequences from 27 photodocumented single cells, including *Lentomonas* morphotypes, as well as cells identified morphologically as *Anisonema glaciale* and *A. trepidum*. Phylogenetic analyses show that ploeotids *sensu lato* (i.e. including *Entosiphon* and *Ploeotia*-like anisonemids) represent at least 8 major molecular lineages. Crucially, *P. vitrea* is not specifically related to *P.* c.f. *vitrea* sensu Lax & Simpson (2013), and instead is closely and strongly related to *S./P. costata* (and *P. oblonga*). This refutes the notion that *Serpenomonas* and *Ploeotia* should belong to different high-rank taxa. We show by SEM that the pellicle structure of *Ploeotia* c.f. *vitrea* is distinctly different from *Ploeotia vitrea*, and based on the combined molecular and morphological data propose a new genus and species, *Olkasia polycarbonata* n. gen. n. sp., for *Ploeotia* c.f. *vitrea* sensu Lax & Simpson (2013). *Anisonema trepidum* and *A. glaciale*-like cells clearly branch outside of Spirocuta—since they are phylogeneti-

cally distinct from the genus *Anisonema*, we propose the new genus *Hemiolia* n. gen. for *Anisonema trepidum* Larsen 1987; and *Liburna* n. gen. for *Anisonema glaciale* Larsen and Patterson 1990.

2.3 Methods

2.3.1 Cultures: Establishment and Nucleic Acid Extraction

Marine intertidal sediments were used to establish crude cultures, with sterile seawater supplemented with 1–3% LB (v/v) as the initial medium (sample sites yielding euglenid cultures are reported in Table A.5). Cultures of *Serpenomonas costata*, *Ploeotia vitrea*, *P. oblonga* and *Keelungia* sp. were obtained by serial dilution of crude cultures, or isolation of single cells with a pipette. After establishment, cultures were maintained in tissue culture flasks in ~10 ml sterile seawater supplemented either with 0.1–1% LB media (v/v) or a sterile barley grain, with transfers every 2–3 weeks. Cells of *Entosiphon* sp. were isolated from freshwater sediment into tap water to form a crude culture. The culture was then supplemented with *Haematococcus* sp. as prey, and transferred every 2–24 months. Cultures were imaged under coverslips with a Zeiss Axiovert 200M and AxioCam M5.

DNA from all cultures was isolated with a Qiagen DNAeasy Blood and Tissue kit, using the tissue protocol, and quantified via spectroscopy with a NanoDrop (Thermo Scientific). We also isolated mRNA from three cultures either using a TRIzol-based extraction, following the manufacturers' instructions (*S. costata* HAK-MF and *P. vitrea* MX-CHA; Thermo Life Sciences), or RNA spin-columns (*Keelungia* sp. culture KM082; Macherey-Nagel NucleoSpin RNA XS). Reverse transcription to cDNA was carried out with a template-switching oligo, as described by Picelli et al. (2014).

2.3.2 Single Cells: Photodocumentation and DNA Amplification

Preparation of marine sediment samples largely followed Larsen & Patterson (1990). Briefly, sediment was collected from various sites across North America (Table A.5), placed into small trays, and spread out to 1–2 cm thickness. Kimwipe tissue was added on top, then 50x20 mm glass coverslips. A transparent lid was added to reduce evaporation. After 12–72 h under ambient within-laboratory day-night cycle, the coverslips were examined bottom facing up with seawater added, on an inverted microscope (either Zeiss Axiovert 200M under 1000x; or Leica DM IL LED under 400x or 630x total magnification). Individual cells were imaged (with a Zeiss AxioCam M5

or a Sony NEX6, respectively), and then isolated by mouthpipetting with a drawn-out glass pipette. Cells were washed 3–5 times in 2 µl drops of sterile seawater under the microscope before being expulsed into separate 0.2 ml PCR tubes containing 9.5 µl PCR-grade distilled water. Cells were lysed by up to 10 freeze-thaw cycles (-80˚C and RT) and DNA was amplified by multiple displacement amplification (MDA), using an Illustra GenomiPhi v3 kit (GE Healthcare), following the manufacturer's instructions, but with the 30˚C isothermal amplification time extended to 2 h. Success of the MDA reactions was assessed by running 1 µl of the product on a 0.5% agarose gel.

2.3.3 PCR Amplification and Sequencing

SSU rDNA fragments were amplified from DNA from cultures and cells by PCR, using a variety of different euglenid/ploeotid-biased primers (Table A.1). Primers biased towards ploeotids (including *Hemiolia* and *Liburna*) were designed by modifying previously published euglenid-biased primers (Busse et al. 2003, Lax & Simpson 2013), or scanning alignments of existing SSU rDNA sequences for suitable conserved sites. Amplifications were carried out with Invitrogen Recombinant *Taq*, 2 mM $MgCl_2$, and 0.2 mM dNTPs. Initial denaturing was done at 95˚C for 3 min; then 35 cycles of: denaturing at 95˚C for 30 sec, annealing at 50–64˚C (see Table A.1 for primer combinations) for 30 sec, elongation at 72˚C for 2 min; and a final elongation step at 72˚C for 10 min. Products were visualised with gel-electrophoresis on a 1% agarose gel, and either sent directly for Sanger sequencing (Génome Québec) or first gel-extracted with a Qiagen Gel-Extraction kit. Almost full-length SSU rDNA sequences were obtained for most cells/cultures by amplifying at least two overlapping fragments. Raw reads were quality-checked by eye and automatically assembled de novo using Geneious R10 (Kearse et al. 2012), then queried against the NCBI GenBank nr database to identify any contaminant sequences. For one cultured strain, *P. vitrea* MX-CHA, only a partial SSU rDNA sequence was obtained by PCR amplification, but a full-length SSU rRNA was extracted from transcriptome data from the same strain. The methods for this Illumina sequencing will be reported elsewhere (Lax et al. unpublished).

2.3.4 Alignment and Phylogenetic Analyses

The new sequences were added to a seed alignment including 34 phagotrophic euglenid sequences available on GenBank, a phylogenetically broad selection of primary osmotrophic and phototrophic euglenids, and symbiontids. Representative se-

quences from diplonemid and kinetoplastids were included as outgroups. To exclude potential long-branch attraction artifacts we also created datasets without *Entosiphon* and/or without outgroups (four datasets total). To exclude potential reduced resolution due to short sequences, we created a fifth dataset without sequences <1000 bp. In a sixth dataset, we excluded CARR5, SMS7, WF2_3, *Heteronema/Teloprocta scaphurum*, and 13 other sequences, since they were identified as rogue taxa by RogueNaRok (Aberer et al. 2013) under the RNR algorithm. The base dataset was aligned with MAFFT E-INS-I (Katoh & Standley 2013), checked manually with AliView (Larsson 2014), and masked by eye with SeaView (Gouy et al. 2010) to exclude ambiguous positions for all five taxon selections. This yielded a 1276 nt trimmed alignment for the datasets with all taxa included (156 taxa) and all taxa minus outgroup (123 taxa). A 1360 nt trimmed alignment was generated for the datasets without *Entosiphon* (151 taxa); without *Entosiphon* and outgroups (118 taxa); and without *Entosiphon*, short sequences; and rogue taxa (100 taxa). The 'all-taxa' dataset was also automatically trimmed with trimAl to test whether any major results could stem from user bias in site masking (-st 0.001 -gt 0.89; 1354 sites; Capella-Gutiérrez et al. 2009). For each dataset, Maximum Likelihood phylogenies were inferred with RAxML under the GTR + Γ model (Stamatakis 2014), with robustness assessed with 1000 bootstraps for each analysis. We also carried out Bayesian analyses on the main dataset and on the dataset with *Entosiphon* and with outgroups. MrBayes (Ronquist et al. 2012) was used under the GTR + Γ model, running 4 chains (default heating parameters) for 5,000,000 generations each, with trees sampled every 1000 generations and the first 25% discarded as burn-in. Convergence was confirmed by assessing that PRSF values (Potential Scale Reduction Factor) approached 1.0. The 'all-taxa' ML-derived trimAl tree was very similar to the dataset masked by hand, with none of the differing bipartitions receiving more than 50% BS support in either analysis.

Potential group I introns were identified manually by looking for conspicuous insertions in the alignment and then examining those sequences with RNAweasel (`http://megasun.bch.umontreal.ca/RNAweasel/` Accessed Nov. 5 2018). Conspicuously, RNAweasel did not identify possible introns in isolate ABF1 (*Serpenomonas costata*) or culture SJB2 (*Ploeotia vitrea*; despite these having 406 and 510 bp, and 402 bp insertions, respectively). We subsequently generated cDNA-derived SSU sequences for HAK-MF (*S. costata*), MX-CHA (*P. vitrea*), and KM082 (*Keelungia* sp.; see above). Sequences from both DNA and cDNA were aligned and compared, to confirm that the insertions were excised from the rRNA. To investigate the presence of homing endonuclease (HE), we blasted the intron sequences using BLASTx against

the GenBank nr database.
Pairwise sequence identities were calculated from whole (unmasked) sequence alignments, but with intron sequences excluded. Partial sequences were excluded from reporting, unless specifically noted (all scores can be found in Table A.3).

2.3.5 Scanning Electron Microscopy

50–100 µl of cells from ~2 week old cultures of *Serpenomonas costata* strain HAK-MF, *Ploeotia oblonga* strain CAS1, and *Ploeotia vitrea* strains BoP4.1PV and MX-CHA were transferred onto poly-L-lysine-coated 12 mm round coverslips, and immediately fixed with a drop of 25% glutaraldehyde and OsO_4-vapor, for 1 h. After fixation, the coverslips were washed 3x in filtered seawater or dH_2O, and then subjected to a dehydration-series of ethanol-water mixtures, as follows: 30%, 50%, 70%, 80%, 90%, 95%, 100% (3x). This was followed by critical-point drying with CO_2 on a Leica EM CPD300, then a ~15 nm Au-Pd coat was added with a Leica EM ACE200 sputter-coater. Samples were imaged on Hitachi S4700 or Zeiss LEO 1455VP scanning electron microscopes.

Single cells of *Olkasia polycarbonata* were isolated by pipette from fresh samples, and dropped onto poly-L-lysine coated coverslips with ~50 µl of filtered seawater and a drop of 25% glutaraldehyde. Fixation, dehydration and imaging followed the same procedure described above. In order to associate these cells with a molecular identity, some cells of *O. polycarbonata* designated for molecular work (see above) were isolated from the same samples at the same time (i.e. presumably from the same populations).

2.4 Results

2.4.1 Studied Isolates

The organism codes, the assigned taxa and GenBank accession codes for all studied cultures/cells can be found in Table A.1.

Ploeotia vitrea

Three cultures of *Ploeotia vitrea* were established, with an additional single cell isolated from a sediment sample (isolate STS2, Table A.1). Cells are oval, with a pointed posterior end and a conspicuous hook-shaped feeding apparatus that extends almost

down the complete length of the cell (Figures 2.1a–b, 2.2a). The 10 pellicle strips are roughly evenly spaced and raised at their edges to form characteristic keels, or 'double-raised ridges', that are readily visible with light microscopy when viewed in grazing optical section or cross-section (Fig. 2.1c). SEM confirms that the small grooves at the connections between strips run along the spines of these keels (Fig. 2.2a–d). The central ventral strip is narrower than other strips (Fig. 2.2b). Every second strip is slightly shorter than the adjacent strips, such that the keels of the strips come together in a 5-point star at the posterior end of the cell (Fig. 2.2a, d). Movement was typical of ploeotids: cells glide with their posterior flagellum attached to the surface and trailing behind, whereas the anterior flagellum sweeps from side to side in front of the cell (see Video 1). Feeding on bacteria was observed in cultures: if the anterior flagellum encounters a suitable prey, the whole cell pulls close to it with the anterior flagellum attached to the bacterium. The *P. vitrea* cell then tips itself over its prey. Measurements of cultured cells are listed in Table 2.1. In addition to bacteria, culture MX-CHA was tested and found to be able to grow on *Phaeodactylum* sp. (a pennate diatom), in which case cells often contained pigmented ingesta 3.7–6.7 μm in diameter (n = 31), and some cells contained whole diatoms (20.5–25.9 μm long, 4.1–4.2 μm wide; n = 3).

Table 2.1: Morphological measurements of ploeotid cultures, with mean length and width (including standard deviations), and mean relative anterior and posterior flagellum lengths, derived from 30 cells each.

species	strain	length	width	ant. flagellum	post. flagellum
Ploeotia vitrea	BoP3.3P1 [1]	17.5–24.9 μm	11.8–14.1 μm	0.6x	2.65x
	SJB2	18.3 μm (±1.7)	12.5 μm (±2.1)	0.8x	2.75x
	MX-CHA	20.4 μm (±1.6)	14.6 μm (±1.9)	0.7x	2.25x
Ploeotia oblonga	CAS1	20.4 μm (±2.2)	14.6 μm (±2.1)	0.7x	2.3x
Serpenomonas costata	BOP4.1N3	18.4 μm (±1.1)	10.9 μm (±1)	0.7x	2x
	HAK-MF	19.2 μm (±2)	12 μm (±1.7)	0.7x	2.1x
	KM040	19.3 μm (±0.8)	11.1 μm (±0.9)	0.7x	2.4x
	KM057	20.6 μm (±1.2)	14.3 μm (±1.3)	0.7x	1.6x
Keelungia sp.	KM082	10.7 μm (±0.8)	6.7 μm (±0.8)	1.25x	3x
Entosiphon sp.	ESC [2]	21.2 μm (±1.7)	10.1 μm (±1.4)	1x	2.9x

Ploeotia oblonga

We established one culture of *Ploeotia oblonga*, strain CAS1. Cells are oblong, with the hook-shaped feeding apparatus extending down almost the full length of the cell (Figures 2.1f–g). The 10 pellicle strips are parallel to the cell outline, but like in *Ser-*

[1] Measurements for BoP3.3P are based on only 2 cells, since the culture was lost before more measurements could be taken.

[2] Measurement for ESC are based on nine cells. Individual measurements in Table A.1.

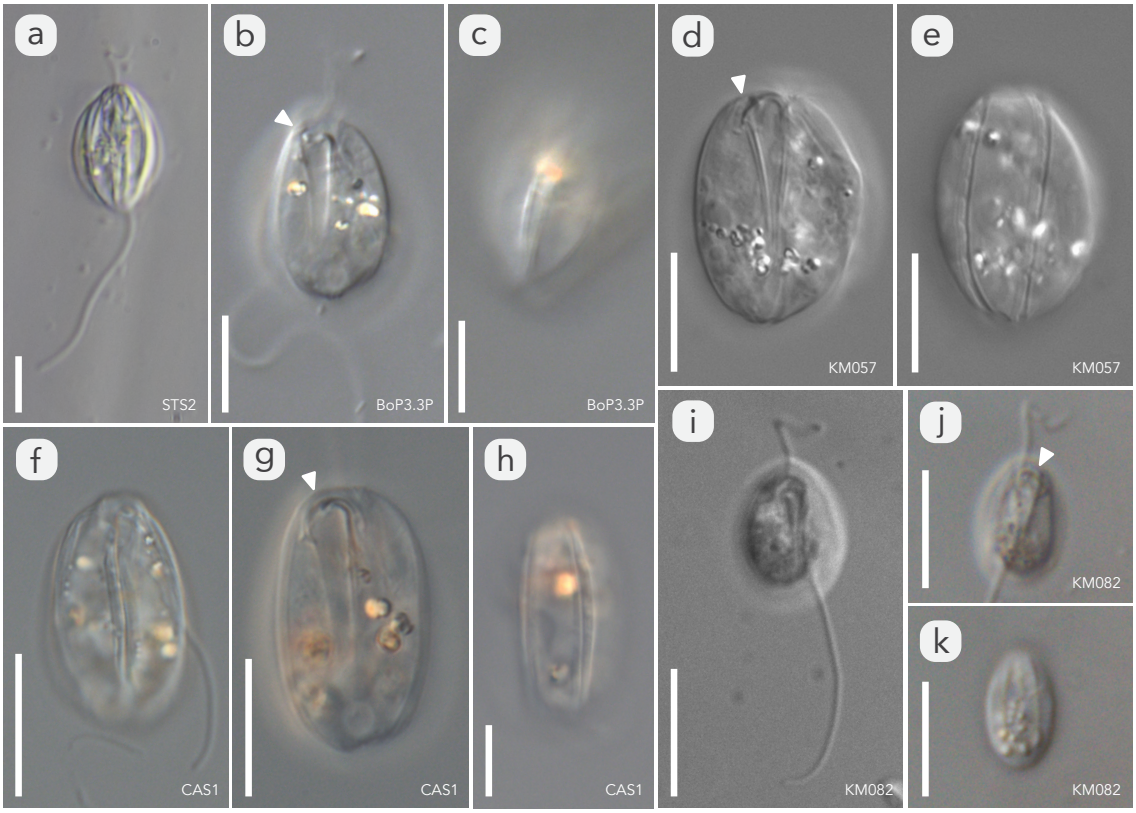

Figure 2.1: Light micrographs of ploeotid taxa derived from cultures and single-cells. a–c) *Ploeotia vitrea*. c) shows pellicle strip arrangement at posterior end. d–e) *Serpenomonas costata*. Note undulating edges of pellicle strips in e). f–h) *Ploeotia oblonga*. Pigmented ingesta are ~1.5 µm in diameter. i–k) *Keelungia* sp. strain KM082. Arrowheads in all images show feeding apparatuses. Scale bars are 10 µm. Isolate names are shown in image. All images were acquired with differential interference contrast optics.

penomonas costata (see below), alternate between narrow and broad strips (Figures 2.1f, h and 2.2e), with the central-most ventral strip being narrow (Fig. 2.2f). In SEM, the fine structure of the pellicle is revealed to be similar to *P. vitrea* in having the boundaries between adjacent strips raised on keels, though these are lower and broader than in *P. vitrea* (Fig. 2.2g). The keels come together in a 5-pronged star-shaped pattern at the posterior end of the cell, similar to *P. vitrea* (Fig. 2.2h). Movement of cells is similar to *P. vitrea* and *S. costata* (Video 1 and 2). Cells of CAS1 are capable of ingesting a coccoid alga ~1.5 µm across (bright inclusions in Fig. 2.1f–h). Morphological measurements are listed in Table 2.1.

Serpenomonas costata (Ploeotia costata)

We established four cultures of *Serpenomonas costata*, and isolated one cell from a sample (isolate ABF1). Cells are oval and have a conspicuous hook-shaped feeding

apparatus (Fig. 2.1d). The 10 pellicle strips alternate in size, such that cells appear to only have 5 strips separated by deep grooves when viewed with light microscopy (Fig. 2.1e), but each groove actually houses most of another narrow pellicle strip, as shown by SEM (Fig. 2.3a–c). Characteristically, SEM shows that the visible part of each narrow strip has an undulating edge that extends over the groove (Fig. 2.3c). Movement of cells was very similar to that of *Ploeotia vitrea* (Video 2). Feeding on bacteria was observed in culture and was the same as in *P. vitrea* (see above). Measurements of individual cultures are listed in Table 2.1. In addition to bacteria, strain KM040 was tested and found to be able to grow on *Phaeodactylum* sp., with most cells containing pigmented ingesta ranging from 3.5–6.6 μm in diameter (n = 16). A few cells (all moribund at time of observation) contained whole diatoms (21.1–33.6 μm long, 3.9–7.6 μm wide; n = 3). Sequence identity within clade A (see below; without intron sequences) is 97–99.4%, and is 98–99.5% within clade B. Sequence identity between members of clades A and B is 73.3–75.8% (Table A.3).

Keelungia sp. KM082

We established a culture of *Keelungia* sp. (strain KM082) and sequenced its full-length SSU-rRNA gene. Cells are oblong to ovoid in profile, not flattened, and 9–13 μm long and 4.6–8 μm wide (Fig. 2.1i). The hook-shaped feeding apparatus is conspicuous, broad on the anterior end and tapers considerably while extending 3/4 or more down the cell (Fig. 2.1i, j). The anterior flagellum is about 1.25x cell length, whereas the thicker (but tapering towards the distal end) posterior flagellum is 3x cell length and trails behind (Fig. 2.1i). Three faint pellicle strip joints can be seen on both the ventral and the dorsal side, with four laterally (total 10; Fig. 2.1k). When tested, KM082 was able to feed on *Phaeodactylum* sp. material, with cells containing 2 to more than 10 small pigmented ingesta (1–2.9 μm in diameter; n = 5). Morphological measurements can be found in Table 2.1.

Olkasia polycarbonata n. gen. n. sp.

We generated SSU rDNA data for six isolated cells from two different locations. Additionally, we used SEM to image cells that were isolated at the same times as two cells that were processed for sequencing, namely UB41 and UB58 (see Methods). Cells studied by light microscopy are oblong, 55–62.8 μm long by 30.4–37.2 μm wide, ventrally flattened, and have a conspicuous chisel-shaped feeding apparatus that extends down the whole length of the cell (Figures 2.4a, c–d and 2.5a–d; morpho-

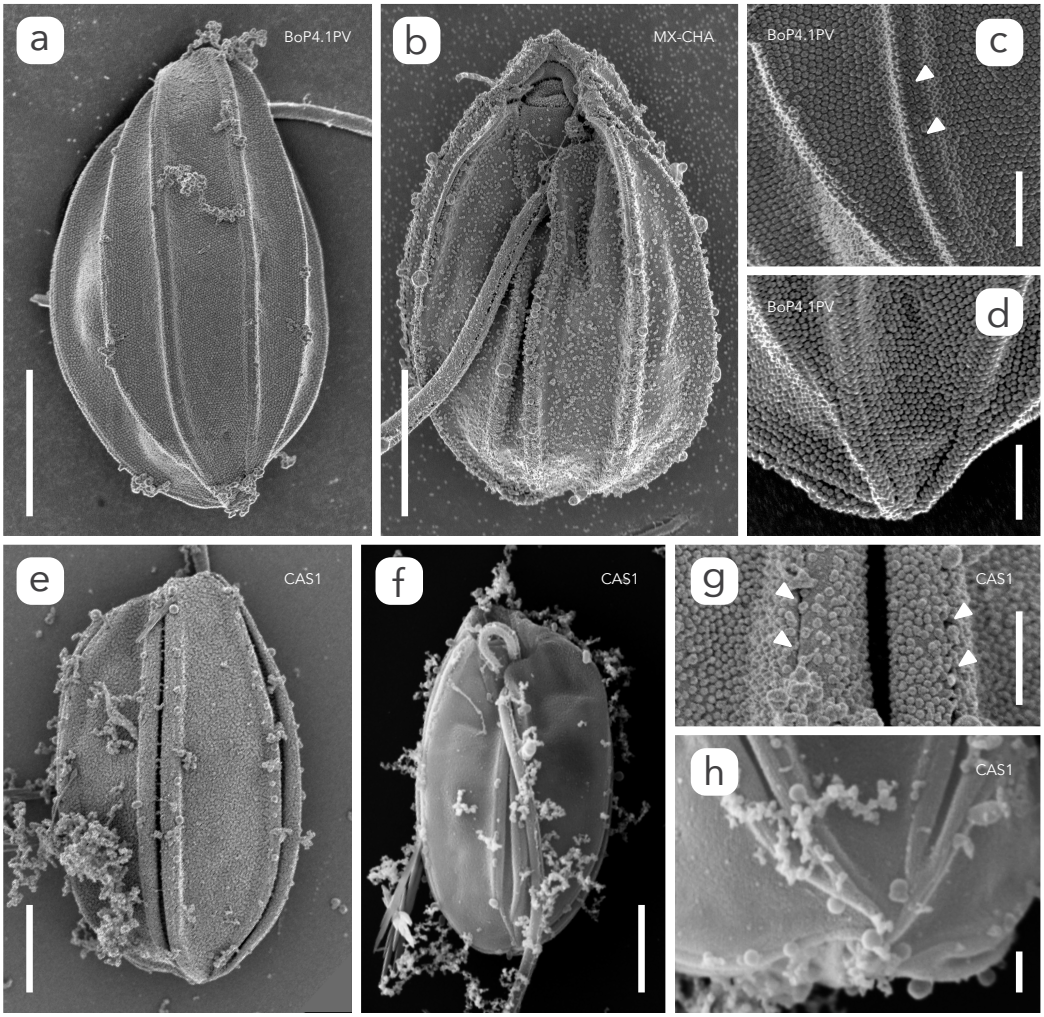

Figure 2.2: Scanning electron microscopy images of *Ploeotia vitrea* and *Ploeotia oblonga*. a–d) *Ploeotia vitrea* strains BoP4.1PV (a, c, d) and MX-CHA (b), showing the dorsal (a) and ventral sides (b). The pellicle strip boundaries raised on keels are shown in (c) and at the posterior end in (d). e–h) *Ploeotia oblonga* strain CAS1, with the dorsal (e) and ventral (f) sides shown. Details of the pellicle strip arrangement are shown in (g) and the posterior end in (h). Arrowheads denote joints between two pellicle strips. Scale bars are 5 µm for (a–b) and (e–f), and 1 µm for (c–d) and (g–h).

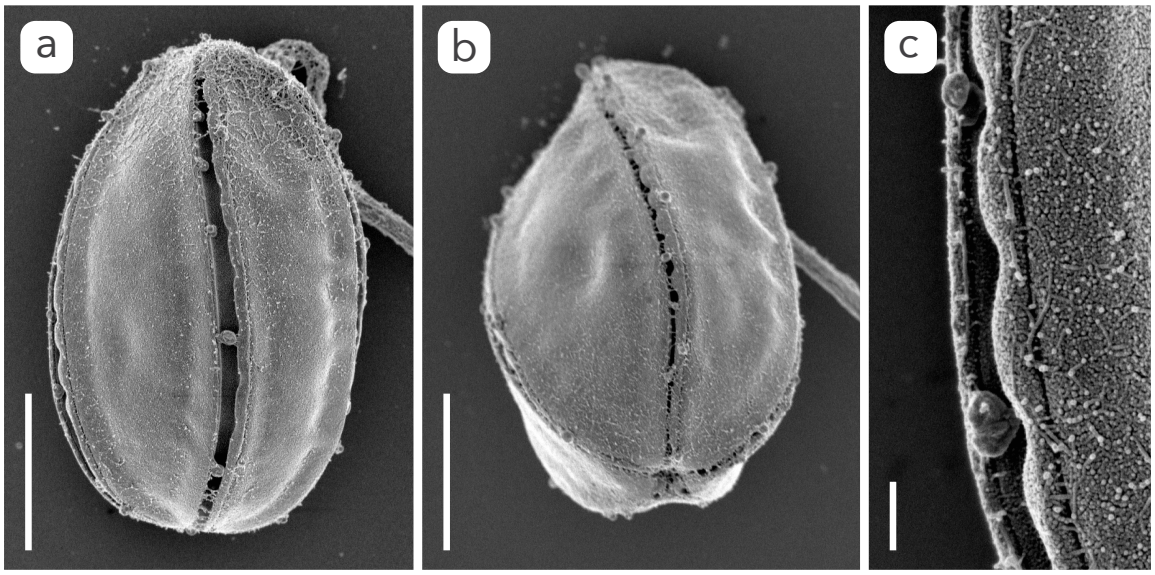

Figure 2.3: Scanning electron microscopy images of *Serpenomonas costata* strain HAK-MF. a) dorsal side with pellicle strips clearly visible. b) dorsal posterior, with star-shaped pellicle arrangement. c) joint between a wide and narrow pellicle strip. Note the characteristic undulating edge within the narrow strip. Scale bars are 5 µm for (a–b), and 1 µm for (c).

logical measurements of the six cells isolated for molecular work are listed in Table A.2). With SEM, the anterior end of the feeding apparatus can be seen ventrally, and appears 'capped' (Fig. 2.5d, e). The pellicle is composed of 10 roughly equal-size strips that can be clearly seen with light microscopy (Fig. 2.4b–c, e). With SEM, the strips appear as S-shaped in cross-section (especially on the dorsal side; Fig. 2.5a) and overlapping with each other, such that the joints between strips face laterally, rather than running along the spine of each ridge as in *P. vitrea* (Fig. 2.5b). The posterior flagellum is 1.8x cell length, and conspicuously thickened. Cell movement is similar to that of most other ploeotids: Cells move their anterior flagellum (0.9x cell length) with a sweeping motion, and jerk back when under duress (Video 3). Structures resembling discharged extrusomes with diameter ~100 nm were observed in some SEM preparations (Fig. 2.5f). Efforts to establish cultures were unsuccessful. Most observed cells contained algal material in the form of rounded ingesta 2.1–3.6 µm in diameter (Fig. 2.4a–c). Sequences in clade A (see below) are 97.7–99.8% identical to each other, whereas in clade B 97.8% of sites are identical. Sequence identity between clade A and B is 90.6–93.5% (Table A.3).

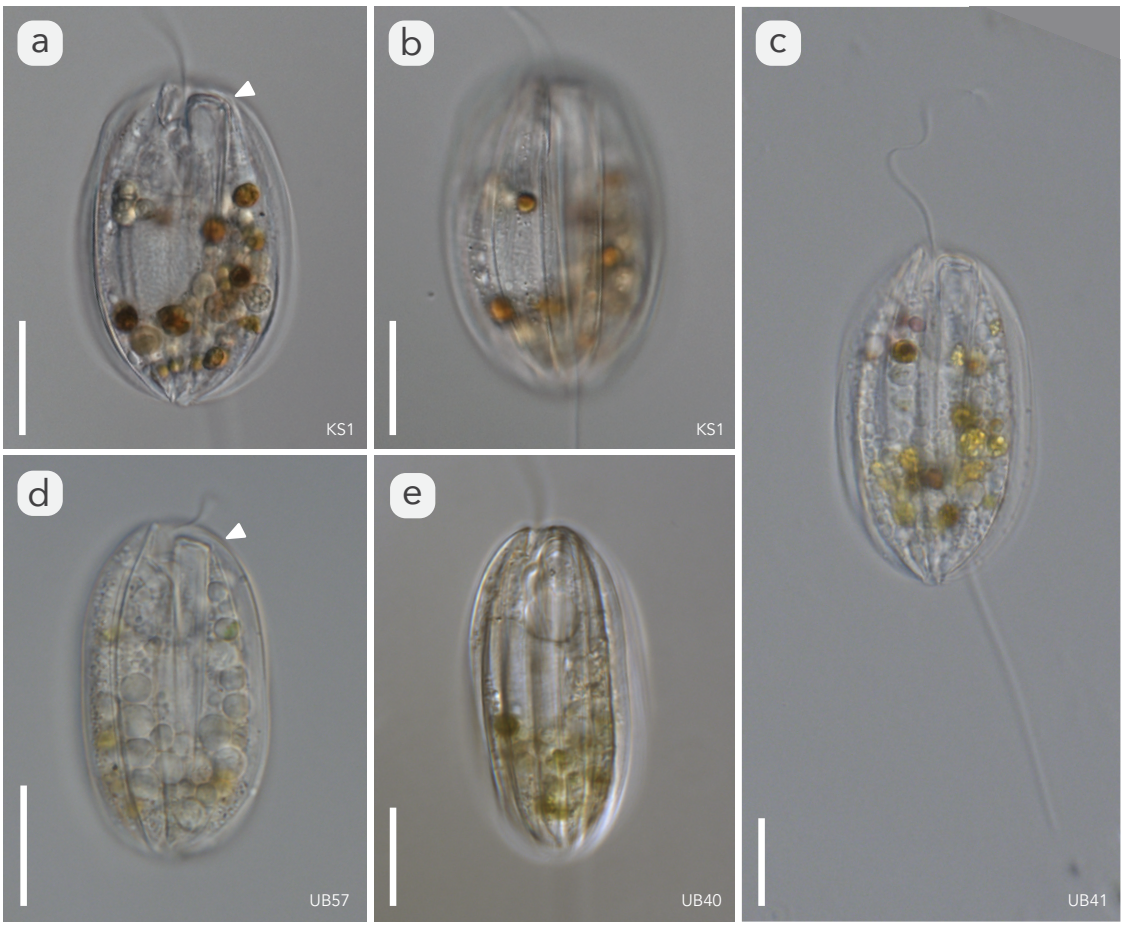

Figure 2.4: Light micrographs of *Olkasia polycarbonata* n. gen. n. sp., with strain names in image. a–c) Cells from clade A, with pigmented ingesta 1.2–5.2 µm in diameter; d–e) Cells from clade B. Arrowheads show feeding apparatus. All scale bars are 20 µm, all images were acquired with differential interference contrast optics.

Lentomonas azurina and *Lentomonas corrugata*

We generated SSU-rRNA gene sequence data from four isolated single cells belonging to the morphospecies *L. corrugata* and one from *L. azurina* (Cell STS5). Both morphospecies are elliptical in profile, ventrally flattened and dorsally convex (Fig. 2.6a, c). The ratio of length/width for *L. corrugata* cells is 1.14–1.36 (average 1.24), whereas it is 1.73–1.89 (average 1.78) for cells identified as *L. azurina* (STS5 and two cells for which no molecular data could be acquired; see Table A.2). The dorsal side shows seven strongly corrugated pellicle strips, including two lateral ones (Fig. 2.6a, c), while the ventral side has three flatter strips (total 10). The feeding apparatus is hook-shaped, oblique, and extends to almost the full length of the cell (Fig. 2.6b, d). Cells pull back frequently during gliding (Video 4). 97.2–99.8% of sites are identical

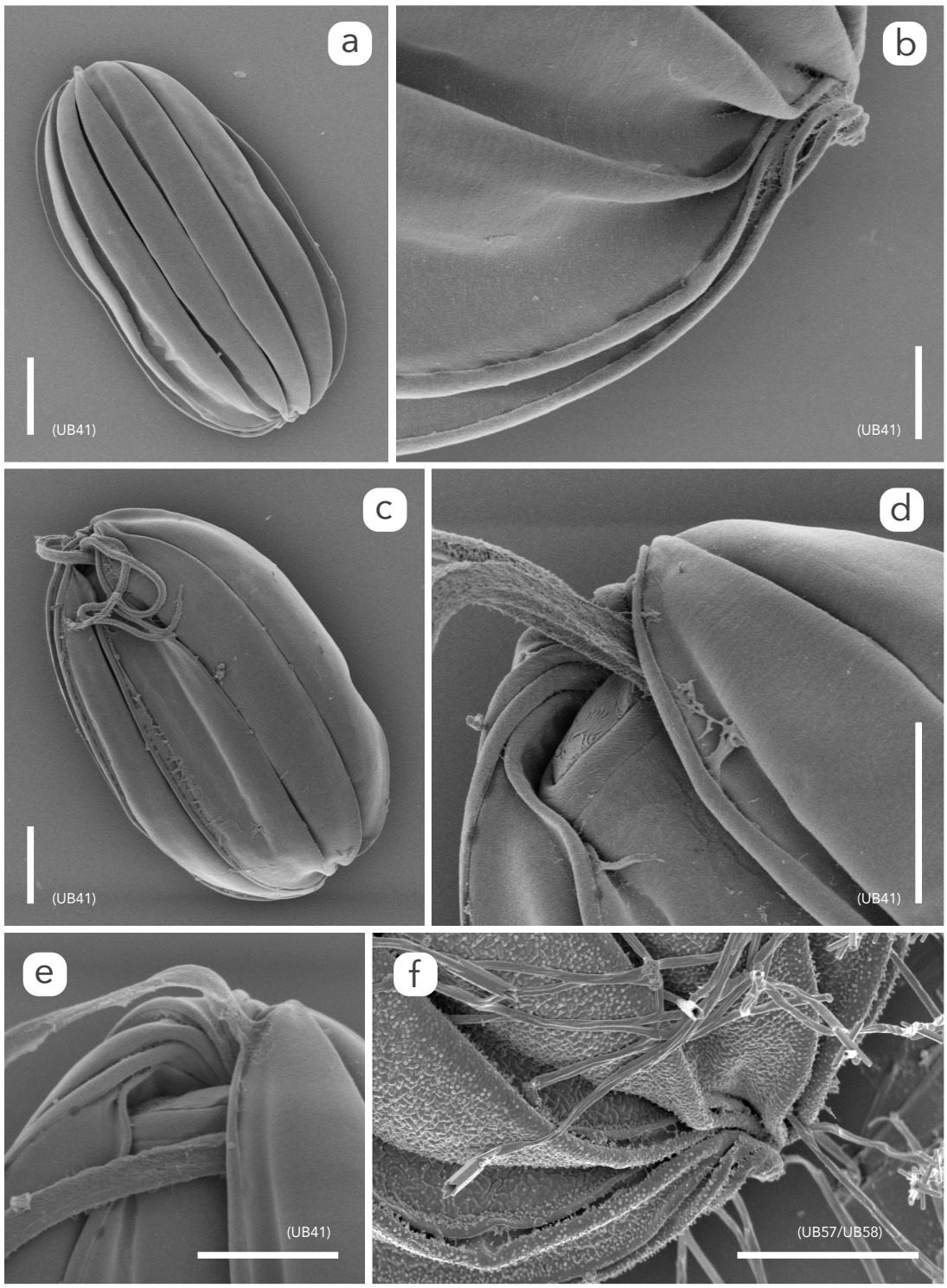

Figure 2.5: Scanning electron microscopy images of single cells of *Olkasia polycarbonata* n. gen. n. sp. Cells in a–e) were isolated from the same population as UB41 in clade A; Cell in f) was isolated from same population as UB57 and UB58 in clade B. a) dorsal side showing overall pellicle arrangement. b) detail of posterior dorsal end of a different cell, showing pellicle strip joints. c) ventral side. d) detail of ventral anterior, showing feeding apparatus. e) detail of ventral anterior pellicle strip joints. f) close-up of dorsal posterior, showing discharging extrusomes. Scale bars are 10 µm for (a and c), 5 µm for (d– f), 2 µm for (b). (a–e) are images of part of the Hapantotype of *Olkasia polycarbonata*.

between *L. corrugata* and *L. azurina* STS5, while there is 95–99.8% identity amongst different *L. corrugata* cells (Table A.3).

Hemiolia trepidum n. gen. n. comb.

We generated SSU rDNA sequences from six cells identified morphologically as *Anisonema trepidum* (Larsen 1987). This morphotype has an oblong cell shape and is moderately flattened (Fig. 2.6e–h). Three to four faint pellicle striations are sometimes seen on the dorsal side (Fig. 2.6h). Cells glide rapidly in relatively straight lines, often with occasional stops when cells sit with only the anterior flagellum beating for 1–2 seconds, and then resume movement in the same direction. Characteristically, the anterior flagellum is held to the right-hand side of the cell, performing a trembling motion with the distal quarter of the flagellum. The 'jerking-back' motion that is common in *Anisonema acinus* is less frequent and is always followed by an abrupt change in direction (Video 5). Lengths for the 4 cells observed were 12.4–22.7 μm (average 16.8 μm), and widths 7.3–9.9 μm (average 8.5 μm). Anterior flagella were typically 1.5x cell length, and posterior flagella were 3.3x cell length. The feeding apparatus could not be observed by light microscopy. Three of the isolated cells had ingested whole diatoms (e.g. Fig. 2.6e; ingested cell is 10.3 x 4.3 μm). Individual measurements of single cells are listed in Table A.2.

Liburna glaciale n. gen. n. comb.

SSU rDNA sequences were generated from five cells identified as *Anisonema glaciale* (Larsen & Patterson 1990). Like *Hemiolia trepidum*, these cells have an oblong cell shape and are moderately flattened (Fig. 2.6i–l), but are substantially larger in size, averaging 26.1 μm in length (25.4–27.1 μm), and 12.4 μm in width (11.4–13.9 μm). Dorsally, 5–6 faint pellicle striations were sometimes observed (Fig. 2.6l). Like *Hemiolia trepidum*, *Liburna* exhibits the movement pattern of gliding rapidly in straight lines, with the anterior flagellum held to the right side (Video 6). The anterior flagellum is approximately 1.4x cell length, while the posterior flagellum is ~3x cell length. A feeding apparatus was not observed by light microscopy. Three out of five cells isolated had ingested whole diatoms (Fig. 2.6i–k; two were measurable: 19.6 x 4.3 μm and 17 x 4 μm). Measurements of single cells can be found in Table A.2.

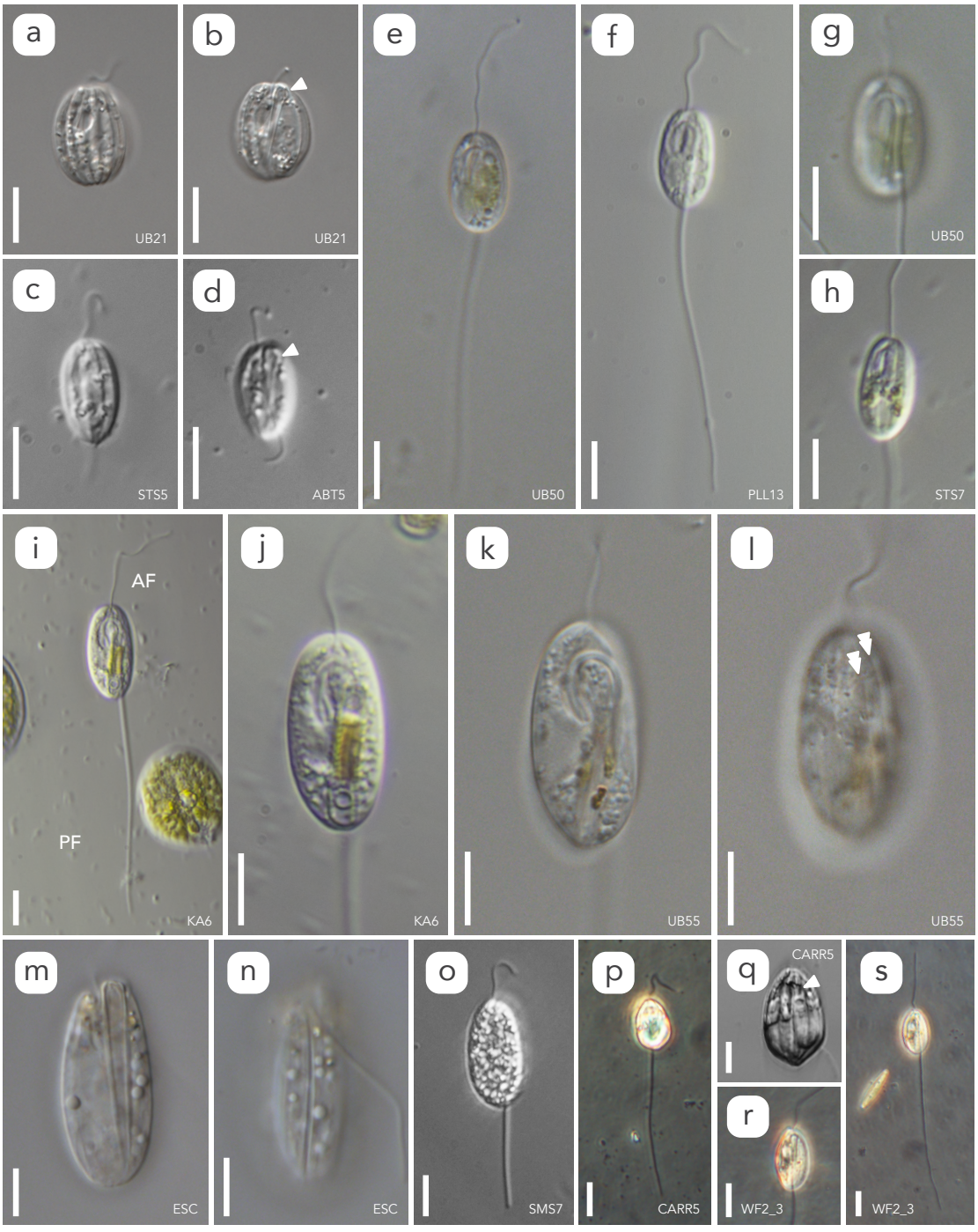

Figure 2.6: See next page.

Figure 2.6: Light micrographs of ploeotid single cell isolates and cultures. a–b) *Lentomonas corrugata* isolate UB21, showing the dorsal corrugated appearance (a) and arrowhead pointing to feeding apparatus (b). c–d) *Lentomonas azurina*, general view and dorsal side in (c) (isolate STS5), arrowhead pointing to feeding apparatus in (d) (isolate ABT5, no molecular data). e–h) *Hemiolia trepidum* n. gen n. comb., several isolates. General appearance in isolate UB50 (e) and PLL13 (f), with details of the proximal posterior flagellum in UB50 (g), and pellicle striation in STS7 (h). Note the 10.3 x 4.3 µm ingested diatom in (e). i–l) *Liburna glaciale* n. gen n. comb., several isolates. General view of isolate KA6 (i), and detail of cell body with flagellar pocket and 19.6 x 4.3 µm ingested diatom visible (j). Detail of flagellar pocket in isolate UB55 and 17 x 4 µm ingested diatom (k) and pellicle striations (arrowheads) in (l). m–n) *Entosiphon* sp. strain ESC, chisel-shaped feeding apparatus in m), and pellicle striations in n). o) Unidentified ploeotid isolate SMS7, general appearance. The posterior flagellum of this cell was truncated at time of imaging. p–q) *Liburna*-like isolate CARR5, general appearance (p) and close-up of pellicle striations (q) (arrowhead). r–s) *Hemiolia*-like isolate WF2_3, close-up of cell body (r) and general appearance (s). All scale bars are 10 µm. Images were acquired with differential interference contrast optics, except for (p, r, s) where phase contrast optics were used.

Entosiphon sp. ESC

A partial SSU-rRNA sequence of *Entosiphon* sp. strain ESC was acquired through transcriptome sequencing that will be reported elsewhere (Lax et al. unpublished, Chapter 3). Cells are oblong and elongated, with a conspicuous moving feeding apparatus with strong rods, extending down the whole length of the cell (Fig. 2.6m). 12 clearly visible pellicle strips run straight down the length of the cell. Narrower pellicle strips alternate with broader strips, at least on the dorsal side (Fig. 2.6n). Strain ESC feeds on material from *Haematococcus* sp. in culture (see bright inclusions at anterior in Fig. 2.6m). Morphological measurements can be found in Table 2.1.

Unidentified ploeotid SMS7

A partial SSU rDNA was sequenced. The 16.2 x 8.2 µm cell is oblong and covered in refractile granules (Fig. 2.6o), making observation of internal or pellicle structures impossible with available optics. The cell glides on the thick posterior flagellum (5x cell length). Beating of the thick anterior flagellum (~1.5x cell length) during movement was similar to *Liburna* and *Hemiolia*, but more active and in a broader arc.

Unidentified ploeotid CARR5

A partial SSU rDNA was sequenced. The 31.9 x 20 µm cell is roughly pyriform, with a pointed posterior end (Fig. 2.6p, q). The anterior flagellum appears thick (~1x cell length), whereas the very thick posterior flagellum (4.5x cell length) tapers slightly. Four strongly developed pellicle strip joints can be seen on the dorsal, and four on the ventral side (Fig. 2.6q, number of lateral strips unclear).

Unidentified ploeotid WF2_3

A partial SSU rDNA was sequenced. The 36.2 x 20.6 μm cell is oblong (Fig. 2.6r, s) and glides on a thickened, tapering posterior flagellum (5x cell length; Fig. 2.6s). The anterior flagellum (~1x cell length) is held on one side like in *Hemiolia*. The ingestion organelle extends down half the cell length (Fig. 2.6r). Since this cell was only imaged using phase contrast optics, no pellicle strip joints could be seen.

2.4.2 Phylogeny

We conducted six separate phylogenetic analyses, differing in taxon sampling: 1) No *Entosiphon*, no outgroup (main dataset); 2) No *Entosiphon*, with outgroup; 3) With *Entosiphon*, no outgroup; 4) With *Entosiphon*, with outgroups; 5) No *Entosiphon*, with outgroups, no partial sequences (<1000 bp, 9 sequences excluded); 6) No *Entosiphon*, no outgroup, and no 'rogue taxa' (see Methods). All datasets were subjected to Maximum Likelihood (ML) analyses under the GTR + Γ model, with robustness estimated from 1000 bootstrap replicates. Datasets 1 and 3 were also subjected to a Bayesian analysis under the same model (see Methods for further details).

The Euglenida (with symbiontids) grouping consists of several well-supported clades, with little robust phylogenetic structure linking them. These are: a) The clade Spirocuta, containing phototrophic euglenids (Euglenophyceae), primary osmotrophic euglenids (Aphagea) and phagotrophic euglenids with a flexible pellicle (Table A.4); b) Petalomonadida (fully supported in all analyses); c) Symbiontida (fully supported in all analyses); and d) eight clades of ploeotid sequences, as described below (Figures 2.7 and 2.8).

The new genus *Olkasia* is represented by *Olkasia polycarbonata* n. sp. (=*Ploeotia* c.f. *vitrea* sensu Lax and Simpson 2013). In all analyses, six novel sequences branched with maximum support with the one previously reported sequence (Lax & Simpson 2013). Two strongly supported subgroups were recovered within this clade (clade A and clade B, 95–100% BS, 1 pp). In four out of the five datasets that included unidentified ploeotid SMS7, it branched sister to *Olkasia*, although with negligible support (25–41% BS and 0.55 pp; Figures 2.7 and 2.8). The novel genus *Hemiolia* includes six cells identified as *H. trepidum* (basionym *Anisonema trepidum*), always on a long and maximally supported branch (Fig. 2.7). The partial sequence of unidentified ploeotid WF2_3 branched as sister to this clade in three of the four analyses where it was included, but with negligible support (e.g. Fig. 2.7). Novel genus *Liburna* was composed of five cells identified as *L. glaciale* (ba-

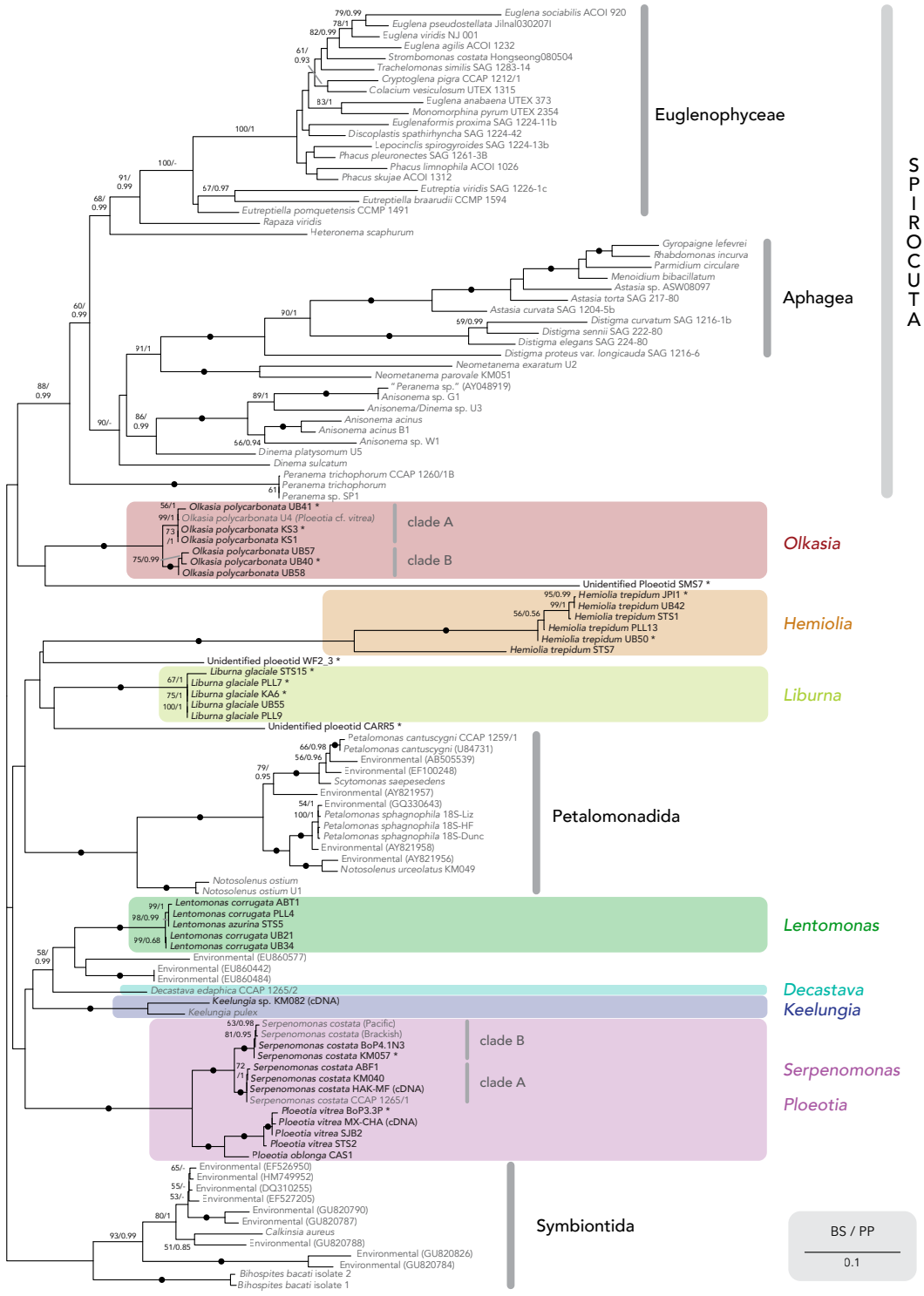

Figure 2.7: Maximum Likelihood phylogeny of the SSU-rRNA gene of euglenids under the GTR+Γ model, with posterior probabilities derived from Bayesian analysis under the same model. Major groups of ploeotids are shown, with sequences acquired in this study bolded, an asterisk (*) on taxa names denoting partial sequences. This phylogeny is unrooted; for rooted phylogenies with inferred outgroups included see Figures 2.8 and A.1. Maximum bootstrap support (100%) and posterior probability (pp of 1) is shown with a black circle. Support values below 50% and 0.9 pp not shown, and nodes with a dash (-) are not resolved in Bayesian analysis (polytomy).

sionym *Anisonema glaciale*), always forming a maximally supported clade. Branch lengths between individual sequences were short. The partial sequence of unidentified ploeotid CARR5 fell sister to the *Liburna* clade in five out of the six analyses when it was included, albeit with poor-to-no support (19–47% BS; e.g. Fig. 2.7). *Lentomonas* included four novel sequences of *L. corrugata* and one sequence of *L. azurina*, forming a clade with maximum support in both ML and Bayesian analyses. The *L. azurina* cell STS5 branched in among the *L. corrugata* sequences, and thus no phylogenetic separation was observed between the two morphotypes. *Keelungia* formed a maximally supported clade consisting of *Keelungia pulex* (Chan et al. 2013) and our sequence of *Keelungia* sp. strain KM082. The *Decastava* clade consisted of only one sequence, from *Decastava edaphica* (Cavalier-Smith et al. 2016). The *'Ploeotia + Serpenomonas'* clade (which we equate with the taxon Ploeotiidae) consisted of individual subclades of sequences belonging to *P. vitrea*, *P. oblonga*, and *S. costata*. As expected, our five new *Serpenomonas costata* and the three previously available sequences formed a maximally supported clade in all analyses (e.g. Figures 2.7 and 2.8, and Fig. A.1). Within this clade, two maximally supported sub-clades were recovered, separating isolates with group I introns in their SSU rDNA sequences (clade A, includes strain CCAP 1265/1), from a clade without any group I introns (clade B, includes strain KM057; Fig. 2.9a). Intriguingly, clade A is composed of strains isolated from both coasts of North America and Europe, whereas clade B isolates are Asian and Caribbean. Sister to *Serpenomonas*, we recovered *Ploeotia vitrea* (four new sequences) and *Ploeotia oblonga* (one new sequence), the latter branching sister to *P. vitrea* with strong to full support (e.g. Figures 2.7 and A.1). Some *P. vitrea* sequences included group I introns (see below, and Fig. 2.9). *Entosiphon* consists of a tight cluster of similar sequences. This clade was extremely long branching and therefore only included in some of our analyses to limit long branch attraction artifacts, as in several recent studies (Lax & Simpson 2013, Cavalier-Smith et al. 2016, Paerschke et al. 2017). The new sequence of *Entosiphon* sp. strain ESC was sister to a clade containing *E. oblongum* and *Entosiphon* sp. strain TCS-2003, with moderate support (75% and 81%).

Relationship Between Clades

The exact relationships between the individual clades of ploeotids were poorly resolved and often differed between datasets and analyses (Table A.4). We recovered *Olkasia* branching as a sister to Spirocuta in all our datasets, although always with negligible support (Table A.4). Rogue taxon SMS7 branched with *Olkasia* in four of

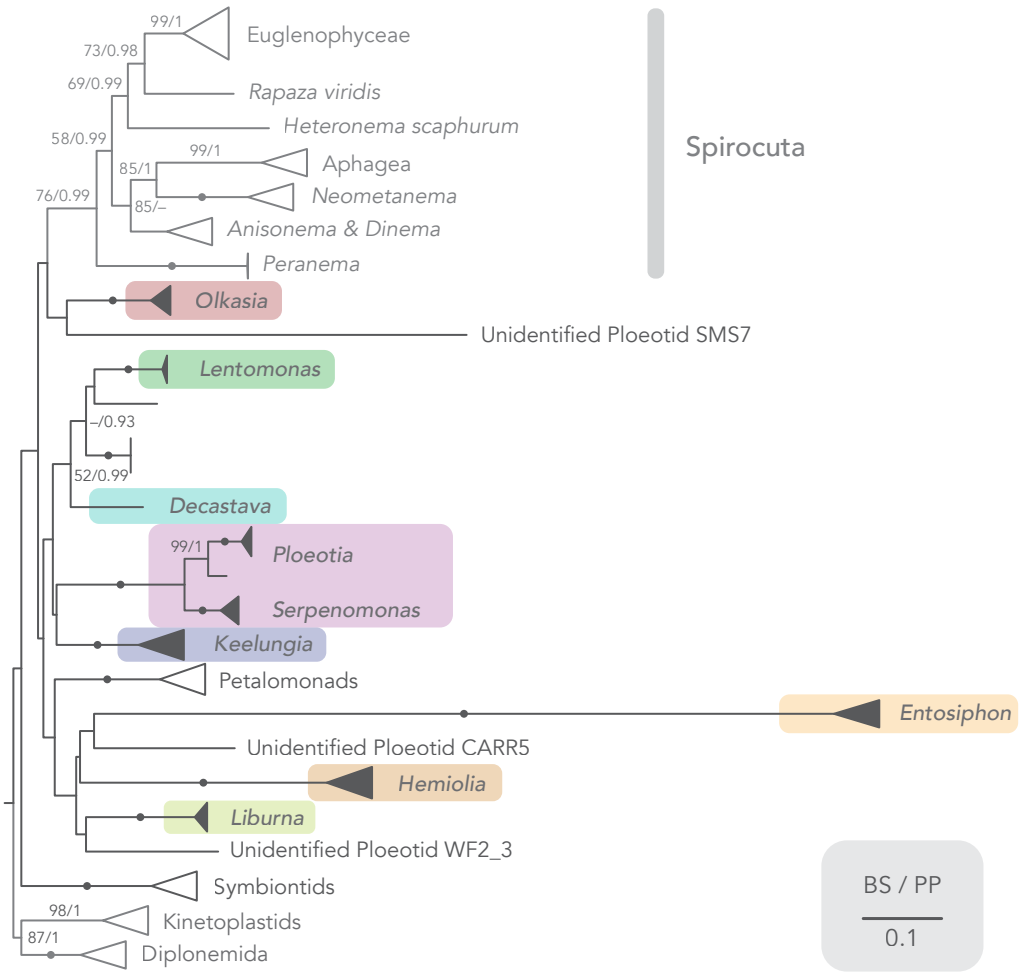

Figure 2.8: Summary view of euglenid phylogeny including *Entosiphon*, using the SSU-rRNA gene with maximum likelihood and Bayesian analyses (GTR+Γ model). Major groups of ploeotids are shown as collapsed, filled triangles. Asterisk (*) on taxa names denotes partial sequences. This tree is rooted on an outgroup of kinetoplastid and diplonemid taxa. Maximum bootstrap support (100%) and posterior probability (pp of 1) is shown with a black circle, while support values below 50% and 0.9 pp are not shown.

these analyses (Figures 2.7 and 2.8). *Hemiolia* and *Liburna* formed a group to the exclusion of all other identified genera in five out of our six analyses, though with negligible support. In four cases, *Hemiolia* and *Liburna* (with or without *Entosiphon*—see below), form the sister group to Petalomonads, with little support (Table A.4). *Lentomonas* and *Decastava* formed a very poorly to moderately supported branch that also contained three partial-length environmental sequences when the latter were included (Fig. 2.7). The placement of *Keelungia* was unstable; in two analyses without *Entosiphon* it branched together with *Lentomonas* and *Decastava*, whereas in analyses including *Entosiphon* it grouped with *Ploeotia* and *Serpenomonas* (Fig. 2.8). In the analysis omitting short sequences, *Keelungia* branched with *Hemiolia*. None of these

relationships was supported (2.7). Ploeotiidae (*Ploeotia* + *Serpenomonas*), either alone or with *Keelungia*, formed a branch sister to the *Lentomonas* + *Decastava* + environmental sequences clade in five out of six analyses, with no support. In both analyses including *Entosiphon* support values were further reduced across the whole tree, likely owing to their long-branching nature (Fig. 2.8, Table A.4). *Entosiphon* formed a grouping including *Hemiolia* and *Liburna* in both, and this assemblage was in turn sister to petalomonads (see above).

2.4.3 Group I Introns

SSU-rRNA sequences derived from cDNA generated for *Serpenomonas costata* HAK-MF and *Ploeotia vitrea* MX-CHA were shorter than their DNA-derived counterparts. Alignment of RNA and DNA sequences showed a 494 bp intron in HAK-MF, at the same position as previously reported for *Serpenomonas costata* CCAP 1265/1 (Busse & Preisfeld 2003b). At the same position, isolates KM040 (*S. costata*, culture), ABF1 (*S. costata*, single cell), SJB2 (*P. vitrea*, culture), and STS2 (*P. vitrea*, single cell) also all showed similarly long introns. A 406 bp long intron was found only in ABF1 (*S. costata*, single cell) at a second position. A third intron site was found in HAK-MF (*S. costata*, culture), ABF1 (*S. costata*, single cell), MX-CHA (*P. vitrea*, culture), and SJB2 (*P. vitrea*, culture). Sizes ranged from 388 and 406 bp in *P. vitrea*, to 510 bp in *S. costata*. This last intron was confirmed with RT-PCR in both HAK-MF and MX-CHA. Introns at the second and third site had direct repeats at one end and in the opposite flanking region (Fig. 2.9b). The SSU rDNA sequence of *Keelungia* sp. strain KM082 had five insertions that were missing from cDNA-derived SSU-rRNA sequences (Fig. 2.9a). These insertions were scattered along the whole length of the sequence, ranged from 158–170 bp, and were always found in conserved regions. All of these introns had 2–4 bp long direct repeats in one end of the intron and in the opposite flanking region, like in *Serpenomonas* and *Ploeotia* (Fig. 2.9b).

Original light microscopy images for all taxa, and, where available, scanning electron images are deposited under Dryad accession dryad.k08pc1r, as are untrimmed alignments, trimmed alignments, and treefiles for all phlylogenetic analyses. The new SSU rDNA sequences reported here are deposited on GenBank under accessions MK239274–MK239309 and MK213404–MK213407.

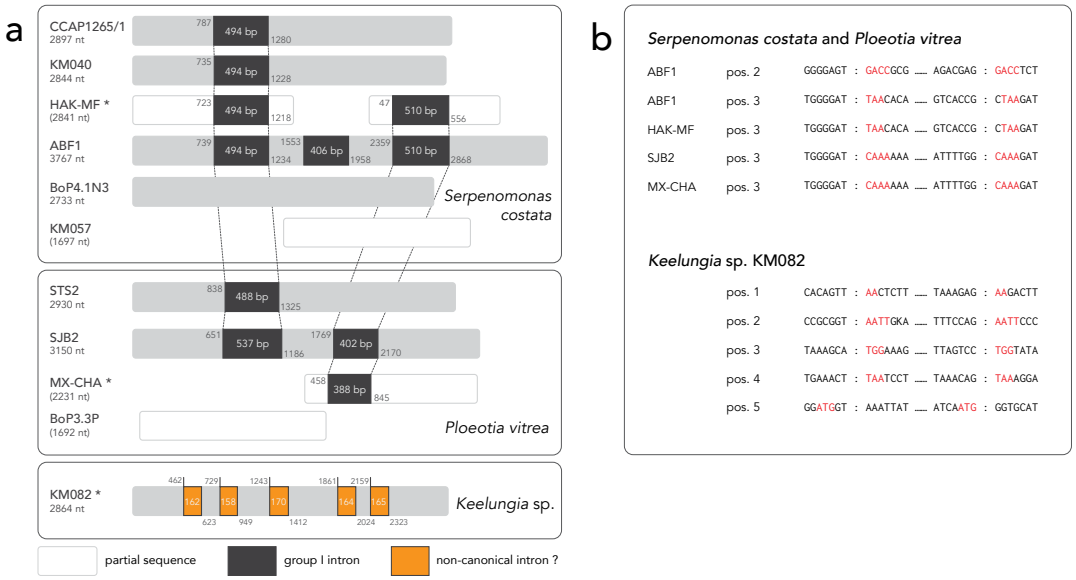

Figure 2.9: Introns in *Serpenomonas*, *Ploeotia* and *Keelungia* sp. a) Positions of group I intron (dark grey) and non-canonical introns (orange) in the SSU-rRNA sequence of ploeotids. For taxa with white boxes, only partial SSU sequences were acquired. Lengths of introns are within boxes, with their start and end positions noted on either side. Intron boxes connected by vertical dotted lines mark that they appear at the same position within the SSU-rRNA. b) Intron:exon boundaries of non-canonical introns, at different positions (see a). Direct repeats are marked in red, and ":" shows the intron:exon boundary. Strains with * have had their SSU-rRNA sequence derived from cDNA as well as DNA.

2.4.4 Taxonomic Summary

The following nomenclatural acts are deposited under ZooBank LSID:
urn:lsid:zoobank.org:pub:5C4D3829-EA77-4E0B-AFF6-24D4EBCC363D

Olkasia gen. nov. Lax, Lee, Eglit and Simpson (ICZN)

Description: Free-living, inflexible, biflagellate, heterotrophic euglenid, oblong in profile, ventrally flattened, with a conspicuous chisel-shaped feeding apparatus. The 10 pellicle strips are similarly sized, and S-shaped in cross-section (especially on the dorsal side), and overlap slightly with each other. Cells glide on their thickened posterior flagellum; and 'jerk back' when stressed. The anterior flagellum sweeps from side to side.

Type species: *Olkasia polycarbonata* Lax, Lee, Eglit and Simpson (description see below)

Etymology: From ‘olkas’ (Greek), a large trading barge used in Hellenistic times. Refers to the large size of the type species relative to *Ploeotia* (*Ploeotia* is from ‘ploion’ = boat in Greek). Feminine.

ZooBank Accession: LSID urn:lsid:zoobank.org:act:DBCD8D51-3766-4521-A1CD-4D91A79CD317

***Olkasia polycarbonata* sp. nov. Lax, Lee, Eglit and Simpson (ICZN)**

Description: Oblong cells, 54–63 by 30–37 μm, ventrally flattened.

Type material: The name-bearing type (hapantotype) is an SEM-stub with five osmium-fixed and sputter-coated single cells isolated by hand from the type locality. Deposited with the American Museum of Natural History, New York, as AMNH_IZC 00343283. Type locality: Horseshoe Island Park, Halifax, Nova Scotia, Canada (N44° 38’23.4”, W63°36’45.2”), oxic intertidal sediment.

Etymology: After ‘polycarbonate’, the transparent thermoplastic polymer [originally ‘poly’ = many (Greek), ‘carbo’ = coal, charcoal (Latin)]. Polycarbonate is often used instead of glass for windows, windscreens, etc. Refers to previous appellation of the organism as *Ploeotia* c.f. *vitrea* (i.e. similar to *P. vitrea*), from ‘vitrum’ = glass (Latin).

Gene sequence: The partial SSU rDNA sequence of a single cell (UB41) collected from the same sample as the hapantotype, at the same time, has the GenBank accession number MK239294.

ZooBank Accession: LSID urn:lsid:zoobank.org:act:22D0646E-BEA5-4556-AEF9-824A5C7F83EB

***Hemiolia* gen. nov. Lax, Lee, Eglit and Simpson (ICZN)**

Description: Free-living, inflexible, biflagellate euglenids, oblong in profile, moderately dorso-ventrally flattened. Cells glide rapidly on their thickened, >3x cell length posterior flagellum and with the anterior flagellum held to the right side and trembling. Movement occasionally arrests for a couple of seconds, then continues

in same direction. Feeding apparatus not observed with light microscopy. Pellicle strip margins difficult to observe by light microscopy. Phylogenetically more closely related to *Hemiolia trepidum* than to *Liburna glaciale*.

Type species: *Anisonema trepidum* Larsen 1987 (=*Hemiolia trepidum*, comb. nov.)

Etymology: 'Hemiolia' (Greek) was a type of fast, light attack and scouting ship with one and a half banks of oars per side, used by pirates and navies in the Hellenistic era (from 'hemiolis'—one and a half). Refers to the speed and small relative size of cells (see *Liburna*, below). Feminine.

ZooBank Accession: LSID urn:lsid:zoobank.org:act:0A97EDBC-808D-4DDB-BE94-DFC83907D046

Transfer of existing species to *Hemiolia*

Hemiolia trepidum (Larsen 1987) Lax, Lee, Eglit and Simpson comb. nov.

Basionym: *Anisonema trepidum* Larsen, 1987 (594-595, Fig. 6)

***Liburna* gen. nov. Lax, Lee, Eglit and Simpson (ICZN)**

Description: Free-living, inflexible, biflagellate euglenids, oblong in profile, moderately flattened dorso-ventrally. Cells glide rapidly on their thickened, hook-shaped posterior flagellum (3x cell length). The anterior flagellum is held to the right side and trembles. Feeding apparatus not observed with light microscopy. Pellicle strip margins difficult to observe by light microscopy; apparently >10 strips. Type species is larger than that of *Hemiolia*, with which it shares most characteristics. Phylogenetically more closely related to *Liburna glaciale* than to *Hemiolia trepidum*.

Type species: *Anisonema glaciale* Larsen and Patterson 1990 (=*Liburna glaciale*, comb. nov.)

Etymology: 'Liburna' (Latin) was a fast attack ship with two banks of oars widely used in the Roman navy from the late Republic onwards. Refers to the speed and size of cells (see *Hemiolia*, above). Feminine.

ZooBank Accession: LSID urn:lsid:zoobank.org:act:691C514A-78C8-4E39-9D4D-2F793CFA2D42

Transfer of existing species to *Liburna*

Liburna glaciale (Larsen and Patterson 1990) Lax, Lee, Eglit and Simpson comb. nov.

Basionym: *Anisonema glaciale* Larsen and Patterson, 1990

2.5 Discussion

2.5.1 Molecular Phylogenetics of Phagotrophic Euglenids

Ever since the first sequence data from a phagotrophic euglenid was acquired, taxon sampling has slowly increased, and gradually advanced our understanding of euglenid phylogenetics (Montegut-Felkner & Triemer 1997, Busse et al. 2003, Breglia & Leander 2007, Lax & Simpson 2013, Cavalier-Smith et al. 2016). It has become increasingly clear that flexible taxa with >12 pellicle strips indeed form a major derived clade within euglenids, as had been inferred primarily from morphology (Leander et al. 2001a,b)—this taxon, Spirocuta (or Helicales), includes a variety of phagotrophic euglenids in addition to phototrophs and osmotrophs (Lee & Simpson 2014b, Cavalier-Smith 2016, Paerschke et al. 2017). Within Spirocuta, *Neometanema* was established as representing the sister group to osmotrophs (Lax & Simpson 2013, Lee & Simpson 2014b), whereas outside Spirocuta, petalomonads have consistently been recovered as a clade (e.g. Kim et al. 2010, Lee & Simpson 2014a, Cavalier-Smith et al. 2016). Also, symbiontids were defined as a significant monophyletic group, either within euglenids, or sister to them (Breglia et al. 2010, Lax & Simpson 2013, Cavalier-Smith 2016). Within this context, the molecular examination of ploeotids stands out as having raised more questions than it answers. Incremental improvements in taxon sampling supported the notion that ploeotids likely represented several distinct clades with difficult-to-resolve relationships (Chan et al. 2013, Lax & Simpson 2013, Chan et al. 2015, Cavalier-Smith et al. 2016), but it remained essentially unknowable how many major clades there are. With this, any inferences about the deep-level phylogeny and evolution of euglenids were inevitably based on unreliably supported relationships and conjectural extrapolation to incorporate taxa with no molecular data (e.g. Cavalier-Smith 2016). It is this profound uncertainty that

we aimed to reduce by increasing the taxonomic breadth of sequence data available for ploeotids.

2.5.2 *Ploeotia, Serpenomonas* and *Olkasia*

We resolved a central problem in the molecular biodiversity of ploeotids, by sequencing SSU rDNA from *Ploeotia vitrea*, type species of *Ploeotia*. The phylogenetic position of *P. vitrea* was surprising given recent studies, being remote from '*Ploeotia* c.f. *vitrea*' (Lax & Simpson 2013) but close to *Serpenomonas*, which was recently inferred to differ markedly from *Ploeotia* on morphological grounds (Cavalier-Smith et al. 2016, Cavalier-Smith 2016). The major consequences for the systematics of euglenids are discussed below.

The previously undescribed morphospecies identified as *Ploeotia* c.f. *vitrea* by Lax & Simpson (2013) was considered to be similar to *Ploeotia vitrea*, but much larger, based on light microscopy observations. However, our molecular phylogenetic analyses including *P. vitrea* clearly show that it is not closely related to this *Ploeotia* c.f. *vitrea*. In addition, our investigation with SEM revealed that the organization and linking between strips is distinctly different in the two taxa (compare Fig. 2.2a, c with Fig. 2.5a, b). Based on the strongly supported phylogenetic placement outside the *Ploeotia-Serpenomonas* clade, as well as the morphological differences, we propose a novel genus and species, *Olkasia polycarbonata*, for *Ploeotia* c.f. *vitrea* sensu Lax and Simpson (2013).

Surprisingly, our phylogenetic analyses strongly placed *P. oblonga* as sister to *P. vitrea* (Fig. 2.7). *Ploeotia oblonga* was previously considered to be closely allied to *Serpenomonas costata* based on light microscopy, in particular the alternating wide and narrow pellicle strips (e.g. Al-Qassab et al. 2002, Fig. 4k in Patterson and Simpson 1996). SEM-investigation confirmed that the arrangement of alternating pellicle strips is similar to *S. costata*, but showed that the strip connections were located on raised keels, similar to *P. vitrea* (whereas these boundaries are folded over in *S. costata*; Farmer & Triemer 1988).

Serpenomonas costata/Ploeotia costata is closely related to *Ploeotia vitrea* and *P. oblonga*, such that it is a subjective decision whether to assign this taxon to the genus *Ploeotia* (Farmer & Triemer 1988, Larsen & Patterson 1990) or to a separate genus, *Serpenomonas* (Triemer 1986, Cavalier-Smith 2016). We recommend *Serpenomonas*, since we anticipate that future researchers will favour a larger number of narrower genera: Currently broadly 'defined' morphospecies will likely be subdivided into

separate species taxa based on their divergent molecular sequences (e.g. *S. costata* or *O. polycarbonata*, Fig. 2.7). It is also likely that many more morphospecies of ploeotids will be found with more comprehensive taxon sampling. These increases in species will likely push genera to be less encompassing.

2.5.3 Some *Anisonema* Species are 'Ploeotids'

Our analyses, and several previous investigations, place *Anisonema* within Spirocuta with strong support (e.g. Lax & Simpson 2013, Paerschke et al. 2017). This is based on sequences from several different *Anisonema* isolates (Busse et al. 2003, Lax & Simpson 2013), most of them very similar to (or the same as) *Anisonema acinus*, which represents the type for the genus. Nonetheless, *Anisonema* has had a broad morphological identity—cells with widely spaced and almost longitudinal pellicle strips, gliding on the posterior flagellum, with the feeding apparatus not seen by light microscopy (Larsen 1987, Larsen & Patterson 1990, Lee & Patterson 2000)—and it has been unclear whether it represents a natural group (Larsen 1987, Lee & Patterson 2000). In our analyses, cells of *Hemiolia trepidum* (formerly *A. trepidum*) and *Liburna glaciale* (formerly *A. glaciale*) robustly fall outside Spirocuta (Figures 2.7 and 2.8). *Anisonema* species differ in their flexibility, and while *A. acinus* is often reported to be slightly flexible when compressed or stressed (Lee et al. 2005), both *Hemiolia* and *Liburna* are more rigid (Larsen 1987), consistent with their phylogenetic placement.

Hemiolia and *Liburna* share several unusual characteristics: Unlike *A. acinus*, both are fast gliders with their anterior flagellum held to the right side, trembling rapidly (hence 'trepidum' = trembling), and often freezing for a few seconds before resuming movement (hence 'glaciale' = ice; Larsen 1987, Larsen & Patterson 1990). *Hemiolia* and *Liburna* mostly jerk back when stressed and always change direction when doing so, hold their cell bodies close to the surface, and exhibit smooth gliding (see Videos 5 and 6), while *A. acinus* (for example) often moves erratically, frequently jerking backwards (e.g. Larsen 1987).

Within the ploeotid context, the particular locomotion behaviour of *Hemiolia* and *Liburna* is again distinctive, as is their feeding apparatuses not being visible by light microscopy (Larsen & Patterson 1990, Lee & Patterson 2000). With these similarities, it is perhaps surprising that *Hemiolia* and *Liburna* show considerable SSU rDNA sequence divergence between them, and while *Hemiolia* and *Liburna* branch together in all our analyses, this relationship is never supported statistically (Figures

2.7 and 2.8). For this reason, and bearing in mind recent splitting approaches to ploeotid taxonomy (Cavalier-Smith 2016), we propose to make *Anisonema trepidum* the type of a novel genus, *Hemiolia*, and *Anisonema glaciale* the type of a second new genus, *Liburna*. It is possible that *Hemiolia* and *Liburna* may have differing pellicle arrangements: It has been estimated previously that there are 6–7 strips in *Hemiolia trepidum*, and 12–15 in *Liburna glaciale* (Ekebom et al. 1995, Lee & Patterson 2000). A large number of strips in *Liburna* is consistent with our observations of 5–6 strip joints on the dorsal side of *L. glaciale*. Nonetheless, the strips are difficult to observe in both taxa, and accurate counts will likely require electron microscopy data. At present *Hemiolia* and *Liburna* (or at least their type species) are most reliably distinguished by size (Table A.2; Larsen & Patterson 1990, Lee & Patterson 2000).

The phylogenetic placement of *Liburna* and *Hemiolia* demonstrates that the morphological boundaries in the genus *Anisonema* are dubious, as long suspected (Larsen & Patterson 1990). In the future it is possible that other morphotypes currently in *Anisonema* will also need to be moved to other genera. Additional molecular phylogenetic studies are needed to increase taxon sampling in this group, and to properly place its current members in a phylogenetic context.

2.5.4 Higher-Order Systematics

Most work on ploeotids has examined their diversity and/or phylogeny, and until recently there were essentially no treatments of their higher systematics. However, Cavalier-Smith (Cavalier-Smith 2016, Cavalier-Smith et al. 2016) proposed a system of 18 higher taxa from the level of family up to infraphylum, that grouped together genera of ploeotids, or some ploeotids with other subgroups of euglenids. This was based on the sequence data available at the time, plus some inferences from morphology. Several unusual aspects of this system are important for context: 15 of the taxa were new, all the 'families' contained a single genus, one new taxon was explicitly envisaged as a paraphyletic group, and none of the other taxa for which there were data from >1 ploeotid genus corresponded to a strongly supported clade in Cavalier-Smith's (2016) own SSU rDNA phylogenies. Startlingly, none of the proposed higher taxa containing >1 ploeotid genus form even a weakly supported clade in our analyses with their current compositions.

The monophyly of some of the proposed major taxa is strongly disconfirmed by our phylogenies. In particular, the class Ploeotarea and order Ploeotiida group *Ploeotia* with *Lentomonas* only, while class Stavomonadea groups *Serpenomonas* with peta-

lomonads, *Keelungia* and *Decastava* (Cavalier-Smith 2016). This dichotomy sharply conflicts with the strongly supported close relationship between *Ploeotia* and *Serpenomonas* to the exclusion of *Lentomonas*, petalomonads, *Keelungia* and *Decastava*. This situation stems mainly from Cavalier-Smith (2016) assuming that *Ploeotia vitrea* is closely related to *Ploeotia* c.f. *vitrea* sensu Lax and Simpson (2013; now *Olkasia polycarbonata*), which proves not to be the case. The assignation of *Lentomonas* to Ploeotarea and Ploeotiida was made without molecular data, apparently on the basis of similar pellicle and feeding apparatus structure (see Diagnosis of Ploeotarea in Cavalier-Smith 2016). Given the closer relationship of *Ploeotia* with *Serpenomonas*, these similarities at best identify a broader group than *Ploeotia* plus *Lentomonas*. The close relationship between *Serpenomonas* and *Ploeotia* also argues against assigning *Serpenomonas* to its own subclass Heterostavia, order Heterostavida and family Serpenomonadidae, especially since the heterostavous condition is also seen in *Ploeotia oblonga* (see below).

Meanwhile, another set of taxa in Cavalier-Smith's (2016) system were not strongly contradicted by our analyses, but nonetheless were never recovered as clades, irrespective of whether the three new genera we propose are considered. These unsupported taxa include the subclass Homostavia (uniting *Decastava* and *Keelungia* with petalomonads), order Decastavida (uniting *Decastava* and *Keelungia*), and superclass Rigimonada (uniting all ploeotids and petalomonads, to the exclusion of *Entosiphon*), though the latter was actually envisaged as paraphyletic (Cavalier-Smith 2016). We also did not recover *Entosiphon* as a sister clade to all other euglenids (excluding symbiontids, Fig. 2.8) as implied by the division of euglenids into two infraphyla: Entosiphona (*Entosiphon* only), and Dipilida (all other euglenids). *Entosiphon* represents an extremely long branch in SSU rDNA phylogenies, and this marker is widely recognised as unreliable for inferring the placement of this genus (Cavalier-Smith et al. 2016, Paerschke et al. 2017, Busse et al. 2003, von der Heyden et al. 2005). Unfortunately, our study confirms that SSU rDNA is currently the only marker with taxon sampling anywhere near rich enough to credibly address the deepest phylogenetic divisions within euglenids. Thus, while the *Entosiphona*/Dipilida distinction draws mainly on Hsp90 phylogenies and morphological arguments, rather than SSU rDNA analyses (Cavalier-Smith 2016, Cavalier-Smith et al. 2016), the only ploeotids in those Hsp90 phylogenies are *Entosiphon* and *Decastava*.

As a result of our investigations, the entire systematisation above the level of family involving ploeotid taxa proposed by Cavalier-Smith (2016) seems to be without utility, and in our opinion, should be overlooked. The system was proposed on the

basis of weak phylogenetic evidence in the first place, and—with the benefit of additional data—proves to be largely (likely entirely) incompatible with the widely held ideal of higher taxa being monophyletic where possible. Even with the improved taxon sampling of our study, the deep-level phylogenetic relationships amongst euglenids remain unresolved, and any replacement proposal for supra-familial taxa (or dramatic alteration of the concepts applying to existing names) runs an unacceptably high risk of a similar incompatibility. Systematics should wait until repeatable, strongly supported molecular phylogenetic results are obtained; for euglenids these are likely to be available soon, through low-cost transcriptomics (see below).

Serpenomonas and *Ploeotia* turn out to be very closely related and genetically similar, such that it is a more-or-less subjective decision whether to regard them as separate genera (see above). We anticipate that assigning them to the same family will be uncontroversial. We propose Ploeotiidae Cavalier-Smith 2016, rather than Serpenomonadidae Cavalier-Smith 2016, since *Ploeotia* Dujardin 1841 has priority over *Serpenomonas* Triemer 1986. This would minimise confusion if *Ploeotia* and *Serpenomonas* are considered synonymous in the future.

The presence or absence of an unpaired U nucleotide in conserved helix 44 of the V9 region of SSU rRNA has recently been identified as a character that is potentially relevant to euglenid phylogenetics (and systematics), and specifically provides some evidence for a Spirocuta + *Entosiphon* grouping (Paerschke et al. 2017). This unpaired U is present in more basal euglenids like petalomonads, *Serpenomonas costata*, *Keelungia*, and *Olkasia*, but not in Spirocuta or *Entosiphon* (Paerschke et al. 2017). As expected, all new sequences of *S. costata*, *Ploeotia vitrea* and *P. oblonga*, *Keelungia*, and *O. polycarbonata* have the unpaired U (data not shown). This nucleotide is absent from *Entosiphon* sp. ESC (as expected), but also not present in *Lentomonas*, *Hemiolia* and unidentified ploeotid WF2_3. Meanwhile, *Liburna* sequences have two unpaired Us at this position. We conclude that the absence of this unpaired base is more widespread in euglenids than was previously supposed. While this character might still represent a synapomorphy of a group including Spirocuta + *Entosiphon*, it is more likely that it has a complex evolutionary history. The unresolved branching order of euglenids currently inhibits any further analysis past speculation.

2.5.5 Phylogenetic Structure within Morphospecies

At present, most heterotrophic euglenid species are defined by morphology applied at the light microscopy level, with the majority of broad morphospecies distinctions

based on criteria like cell shape and size, surface structure, flexibility and whether the feeding apparatus is visible (e.g. Lee 2012). We now have sufficient molecular sampling for a few ploeotid morphospecies; to investigate their sequence divergence and phylogenetic structure more closely. *Serpenomonas costata* is split into two maximally supported, distinct subgroups (Fig. 2.7): clade A (taxa from North America/Europe, includes the type strain CCAP 1265/1), and clade B (taxa from Asia/Caribbean, includes strain KM057). Intriguingly, taxa in clade A all have group I introns, whereas none were found in clade B (see below; Fig. 2.9). This finding is largely consistent with Chan et al. (2015), who reported *S. costata* sequences isolated from Taiwan. This tentative biogeographical separation could be an artefact of still-modest taxon sampling, but is worth further investigation. Sequence similarity within both clades was high, was only 73.3–75.8% between clade A and B isolates (considering full-length sequences without introns). *Olkasia* is also split into moderately supported A and B clades (Fig. 2.7). The two *Olkasia* clades are certainly not biogeographically distinguished since both have been found at one site in Nova Scotia, Canada. Sequence identities within clades A and B were high (97.9 and 99.1%), while identity was 90.6–93.5% between clades (Table A.3). It is probable that *Serpenomonas costata* will be split up into at least two nomenclatural species in the future, and possible that *Olkasia polycarbonata* will be as well.

Conversely, *Lentomonas corrugata* and *L. azurina* sequences do not form separate clades in SSU-rRNA gene phylogenies. There is no clear distinction in their SSU-rRNA gene identities: 97.2–99.8% of sites between *L. corrugata* and *L. azurina* are identical, while there is 95–99.8% identity amongst different *L. corrugata* cells. *Lentomonas azurina* was originally distinguished from *L. corrugata* by being substantially more slender (Patterson & Simpson 1996), and a sharp distinction in length:width ratio was observed in the cells we examined (Table A.2). It is possible that the *L. corrugata* morphospecies gave rise to the *L. azurina* morphospecies, or they could also represent a single species, with *L. azurina* representing a rarer life cycle stage of *L. corrugata*. It is also unclear how much morphological variation is seen within clones of *Lentomonas* morphotypes, especially in length:width ratio. It is also possible that our STS5 cell represents an aberrant or sick cell, not representative of 'true' *L. azurina*. More comprehensive molecular sampling of this group, especially the less commonly observed *L. azurina*, is needed to resolve these issues, ideally from different geographic locations, and with the use of cultures.

2.5.6 Pellicle Strip Architecture

Broadly speaking, differences in the organisation of individual pellicle strips seem to distinguish the major phylogenetic groupings of ploeotids, which is consistent with patterns seen in other euglenids, especially phototrophs (Leander et al. 2007, Yubuki & Leander 2012). For example, taxa with straight, bifurcating keels (strip boundaries on raised keels) fall within *Ploeotia*. The closely related taxon *Serpenomonas* also has bifurcations, though they are at the edges of a trough-like structure made by the narrow strips. In contrast, *Lentomonas* has raised bifurcating ridges (dorsal side only) with strip joints located on one side of the ridges (i.e. not straight, like in *Ploeotia*). Nonetheless, pellicle strips need to be examined carefully, and while light microscopy is easily accessible, scanning electron microscopy helps greatly in understanding the organization of strips (Esson & Leander 2006). One example is *P. oblonga*: it has previously been thought that this morphospecies is most closely allied with *S. costata* (Larsen & Patterson 1990, Lee & Patterson 2002, Lee 2008), since *P. oblonga* resembles *S. costata* in having alternating thin and wide strips. Our SEM analyses revealed that the strip boundaries are considerably different in *P. oblonga*, being raised on keels like in *P. vitrea* (Fig. 2.2).

It is likely that pellicle strip structure has a complex evolutionary history within ploeotids, for example the alternation of narrower and wider strips likely arose at least three times independently: in *Ploeotia oblonga*, *Serpenomonas*, and, arguably, *Entosiphon* (it is also possible it arose twice, and was subsequently lost in *Ploeotia vitrea*). This alternation in strip width could be explained by pellicle strip morphogenesis during cell division. Cavalier-Smith (2017) reinterpreted data from Triemer & Fritz (1988), Leander & Farmer (2001) and Leander et al. (2007). He inferred that during division of *S. costata,* the narrow strips mature into wide strips, and ten new narrow strips inserted at their flanks (Cavalier-Smith 2017). If so, this process may help explain the occurrence of heteromorphic pellicle strips in multiple taxa of ploeotids including *P. oblonga*, since it is—in principle—a simple difference in pellicle strip development timing (heterochrony). This phenomenon has been inferred to explain pellicle whorl patterns in phototrophic euglenids (Esson & Leander 2006, Leander et al. 2007, Esson & Leander 2008). Further cell-developmental investigations of *Serpenomonas*, *Entosiphon* and/or *P. oblonga* (and other taxa with heteromorphic strips) would be illuminating.

Our SEM observations show that *Olkasia* has S-shaped pellicular strips, with the joints between strips lying under an overhang. Of all ploeotid groups examined so far, this is arguably the most similar to the strip morphology of typical spirocutes

(Leander et al. 2001a). In addition, the chisel-shaped feeding apparatus of *Olkasia* differs from the hook-shaped apparatus seen in the other ploeotids sequenced to date, in which the feeding apparatus is clearly visible, other than the special case of *Entosiphon* (Lee 2008, Chan et al. 2013, Cavalier-Smith et al. 2016). Chisel-shaped feeding apparatuses are common in phagotrophic spirocutes, as well in some ploeotids for which there are no sequence data (Larsen & Patterson 1990, Lee 2008). These similarities are worth further examination, especially bearing in mind that *Olkasia* branches as a sister group to spirocutes in our phylogenetic analyses (with negligible support), as well as in some other recent analyes (Cavalier-Smith 2016), although not in all (Paerschke et al. 2017).

2.5.7 Eukaryotrophy in Ploeotids

It has been proposed that phagotrophic euglenids were ancestrally bacterivorous and that the development of a flexible pellicle in an ancestor of Spirocuta was needed for phagotrophic euglenids to become effective eukaryotrophs (Leander et al. 2001a, Leander 2004). This eukaryotrophy subsequently enabled a phagotrophic euglenid to participate in secondary endosymbiosis as the host (Leander et al. 2001a). This association has been questioned, however (Lax & Simpson 2013, Cavalier-Smith 2016), and our study provides further evidence that at least some rigid euglenid taxa (like ploeotids) can be effective eukaryotrophs. All of the *Olkasia* cells we isolated contained ingested algae (Fig. 2.4), as did the cell studied by Lax & Simpson (2013). Further, we observed ingested algal material in *Ploeotia vitrea*, *P. oblonga*, *Entosiphon* sp. (Figures 2.1f–h and 2.6m, n), and *Keelungia* sp., and *Serpenomonas costata*, with the latter already reported to be able to ingest algae and other eukaryotic cells (Linton & Triemer 1999). In this study we also demonstrated that some taxa formerly assigned to *Anisonema* branch outside Spirocuta. These taxa, *Hemiolia* and *Liburna*, seem to include accomplished eukaryotrophs: Several of our isolated cells of *Hemiolia* and *Liburna* contained whole ingested pennate diatoms up to 19 μm in length (Fig. 2.6e, i–k). Both *Hemiolia* and *Liburna* have previously been reported to contain 'small granules' (Lee & Patterson 2000, Lee 2008), which might be ingested prey, as mentioned as a possibility in the original description (Larsen 1987). We conclude that the ability to ingest smaller eukaryotes is widespread among ploeotids *sensu lato*.

2.5.8 Group I and Non-Canonical Introns

A group I intron was discovered in *Serpenomonas* costata strain CCAP 1265/1 by Busse & Preisfeld (2003b). Chan et al. (2015) subsequently sequenced two additional *Serpenomonas* isolates that lacked group I introns. With further sampling we found group I introns across *Serpenomonas* clade A, which includes CCAP 1265/1, but did not find any in clade B, to which Chan et al.'s sequences belong (Fig. 2.9a; see above). Interestingly, we also found a group I intron of similar length at the same position in most of our *Ploeotia vitrea* isolates (Fig. 2.9a).

Gain and loss of a group I intron is known as the Goddard-Burt cycle (Goddard & Burt 1999, Haugen et al. 2005), and is initiated with a cutting site created by a homing endonuclease (HE). This evolutionary process means that soon after gain of the intron, the HE-gene is still present, but is subsequently truncated prior to loss. Homing endonuclease genes were absent from the introns in *S. costata* clade A and *P. vitrea*, suggesting that these are in the loss phase of the cycle. A phylogenetic analysis based on the dataset of Busse & Preisfeld (2003b) showed that the *P. vitrea* group I intron is not specifically related to that in *S. costata* (data not shown). It is thus likely that these introns invaded *Serpenomonas* and *Ploeotia* independently (and *S. costata* clade B may never have been invaded), which is consistent with the genetic divergence between *Ploeotia* and *Serpenomonas*.

We were not able to detect any additional group I or II intron sequences at other positions with RNAweasel (see Methods), but the SSU rDNA alignment showed conspicuous insertions at two other positions in some *Serpenomonas costata* and *Ploeotia vitrea* isolates, and we confirmed in two cases that these insertions are absent from the rRNA (Fig. 2.9a). We inferred the secondary structures of these introns and found at least 8 of the 10 conserved core domains (P2–P9; data not shown) that are characteristic of group I introns (Haugen et al. 2005). In the phylogenetic analysis with group I introns mentioned above, sequences from these positions formed a clade that was clearly separate from the introns at position 1.

In addition, we recorded five 158–170 bp long inserts in *Keelungia* sp. strain KM082, which did not share positions with the introns in *S. costata* and *P. vitrea* (Fig. 2.9a). These insertions were not present in cDNA-derived SSU-rRNA sequences, suggesting that they are also introns. While group I introns shorter than 200 bp have been found (Zhou et al. 2007), our RNAweasel analysis again did not identify them as group I or II introns or other related self-splicing entities, and we were not able to construct secondary structures. Additional examination showed that the introns at these positions have similarities to the non-canonical introns found in several

protein-coding genes of euglenozoans, mainly *Euglena gracilis* (Henze et al. 1995, Breckenridge et al. 1999, Milanowski et al. 2014, Guminska et al. 2018) and diplonemids (Gawryluk et al. 2016). Data for phagotrophic euglenids are sparse, but there are several short, non-canonical introns in the Hsp90 gene of the spirocute *Peranema trichophorum* (Breglia & Leander 2007). In addition to possessing a stable secondary structure bringing both splicing sites together, these introns have short 2–4 bp direct repeats in one end of the intron and in the opposing end of the flanking region (Henze et al. 1995). We speculate that these features in KM082 *Keelungia* sp. might represent non-canonical introns. Non-canonical introns are common in protein coding genes of euglenids (Henze et al. 1995, Breglia & Leander 2007, Guminska et al. 2018), but have not—to our knowledge—been reported in rRNA genes. More research into the distribution, structure, and splicing of these putative non-canonical introns is merited.

2.5.9 Have We Found all Major Clades of Ploeotids?

We here report sequences from ten different morphospecies of ploeotids, many of which were formerly assigned to genus *Ploeotia*, whereas there are at least 22 described morphospecies in *Ploeotia* alone (for example see Larsen & Patterson 1990, Patterson & Simpson 1996, Al-Qassab et al. 2002). It is possible that some of these other species represent additional major clades of ploeotids, beyond the eight identified thus far. This is especially true for taxa with arrangements of pellicle strips that differ from those of the identified major clades and/or where the fine structure of strip joints is unknown, like *Ploeotia adhaerens* or *P. scrobiculata* (Larsen & Patterson 1990). Until molecular data are obtained from these morphospecies, their phylogenetic placement will remain highly uncertain: Their assignation to *Ploeotia* must be seen as provisional, but should be retained until clear evidence of their actual phylogenetic affinities is available.

2.5.10 The Need for Multigene Phylogenies

In all of our analyses ploeotids appear as paraphyletic, forming the base of the euglenid tree, if the question of the position and identity of symbiontids is overlooked (Figures 2.7 and 2.8). On its own, this suggests that a ploeotid-like organism could indeed be the ancestor from which all other euglenids arose, however the branching order amongst the major ploeotid clades was very poorly resolved and supported in our study. Although we now have a clearer picture of the number and diversity

of the major groups of ploeotids, it is obvious that SSU rDNA sequences alone will not resolve relationships between those groups. Multigene phylogenetics are likely needed to resolve deep euglenid phylogeny. This approach has been used successfully in a range of protist groups, including phototrophic euglenids (e.g. Karnkowska et al. 2015). In addition to using culture-based approaches, methods like single-cell transcriptomics have the potential to rapidly increase taxon sampling for phylogenomic studies, especially for taxa that are difficult to culture (Kolisko et al. 2014, Lax et al. 2018). Our identification of major groups of ploeotids is an important step in establishing what taxon sampling is appropriate for future phylogenomic analyses.

2.6 Acknowledgments

We thank the Tula Foundation for supporting field work on Calvert and Quadra Island, CARMABI for enabling field work in Curaçao, Noriko Okamoto and Patrick Keeling (University of British Columbia) for help with sampling and use of their microscope while in the field, and to Michelle Leger for providing the source samples for cultures MX-CHA and ESC. We thank Ping Li and Pat Scallion (both Dalhousie University) for assistance with SEM-imaging, Michael Gray (Dalhousie University) for feedback on introns in euglenids, and Claire Burnard for some sampling assistance. The research was supported by NSERC grant 298366-2014 to AGBS, and by CIfAR. WJL was supported by grants from the National Institute of Biological Resources, Korea (NIBR201801202) and the National Research Foundation of Korea (NRF-2018R1A2A3075567).

Chapter 3

Towards a Resolved Tree of Euglenids: A Single-cell and Phylogenomics Approach

Lax G, Kolisko M, Eglit Y, Burger G, Yubuki N, Karnkowska A, Soukal P, Hampl V, Keeling P, Simpson A (in preparation).
Towards a resolved tree of euglenids: A single-cell and phylogenomics approach.

Authors' contributions: GL and AS conceived project idea, GL isolated, identified, imaged and generated SSU rDNA molecular data from single cells. GL generated single-cell transcriptome data with help from MK and PK. YE isolated cultures, NY, AK, and GB provided some mass-culture transcriptome data. GL conducted phylogenetic analyses, and wrote manuscript with input from AS and other authors.

¶

3.1 Introduction

EUGLENIDA (Discoba; Euglenozoa) is a major and diverse group of microbial eukaryotes that inhabit freshwater, soil and marine environments. Euglenids

show a variety of trophic modes, including photoautotrophy, osmotrophy, and phagotrophy (Leander et al. 2017). The photoautotrophic clade (green euglenids; Euglenophyceae) arose through a secondary endosymbiosis of a phagotrophic euglenid host with a pyramimonadalian green alga (Jackson et al. 2018, Turmel et al. 2008). Irrespective of trophic mode, almost all euglenids are unicellular flagellates with one or two flagella. The main morphological apomorphy is the euglenid pellicle, a system of overlapping proteinaceous strips beneath the cell membrane that run longitudinally or spirally along the cell (Leander et al. 2017). The strips can slide actively against each other in many species, enabling some cells to undergo squirming or peristalsis-like deformations called 'euglenoid motion' or 'metaboly' (Leander et al. 2001b, 2017). Different taxa can have dramatically differing numbers of pellicle strips; from a minimum of four up to more than 100. Consequently, euglenids exhibit a vast diversity in morphology, ranging from rigid cells with few but sometimes elaborately shaped strips (e.g. some species of *Petalomonas*), to large and highly flexible cells with dozens of similar strips (e.g. many photoautotrophic euglenids). Almost all phagotrophic euglenids exhibit some form of flagellar gliding to move across surfaces, except apparently sessile *Dolium* (Larsen & Patterson 1990, Leander et al. 2017). Some taxa glide on their anterior flagellum (e.g. *Petalomonas, Notosolenus, Urceolus, Peranema*), whereas others glide on their posterior flagellum (e.g. *Anisonema, Dinema, Ploeotia, Lentomonas*).

3.1.1 General Structure of Euglenid Phylogenies

Evolutionary relationships among euglenids are poorly understood at this moment, particularly at deeper phylogenetic levels. Photoautotrophic euglenids (Euglenophyceae) and primary osmotrophs (Aphagea) both represent constrained clades with reasonably- to well-resolved internal relationships (Karnkowska et al. 2015, Busse & Preisfeld 2003a, Preisfeld et al. 2001). By contrast, phagotrophs are a sprawling, paraphyletic assemblage (Lax et al. 2019, Paerschke et al. 2017, Cavalier-Smith 2016): As such, deep-level euglenid phylogeny and evolution is fundamentally about understanding the relationships among phagotrophs. To date, almost every molecular phylogeny encompassing all euglenids has examined the small subunit of ribosomal DNA (SSU rDNA, but see below). Even this most-studied marker has long suffered from undersampling of phagotrophic taxa, to the extent that it was unclear what the major phylogenetic groups of phagotrophic euglenids were (Lax et al. 2019). Sampling efforts in recent years have substantially increased the number of phagotroph

SSU rDNA sequences available (Lax et al. 2019, Schoenle et al. 2019, Cavalier-Smith et al. 2016, Chan et al. 2015, 2013), yet much more diversity is known from morphological studies that remains without molecular data. One further issue is the divergent nature of the euglenid SSU rDNA (Busse & Preisfeld 2003a, Busse et al. 2003, Łukomska-Kowalczyk et al. 2016, Cavalier-Smith et al. 2016) that complicates PCR-amplification with standard eukaryotic primers as well as subsequent phylogenetic analyses. Two studies used the Hsp90 gene to infer phylogenies of euglenids, but suffer even more from taxonomic undersampling (Breglia & Leander 2007, Cavalier-Smith et al. 2016).

3.1.2 Relationships within Spirocuta

Single-gene phylogenies have helped to resolve some relationships: in particular, confirming the monophyly of Spirocuta (Lee & Simpson 2014b, Cavalier-Smith et al. 2016, Paerschke et al. 2017). This clade was originally inferred mainly from morphological studies (Leander et al. 2001a, Leander & Farmer 2001) and includes taxa with more than 14 pellicle strips, many of which are highly flexible. It encompasses the photoautotrophs, primary osmotrophs, and a large number of flexible phagotrophs (Leander et al. 2017). The spirocute phagotrophs can be divided into two morphological catergories: Anisonemids and peranemids. Anisonemids glide on their posterior flagellum and can be modestly flexible (e.g. *Anisonema* and *Dinema*), whereas peranemids often only have a single (anterior) emergent flagellum that they glide on and are usually highly metabolic (e.g. *Peranema, Jenningsia, Urceolus, Heteronema*). While Spirocuta is now robustly supported by several molecular analyses of the SSU rDNA and Hsp90 (Paerschke et al. 2017, Cavalier-Smith et al. 2016, Lax et al. 2019), relationships of phagotrophs within the clade are currently still unresolved, as is the sistergroup to phototrophic euglenids. As for other euglenid groups, taxon sampling within Spirocuta remains poor at best for SSU rDNA, or any other marker gene.

3.1.3 Relationships at the Base of the Euglenid Tree

The relationships of non-spirocute euglenids are even less clear. These taxa, all of which are phagotrophs, can be subdivided into two main morphological categories: petalomonads and ploeotids. Ploeotids are all rigid cells with a few (usually 10) pellicular strips that glide on their posterior flagellum and are represented by the genera *Ploeotia, Serpenomonas, Entosiphon, Keelungia, Decastava, Lentomonas, Olkasia, Liburna*, and *Hemiolia* (Lax et al. 2019). Ploeotids have always been recovered as a

paraphyletic group in SSU rDNA phylogenies (Lax et al. 2019, Cavalier-Smith et al. 2016, Paerschke et al. 2017, Schoenle et al. 2019), though the relationships between ploeotid genera vary tremendously across analyses. The only relationships recovered with strong support in all analyses is the *Ploeotia* + *Serpenomonas* grouping (Ploeotiidae; Lax et al. 2019), although several analyses also show a poorly-to-moderately supported *Keelungia* + *Decastava* grouping (Paerschke et al. 2017, Schoenle et al. 2019, Lax et al. 2019). All molecular sequences currently available for petalomonads on the other hand, fall within a single clade (Lee & Simpson 2014a, Cavalier-Smith et al. 2016). All petalomonads are rigid, with 4–10 pellicle strips, and glide on their anterior flagellum. This group is known to be particularly rich in species and show considerable morphological variation in pellicle structure (Huber-Pestalozzi 1955, Larsen & Patterson 1990, Lee & Patterson 2000). The majority of species fall within genera *Petalomonas* and *Notosolenus*, which together with *Scytomonas* and *Biundula* (recently proposed to be distinguished from *Petalomonas*), have at least some molecular data. Despite their species richness, molecular sampling for petalomonads remains low and, again, is largely restricted to the SSU rDNA (e.g. Schoenle et al. 2019, Cavalier-Smith et al. 2016). Further a phylogenetic study with just two *Notosolenus* species recovered *Notosolenus* as non-monophyletic (Lee & Simpson 2014a), highlighting potential conflicts with morphologically defined genera and phylogenetic clades. Several other genera, including *Sphenomonas, Dolium, Calycimonas*, and *Atraktomonas* are potential members of this group, based on their morphology, but lack molecular study (Lee & Simpson 2014a). Morphological cladistic and early molecular phylogenetic analyses placed petalomonads as the deepest branch (or branches) within euglenids (Leander et al. 2001a, Breglia & Leander 2007, Montegut-Felkner & Triemer 1997, Müllner et al. 2001, Preisfeld et al. 2000, 2001), but recent molecular phylogenetic analyses have usually placed them among ploeotids, albeit with weak statistical support (Lax et al. 2019, Lax & Simpson 2013, Chan et al. 2013, Schoenle et al. 2019, Cavalier-Smith et al. 2016).

3.1.4 Symbiontida

Symbiontids are an enigmatic group of euglenozoans that inhabit low-oxygen saline environments (Yubuki et al. 2009, Breglia et al. 2010, Edgcomb et al. 2010), and are represented by a large number of environmental SSU rDNA sequences that fall into several distinct clades (Yubuki & Leander 2018). Symbiontids host an array of episymbiotic bacteria (Edgcomb et al. 2010), in some cases including magneto-

tactic Deltaproteobacteria that enable the host cells to align along the geomagnetic field (Monteil et al. 2019). Three SSU rDNA sequences from two described symbiontid taxa are available (*Calkinsia aureus, Bihospites bacati*; Yubuki et al. 2009, Breglia et al. 2010), as well as ultrastructural data from three species (*C. aureus, B. bacati, Postgaardi mariagerensis:* Yubuki et al. 2013). Yet the exact placement of this group within Euglenozoa is currently unresolved, as symbiontids clearly possess euglenozoan traits, but do not show the defining synampomorphies of euglenids, kinetoplastids, or diplonemids (Yubuki & Leander 2018, Yubuki et al. 2009, 2013, Simpson et al. 1997). SSU rDNA based phylogenetic analyses place them either within euglenids or as sister to them (e.g. Breglia et al. 2010, Cavalier-Smith 2016), and ultrastructural data shows *Bihospites* to have a cell surface organised by S-shaped folds reminiscent of the euglenid pellicle (Yubuki et al. 2013, Breglia et al. 2010). These conflicting results have lead researchers to treat them either as derived euglenids (e.g. Breglia et al. 2010), or as a distinct euglenozoan clade (e.g. Simpson et al. 1997, Cavalier-Smith 2016).

3.1.5 Increasing Taxon Sampling for SSU rDNA and Multigene Analyses

In an effort to increase taxon sampling among phagotrophic euglenids, especially spirocutes and petalomonads, we generated 73 novel SSU rDNA sequences using single-cell molecular approaches. Guided by this improved understanding of euglenid diversity, we generated 29 single-cell and mass-culture derived transcriptomes for a broad sampling across phagotrophic euglenid diversity. We used this data to enable the first multi-gene phylogenetic analyses of euglenids encompassing phagotrophs, and thus to address several of the outstanding questions of euglenid phylogenetics, including relationships within Spirocuta (e.g. identifying the closest living relatives of Euglenophyceae), the branching order near the base of the euglenid tree, and the placement of symbiontids.

3.2 Methods

3.2.1 Single-Cell Isolation & Photodocumentation

Marine and freshwater sediment and detritus samples were prepared as reported by Larsen & Patterson (1990). Briefly, sediment and detritus was collected from vari-

ous sites in Eastern and Western Canada, and Curaçao (Table B.1) and placed into small trays, spread out to a height of 1–2 cm. A tissue paper (Kimwipe) was then placed on top and 50 x 20 mm coverslips were then added to the surface. Containers were incubated for 12–72h under ambient sunlight, following a day-night cycle. Then, coverslips were examined, bottom facing up, on an inverted microscope (Zeiss Axiovert 200M under 1000x total magnification or Leica DM IL LED under 400x or 630x). After imaging (Zeiss AxioCam M5 or Sony NEX6, respectively), individual cells of interest were isolated using a drawn-out glass pipette, and washed in 1–5 µl sterile seawater, tapwater, or distilled water. For cells destined for single-cell transcriptomics, extra care was taken to gently aspirate and wash them in order to minimise stress.

3.2.2 Single-Cell SSU rDNA Sequencing

Isolated cells were placed in 0.2 ml PCR-tubes containing 9.5 µl of distilled water and stored at -80˚C. To lyse the cells, the tubes were subjected to 5–10 freeze thaw cycles (between -20˚C and room temperature), with cells with fewer pellicle strips undergoing more cycles. Genomic DNA was amplified using an Illustra GenomiPhi v3 kit (GE Healthcare), following the manufacturer's recommendations, but using a 2h-long isothermal amplification step at 30˚C. Success of the MDA reactions was assessed by running 1 µl of product on a 0.5% agarose gel.

Nearly full-length SSU rDNA sequences were obtained by amplifying and sequencing two overlapping fragments of the gene, using PCR with a variety of euglenid- and Spirocuta-biased primers (Table B.2). PCR reactions contained 2 mM $MgCl_2$ and 0.2 mM dNTPs, and were carried out with recombinant *Taq* polymerase (Invitrogen). Initial denaturation was done at 95˚C for 3 min and followed by 35 cycles of: denaturing at 95˚C for 30 sec, annealing at 50–64˚C for 30 sec (see Table B.2 for primer sequences, combinations and melting temperatures), and elongation at 72˚C for 2 min; followed by a final elongation step at 72˚C for 10 min. PCR products were visualized with gel-electrophoresis on a 1% agarose gel. Appropriately-sized PCR products were either directly sequenced using Sanger technology (Génome Québec), or first gel-extracted using a Qiagen Gel-Extraction kit. Raw reads were visually checked and assembled de novo in Geneious R10 (Kearse et al. 2012), and resulting sequences queried against the NCBI GenBank nt database using blastn to identify contaminant sequences.

The SSU rRNA sequences present (as cDNA) within the single-cell transcriptome data (see below) were extracted using barrnap version 0.9 (`https://github.com/tseemann/barrnap/`). In addition, SSU rDNA sequences were derived from sequencing of single amplified genomes (SAGs) of four phagotrophic euglenids (cells BP3, SDB1, SDB4, UB10). These cells were isolated and the genomic DNA amplified by MDA as described above. Then, the MDA products were used to construct Nextera XT libraries, which were sequenced on Illumina MiSeq runs with 2x250bp paired-end sequencing. After adapter- and quality-trimming with trimmomatic version 0.39 under default parameters (Bolger et al. 2014), reads were assembled with SPAdes version 3.13.0 (–sc flag; Bankevich et al. 2012) and SSU rDNA sequences extracted using barrnap.

3.2.3 Single-Cell Transcriptomics

Cells that were isolated for single-cell transcriptomics were placed in 0.2 ml PCR-tubes containing 2 µl of lysis buffer (0.2% Triton-X with added RNAse Inhibitor), and rapidly frozen in liquid nitrogen or at -80˚C. To ensure lysis of cells, tubes were subjected to 1–5 freeze-thaw cycles (between -80˚C and room temperature), with cells with fewer pellicle strips undergoing more cycles. Reverse transcription followed the SmartSeq2 protocol reported in Picelli et al. (2014). Briefly, a template-switching oligo (TSO) enables cDNA generation and subsequent amplification of all products using a limited-cycle PCR (18–21 cycles depending on the sample, Table B.3). Amplified cDNA products were subsequently purified using magnetic beads (Agilent Ampure XP), and quantified with a Qubit HS DNA assay. Then, a sample of the cDNA was cloned into *E. coli* and 10–16 clones were Sanger-sequenced to allow a preliminary assessment of the proportion of contaminant sequences among the cDNA of each cell sample. After library generation with Nextera XT, samples were sequenced on Illumina HiSeq or MiSeq systems, using 2x250 bp paired-end sequencing (see Table B.3 for multiplexing information).

3.2.4 Transcriptomes Derived from Mass Cultures

For bulk RNA-extraction, cultures of *Ploeotia vitrea* strain MX-CHA and *Notosolenus urceolatus* KM049 (Lee & Simpson 2014a) were mass-cultured in 150 mm Petri dishes. For cell harvesting, most of the medium was discarded, and the Petri dish was scraped thoroughly with a sterile cell scraper. The remaining liquid (which contained the dislodged euglenids) was then transferred to a 50 ml tube and RNA was

extracted using TRIzol (ThermoFisher), following the manufacturers' instructions. Purity and quantity of the RNA was assessed with a NanoDrop spectrophotometer (ThermoFisher).

A culture of *Rhabdomonas costata* strain PANT2 was grown in cerophyll media, and transferred every two weeks. RNA-extraction was done three times: 1) Direct isolation of mRNA with polyA-selection (Dynabeads mRNA Direct Kit), 2) Isolation of mRNA (Dynabeads mRNA Direct Kit) starting from isolated total RNA (GeneAll Hybrid-R Total RNA Purification Kit), 3) Isolation of total RNA (GeneAll Hybrid-R Total RNA Purification Kit). Library construction for all three extractions was done separately with a TruSeq Stranded mRNA kit, and sequenced separately on an Illumina MiSeq system with 2x150bp paired-end reads. Raw reads from all sequencing runs were assembled together with Trinity. The assembled transcriptome was translated to protein sequences with TransDecoder (`https://github.com/TransDecoder/`), and had its redundancy reduced with CD-HIT (default parameters; Fu et al. 2012).

RNA from a culture of *Rapaza viridis* (Yamaguchi et al. 2012) was extracted with an Ambion RNAqueous Micro Kit following the manufacturers' instructions, a library was prepared with a TruSeq Stranded mRNA kit, and sequenced on an Illumina HiSeq 2000 system with 2x150bp paired-end reads. The assembly was done with Trinity under default parameters and translated to protein sequences with TransDecoder.

Cultures of *Petalomonas cantuscygni* strain CCAP 1259/1 and *Entosiphon sulcatum* were maintained as described previously (Roy et al. 2007). For both cultures, cells were collected by centrifuging and RNA was extracted with a homemade TRIzol substitute (Rodríguez-Ezpeleta et al. 2009). Any residual DNA was removed either via RNAeasy column purification (Qiagen) or by digestion with RNase-free *DNase I* (Roche), followed by a phenol-chloroform extraction. The extracted RNA was used to construct a library using an Illumina Stranded TruSeq RNA library kit, and sequenced on MiSeq- and HiSeq-systems with 2x250 bp and 2x150 bp paired-end reads, respectively. Reads from both runs for each organism were adapter- and quality-trimmed with Cutadapt (Martin 2011), then co-assembled with rnaSPAdes under default parameters.

3.2.5 SSU rDNA Phylogenetics

Newly obtained SSU rDNA sequences were appended to a dataset based on that used by Lax et al. (2019, see Chapter 2). This 'SSU-base' dataset has a phylogenetically broad sampling across Aphagea, Euglenophyceae, and Symbiontida, and contains

all SSU rDNA sequences of phagotrophic euglenids publically available as of October 2019, except from *Entosiphon* (see below). Additionally, it contains representative kinetoplastid and diplonemid sequences as outgroups. The 'SSU-base' dataset omits sequences of the phagotrophic euglenid genus *Entosiphon*, since this taxon is extremely long-branching in SSU rDNA phylogenies (Lax et al. 2019, Paerschke et al. 2017). It also omits two SSU rDNA sequences reported in the current study, 'PP6 *Heteronema* sp.' and 'CB1 RNAseq *Chasmostoma nieuportense*', because of their divergent nature and resulting long branches in phylogenies. Several additional SSU rDNA datasets were derived from 'SSU-base', as follows. (i) A dataset with PP6, CB1, and five *Entosiphon* sequences added, was named 'SSU-LB'. (ii) Additional long-branching sequences (in addition to PP6, CB1 and *Entosiphon*) as identified by a custom script that calculates tip-to-tip distances between taxa (ranked from largest to smallest; the taxa removed were significantly longer than others) were deleted from 'SSU-base', producing 'SSU-noLB'. (iii) Removing sequences shorter than 1000 bp from 'SSU-base' produced the 'SSU-noShort' dataset. (iv) A webserver-based RogueNaRok analysis under the RNR-algorithm and strict consensus settings (Aberer et al 2013) identified several rogue taxa that were then removed from 'SSU-base', resulting in the 'SSU-noRogues' dataset. (v) We removed all non-euglenid taxa (including Symbiontida) from 'SSU-base', generating the 'SSU-euglenids-only' dataset. (vi) A dataset composed of the 'SSU-base' dataset with 17 added Discoba outgroup sequences was generated, called 'SSU-Discoba'. (vii) A Symbiontid-only dataset ('SSU-Symbiontids') with environmental sequences published or included in recent studies (Monteil et al. 2019, Yubuki & Leander 2018).

The overall dataset was aligned using MAFFT E-INS-i version 7.407 (Katoh & Standley 2013), and manually checked for misaligned sequences, which were corrected by eye in AliView version 1.26 (Larsson 2014). Masking of ambiguous positions was carried out manually in SeaView version 4.5.4 (Gouy et al. 2010), except for 'SSU-Symbiontids' where trimming was done with trimAl version 1.4 (-gt 0.8 -st 0.001; Capella-Gutiérrez et al. 2009). This yielded a 1090 site alignment for 'SSU-base' with 233 taxa ('SSU-LB': 240 taxa; 'SSU-no short': 215 taxa; 'SSU-noRogues': 224 taxa; 'SSU-euglenids-only': 165 taxa), while 'SSU-Discoba' had 1069 sites and 250 taxa, and 'SSU-Symbiontids' 903 sites and 202 taxa.

Each masked dataset was then analysed using Maximum Likelihood (ML) in RAxML version 8.2.6, under the GTR + Γ model, with 1000 non-parametric bootstrap replicates to assess robustness (Stamatakis 2014). In addition, Bayesian analyses were carried out on 'SSU-base' and 'SSU-LB' with MrBayes version 3.2.7a (Ronquist

et al. 2012) under the GTR + Γ model, running 4 chains with default heating parameters for 50,000,000 generations each. Trees were sampled every 10,000 generations with the first 25% discarded as burn-in. PSRF values (Potential Scale Reduction Factor) approaching 1.0 were used to confirm convergence of chains.

3.2.6 Assembly of Transcriptomic Data

Raw reads from the single-cell transcriptomes, as well as the transcriptomes derived from cultures were corrected using rcorrector version 1.0.4 (Song & Florea 2015), quality- and adapter-trimmed with Trimmomatic version 0.39 with default parameters, and assembled with rnaSPAdes version 3.13.1, under default parameters (Bushmanova et al. 2018). In some cases—due to assembly errors in rnaSPAdes (Table B.3)—corrected and trimmed reads were subsequently re-assembled with Trinity version 2.4.0 (Haas et al. 2013). Finished transcriptome assemblies were subjected to WinstonCleaner (`https://github.com/kolecko007/WinstonCleaner/`), to reduce cross-contamination between samples sequenced on the same MiSeq- or HiSeq-run. General assembly metrics were determined on 'clean' assemblies with transrate version 1.0.3 (Smith-Unna et al. 2016), and a proxy for assembly completeness was assessed with BUSCO version 3.0.2 (Simão et al. 2015). All metrics, BUSCO scores, assembly strategies and multiplexing information for individual samples can be found in Table B.3.

3.2.7 Phylogenomics – Dataset Construction

All new transcriptome assemblies were added to the 104-taxa, 351-gene eukaryote-wide dataset from Lax et al. (2018, see Chapter 4) using a previously published pipeline (Brown et al. 2018). We also included publically available transcriptome and genome data for additional kinetoplastid, diplonemid and euglenid taxa, using the same pipeline (Table B.4). After addition of new taxa, 351 single-gene alignments were generated with MAFFT L-INS-i version 7.0, trimmed with BMGE version 1.0 (-m BLOSUM30 -h 0.5 -g 0.2; Criscuolo & Gribaldo 2010), and single-gene phylogenies estimated for each with IQ-TREE version 1.5.5 (Nguyen et al. 2015) under the LG + C10 + F + Γ model and 1000 ultra-fast bootstrap replicates (Minh et al. 2013). All trees were manually checked for contaminant, paralogous, long-branching, or otherwise aberrant sequences, which were then removed from the dataset using a custom script. After re-aligning, re-trimming, and re-inferring single-gene phylo-

genies, trees were checked one more time, with aberrant sequences again being removed.

A final dataset was constructed by filtering all 351 genes by taxon-completeness (threshold of $\geq$70% taxa present), then choosing the genes whose trimmed alignments lengths exceeded 250 aa. These final 20 genes were reasonably long and thus provided more data to estimate phylogenies, and also enabled careful curation (e.g. checking single-gene trees for paralogous and contaminant sequences). Additionally this approach enabled recovery of additional transcripts from samples that our pipeline was not able to capture. To this extent we manually recovered potential sequence candidates from 'missing' transcriptomes using blastx and *Euglena gracilis, Eutreptiella gymnastica* or *Trypanosoma grayi* query sequences for a given gene, discarding sequences shorter than 50 aa. A subsequent search using blastp against NCBI's nr database was used to eliminate contaminants from prey and other unrelated organisms, then the additional candidate sequences were included in the selected 20 gene alignments and single-gene trees were inferred again. Sequences identified by inspection of trees as paralogs or contaminants were excluded.

The final 'base' dataset contained 20 genes and retained 6289 aa sites after masking of 25 phagotrophic euglenids, 4 phototrophic euglenids including *Rapaza viridis*, 1 primary osmotrophic euglenid (*Rhabdomonas costata*), 2 symbiontids, 8 kinetoplastids, and 6 diplonemids. Nine additional discobid taxa acted as an outgroup to Euglenozoa, for a total of 55 taxa. A full list of data used and their sources can be found in Table B.4.

3.2.8 Phylogenomics – Analyses

The 'base' dataset was analysed in IQ-TREE with Maximum Likelihood methods, using the site-heterogeneous mixture model LG + C60 + F + Γ and 1000 Ultra-fast bootstrap replicates (UFB). Using the obtained tree as a guide, we derived a posterior mean-site frequency model (PMSF; Wang et al 2017) in IQ-TREE under the fitted LG + C60 + F + Γ model. This enabled us to run an additional 200 non-parametric bootstraps under PMSF. A Bayesian analysis was performed in PhyloBayes version 4.1 (Lartillot et al. 2009) under the CAT + GTR model, with 4 parallel chains run to convergence (assessed with command bpcomp, and a burnin of the first 500 trees; analysis is not yet finished since chains have not yet converged, posterior probability values from current tree with maxdiff = 0.143115).

In addition to our 'base' analysis we conducted several subsequent analyses. (i) We removed long-branching taxa (*Petalomonas cantuscygni, Percolomonas cosmopolitus,* kinetoplastid SAG D1, and *Sawyeria marylandensis*), as identified by a custom script that calculates tip-to-tip distances between taxa (see above), resulting in the 51-taxa 'noLB' dataset. (ii) A RogueNaRok analysis under the RNR algorhithm and strict consensus settings identified the following as rogue taxa: SAG D1, *Diplonema papillatum, Tsukubamonas globosa, Keelungia* sp. KM082, *Jenningsia fusiforme* ABIC1, and *Dinema litorale* UB26. These taxa were removed from the base dataset, resulting in the 49-taxa 'noRogue' dataset. (iii) To explore any influence on topology and robustness from outgroup rooting, we generated a 'Euglenozoa-only' dataset that had all non-euglenozoan taxa removed. (iv) We excluded the outgroups, kinetoplastids, diplonemids and symbionts to generate a 30-taxa 'Euglenida-only' dataset. Phylogenetic trees for these additional datasets were estimated under the $\mathrm{LG}+\mathrm{C60}+\mathrm{F}+\Gamma$ model with 1000 UFB replicates.

In a final analysis, we reduced the number of sites in our 'base' dataset by incrementally removing fast-evolving sites in 4%-steps (using the assignment of per-site-rates in IQ-TREE with the -wsr flag), until 52% of the data remained (done using a custom script). At each step, a $\mathrm{LG}+\mathrm{C20}+\mathrm{F}+\Gamma$ phylogeny with 1000 UFB replicates was run.

3.3 Results

3.3.1 SSU rDNA Phylogenetics

Overview SSU rDNA Phylogenies

We obtained 73 new SSU rDNA sequences, with most of the sequences (45) from the Spirocuta clade (Fig. 3.1). The majority of organisms (60) were isolated from marine environments, 12 from freshwater, and one from a brackish environment (Table B.1). The SSU rDNA phylogeny now contains numerous clusters of phagotrophic euglenids that sometimes correspond to a nominal genus (or occasionally two or more genera). Frequently, however, cells assigned to a particular genus by morphology do not form a single related cluster. For convenience, we have arbitrarily divided phagotrophic euglenids in the SSU rDNA tree into 32 clusters, labelled A–Y and α–η. These mostly represent SSU rDNA clades with at least moderate support, with a couple of exceptions for convenience. In all, 24 of these labelled clusters include new sequences, and in most cases, representative cells from these clusters are

shown in Figures 3.2–3.4. Measurements, locomotion mode, and assigned groups of individual isolated cells can be found in Table B.5, and videos of representatives from clusters can be found in Videos B.7 in the Appendix.

At the broad scale, euglenids can be divided into three main categories: Spirocuta, petalomonads, and ploeotids. Spirocuta includes Euglenophyceae, Aphagea, 'Anisonemids', and 'Peranemids' (the latter two assemblages are not recovered as monophyletic—see below), plus *Neometanema* and some *Heteronema* (notably *Heteronema vittatum*). These four phagotrophic assemblages are shaded with different colours in Figures 3.1–3-5. Petalomonads (orange) are represented by *Petalomonas, Notosolenus, Sphenomonas* and *Scytomonas*. The 'ploeotids' category includes several taxa whose exact relationships are unresolved; *Olkasia, Entosiphon*, Ploeotiidae (*Ploeotia* and *Serpenomonas*), *Liburna, Hemiolia, Lentomonas, Keelungia*, and *Decastava*. These are shaded with 4 different colours in Figures 3.1–3.5 to reflect phylogenetic relationships inferred in in the multigene analysis (see below). Both Spirocuta (40–100% BS, 0.99 pp) and petalomonads (49–100% BS, 1 pp) are always recovered as clades in our SSU rDNA analyses, whereas 'ploeotids' are always inferred to be paraphyletic (see below).

'Peranemids'

Here we include groups A–B, H–L, α, δ, and ε, represented by cells assigned morphologically to the genera *Peranema, Jenningsia, Urceolus*, and *Chasmostoma*, as well as some *Heteronema*, plus *Teloprocta*. 'Peranemids' are always recovered as paraphyletic in our analyses (Fig. 3.1). All sequences of *Peranema* fall into one maximally supported clade (Clade I in Fig. 3.1), usually at the base of Spirocuta. Cells identified as *Urceolus* fell into up to three clusters: a clade of six *Urceolus* sequences (Clade K) was always fully supported and corresponds to the *Urceolus* clade included in our phylogenomic analyses (see below), while cells *Urceolus* sp. ABLN1 and *Urceolus* c.f. *costatus* WBF1 (groups α and ε) usually branch separately, though with little support for any particular placement. In analyses 'SSU-noShort' (Fig. B.3), 'SSU-noLB' (Fig. B.2), 'SSU-Euglenids' (Fig. B.5), and 'SSU-LB' (Fig. B.1) these two *Urceolus* isolates branched together with negligible support (38–40% BS, 4% in 'SSU-LB', Fig. B.1). When the SSU rDNA sequence from the transcriptome of *Chasmostoma nieuportense* CB1 was included (analysis 'SSU-LB', Fig. B.1; group J), it fell sister to *Urceolus* ABLN1 and WBF1, albeit with no support. Sequences of *Jenningsia* fell into two distinct clades, which is also supported by our phylogenomic analyses (Fig. 3.1, see below). The highly supported *Jenningsia I* clade contains *J.*

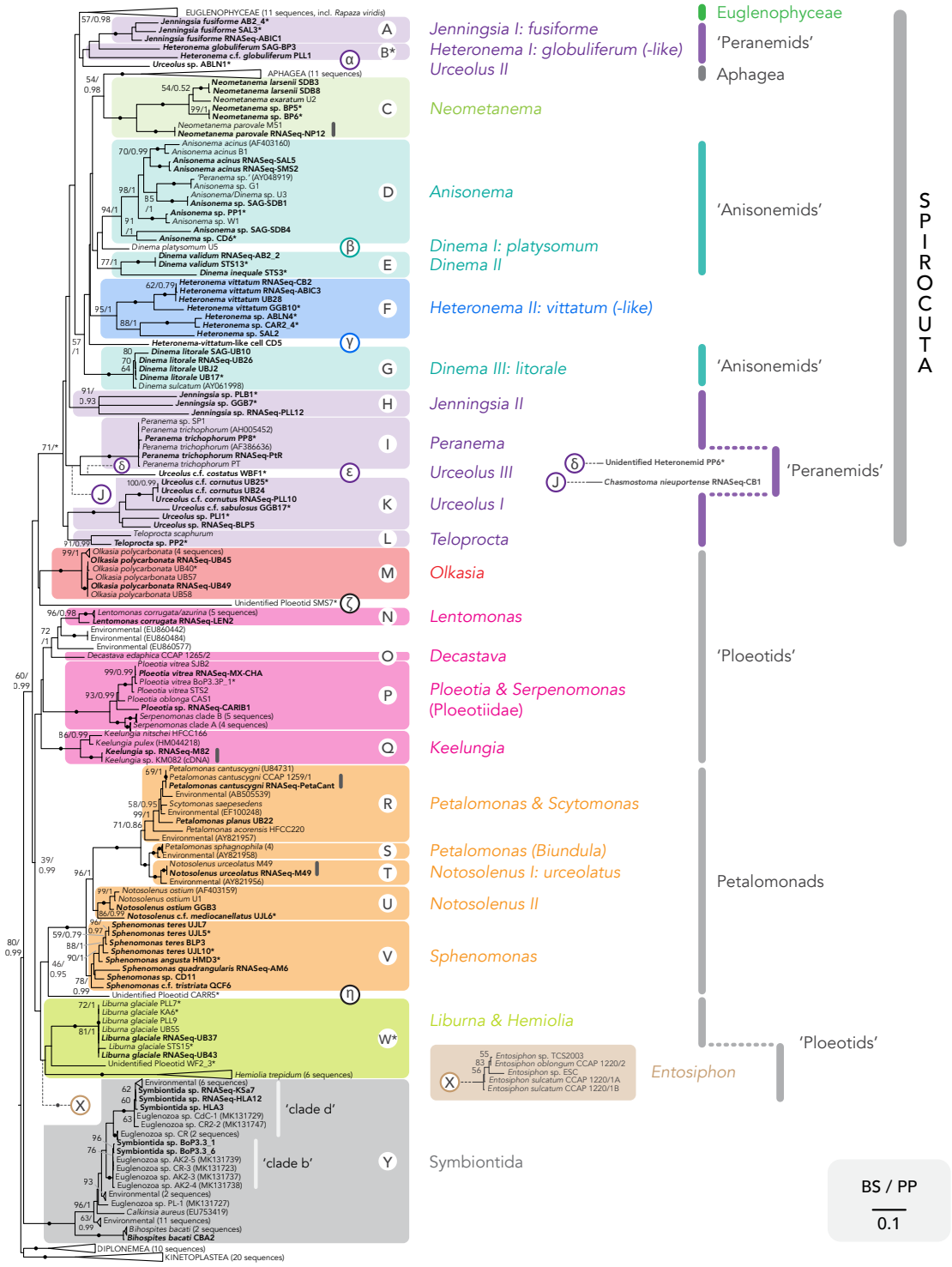

Figure 3.1: Maximum Likelihood phylogeny of the SSU rDNA gene of euglenids, estimated under the GTR+Γ model with 1000 bootstrap replicates. Posterior probabilities were derived from the same model. Novel sequences from this study are bolded and major groups of euglenids are shown. Five *Entosiphon*, CB1 *Chasmostoma*, and PP6 Unidentified Heteronemid were not included in the analysis (total of seven sequences), but their positions determined from a separate analysis ('SSU-LB', Fig. B.1) are marked with dotted branches, an asterisk (*) marks partial sequences, and bold brackets next to two sequences denote they were derived from the same culture. Nodes receiving maximum support for bootstraps (100%) and posterior probabilities (pp of 1) are denoted by filled circles. Support values below 50% and 0.9 pp are not shown.

fusiforme (A: 78–100% BS, 1 pp), whereas the highly supported *Jenningsia II* clade (H: 85–92% BS, 0.93 pp) contained an undescribed morphospecies corresponding to '*Jenningsia macrostoma* Form II' of Lee (Lee 2001; see below). Organisms that until recently would be assigned to the genus *Heteronema* branch in three distinct places. Two sequences from cells identified as *Heteronema globuliferum* and c.f. *globuliferum* were recovered as paraphyletic (*Heteronema* I: group B), whereas a sequence from an unidentified *Teloprocta/Heteronema* species formed a highly supported 'Clade L' with *Teloprocta/Heteronema scaphurum* (74–91% BS, 0.99 pp). When cell PP6 (unidentified heteronemid, group δ) was included it branched with *Urceolus III* WBF1 (group ε; Figures 3.1 and B.1), albeit with no support. By contrast, *Heteronema II*, including *Heteronema vitattum*, always fell within 'Anisonemids' (see below). In almost all of our analyses, clade A (*Jenningsia I*) and paraphyletic group B (*Heteronema* c.f. *globuliferum*) were recovered as an unsupported clade, which in turn sometimes branched as sister to Euglenophyceae, albeit with no support. In other analyses, peranemids excluding *Peranema* (i.e. clades A, B, H, K, L, α, δ, ε) formed a poorly supported clade that branched close to Euglenophyceae. In summary, peranemids exhibited no clear well-supported branching pattern.

'Anisonemids plus' (Anisonemids + *Neometanema* + *Heteronema vittatum*)

This reliably recovered, but weakly supported clade within Spirocuta is composed of primary osmotrophs (Aphagea), plus *Neometanema*, 'Anisonemids' (*Anisonema* and 3 clades of *Dinema*), and *Heteronema II*, including *H. vittatum* (43–65% BS, 1 pp; Fig. 3.1 groups C–G, β, γ). Aphagea are always maximally supported, and *Neometanema* (group C) fall sister to them in most analyses (except Fig. B.2 'SSU-noLB' as no Aphagea were included, where they fell sister to *Anisonema*), with weak-to-moderate support (46–61% BS, 0.98 pp). All 12 sequences of *Anisonema* (6 new) fall within a single, highly supported clade (Clade D; 93–99% BS, 1 pp). *Dinema II* (including *D. validum* and *D. inequale*) is a moderately to highly supported clade (Clade E; 76–82% BS, 1 pp), that often falls sister to *Anisonema*, albeit with negligible support ($<50\%$). The maximally supported *Dinema III* (group G; including *D. litorale* and *D. sulcatum*) has an unresolved position within anisonemids in our phylogenies. Likewise the placement of sequence *Dinema platysomum* (group β) was ambiguous in our analyses, but sometimes branched sister to clade G albeit without support. No definitive position of highly supported clade F, including *Heteronema vittatum* and similar cells (93–96% BS, 1 pp) within 'Anisonemids plus' was recovered in our SSU rDNA phylo-

genies. Similarly, the position of the *Heteronema-vittatum*-like single cell CD5 (clade γ) is unclear from our analyses.

'Ploeotids'

Phylogenetic relationships and morphological characters of several ploeotid taxa was explored in Lax et al 2019 (Chapter 2). Briefly, ploeotids make up much of the backbone of euglenids, and fall into up to 8 distinct groups with exact relationships undetermined with SSU rDNA phylogenies. As Lax et al. (2019), we either recover *Olkasia* (Clade M) and SMS7 Unidentified Ploeotid (Group ζ; 24–63% BS, 0.98 pp) or just SMS7 (30–71% BS, 0.54 pp) as sister to Spirocuta in all analyses. The *Ploeotia/Serpenomonas* Clade P (Ploeotiidae) is maximally supported in all analyses, and includes *Ploeotia* sp. CARIB1, albeit the exact branching order amongst *P. vitrea, P. oblonga* and CARIB1 is unclear. Our single-cell transcriptome of *Lentomonas* c.f. *corrugata* LEN2 falls sister to other *Lentomonas* sequences into highly supported clade N (96–100% BS, 0.98 pp). Relationships between *Lentomonas* (Clade N), *Keelungia* (Clade Q), and *Decastava* (Clade O) remain unresolved in our SSU rDNA phylogenies. The positions of and relationships between *Liburna* and *Hemiolia* (Group W) likewise remain unresolved, as are those of Unidentified Ploeotids CARR5 (Group η) and WF2_3 (within Group W). When *Entosiphon* sequences are included in our analyses, they fall into a fully supported Clade X, but its position among other euglenids remains poorly supported, likely due to its divergent (and thus long-branching) nature (Lax et al. 2019, Paerschke et al. 2017).

Petalomonadida

Petalomonadida (Groups R–V) is recovered as a fully supported clade in all analyses—it includes species assigned to the genera *Petalomonas, Notosolenus, Scytomonas*, and *Sphenomonas* (albeit the former two genera are not monophyletic). Novel sequence *Petalomonas planus* UB22 falls within clade R, containing sequenced *Petalomonas* species (aside from *Petalomonas/Biundula sphagnophila*) and *Scytomonas*, whereas the novel sequence *Notosolenus* c.f. *mediocanellatus* branches in clade U as sister to *Notosolenus ostium*. Our eight single-cell sequences of several *Sphenomonas* morphotypes form a moderately supported clade V (70–79% BS, 0.99 pp), which in turn is always basal to other petalomonads (94–96% BS, 1 pp). *Sphenomonas teres* isolates represent half of the novel sequences and together with *S. angusta* form their own highly

supported subclade (89–92% BS, 1 pp), which is distinct from other *Sphenomonas* morphotypes (*S. quadrangularis, S.* c.f. *tristriata*, *S.* sp. CD11; Fig. 3.4).

Symbiontida

No clear placement for Symbiontida (Group Y) can be recovered among our SSU rDNA analyses, and varies between falling within euglenids and branching between euglenids and Glycomonada (Diplonemea + Kinetoplastea; Fig. 3.1). Three of our new sequences (including both transcriptome-derived SSU rDNA sequences KSa7 and HLA12) fall within environmental 'Symbiontida clade d' (Yubuki & Leander 2018), while two (BoP3.3-1 and BoP3.3-6) branch within environmental 'Symbiontida clade b' (Yubuki & Leander 2018, Orsi et al. 2011, Fig. B.7). Our CBA2 sequence branches with other *Bihospites bacati* isolates (Fig. 3.1).

3.3.2 Morphology

A – *Jenningsia I: fusiforme*

Cells in Clade A are identified as *Jenningsia fusiforme*; they are metabolic, but are elongate-pyriform with rounded posteriors when gliding on their single emergent flagellum. The pellicle is spirally striated, and the feeding apparatus has moderately well-developed rods $\sim$⅕th of cell length.

B – *Heteronema I: globuliferum* (-like)

This cluster includes two cells corresponding to the *Heteronema globuliferum* morphotype: They are both metabolic with strong spiral pellicle striations, and with two emergent flagella. Only the anterior flagellum seems to be used for gliding, while the posterior drags freely under the cell (not adpressed to the cell, like in *Peranema*). Cells default to a pyriform shape, with a rounded/flattened posterior end during gliding. Well-developed oval flagellar pocket, but not feeding apparatus.

C – *Neometanema*

Organisms assigned to *Neometanema* characteristically move via 'skidding': these cells have the distal parts of the posterior flagellum in contact with the surface, while the anterior flagellum beats rhythmically to rapidly move cells forward. Cells have around 20 pellicle strips, are capable of metaboly when distressed, and have a

feeding apparatus visible by light microscopy. Two of our cells were assigned to *N. larsenii*, whereas two more could not be identified to species.

D – *Anisonema*

This phylogenetically diverse clade unites cells with very similar morphology: All have ≥14 pellicle strips, are rigid but are capable of some squirming movement, and glide on their thickened, hook-shaped posterior flagellum, whereas the anterior flagellum flails freely in front of the cell. Cells commonly show 'jerking movement', by contraction of the posterior flagellum. Unlike *Dinema*, their pellicle is not thickened, and the feeding apparatus is difficult or impossible to observe. Many of our isolated cells have ingested algal material.

E – *Dinema II: validum/inequale*

Taxa within Clade E belonged to the morphospecies *Dinema validum* and *D. inequale*, which differ mainly in size (*D. inequale* is the smaller). As with other *Dinema* (see below), they have a thickened pellicle and can be metabolic, but have widely-spaced striations. Cells glide on their thickened, hook-shaped posterior flagellum and often jerk back, whereas the anterior flagellum beats freely. A weakly developed feeding apparatus is present, but sometimes hard to observe.

F – *Heteronema II: vittatum* (-like)

Cells within this clade fell into two phylogenetically separated clusters. These cells are all highly metabolic, but take an oval-pyriform shape when gliding on their posterior flagellum, while the anterior flagellum is held in front of the cell and only the proximal ½ at most is in contact with the surface. Pellicular striations follow a spiral pattern and are usually easy to observe. The feeding apparatus tends to be difficult to observe and is up to ¼ cell length. Cells ABIC3, CB2, GGB10, UB28 have pointed posterior ends when gliding and were identified as *H. vittatum*, whereas ABLN4, CAR2_4, and SAL2 have rounded posterior ends, and appear to represent an undescribed morphospecies. *Heteronema-vittatum*-like cell CD5 (group γ) is morphologically similar to the latter three cells, but branches outside Clade F in our phylogenies.

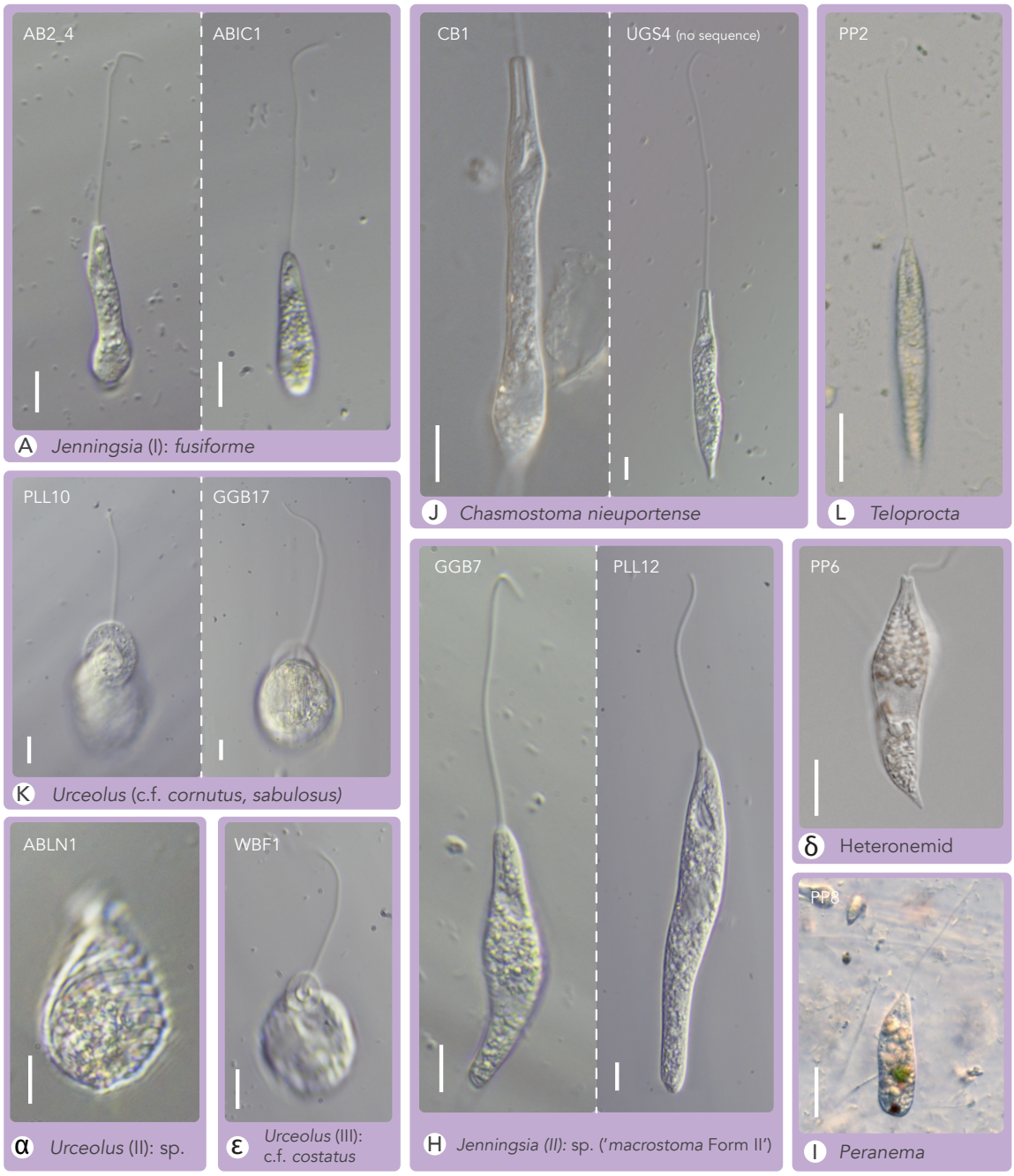

Figure 3.2: Light micrographs of single-cell isolates of peranemid, anterior gliding euglenids. SSU rDNA clades are correlated to this figure by being grouped in the same coloured box and clade name (e.g. 'Clade A'=*Jenningsia I*). A dashed line between micrographs denotes the same morphotype, but different isolate. All scale bars are 10 µm.

G – *Dinema III: litorale*

Cells from this clade were identified as *Dinema litorale* based on their characteristic thickened pellicle with numerous, narrowly spaced strips, combined with some metaboly. *D. litorale* has a thickened, hook-shaped posterior flagellum that is used for gliding, while the anterior flagellum beats freely in front. As with other *Dinema* (see above), the flagellar pocket is well developed, but a feeding apparatus is either absent or difficult to observe. They are elongate when gliding. This group also includes a sequence identified in a previous study as *D. sulcatum* (AY061998), but no micrographs exist of the cell from which it was derived (Busse et al. 2003).

H – *Jenningsia II*

These three single-cell isolates (GGB7, PLB1, PLL12) assigned to *Jenningsia*, share a similar morphology. They are highly metabolic and elongate when gliding on their single emergent anterior flagellum that beats only with its distal ⅓. Their posterior ends are slightly oblique. The nucleus is central, the feeding apparatus is inconspicuous and not well-developed. Pellicular striations are hard to observe but appear to be numerous and spirally arranged. Cells are full of granules. This morphology is in line with that of *Jenningsia macrostoma* Form II *sensu* Lee 2001, which represents an undescribed morphospecies (see discussion).

I – *Peranema*

Taxa within Clade I were assigned to *Peranema*: Cells glide on their anterior flagellum, while the posterior flagellum is adpressed to the ventral surface of the cell and is shorter than the cell. They are highly metabolic and often contain ingested algal material. The majority of cells were assigned to *P. trichophorum*.

J – *Chasmostoma nieuportense*

A single cell in Clade J was assigned to *Chasmostoma nieuportense*: A highly metabolic cell, elongate when gliding but contracts into a twisted ‘spindle-shape’ when distressed. The cell has a characteristic anterior flagellar cavity that is $\sim 1/7^{\text{th}}$ cell length, which houses a single, long emergent flagellum, which is used for gliding. The small club-shaped feeding apparatus is hard to observe, and pellicular striations are very fine, closely spaced, spirally arranged, and numerous.

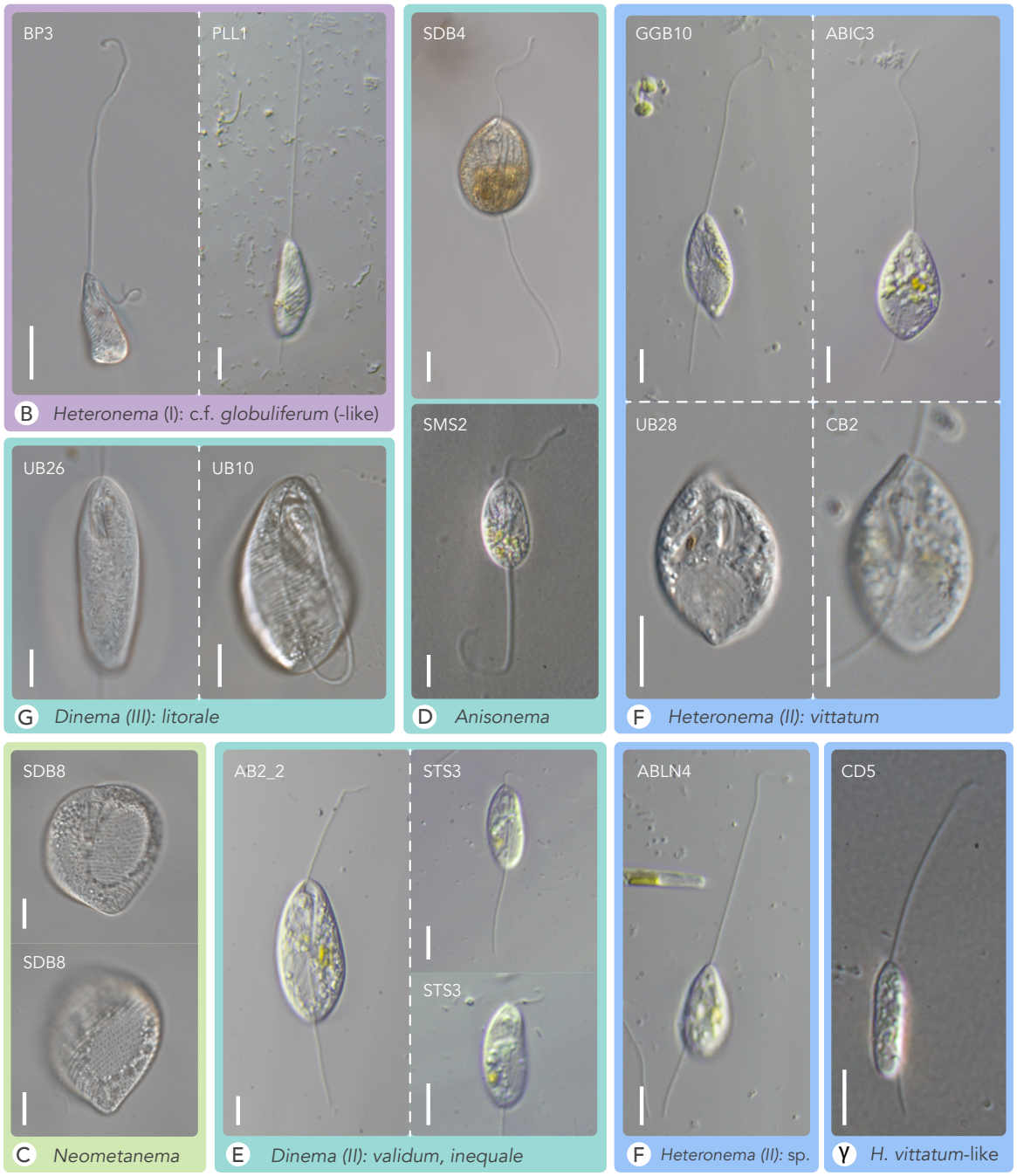

Figure 3.3: Light micrographs of single-cell isolates of anisonemid (and related taxa) and peranemid euglenids. SSU rDNA clades are correlated to this figure by being grouped in the same coloured box and clade name (e.g. 'Clade C'=*Neometanema*). A dashed line between micrographs denotes the same morphotype, but different isolate. All scale bars are 10 µm.

K, alpha & epsilon – *Urceolus*

Cells within these three groups were assigned to four different morphotypes of *Urceolus*, based on their flexibility, a single emergent flagellum that is used for gliding, and a characteristic anterior collar that is in contact with the surface during gliding. The main differences between the morphotypes are the shape and width of the collar, the pellicular striations (wide- vs. narrowly-spaced strips), and the presence of a well-developed feeding apparatus. Some cells have detritus attached to their pellicle (e.g. UB25), whereas cells that are otherwise completely identical do not (e.g. UB24).

L – *Teloprocta*

Our cell PP2 was assignable to *Teloprocta*, as it shows metabolic movement but is elongate fusiform during gliding on its anterior flagellum. Cell PP6 has a very similar morphology, but branched outside Clade L in our SSU rDNA phylogenies (Fig. B.1). The spindle-shape and the pointed posterior in our PP2 and PP6 are more pronounced than in *T. scaphurum*, which has more of a 'sack-shape' (Breglia et al. 2013, Schroeckh et al. 2003).

N – *Lentomonas*

The clade contains ventrally-flattened, rigid cells with 7 corrugate dorsal/lateral pellicle strips (plus 3 ventral strips whose margins are difficult to discern) and that are smaller than *Ploeotia* or *Serpenomonas*. They move on their posterior flagellum, while the anterior beats in front of the cell. An oblique feeding apparatus can be seen with well-developed rods that can extend most of the length of the cell.

P – *Ploeotia & Serpenomonas* (Ploeotiidae)

Organisms within this clade are rigid with 10 pellicle strips arranged in a variety of different ways, in the case of *Ploeotia* the strip margins are located on raised keels (Lax et al. 2019, see Chapter 2). They glide on their posterior flagellum, whereas the anterior flagellum is flailing in front of the cell. Our novel isolate, CARIB1, is assigned to *Ploeotia* based on the strip organisation, but is larger and has a longer posterior flagellum (at 3x cell length) than described species.

Figure 3.4: Light micrographs of single-cell isolates of petalomonads and symbiontids. SSU rDNA clades are correlated to this figure by being grouped in the same coloured box and clade name (e.g. 'Clade V'=*Sphenomonas*). A dashed line between micrographs denotes the same morphotype, but different isolate. The arrowhead in (U) *Notosolenus II* points at the posterior flagellum. All scale bars are 10 µm.

R – *Petalomonas I*

Organisms within this clade are assigned to *Petalomonas* and *Scytomonas*. They are rigid cells with few pellicle strips, and have a single emergent flagellum that they use for gliding. Our cell UB22 was assigned to *Petalomonas planus* (Lee & Patterson 2000) based on its broad fusiform shape with a drawn-out posterior end and the numerous hyaline inclusions throughout the cell.

U – *Notosolenus II*

Like other petalomonads, *Notosolenus* are rigid cells with few pellicle strips, but have two emergent flagella rather than one. They glide on their anterior flagellum (see Video GGB3 in B.7), and a short (often hard to observe) posterior flagellum is trailing under the cell. Cell UJL6 was identified as *Notosolenus* c.f. *mediocanellatus* based on the longitudinal median grooves on both the ventral and dorsal side (Schroeckh et al. 2003, Fig. 3.4 – U).

V – *Sphenomonas*

All cells within this clade were assigned to *Sphenomonas*, into five different morphotypes. All of them are biflagellate, rigid, roughly fusiform cells that are not strongly flattened, and have a large hyaline inclusion towards the posterior that can occupy up to ⅔ of a cell. These features distinguish *Sphenomonas* from other nominal petalomonad genera, e.g. *Notosolenus* (Schroeckh et al. 2003). *Sphenomonas* has only 4 pellicle strips (the lowest number in euglenids), but they are often not distinguishable (e.g. in *S. angusta*, Schroeckh et al. 2003). They glide on their anterior flagellum (1–2x cell length), whereas the posterior flagellum is short and trailing. The identified morphotypes have different cross-sectional profiles, ranging from round (*S. angusta*) to cruciform (*S. quadrangularis*). Cells BLP3, UJL5, UJL7, and UJL10 were identified as *S. teres* based on their elongate-elliptical, almost fusiform shape. HMD3 was identified as *S. angusta* because of its ovate shape and the single ventral median groove running along the whole cell. This taxon has been found in both marine and freshwater environments (Schroeckh et al. 2003, Lee 2008, Lee & Patterson 2000). *S. quadrangularis* Stein (1878) is the type species of the genus, and was identified by its characteristic cruciform shape (Fig. 3.4 – V). QCF6 was identified as *S.* c.f. *tristriata* based on three longitudinal grooves running along the cell (Schroeckh et al. 2003); this morphotype has only been found in freshwater, whereas our cell was isolated from a marine environment (hence c.f. *tristriata*). Cell CD11 could not be assigned

to any morphotype, but is ovate with a minimally flared anterior and has at least one longitudinal groove (Fig. 3.4 – V).

Y – Symbiontida

Our novel sequences of symbiontids fall into three separate clades, although the first two are morphologically similar: rigid, swimming cells with two flagella that are covered in rows of episymbiotic bacteria, with a median furrow. No pellicular striations can be seen. This morphotype corresponds roughly to *Postgaardi* (Simpson et al. 1997, Yubuki et al. 2013, Fenchel et al. 1995). *Bihospites bacati* on the other hand is also covered by episymbiotic bacteria, but has obvious pellicle-like striations and is capable of metaboly, which we observed in our cell CBA2. While *B. bacati* was reported to drag itself along surfaces with its anterior flagellum (Breglia et al. 2010), we could not observe any locomotion as the cell was too distraught.

3.3.3 Phylogenomics

Our 20-gene, 55-taxa multigene analysis included data from 29 transcriptomes of euglenids and symbiontids, 24 of which were derived from single cells. Most of the labelled groups in our SSU rDNA phylogenies and outlined above had at least one representative taxon in the multigene analyses (see correlated colour codes and clade names in Figures 3.1–3.5).

Structure of Discoba/Euglenozoa

Euglenozoa is highly-to-maximally supported in all of our phylogenetic analyses (lowest 96% UFB), and is composed of Euglenida (93–99% UFB, 1 pp) and Glycomonada (53–91% UFB, 0.95 pp) that is, diplonemids are sisters to kinetoplastids, rather than either being more closely related to Euglenida. Within this framework, Symbiontida almost always branch sister to Glycomonada with moderate to high support (84–99% UFB, 1 pp), but in some of the trees of the 'FSR-removal' dataset series, they instead fall within Glycomonada, sister to Kinetoplastea, with very weak support (46–69% UFB, Fig. B.12).

Structure of Euglenida

As in our SSU rDNA phylogeny, Spirocuta forms a major clade within euglenids, and is always fully supported. Ploeotids form either 3 or 4 clades attached sequentially

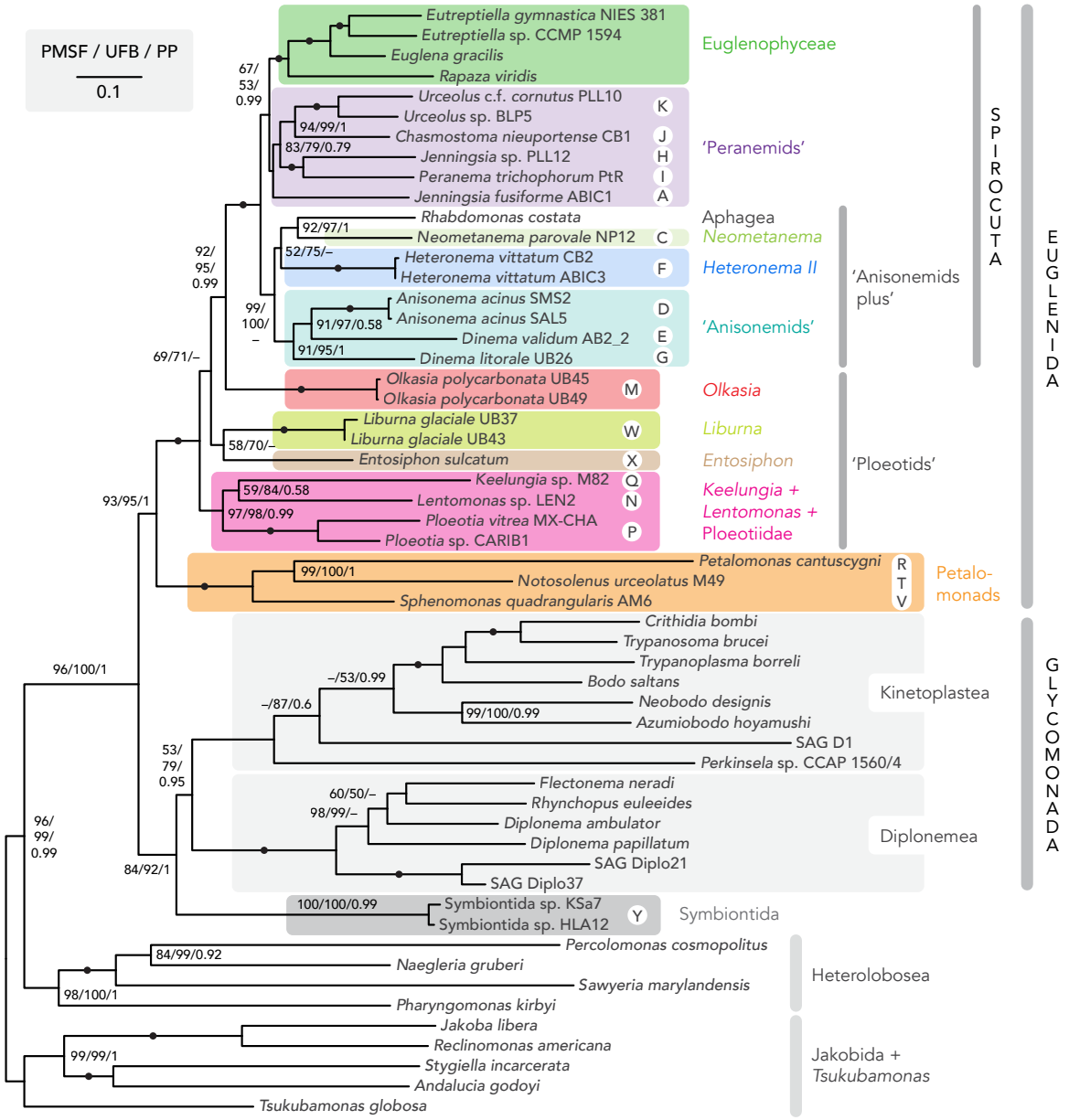

Figure 3.5: Phylogeny of Discoba inferred from 20 genes and estimated in the Maximum Likelihood framework under the LG+C60+F+Γ model, with robustness assessed with 200 'true' bootstrap replicates (PMSF), 1000 Ultrafast Bootstrap replicates (UFB), and Bayesian Propabilities (pp, under CAT+GTR model). Major clades of euglenids are shown in coloured boxes. Nodes receiving maximum support for both bootstrapping methods (100%) and posterior probabilities (pp of 1) are denoted by filled circles. Support values below 50% and 0.9 pp are omitted.

to Spirocuta. *Olkasia* falls sister to Spirocuta in all of our analyses with high support (92–99% UFB, 0.99 pp). *Liburna* and *Entosiphon* sometimes form a weakly supported clade (43–70% UFB), but in other analyses *Liburna* and *Entosiphon* branch apart in different parts along the base of the tree. When branching together, this clade falls sister to the Spirocuta + *Olkasia* clade (with the exception of the 'Euglenida-only' analysis, Fig. B.11), with low support (45–71% UFB). *Ploeotia, Keelungia*, and *Lentomonas* always form a highly supported clade (94–100% UFB, 0.99 pp), that branches sister to the above grouping in all analyses, with low support (see above). At the base of euglenids, taxa *Petalomonas, Notosolenus*, and *Sphenomonas* form maximally supported Petalomonadida clade (Fig. 3.4). Support for the basal position of petalomonads is full in all of our analyses, even when long-branching taxa were removed (limiting petalomonads to *Notosolenus urceolatus* and *Sphenomonas quadrangularis*; Fig. B.8), and rogue taxa are omitted (Fig. B.8). The basal position of petalomonads was also recovered in all trees of our FSR-analysis (lowest 96% UFB with 52% sites retained, Fig. B.12).

Structure within Spirocuta

Spirocuta falls into two main groups: a poorly supported Euglenophyceae + 'Peranemids' grouping (52–67% UFB, 0.99 pp), and highly supported 'Anisonemids plus' grouping (99–100% UFB, but 46% UFB in Fig. B.11 'Euglenida-only'). Euglenophyceae (including *Rapaza viridis*) is maximally supported. 'Peranemids' is recovered as a clade but receives poor support (39–60% UFB, 0.58 pp) due to the poorly-resolved position of *Jenningsia fusiforme* (*Jenningsia I*/group A), which is the deepest branch among 'Peranemids'. The remaining peranemids (*Urceolus, Chasmostoma, Peranema*, and *Jenningsia* sp. PLL12, i.e. *Jenningsia II*/group H) form a clade with low to moderate support (78–85% UFB, 0.79 pp). Within this, *Chasmostoma* formed a highly supported clade with *Urceolus I* (89–99% UFB, 1 pp; other *Urceolus* groups not sampled), while *Peranema* and *Jenningsia* sp. PLL12 (*Jenningsia II*) always formed a maximally supported clade (99–100% UFB, 1 pp).

'Anisonemids plus' is composed of three main groups: a highly supported *Anisonema* + *Dinema* clade (Anisonemids *sensu stricto*; 91–100% UFB, 1 pp), a moderately-to-highly supported Aphagea + *Neometanema* clade (82–99% UFB, 1 pp), and *Heteronema vittatum* (i.e. *Heteronema II*/group F). Placement of the latter is almost always sister to Aphagea + *Neometanema*, but with low support (52–75% UFB). *Dinema* is recovered as paraphyletic, with *Dinema II* (Clade E) branching closer to

Anisonema/Clade D (89–100% UFB, 0.58 pp) than *Dinema III* (Clade G, 91–98%, 1 pp). *Dinema I* (group β) is not represented in our phylogenomic analyses.

3.4 Discussion

3.4.1 New Insights into Euglenid Diversity at and around 'Genus Level'

Have We Sampled All Euglenid Genera?

Together with recent publications, this study considerably increased taxon sampling for the SSU rDNA across phagotrophic euglenids (Lax et al. 2019, Cavalier-Smith et al. 2016). The great majority of morphologically defined genera are now represented by at least one sequence: out of 29 nominal phagotrophic genera (Schoenle et al. 2019, Lax et al. 2019), 24 are sampled. Based on their morphology—mainly pellicle strip architecture and locomotion patterns—most of the missing genera (*Dolium, Atraktomonas, Tropidoscyphus, Calycimonas*, and *Dylakosoma*) are inferred to be petalomonads (see Cavalier-Smith 2016). These missing genera are mostly freshwater organisms (*Dolium* is an exception), while the sampling for phagotrophic euglenids is currently skewed heavily towards marine organisms: Out of the currently 142 available sequences (including this study), 101 are from organisms that were found in marine environments, 34 were from freshwater/terrestrial environments, 4 from brackish environments, and 3 from an unknown source. This bias may partially explain why it is these particular genera have not been sampled. However, while the bulk of described phagotrophic euglenid genera are now represented by at least a single SSU rDNA sequence, this does not necessarily mean that the full diversity of phagotrophs has been captured at the 'genus-level', as elaborated below.

Morphological Genera do not Necessarily Equal Molecular Clades

Early phylogenetic analyses of euglenids often only had a single representative of a given, morphologically-defined genus (e.g. *Dinema, Petalomonas*, etc.; Preisfeld et al. 2000, Busse et al. 2003, Montegut-Felkner & Triemer 1997). The adoption of single-cell methods and increased cultivation efforts has resulted in more species being examined within already sampled genera, in addition to previously unsampled genera being sequenced (e.g. Lax et al. 2019, Schoenle et al. 2019, Cavalier-Smith et al. 2016, Chan et al. 2013). This increase in taxon sampling has shown that many nominal genera are not recovered as clades, but rather as paraphyletic or polyphyletic assemblages (e.g. *Notosolenus* and *Petalomonas*). These findings ultimately lead to taxonomic splitting and establishment of new genera, as was done with *Neometanema* (including several species previously in *Heteronema*, Lee & Simp-

son 2014b), and *Liburna* and *Hemiolia* (previously *Anisonema* proto parte, Lax et al. 2019). With the availability of 45 new SSU rDNA sequences of 9 genera in Spirocuta, this study identifies several further cases of non-monophyly at the traditional genus level, with two prominent cases in *Dinema* and *Heteronema*.

Dinema

Taxa were assigned to the genus *Dinema* on the basis of a thickened pellicle, slightly squirming movement, and by gliding on their posterior flagellum (Larsen & Patterson 1990). They are morphologically similar to closely related *Anisonema*, but differ in that they have a flexible and thickened pellicle, and that the feeding apparatus is more visible by light microscopy (Larsen & Patterson 1990, Lee & Patterson 2000). While all of our sampled *Dinema* are morphologically similar in pellicle thickness, flexibility and gliding pattern, our sequences and previously available sequences fall into three SSU rDNA clades that are not specifically related to each other (E, G and β; Figures 3.1 and B.1). Our multigene-analyses also recover *Dinema* as a paraphyletic group with respect to *Anisonema*, although only two out of three clades are sampled (*Dinema II*/Clade E and *Dinema III*/Clade G). As such, we currently consider *Dinema* to be paraphyletic: This is somewhat in line with the morphological differences among the *Dinema* subclades. *Dinema II* (*D. validum* and *D. inequale*) cells have wider pellicle strips than those of *Dinema III* (*D. litorale* and *D. sulcatum*). Crucially, though, no molecular data are available for the type species of *Dinema, Dinema griseolum* Perty. By some accounts, *D. litorale* might be a senior synonym of *D. griseolum* (Lee & Patterson 2000), as they are almost identical in morphology, and supposedly only distinguishable by the muciferous bodies lining the pellicle strips and the slightly larger size of *D. griseolum* (Lee & Patterson 2000, though *D. griseolum* has only been found in freshwater). This means that *Dinema III* (Clade G) potentially houses the type of *Dinema*, yet any taxonomic decision and establishment of new genera (for *Dinema I* and *II*) would require additional observational studies of *D. litorale* and *D. griseolum*, and phylogenies that include additional SSU rDNA sequences. While *Dinema II* and *III* are well-sampled, *Dinema I* currently only has a single sequence of *Dinema platysomum* available, and has an uncertain placement among anisonemids. We currently cannot recommend any taxonomic changes, as the possibility that the current paraphyly of *Dinema* is due to phylogenetic instability cannot be excluded.

Heteronema

The genus *Heteronema* used to encompass taxa that can squirm, show an ingestion apparatus in light microscopy, and have their posterior flagellum free from the cell body (not adpressed to the body, as in *Peranema*, Larsen & Patterson 1990). This definition did not delineate between modes of locomotion, but included taxa that glide on their anterior flagellum (e.g. *Heteronema/Teloprocta scaphurum, H. globuliferum*), taxa that glide on both their posterior and anterior flagella (e.g. *H. vittatum*), as well as taxa that skid (e.g. *Heteronema/Neometanema exaratum*). The latter category was transferred to the new genus *Neometanema*, based on their morphology and phylogenetic distance to *Heteronema/Teloprocta scaphurum* (Lee & Simpson 2014b). Shortly after, *Teloprocta* was established as a new genus by Cavalier-Smith et al. (2016) on the basis that it glides on its thickened anterior flagellum, which is unlike the description of the type *H. marina* Dujardin, which supposedly glides on its posterior flagellum and is considered an anisonemid. Our phylogenies including novel SSU rDNA sequences of several *Heteronema vittatum* (-like) cells do not place them close to the other *Heteronema* species or *Teloprocta*, but instead with 'Anisonemids', *Neometanema* and Aphagea. These organisms glide on their posterior flagellum with the anterior flagellum flailing in front of the cell (Leander et al. 2017), similar to *H. vittatum*, which seems to also have its anterior flagellum partially in contact with the surface (Video ABLN4 in B.7).

Heteronema vittatum arguably represents a novel genus, based on the phylogenetic distance to other taxa traditionally considered *'Heteronema'*, but this is complicated by conflicting taxonomic approaches to this nominal genus. Dujardin's (1841) original description of the type species, *Heteronema marina* indicates a flexible cell that glides on its thickened posterior flagellum, while the anterior flagellum is thinner, suggesting an anisonemid. The description is otherwise vague and was deemed not sufficient for identification (Larsen & Patterson 1990). Instead, the understanding of the genus *Heteronema* since the late 19th century mainly comes from Stein (1878), when into this genus he placed several flexible, anterior gliders, which are unlike the probable anisonemid posterior glider Dujardin described. That Dujardin's type, *Heteronema marina*, was anisonemid-like was largely ignored (for better or for worse) by euglenid taxonomists until Cavalier-Smith et al. (2016) transferred *H. scaphurum* to newly established *Teloprocta* to reflect the differences in locomotion. Yet this transfer did not solve the mystery surrounding the nature of *H. marina* that Dujardin described: None of the currently sequenced *Heteronema*-like organisms looks similar to the type described in 1841. Our *H. vittatum* are flexible, and appear to glide on

both anterior and posterior flagella (the anterior is partly attached to the substrate and ~2x cell length). Yet Dujardin's description of *H. marina* mentions a thickened posterior flagellum and a freely-beating anterior flagellum that is shorter than a cell length. While Dujardin's description is likely more of a large anisonemid, rather than a *Heteronema vittatum*-like organism, we consider the description to be insufficient to exclude the possibility that *H. vittatum* is a representative of a clade that includes *H. marina*.

The taxonomic identity of *Heteronema globuliferum* is similarly complicated: Our SSU rDNA sequences do not form a clade, but both of them always branch separately from other *Heteronema and Teloprocta* sequences (Figures 3.1 and B.1). Both *H. globuliferum* and *Teloprocta scaphurum* glide on their anterior flagellum and are flexible, yet the anterior flagellum in *H. globuliferum* is considerably longer in relative terms (2.7x vs ~1x). Additionally *H. globuliferum* is consistently smaller and conforms to a 'sack-shape' (Fig. 3.3 B, 17–29 μm), rather than the long spindle shape seen in *Teloprocta* (Fig. 3.2 L, ~60 μm). It is possible that *H. globuliferum* represents its another yet undescribed genus, but the non-monophyly and still-poor sampling of this morphotype render it unwise to establish one on current information.

Jenningsia and *Peranema*

All currently available SSU rDNA sequences of *Peranema* are almost identical and were all identified as *P. trichophorum* (Fig. 3.1), one of the most well-known phagotrophic euglenids. Despite this fame, the exact identity of the genus *Peranema* is surrounded by some confusion (Larsen & Patterson 1990, 1991). Several taxa originally described as *Peranema* (and *Peranemopsis*) have been transferred to *Jenningsia* based on their single emergent flagellum (Lee et al. 1999), but the second flagellum of *Peranema* is hard to observe as it lies recurrently in a ventral groove (rather than free from the cell body as in *Heteronema* sensu lato, Larsen & Patterson 1990). While these definitions of genera could not be tested extensively with molecular tools in the past, we now have an extended set of molecular sequences available for *Jenningsia* (and *Heteronema*, see above).

In all of our phylogenetic analyses *Jenningsia* is not recovered as monophyletic (Figures 3.1 and 3.5), but rather is divided into Clades A and H. This situation is complicated by an existing peculiarity of the species taxonomy of *Jenningsia*. Clade A (*Jenningsia I*) includes the morphospecies *Jenningsia fusiforme*, while Clade H (*Jenningsia II*) corresponds to an undescribed morphospecies, likely '*Jenningsia macrostoma* Form II' of Lee (2001). Lee identified two 'Forms' in his report of *J. macrostoma*:

Form I had strongly developed feeding apparatus rods, while Form II did not (Lee 2001). The two forms also differed in their reported lengths (Form I: up to 80 μm, Form II: 100–114 μm), and Form II was reported to be full of refractile granules, whereas Form I only had a few distributed throughout the cell. Modern approaches to morphology-based taxonomy (e.g. *Serpenomonas costaversata* in Schoenle et al. (2019)) would likely consider them two separate nomenclatural species, rather than two 'Forms' of the same species. We consider Form I to best fit the original description of *J. macrostoma*, based on their size, conspicuous feeding apparatus rods, and sack-like cell shape (Ekebom et al. 1995), and Form II to be an undescribed morphospecies. Consequently, the assignment of cells PLL12, GGB7, and PLB1 to *Jenningsia* is indirect, since they likely represent a new morphospecies in need of deeper investigation and formal description. As such, *Jenningsia* may not be truly polyphyletic, as only one undoubtedly identified *Jenningsia* morphospecies (*J. fusiforme*) has been sequenced, and it is possible that all the several other described species assigned to this genus, including the type *J. diatomophaga* Schaeffer, could be related to *J. fusiforme* (Clade A), rather than to PLL12, GGB7, and PLB1 (Clade H).

Is *Urceolus* Paraphyletic?

Urceolus is a poorly-known genus that has been neglected in the past, with often inadequate species descriptions and poorly delineated species boundaries (Leedale 1967). All 11–12 *Urceolus* morphospecies are capable of euglenoid metaboly, have a single emergent flagellum, and a flattened anterior collar that is in contact with the surface while gliding (Larsen 1987, Larsen & Patterson 1990, Lee 2006, Larsen & Patterson 1991). This collar readily distinguishes *Urceolus* from other phagotrophic euglenid taxa.

Our SSU rDNA phylogenies place six *Urceolus* cells with similar morphology into a well-supported Clade K, whereas two additional sequences ABLN1 (α) and WBF1 (ε) usually branch independently within Spirocuta, but with no clear placement (Figures 3.1 and B.1). The cells in Clade K are all roughly the same size, most have fine, narrow pellicular striations (except GGB17, which has alternating broad and narrow strips), a relatively wide and slightly oval-shaped collar, and a conspicuous feeding apparatus (Fig. 3.2 K). Both single-cell transcriptomes of *Urceolus* in our phylogenomic analyses belong to Clade K (PLL10 and BLP5), and form a well-supported grouping (Fig. 3.5).

In some of our analyses, ABLN1 and WBF1 form a poorly supported clade, which is tentatively supported by their similar morphology. Both organisms are roughly the

same size (~30 x 23 μm), and have a smaller collar, and broader pellicular striations than Clade K cells (Fig. 3.2, α and ε). These cells also hold their shape relatively constant during gliding, whereas the cells in Clade K are much more actively metabolic (Fig. 3.2 K; Videos WBF1 and PLL10 in B.7).

The presence of pellicle 'decoration' was seen as an identifying character for species identification within *Urceolus* by several early authors (e.g. Huber-Pestalozzi 1955, Lemmermann 1913), whereas more recent publications have cast doubt on this distinction (Larsen & Patterson 1990). Within Clade K, both UB24 and UB25 (isolated from the same source material within 24h of each other) share the same morphology, except that UB24 has a smooth pellicle, whereas UB25 has a pellicle decorated with small diatom frustules and other debris. The SSU rDNA sequences are 99.8% identical, suggesting that the presence or absence of debris is not a valid character for species identification. Pellicle striation patterns (fine vs. wide striations), the size of the collar, and the general shape during gliding all seem to be more relevant characters for species identification in *Urceolus* (see Larsen 1987, Larsen & Patterson 1990). For example, UB24, UB25, and PLL10 have very fine pellicular striations and are highly metabolic, and are thus identified here as *Urceolus cornutus* (Larsen & Patterson 1990). Similarly, cell GGB17 was identified as *U. sabulosus* based on the alternation between more prominent ridges and finer striations on its pellicle, which is characteristic for this morphospecies (Larsen 1987).

In our multigene analyses, *Urceolus* forms a well-supported clade with *Chasmostoma* (Fig. 3.5), and in our SSU rDNA analyses a relationship with *Urceolus* c.f. *costatus* WBF1 is occasionally recovered, albeit with low-to-no support (Fig. 3.1). *Chasmostoma nieuportense* (the sole described species) is morphologically similar to *Jenningsia, Peranema,* and some *Heteronema* species in being elongate, metabolic, and showing pellicle striations, but it distinctly differs in having a conspicuous anterior flagellar cavity into which the flagellum can be withdrawn (Lee et al. 1999). Both *Urceolus* and *Chasmostoma* have a single emergent flagellum, are metabolic, and show a similar flagellar movement where at least half of the flagellum is in contact with the surface (see Videos *Urceolus* WBF1 or PLL10 and *Chasmostoma* ABF6 or UGS4). The collar of *Urceolus* and the flagellar cavity of *Chasmostoma* are similar in being pellicle-supported anterior extensions of the cell beyond the emergence of the flagellar canal. It is plausible that the cavity of *Chasmostoma* and collar of *Urceolus* are homologous and a synapomorphy for a *Chasmostoma-Urceolus* clade. It could be envisaged that a flagellar cavity of a *Chasmostoma*-like ancestor flaring outwards formed the collar of *Urceolus*, or vice versa (see Fig. 3.2 J and K).

Petalomonads are Highly Speciose

Petalomonads are one of the most speciose groups of phagotrophic euglenids (Leander et al. 2017, Lee & Simpson 2014a). The majority of described species are assigned to the genera *Petalomonas* and *Notosolenus*, which is reflected in the prior taxon sampling of the SSU rDNA, with almost all morphospecies sampled being identified originally as belonging to one of those genera (*Scytomonas saepesedens* is an exception, Cavalier-Smith et al. 2016). Among petalomonads, the anterior flagellum is in contact with the surface and used for gliding, while the posterior flagellum, where emergent, is short and trails posteriorly (Lee & Simpson 2014a, Leander et al. 2017). The posterior flagellum is present in *Notosolenus* but absent in *Petalomonas*, and in fact this is the main character distinguishing these genera (Larsen & Patterson 1990). It is known that this morphology-based delineation does not correspond to molecular clades (see above). Both *Petalomonas* and *Notosolenus* are currently paraphyletic or polyphyletic, as *N. urceolatus* (2 flagella) branches with *Petalomonas* (and *Scytomonas*), and not with other sequences from *Notosolenus* species (Fig. 3.1; Lee & Simpson 2014b, Cavalier-Smith et al. 2016). Unfortunately, neither type species of *Petalomonas* (*P. abcissa* Stein) and *Notosolenus* (*N. apocamptus* Stokes), nor any particularly similar morphospecies, have been examined using molecular methods. As such, it is difficult to decide on appropriate circumscriptions for these taxa, and to evaluate the appropriateness of the proposal for *Biundula* as a new genus for *Petalomonas sphagnophila* (and three other supposedly related morphospecies without molecular data, Cavalier-Smith 2016, Cavalier-Smith et al. 2016). Our novel sequence of *Petalomonas planus* branches within Clade R (*Petalomonas I & Scytomonas*), while our sequences from *Notosolenus* c.f. *mediocanellatus* and *N. ostium* sit in Clade U (*Notosolenus II*) with previous *N. ostium* sequences. Perhaps surprisingly, these placements do not worsen the phylogenetic entanglement of *Notosolenus* and *Petalomonas*. Yet *N. mediocanellatus* was initially described as *Petalomonas mediocanellata*, and was only later found to have a second, very short emergent flagellum (Schroeckh et al. 2003), highlighting the surrounding ambiguity in confirming the absence of a morphological character (rather than presence) and its relevance for taxonomical decisions.

Most of our novel petalomonad sequences (eight) were identified as belonging to *Sphenomonas*, approximating four of the 14 described nominal species of *Sphenomonas* (Schroeckh et al. 2003), plus one unidentified morphotype. This genus had previously escaped sampling for molecular sequencing, despite their relative abundance in freshwater (Schroeckh et al. 2003), in part because the bulk of recent

studies have examined marine material (e.g. Lee & Simpson 2014b, Schoenle et al. 2019). In our SSU rDNA phylogenies the eight *Sphenomonas* sequences form a single Clade V that is always moderately supported (70–79%). Both SSU rDNA and multigene analyses confidently place *Sphenomonas* as the deepest branch among petalomonads (Figures 3.1 and 3.5). We are able to observe some substructure among *Sphenomonas*: A clade of four *S. teres* sequences form a well-supported clade, with *S. angusta* HMD3 branching sister to it. *S.* c.f. *tristriata* QCF6 branches at the base of *Sphenomonas*, with *S. quadrangularis* and cell CD11 falling in between the *S. teres* clade and the base (Fig. 3.1).

3.4.2 Large-Scale Euglenid Evolution

Overview

The vast majority of euglenid cultures are of phototrophs (Leander et al. 2017), and the few phagotrophic cultures available over time (the majority of them now lost) represented only part of the diversity of phagotrophs (e.g. *Petalomonas, Peranema, Anisonema, Dinema, Entosiphon, Serpenomonas, Ploeotia*; and more recently: *Notosolenus, Neometanema, Keelungia, Decastava*). This scarcity of cultures made investigation of euglenid phylogeny difficult, by skewing SSU rDNA phylogenies heavily towards phototrophic euglenids (e.g. Müllner et al. 2001), and no multigene phylogeny has ever included any phagotrophs (e.g. Karnkowska et al. 2015). By bypassing the strict need for cultures to generate molecular phylogenomic-grade data, our culture-independent single-cell approach was able to generate 24 transcriptomes of phagotrophs, to which we could add 5 phagotroph transcriptomes derived from mass culture.

As in recent SSU rDNA analyses, we recover Spirocuta in our multigene phylogenies, which includes phototrophs, primary osmotrophs and a number of phagotrophs (Fig. 3.5). Unsurprisingly, ploeotids are recovered as paraphyletic: *Olkasia* is sister to Spirocuta (see below), followed by *Entosiphon* and *Liburna* which are often branching together (albeit with little support), followed by a highly supported clade of the remaining sampled ploeotids (*Keelungia, Lentomonas*, Ploeotiidae). Indeed, previous SSU rDNA phylogenies, as well as our own, found that ploeotids consisted not of a single clade, but rather of 5–8 subgroups with poorly supported affinities (Lax et al. 2019, Fig. 3.1). Thus the grouping of *Keelungia, Lentomonas* and Ploeotiidae into a single, well-supported group represents an improvement in phylogenetic structure. Interestingly they all share a similar pellicle structure in that they have ten pelli-

cle strips with bifurcations, which are otherwise only known in *Entosiphon* (Chan et al. 2013, Farmer & Triemer 1994, 1988, Triemer 1986). Recently, *Entosiphon* has been considered both as the deepest-branching euglenid (Cavalier-Smith 2016, Cavalier-Smith et al. 2016), or a shallow-branching ploeotid that is sister to Spirocuta (Paerschke et al. 2017). These inferences were based on SSU rDNA phylogenies, and morphological characters like paramylon structure (Paerschke et al. 2017). None of our analyses recover either of these positions, but place *Entosiphon* in between 'Spirocuta + *Olkasia*', and the '*Keelungia* + *Lentomonas* + *Ploeotia*' clade. The deep branching position of *Entosiphon* in some SSU rDNA phylogenies is likely a phylogenetic artefact caused by the extremely divergent (and thus long-branching) SSU rDNA gene in this taxon, whereas in our multigene analyses *Entosiphon* is neither divergent nor long-branching.

The Deep Phylogenetic Position of Petalomonads

Petalomonads have previously been proposed as the deepest branch or branches among euglenids by some researchers, giving them a special importance in understanding euglenid evolution (Leander et al. 2007, 2001a, Leander & Farmer 2001, Triemer & Farmer 1991). Even with increased taxon sampling, SSU rDNA phylogenies failed to resolve the position of petalomonads (e.g. Lax et al. 2019, Schoenle et al. 2019), although they were sometimes placed close to the base of euglenids (e.g. Chan et al. 2013, Paerschke et al. 2017). Our multigene analyses recover petalomonads as the deepest branch among euglenids (Fig. 3.5), suggesting they are indeed highly relevant for early euglenid evolution. All three petalomonad taxa (*Petalomonas cantuscygni, Notosolenus urceolatus, Sphenomonas quadrangularis*) appear to be longer-branching than other euglenids, and there is a chance their phylogenetic position in our analyses might be due to long-branch attraction (LBA) artefacts. To limit the influence of LBA, we used site-heterogeneous mixture models (LG + C60 + F + Γ) and conducted several additional analyses, including removal of the longest-branching taxa (leaving petalomonads represented by *Notosolenus urceolatus* M49 and *Sphenomonas quadrangularis* AM6), removal of rogue taxa, and a progressive fast-site removal analysis. All of our additional analyses recover the basal position of petalomonads with high-to-full support (lowest is 96% UFB in FSR-analysis when only 52% sites retained), and suggest that this position is robust (Table B.6). Within petalomonads, we recover the same structure as in our SSU rDNA phylogenies (Fig. 3.1), with *Sphenomonas* basal to *Petalomonas* and *Notosolenus*.

Our results support the notion that petalomonads are the deepest branch among euglenids, and are thus pivotal in understanding the evolution of euglenids. For example, based on their low number of pellicle strips, *Sphenomonas* in particular has been suggested to be the deepest petalomonad branch (Leander et al. 2001a, Leander & Farmer 2001). The hypothesis suggested that likely-bacterivorous euglenids with few strips arose first, and through random strip-doubling events gave rise to new taxa with more pellicle strips. This ultimately resulted in the emergence of highly metabolic taxa with $\geq$38 pellicle strips (e.g. *Jenningsia, Urceolus, Peranema*). This flexibility would have enabled a flexible phagotrophic euglenid to engulf a pyramimonadalian green alga, ultimately giving rise to Euglenophyceae (Leander et al. 2001a, Leander & Farmer 2001, Turmel et al. 2008). No molecular sequences of *Sphenomonas* in particular were available prior to this study, and we are now in a much better position to test the hypothesis that *Sphenomonas* is truly the deepest branching euglenid.

Yet despite their morphological diversity and importance for understanding euglenid evolution, petalomonads are surprisingly poorly described in terms of ultrastructure: Few ultrastructural studies have thoroughly investigated petalomonad taxa (Lee & Simpson 2014a), and *Sphenomonas* has not been examined by electron microscopy at all. Since our multigene analyses and other studies place petalomonads close to the base of the euglenid tree, more detailed ultrastructural analyses of *Sphenomonas* and other petalomonads are crucial to further our understanding of early euglenid evolution.

Olkasia is Sister to Spirocuta

A clade of euglenid taxa with more than 12 pellicle strips arranged in a helical pattern has been proposed in the past (Leander et al. 2001a, Leander & Farmer 2001) and has been confirmed by recent SSU rDNA phylogenies of phagotrophic euglenids (Lax et al. 2019, Paerschke et al. 2017, Cavalier-Smith et al. 2016, Lee & Simpson 2014b). It was recently formally established as the taxon Spirocuta (Cavalier-Smith 2016). We recover *Olkasia* as sister to Spirocuta in our multigene-analyses (Fig. 3.5) and in some of our SSU rDNA phylogenies. Support for this position in our multigene phylogenies was high (92–99% UFB), and was always recovered but once (in Fig. B.11 'Euglenida-only' *Olkasia* fell sister to *Liburna* with no support). In some of our SSU rDNA analyses we recover a poorly supported grouping of Unidentified Ploeotid SMS7 and *Olkasia* as sister to Spirocuta (rather than just *Olkasia*), as did Lax et al. (2019) previously, but nothing further is known about SMS7 other than its

appearance in light microscopy. Arguably, the pellicle strip architecture of *Olkasia* is more similar to that of Spirocuta than are other ploeotids, with S-shaped strips that have considerable overhang (Lax et al. 2019). Furthermore, *Olkasia* has a chisel-shaped feeding apparatus, dissimilar to that of most other ploeotids, which possess a hook-shaped feeding apparatus (Cavalier-Smith et al. 2016, Chan et al. 2013, Lee 2008, Larsen & Patterson 1990, Lax et al. 2019). By contrast, chisel-shaped feeding apparatuses are common among Spirocuta (Larsen & Patterson 1990).

Closest Living Relatives to Euglenophyceae

The expansive phototrophic clade of Euglenophyceae arose following a secondary endosymbiotic event involving an unknown phagotrophic euglenid and a pyramimonadalian green alga (Jackson et al. 2018, Turmel et al. 2008). The exact nature of the phagotrophic host is unknown—in the past, *Teloprocta* has been proposed as the closest relative of Euglenophyceae, based its close phylogenetic affinity in some SSU rDNA trees (Cavalier-Smith 2016), as has *Urceolus* (Leander et al. 2001a). The latter proposition was based on the high number of strips ($\geq$38) and other cytoskeletal markers, including a potential paraflagellar swelling and stigma similar to the eyespot in phototrophic euglenids (Leander et al. 2001a). Our SSU rDNA analyses routinely recover some combination of Clade A (*Jenningsia fusiforme*), Group B (*Heteronema* c.f. *globuliferum*), and some *Urceolus* (α, ε) as sister to Euglenophyceae, albeit with no support. None of these sequences have been available before this study. Our multigene analysis includes a clade of *Urceolus*, but in all our analyses 'Peranemids' as a whole form a single clade that branches sister to Euglenophyceae, rather any particular genus-level group. This 'Peranemids' clade includes *Urceolus, Chasmostoma, Peranema*, and two clades of *Jenningsia* (Figures 3.1 and 3.5). This diversification within peranemids is consistent with Euglenophyceae arising relatively early in spirocute evolution. Yet *Jenningsia* is not recovered as monophyletic, as the position of *Jenningsia I* (Clade A) is unstable and in most analyses branches sister to all other peranemids with low to moderate support (78–85% UFB; omitted in Fig. B.9 'noRogues'; and in Fig. B.11 'Euglenida-only' branches sister to 'Anisonemids plus' clade with 47% UFB). As such, *Jenningsia I (fusiforme)* remains a possible candidate for a closer relative of phototrophic euglenids. It is noteworthy that the current multigene sampling does not include any taxa from either *Heteronema I* (Group B) or *Teloprocta* (Clade L). Sampling from these organisms would be preferrable, as they could also be close relatives to Euglenophyceae.

Are Symbiontids an Independent Branch of Euglenozoa?

Some symbiontid morphospecies have been known for several decades (Lackey 1960, Fenchel et al. 1995), yet molecular (SSU rDNA) data was only available for two described morphospecies, *Calkinsia aureus* and *Bihospites bacati*, along with ~190 environmental sequences (Monteil et al. 2019, Yubuki & Leander 2018, Orsi et al. 2011). Phylogenies of the SSU rDNA placed them in a monophyletic group of uncertain affinity within Euglenozoa, recovering them as sister to kinetoplastids (Yubuki et al. 2009), sister to Euglenids (Monteil et al. 2019, Schoenle et al. 2019, Cavalier-Smith 2016), a branch within euglenids (Paerschke et al. 2017, Lax & Simpson 2013), or an unspecified independent branch within Euglenozoa (Breglia et al. 2010, Yubuki & Leander 2018). This uncertain placement has led to competing views on whether symbiontids are derived euglenids (that presumably lost the euglenid pellicle), or a separate clade of Euglenozoa (Breglia et al. 2010, Cavalier-Smith 2016).

Our study is the first to include multigene data for symbiontids (Fig. 3.5). Both examined cells (HLA12 and KSa7) are closely related in our SSU rDNA analyses (Fig. 3.1; 'Symbiontida clade d', as per Yubuki & Leander 2018) and show a similar morphology to *Postgaardi* (Simpson et al. 1997, Fenchel et al. 1995). In our multigene analyses these cells form a fully supported clade that never falls within euglenids, but rather sister to Glycomonada (Fig. 3.5) or to Kinetoplastea (Table B.6). Placements outside euglenids have been proposed in the past based on some SSU rDNA phylogenies (see above) and the view that several ultrastructural features of symbiontids are most reminiscent of diplonemids (Cavalier-Smith 2016). While our analysis strongly argues against symbiontids being derived euglenids, their exact phylogenetic identity remains somewhat elusive: In most phylogenetic analyses they are placed sister to Glycomonada with moderate to high support (84–99% UFB; Fig. 3.5), yet this relationship is unstable in our FSR-analysis (Fig. B.12). The quality of our symbiontid single-cell transcriptomes is comparatively low (Tables B.3 and B.4), which might play a role in their uncertain placement. Therefore, higher quality multigene data and an increased sampling across all Euglenozoa, including across symbiontid diversity, will help to more confidently place this taxon. As no symbiontid has been successfully cultured so far, the availability of high-quality single-cell genomes (see Monteil et al. 2019) or single-cell transcriptomes will likely be key.

Motility Across Euglenids

Euglenids have evolved several mechanisms of locomotion. Euglenophyceae tend to be swimmers, either with a single flagellum or multiple flagella (Leander et al. 2017), whereas phagotrophic euglenids are overwhelmingly surface-associated and glide on their flagella (see Videos B.7). Rather than dragging, gliding smoothly transports cells across surfaces, and while the underlying molecular mechanisms are not entirely understood, it seems to be powered by some surface motility machinery located on the anterior flagellum in *Peranema* (Saito et al. 2003). Multiple forms of gliding exist in euglenids: Some taxa glide on their anterior flagellum (e.g. *Petalomonas, Notosolenus, Jenningsia, Urceolus, Peranema*; Videos B.7), whereas others glide on their posterior flagellum (e.g. *Anisonema, Dinema, Ploeotia, Lentomonas*; Videos B.7). *Neometanema* have evolved to 'skid' along surfaces, with the posterior flagellum in loose contact with the surface while the beating (i.e. motion) of the anterior flagellum powers the cell (Lee & Simpson 2014b). Anterior gliding is present in two quite distinct, large groups of phagotrophs, as anterior-gliding petalomonads appear at the base of the tree, while peranemids lie within Spirocuta (Figures 3.5 and 3.1). The flagellar movement pattern seems to differ between both, with petalomonads moving only the absolute distal tip of the flagellum (mostly up-and-down) with the remainder staying attached to the substrate (e.g. *Notosolenus*; Video GGB3 in B.7). Different peranemid taxa, by contrast, show varying lengths of their anterior flagellum staying in contact with the surface, with some taxa beating as much as the distal ½ of their flagellum (e.g. *Urceolus*; Video WBF1 in B.7).

Petalomonas

Jenningsia

Posterior gliders are represented by ploeotids close to the base of euglenids, and by anisonemids within Spirocuta. Gliding patterns in those two groups appear to be very similar—in both the posterior flagellum is fully in contact with the surface, while the anterior flagellum beats in front of the cell, in some cases quite wildly (e.g. *Anisonema, Olkasia, Ploeotia*). *Heteronema vittatum* presents a special case of anterior gliding, as the posterior flagellum appears to be in contact with the substrate as well. Yet, the long anterior flagellum (~2x cell length) is attached to the surface with its proximal half and moves the cell forward, while the distal half 'sweep-beats' regularly (Videos ABLN4 and ABIC3 in B.7). All other taxa in the 'Anisonemids plus' group are strictly posterior gliders (*Dinema, Anisonema*), 'skidders' (*Neometanema*), or swimmers (primary osmotrophs), suggesting that the anterior-supported gliding strategy in *H. vittatum* has evolved separately from both petalomonad and peranemid anterior gliding.

Ploeotia

H.vittatum

Based on our phylogenies, it would be parsimonious to infer that the posterior gliding strategy was ancestral in Spirocuta, having been inherited from their ploeotid ancestors, and therefore anterior gliding in peranemids evolved from posterior gliding ancestors and independently of petalomonads. The ancestral state of gliding in euglenids, on the other hand, can currently not be inferred. We simply do not know what the ancestral state of motility in other Euglenozoa was, and in combination with the exact phylogenetic placement of symbiontids still unresolved, it remains unclear whether ancestral euglenids were anterior or posterior gliders.

3.4.3 Methodology

Culture vs. Single-Cell Transcriptomes vs. SAGs

We mainly used a single-cell approach to gather both phylogenomic and single-gene data, supplemented by some of the few phagotrophic euglenid cultures currently available (see Lax et al. 2019). The single-cell approach enables a broad molecular sampling of euglenid diversity without the need to establish cultures. Phagotrophic euglenids, in particular, have historically been difficult to cultivate and maintain long-term (see Leander et al. 2017). Circumventing this culturing bottleneck with a single-cell approach enables the rapid gathering of molecular data, and when used in conjunction with high-quality light microscopy, links morphological information with molecular sequences. This linkage can be particularly important when morphologically-defined taxa do not correspond to clades (e.g. *Dinema* and *Heteronema*; *Anisonema* before the separation of *Liburna* and *Hemiolia* in Lax et al. 2019). Yet, the availability of cultures is absolutely crucial to further our understanding of the biology of euglenids through ultrastructural studies and other detailed examinations (e.g. prey preference, presence/absence of specific proteins for locomotion), especially when considering certain traits like the euglenid pellicle.

We employed a single-cell transcriptomics approach (SmartSeq2, Picelli et al. 2014) to acquire multigene data for phagotrophic euglenids. This method has been used successfully to investigate the phylogenies of various protist groups, including Amoebozoa (Kang et al. 2017), Holozoa (Hehenberger et al. 2017) and Retaria (Krabberød et al. 2017a), as well as to determine the position of Hemimastigophora within eukaryotes (Lax et al. 2018). By contrast, the performance of single-cell amplified genomes (SAGs) to infer eukaryote phylogenetics has a checkered history. Due to their size, complexity, and large stretches of non-coding sequences, many eukaryotic genomes are intrinsically difficult to sequence and assemble (Keeling & del Campo

2017). As a result, there have been few studies where SAGs have been used successfully for multigene phylogenetics, i.e. where more than just the SSU rDNA operon was analysed (e.g. Yoon et al. 2011, Ahrendt et al. 2018). Some of the SSU rDNA sequence data examined in this study did come from SAGs (cells BP3, SDB1, SDB4, UB10; Fig. 3.1), but the assemblies did not provide sufficient data for any kind of multigene analysis. The recently published draft genome of *Euglena gracilis* seems to be considerably expanded and complex due to a large proportion of non-coding sequence (Ebenezer et al. 2019). Other euglenid genomes might be similarly complex and accordingly difficult to sequence and assemble, and this may explain the poor quality of our phagotrophic euglenid SAGs. In addition, the SmartSeq2-based transcriptomics method uses a poly-A selection step that reduces bacterial contamination (Picelli et al. 2013, 2014, Kolisko et al. 2014), an issue that can complicate subsequent analyses of SAGs (Yoon et al. 2011). Therefore, our study supports the view that single-transcriptomics approaches tend to generate higher quality data at a lower price-point than single-cell genomics—at least when used for phylogenomic or multigene phylogenetic analyses (Kolisko et al. 2014, Lax et al. 2018).

Our single-cell transcriptomes varied considerably in terms of BUSCO-scores and phylogenomic-pipeline coverage (Tables B.3 and B.4), with the lowest-quality assembly—*Olkasia polycarbonata* UB45—only recovering 16 BUSCOs (complete and duplicate and fragmented) and 13.7% of sites of a previously published 351-gene phylogenomic pipeline (Brown et al. 2018). By contrast, the highest-quality assembly (*Peranema trichophorum* PtR) included 148 BUSCOs and 57.3% of the phylogenomic marker sites. Rather than using a comprehensive set of 100+ marker genes, as is routinely done now for phylogenomic studies (Gawryluk et al. 2019, Lax et al. 2018, Kang et al. 2017, Hehenberger et al. 2017, Brown et al. 2013), we used 20 long marker genes (>250 aa after trimming) that were present in most of our euglenid taxa. The lower number of selected genes resulted in a high-quality dataset that could be manually checked for contaminant, paralogous, and otherwise aberrant sequences several times, and with a greater potential to detect such sequences (see Methods).

Does Phylogenomics Resolve a Tree of Euglenids?

Previously published multigene phylogenies of euglenids or Euglenozoa only included phototrophic euglenids, and thus no phagotrophic euglenids whatsoever (e.g. Yazaki et al. 2017, Karnkowska et al. 2015, Simpson et al. 2006, Hampl et al. 2009). This relative neglect of phagotrophs reflects historical trends: In fact, an EST project

of *Peranema trichophorum* represented the only bulk data of nuclear coding regions reported from a phagotroph prior to our work (Maruyama et al. 2011). This study is thus the first phylogenomic analysis to include phagotrophic euglenids, and thereby capture most of the known picture of euglenid diversity (Fig. 3.5).

Our dataset enables a more detailed study of the deep relationships among euglenids than previous examinations. The SSU rDNA gene has repeatedly been shown to be inadequate for this task as it is considered highly divergent in some taxa and offers poor resolution, especially at the backbone of the tree (e.g. Lax et al. 2019, Lax & Simpson 2013, Paerschke et al. 2017, Breglia & Leander 2007). We recover several well-supported groupings like Spirocuta, 'Anisonemids plus' (Anisonemids + *Neometanema* + Aphagea + *H. vittatum*), a clade of several ploeotid taxa, and petalomonads (Fig. 3.5).

While our study provides an overview of broad relationships among euglenids, there are some known gaps in our dataset. Within Spirocuta, *Teloprocta* and *Heteronema globuliferum* are not yet sampled. Also, *Urceolus* appears as non-monophyletic in SSUr-DNA trees (Fig. 3.1), and we currently only have sampling for one clade of this genus. The availability of these taxa will further test important relationships within Spirocuta; for example, whether 'Peranemids' are truly monophyletic. Our analyses identify ploeotid taxa as making up most of the backbone of the euglenid tree, but still with overall low support for much of the branching order among the different ploeotid groups (apart from *Olkasia*). This might partly result from missing taxa, as for example, both *Decastava* and *Hemiolia* have not been sampled for phylogenomics. Lax et al. (2019) also found several ploeotid cells that show no strong affinity to any particular named clade of ploeotids in SSU rDNA phylogenies (also see Fig. 3.1: CARR5, SMS7, WF2_3). It is possible that these taxa represent important parts of the diversity of phagotrophs that are relevant to the early evolution of euglenids, and could help in improving resolution in future phylogenomic analyses. Also, as noted above, our multigene dataset only includes one of several known major clades of symbiontids.

The phylogenomic dataset reported here provides a robust basis to investigate several questions in euglenid research, including the evolution of the euglenid pellicle and patterns of motility across taxa (see above). Currently unsampled taxa could be easily added to this dataset and will provide additional insight.

Chapter 4

Hemimastigophora is a Novel Supra-kingdom-level Lineage of Eukaryotes

A version of this chapter has been 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**:410–414.
doi:10.1038/s41586-018-0708-8

* contributed equally

Authors' contributions: YE isolated the organisms, and cultivated *Hemimastix kukwesjijk*. YE and GL undertook the microscopy. GL performed the single-cell transcriptomics. YE, GL and EMB analysed the rDNA and environmental sequence data. GL, LE, YE, and AGBS assembled the phylogenomic datasets. GL, LE and AJR performed phylogenomic analyses. LE and YE performed the gene presence analyses. GL, YE and AGBS wrote the manuscript, with input from all co-authors.

¶

4.1 Introduction

ALMOST all eukaryote lifeforms have now been placed within five to eight supra-kingdom-level groups using molecular phylogenetics (Burki 2014, Worden et al. 2015, Burki et al. 2016, Simpson & Eglit 2016). The 'phylum' Hemimastigophora is probably the most distinctive morphologically defined lineage still awaiting such a phylogenetic assignment. First observed in the 19th century, hemimastigotes are free-living predatory protists with two rows of flagella and a unique cell architecture (Klebs 1893, Foissner et al. 1988, Foissner & Foissner 1993), but until now there have been no molecular sequence data and no cultures available for this group. Here we report phylogenomic analyses based on high-coverage cultivation-independent transcriptomics that place Hemimastigophora outside all established eukaryote supergroups. Remarkably, they instead comprise an independent, supra-kingdom-level lineage, most likely sister to the 'Diaphoretickes' half of eukaryote diversity (i.e. supergroups Archaeplastida, Sar, Cryptista and others). The previous 'phylum' rank of Hemimastigophora understates their evolutionary distinctiveness: The group has considerable importance for investigations of the deep-level evolutionary history of eukaryotic life, ranging from the origins of fundamental cell systems, to the root of the tree. We have also established the first culture of a hemimastigote (*Hemimastix kukwesjijk* n. sp.), which will facilitate future genomic and cell biological investigations into eukaryote evolution and the last eukaryotic common ancestor.

4.2 Methods

4.2.1 Cell Isolation and Transcriptomics

Soil from mixed-species woodland in the Bluff Wilderness Trail in Nova Scotia, Canada (44.6610154 N, 63.7674669 W; April 17th 2016) was kept hydrated with distilled water in a Petri dish until hemimastigotes were observed about 4 weeks later. Single *Spironema* and *Hemimastix* cells were isolated with drawn-out micropipettes, photodocumented by differential interference contrast (DIC) light microscopy (using a Zeiss Axiovert 200M/AxioCam ICc5 microscrope and camera system; Carl Zeiss AG, Germany), and subjected to 'single-cell' transcriptomics using the SmartSeq2 protocol (Picelli et al. 2014) with modifications. Briefly, four (*Spironema*) or two (*Hemimastix*) cells were individually picked into 0.2% Triton-X lysis buffer, immediately frozen in liquid nitrogen, then thawed and re-frozen three times. The remaining

procedure followed the original protocol, with 20 (*Spironema*) or 18 (*Hemimastix*) PCR cycles. cDNA quantity and quality was assessed i) by Qubit dsDNA HS assay (Thermo-Fisher, cat. #Q32851), and ii) by PCR, and cloning of cDNA fragments into StrataClone SoloPack competent cells (Agilent Technologies), with 12 clones each Sanger-sequenced. After library preparation with Illumina Nextera XT, sequencing was carried out on an Illumina MiSeq with 2x250 bp dual reads, with the libraries multiplexed on the same run.

4.2.2 Cultivation of *Hemimastix kukwesjijk*

To cultivate *Hemimastix kukwesjijk* strain BW2H, three cells were picked and washed with a micropipette, then transferred to a prey, *Spumella* sp. (strain BW2S), which was cultured by serial dilutions from the same sample. Cultures were maintained in 15 ml tubes containing ~4 mL of 25%-strength ATCC medium 802, with one sterilised barley grain, angled for aeration and transferred weekly. Cells were examined by light microscopy as described above.

4.2.3 Scanning Electron Microscopy of *Hemimastix kukwesjijk*

Cells from a 10 day-old culture of strain BW2H were fixed for 30 min in OsO_4 vapour alone (at room temperature) or OsO_4 vapour simultaneously with 2.5% glutaraldehyde (on ice), and filtered onto 2 μm Isopore Membrane filters (Millipore). These were washed in distilled water and dehydrated in an 50%-70%-80%-90%-95%-100% ethanol series, critical-point-dried in CO_2 and sputter-coated by 10 nm of gold-palladium. Cells were imaged using a Hitachi S-4700 SEM at 3kV.

4.2.4 SSU rDNA Analyses

A single cell of *Spironema* c.f. *multiciliatum* was isolated and washed by micropipetting and photodocumented (see above), then its genomic DNA was amplified using multiple displacement amplification (Illustra GenomiPhi V3 DNA amplification kit, GE Healthcare). Total genomic DNA was extracted from *Hemimastix kukwesjijk* culture BW2H (also including the prey *Spumella* sp., strain BW2S) using a Qiagen DNeasy kit. Partial SSU rDNA sequences were PCR-amplified from *Spironema* c.f. *multiciliatum*, and *Spumella* sp. BW2S, using primers 82F (5'-GAAACTGCGAATGGCTC-3') and 1498R (5'-CACCTACGGAAACCTTGTTA-3'), with annealing temperatures of 58 ºC and 55 ºC, respectively. A partial *Hemimastix* SSU rDNA sequence was PCR-

amplified from strain BW2H using exact-match primers Hemi2-342F (5'-ACTTTCGATTGTAGGATAGA-3') and Hemi2-1103R (5'-AAAACTTGCGATTTCTCTGG-3') with an annealing temperature of 55 ºC. All amplicons were directly Sanger-sequenced at Génome Québec. The SSU rDNA of *Spumella* sp., strain BW2S was 99% identical to *Spumella* strain 187hm (GenBank accession ID DQ388550).

The SSU rRNA sequences for the two hemimastigotes were extracted from the transcriptome data (see above) and compared to the SSU rDNA sequences obtained independently from genomic DNA, to ensure mutual identity (though the rDNA sequence of *H. kukwesjijk* did differ from the transcriptome-derived rRNA sequence in having a 395 bp intron). The transcriptome-derived SSU rRNA sequences (and environmental clone AY689797, retrieved from GenBank via megablast) were then added via profile alignment using MUSCLE (Edgar 2004) to a global eukaryotic alignment of SSU rRNA genes (111 taxa total). Following manual inspection of the alignment, poorly aligned sites were masked using Gblocks (Castresana 2000) with subsequent manual correction (1252 sites retained), and a phylogeny was estimated in RAxML under the GTR + Γ model (Stamatakis 2014) with a 1000 replicate bootstrap analysis (Fig. C.3).

4.2.5 Environmental SSU rRNA/rDNA Sequence Comparisons

Eukaryotic environmental SSU rRNA/rDNA-derived sequences were acquired from VAMPS (Huse et al. 2014, V9), TARA oceans (de Vargas et al. 2015, V9), BioMarKs (Biomarks Consortium 2011, V4), a neotropical soil study (Mahé et al. 2017, V4), a high-arctic Fjord water column study (Marquardt et al. 2016, V4), and a soil metatranscriptome dataset (Geisen et al. 2015) and queried in a BLAST (Altschul et al. 1990) analysis with the appropriate (V4 or V9) section of the *Spironema* and *Hemimastix* SSU rRNAs, at a 85% identity cut-off (top 500 hits). The corresponding short reads from the datasets were first aligned to the eukaryote reference alignment (see above) using PaPaRa (Berger & Stamatakis 2011) version 2.5 and then placed on the SSU rRNA gene tree (Fig. C.3) using pplacer (Matsen et al. 2010) version 1.1. Chimeric reads were identified manually with BLAST against nt/nr and discarded (all cases were from VAMPS V9 datasets). Reads were also discarded if the top 100 BLAST hits were all to a single taxonomic group (e.g. ciliates). Surviving reads were assigned to Hemimastigophora if they were placed on a particular branch within Hemimastigophora with a likelihood-to-weight ratio (LWR) > 0.5, or if they had an accumulated LWR > 0.9 across the multiple branches within Hemimastigophora.

4.2.6 Phylogenomic Dataset Assembly

To perform phylogenomic analyses of eukaryotes that included hemimastigotes, we used the single-cell transcriptomes derived from *Hemimastix* (2 cells) and *Spironema* (4 cells), as described above. Raw reads from the Illumina sequencing were quality-trimmed, and the adapters clipped, with Trimmomatic version 0.3244 (default parameters), then assembled with Trinity version 2.0.2 (default parameters, Grabherr et al. 2011). Assemblies were cleaned of sequencing cross-contamination using a custom script. Marker genes of interest were extracted using a previously reported pipeline (Brown et al. 2018) and appended as translated peptide sequences to a 396-taxon 351-gene eukaryote dataset (Brown et al. 2018). This dataset was pruned to 107 taxa that broadly represented all major eukaryotic groups for which data were available, while excluding extremely 'long-branching' species and, where possible, species with poor sampling of this gene set. The 351 single-gene dataset were aligned individually with MAFFT-L-INS-i version 7.0 (Katoh & Standley 2013), and trimmed with BMGE version 1.0 (-m BLOSUM30 -h 0.5 -g 0.2, Criscuolo & Gribaldo 2010). From the resulting files, single-gene trees were generated with IQ-TREE version 1.4.4 (Nguyen et al. 2015) under the $\mathrm{LG}+\mathrm{C20}+\mathrm{F}+\Gamma$ model with a 1000-replicate ultra-fast bootstrap approximation (UFboot) to estimate branch support (Minh et al. 2013). These trees were manually checked for sequences corresponding to probable paralogs, contaminants, or lateral- or endosymbiotic gene transfers, which were then removed from the datasets. The tree estimation and manual checking was then repeated, and any additional suspect sequences removed. Three taxa with limited remaining data ($<10\%$ of sites) were then excluded, leaving 104 taxa for initial phylogenomic analysis.

4.2.7 Quality of Hemimastigote Transcriptomes

It was particularly important to assess the quality of the data from *Spironema* and *Hemimastix*, both because they were the subject of the study, and because they were derived using single-cell methods from crude enrichments. The transcriptome from *Spironema* included 290 of the 351 genes in the phylogenomic dataset (82.6%) and 77.6% of the sites retained after trimming. The transcriptome from *Hemimastix* included $280/351 = 79.7\%$ of genes, 72.1% of sites. In other words, they were both quite data-rich from a phylogenomic perspective, comparing well to many transcriptomes from cultivated non-model protists (Table C.2). Some 247 of the 351 gene alignments (70.4%) included both taxa. The *Spironema* and *Hemimastix* sequences

formed a clade in 168 of the 247 single-gene trees inferred for these data (=68%), which is consistent with a specific relationship between them while also being quite genetically distinct (and bearing in mind that some of the individual genes in the dataset carry relatively little phylogenetic signal). There was no particular pattern to the relationships between each hemimastigote and other eukaryotes in the remaining 32% of trees. In summary, the single-gene trees indicate that there was little-to-no contamination from other eukaryotes in the analysed hemimastigote data. Furthermore, the *Spironema* and *Hemimastix* sequences always differed in these 247 alignments, confirming that no cross-contamination between the two had carried through to the final dataset.

4.2.8 Phylogenomic Analyses

The 351 individual gene alignments with 104 retained taxa (see above) were concatenated, and trimmed with BMGE (-m BLOSUM30 -h 0.42 -g 1), yielding a 104-taxon, 93798 aa dataset. To allow more complex analyses, we then excluded 43 phylogenetically redundant taxa, followed by re-trimming with BMGE (as above), to generate a 61-taxon dataset with 93903 amino acid sites. Taxa were selected for retention in the 61-taxon dataset such that eukaryote diversity remained reasonably evenly sampled, and that all major taxa included in the 104 taxa dataset were still represented. Where there was a choice, species with high gene coverage were retained in preference to more poorly sampled species, and shorter-branching species were retained over longer-branching species. Phylogenies for both datasets were inferred by Maximum Likelihood (ML) using IQ-TREE under the LG + C60 + F + Γ mixture model, with robustness assessed by ultra-fast bootstrap approximation (UFboot; 1000 replicates). The 61-taxon dataset was also subjected to a 'full' bootstrap analysis with 200 replicates under the Posterior Mean Site Frequency (PMSF) model implemented in IQ-TREE. PMSF is a site-heterogeneous mixture model that can closely approximate complex mixture models such as LG + C60 + F + Γ while reducing computational time several-fold (Wang et al. 2017), making full bootstrapping practical for our ~60 taxa datasets. The ML tree inferred for this dataset under the LG + C60 + F + Γ model (see above) was used as the guide tree for the PMSF analysis. The 61-taxa dataset was also subjected to Bayesian analysis with PhyloBayes (Lartillot et al. 2009) version 4.1 under the CAT + GTR model (Lartillot & Philippe 2004), with default priors and Markov chain Monte Carlo settings. Four independent Markov chain Monte Carlo chains were run for ~10000 generations. Three chains converged (maxdiff < 0.13;

burn-in = 3000). Their consensus tree shows Hemimastigophora sister to (other) Diaphoretickes with maximal support (ie. consistent with the ML tree), while the unconverged chain yielded the topology where Hemimastigophora is sister to the Sar + *Telonema* + Haptophyta + Centrohelida grouping.

Several further sets of analyses were conducted on derivatives of the 61-taxa dataset. First, we used a custom script to calculate average tip-to-tip distances for each taxon and identify 'long-branching' outliers (i.e. taxa whose average tip-to-tip branch length was longer than three standard deviations from the center of the distribution of average branch lengths). Removing the three identified outliers (*Bodo, Diplonema, Tetrahymena*) yielded the '58 taxa, no longest branches' (58-nLB) dataset. This was analysed using ML, as per the main 61-taxa analysis (IQ-TREE with LG + C60 + F + Γ, with 1000-replicate UFboot, and 200 bootstraps using PMSF with LG + C60 + F + Γ ML tree for the 58-nLB dataset as guide tree).

Second, we deleted the three most data-poor taxa, each of which had site coverage < 30% (Telonema, Gromia, and Picozoan PB58411a), resulting in a '58 taxa, no data-poor species' (58-nDP) dataset. This was analysed using ML as per the main 61-taxa analysis, except that the PMSF bootstrap analysis was based on 100 replicates.

Third, we recoded the main 61-taxa dataset into four distinct categories of amino acids (SR4 scheme, Susko & Roger 2007), to address possible compositional heterogeneity. The resulting 61-SR4 dataset was analysed with IQ-TREE under a GTR + R6 + F model, with 500 real bootstrap replicates.

Fourth, we used IQ-TREE's assignment of per-site rates (-wsr flag) for the main 61-taxa dataset, and progressively removed the fastest evolving sites in 10 steps, with approximately 4% of the sites removed in each step. This yielded 10 'Stepwise Fastest Sites Removed' (61-SFSR) datasets. To exclude the influence of the position of Hemimastigophora in the guide trees for subsequent PMSF analyses, we deleted the two hemimastigotes from the full dataset and the 10 SFSR datasets (i.e. 11 total) with phyx version 0.1 (Brown et al. 2017), and pruned these two species from the ML tree from the 61-taxa dataset. The pruned tree was then used as the guide tree to calculate PMSF profiles ('PMSF-nHEMI') under LG + C60 + F + Γ. For each of the original 11 datasets (i.e. including hemimastigotes) we then inferred support for important bipartitions under this LG + C60 + F + Γ PMSF model using a 1000 replicate UFboot analysis, and plotted these against the percentage of sites remaining (Fig. C.8). Note that this method of generating the PMSF model (i.e. PMSF-nHEMI) and evaluating statistical support differs from the main analyses (e.g. 61-taxa; 58-

nLB; 58-nDP), and the support values cannot be directly compared between these analyses and the 61-SFSR analyses.

4.2.9 Identification of Non-Universal Ancient Genes

In order to search the hemimastigote transcriptome data for gene innovations that potentially originated early in the evolution of crown eukaryotes (and thus may also represent synapomorphies informing the relationships between major supergroups), we collated a set of gene systems reported in the literature to include genes with widespread but not universal distributions across major eukaryote groups. Specific genes were selected on the basis of being present in more than one species-rich 'supergroup' of eukaryotes, for example both Obazoa and Amoebozoa (see Table C.3); for this purpose, Metamonada and Discoba were considered distinct supergroups. Sequences were retrieved from GenBank, or from the literature, and used as BLASTp queries against both hemimastigote transcriptomes, translated into amino acid sequences using a custom script (default genetic code). Where genes were not identified with BLASTp, Hidden Markov Model (HMM) profiles were obtained either from the Pfam database, or the literature (as indicated in Table C.3), or were built de novo from the alignments in the corresponding literature using hmmbuild, and then scanned for in both hemimastigote transcriptomes using hmmscan (both hmmbuild and hmmscan from the Hmmer-3.1b2 package, Eddy 2011). Genes that were retrieved in only one of the hemimastigote transcriptomes were used as BLASTp queries against the other. Hemimastigote candidate orthologues were verified by reciprocal BLASTp against the nr database, and, where appropriate, domain annotation databases (InterProScan, SMART), and added to pre-existing alignments from corresponding references (as shown in Table C.3) via profile alignment using MUSCLE in Seaview version 4.6 (Edgar 2004, Gouy et al. 2010). Where phylogenies were necessary to further confirm identity (particularly in the case of multigene families), the alignments were trimmed using BMGE version 1.148 (-m BLOSUM30, Criscuolo & Gribaldo 2010), and phylogenies estimated in IQ-TREE version 1.5.5 (Nguyen et al. 2015) under the LG4X model. An alignment for HPS1 was not available in the original publication and was instead assembled from sequences from GenBank and publicly available transcriptomes, and aligned via MAFFT-L-INS-i (Katoh & Standley 2013). Because of the large size of the myosin gene family and the level of divergence between various paralogues, myosin homologues were instead aligned with

MAFFT-E-INS-i and trimmed less conservatively (BMGE; -m BLOSUM30 -b 2), with the corresponding phylogeny estimated under the LG + C60 + F + Γ model.

4.2.10 Identification of *Spironema* c.f. *multiciliatum*

The elongate shape of these cells (Figures 4.1a and C.1a), and restriction of the 'main row' of flagella to the anterior portion identified our organism with *Spironema*, rather than *Hemimastix* (broad and flattened), *Paramastix* (globular) or *Stereonema* (elongate but main flagellar rows about half the length of the cell; see (Foissner & Foissner 1993, 2002, Zolffel & Skibbe 1997). There are three previously described species of *Spironema: Spironema terricola, Spironema goodeyi*, and *Spironema multiciliatum*. The shape and size of our specimens is inconsistent with *S. terricola* and *S. goodeyi*, both of which are very long and thin (Foissner & Foissner 1993). In addition, neither of these species has any posterior flagella. Our cells are similar in shape to *Sp. multiciliatum* (Klebs 1893). The number of flagella in the 'main row' and the presence of a few difficult-to-observe flagella towards the posterior end are also broadly consistent with a previous account of *S. multicilliatum*, in which such posterior flagella were in some cells (Klebs 1893, Foissner & Foissner 1993). However, our cells are 23–31 μm (mean: 27.4 stdev: 3.45; n = 7; see main text) which is markedly longer than the 18 μm reported for *S. multiciliatum*. Thus, we determined that our specimens are similar, but not identical, to *S. multiciliatum*.

4.2.11 Data Availability

Raw reads of *Spironema* and *Hemimastix* transcriptomes are deposited on GenBank under accession IDs SRR6032743 and SRR6032744, respectively. The assembled *Hemimastix* and *Spironema* transcriptomes, 351 individual gene alignments (104-taxa), concatenated and trimmed alignments and tree-files for the 104-taxa, 61-taxa, 58-nLB, 58-nDP, 61-SR4 and 61-SFSR datasets, alignments and tree files for non-universal ancient genes, raw LM and SEM images, and the SSU rDNA alignment and tree-files are deposited on Datadryad dryad.n5g39d7. The partial SSU rDNA gene sequence of *Hemimastix kukwesjijk* strain BW2H is deposited on GenBank under accession ID MF682191. *Hemimastix kukwesjijk* has been deposited in the ZooBank database (`http://zoobank.org/`) with LSID urn:lsid:zoobank.org:pub:4BA2A83C-8363-4EBE-A9C7-097CA470F9FB.

4.3 Results and Discussion

Two undescribed species of the rarely observed protist group Hemimastigophora (one *Spironema*, one *Hemimastix*) were identified in enrichments from soil. Isolated single cells of *Spironema* c.f. *multiciliatum* were spindle-shaped with a thin 'tail', 23–31 μm long by 4–7.5 μm wide (mean ± stdev = 27.4 ± 3.5 x 5.4 ± 1.6 μm; n = 7), with an oval nucleus, and two rows of 6+ flagella clustered in the anterior quarter, plus 2–3 flagella per row more posteriorly (Figures 4.1a and C.1a). Cells of *Hemimastix kukwesjijk* n. sp. (description below) were oval in profile with a blunt anterior projection (capitulum), and two flagellar rows along their whole length (Figures 4.1b and C.1). In cultivation (strain BW2H), live cells were 16.5–20.5 μm long by 7–12.5 μm wide (18.3 ± 1 x 9.9 ± 1.2 μm; n = 61), with a subcentral rounded nucleus and posterior contractile vacuole (Fig. 4.1c). Each row of 17–19 flagella (mean 18.4; n = 25) lay in a channel between the two thick thecal plates. The anteriormost 9–10 flagella were closely spaced, while the rest emerged from separate notches in the underlying plate (Fig. 4.1b and 4.1e). The capitulum was bordered by the overlapping anterior ends of the flagellar rows, with the adjacent plate margins housing extrusomes (undischarged: Fig. 4.1f; discharged: Fig. C.2c). Cells fed on a small stramenopile (*Spumella* sp.) after attachment at the capitulum, and enclosure by the anterior flagella (Fig. 4.1d, Figures C.1h–k and C.2a–b).

We determined SSU rRNA (small subunit ribosomal RNA) sequences from both hemimastigotes and used these to probe published environmental sequence datasets to determine 1) the group's distribution across habitats, and 2) whether they matched a known environmental clade. Unlike some other recently characterized lineages (e.g. Yubuki et al. 2015), hemimastigotes do not appear to belong to a previously identified environmental clade. One unclassified long-read clone from freshwater sediment (AY689797) was phylogenetically related to *Spironema* (Figures 4.2 and C.3). An additional 37 short reads were detected among V4 or V9 amplicon datasets or soil metatranscriptomes (Fig. 4.2; Table C.1). Many of the V4 and V9 amplicons derived from soil or freshwater, consistent with most light microscopy accounts (Foissner & Foissner 1993). However, nearly half came from marine sediment or water column samples (Fig. 4.2), and one *Hemimastix*-like V4 amplicon was among the 25 most abundant OTUs (operational taxonomic units) in a fjord sediment dataset (Table C.1).

To place hemimastigotes in the tree of eukaryotes, we generated transcriptomes from isolated single cells of both *Spironema* and *Hemimastix*, and assembled 351-

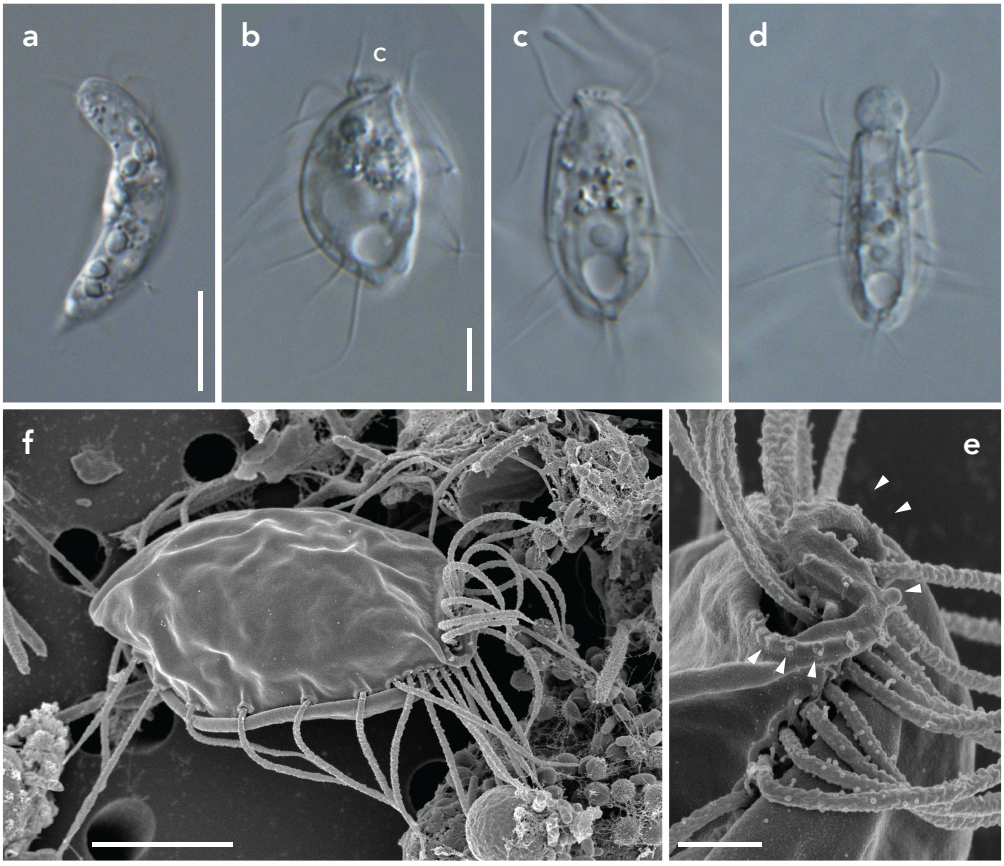

Figure 4.1: Micrographs of studied hemimastigotes. (a) *Spironema* c.f. *multiciliatum*: cell 1 (of 4) isolated for transcriptomics. (b-f) *Hemimastix kukwesjijk*: (b) Cell 1 (of 2) isolated for transcriptomics; note capitulum (c). (c-d) Cells from culture (strain BW2H) – note contractile vacuole at posterior, and nucleus (in c), and feeding on prey with capitulum (in d). (e) General view of cell (strain BW2H), anterior with capitulum to right. (f) Detail of capitulum, showing caps of undischarged extrusomes (arrowheads) and close-spaced flagella in anterior part of flagellar rows. (a-d) DIC; (e-f) SEM. Scale bars: (a) 10 µm, (b-e) 5 µm, (f) 1 µm (bar in b for b-d). □

gene datasets with a broad sampling of eukaryote taxa (initially 104 taxa; reduced to 61 taxa for computationally intensive analyses). The transcriptomes proved to be high-coverage (*Spironema*: 290/351 = 82.6% of genes and 77.6% of sites represented; *Hemimastix*: 280/351 = 79.7% of genes, 72.1% of sites). Maximum Likelihood (ML) analyses of both 104-taxa and 61-taxa datasets agreed with other recent phylogenomic studies (Burki et al. 2016, Brown et al. 2013, Zhao et al. 2012) in dividing previously known eukaryotes into three clans; Diaphoretickes, Discoba, and an 'Amorphea+' assemblage (Figures 4.3 and C.4). Diaphoretickes contained Sar (plus *Telonema*), Haptophyta plus Centrohelida, and Cryptista plus Archaeplastida (and Picozoa) as major subgroups. The 'Amorphea+' grouping contained Obazoa and Amoebozoa, as well as CRuMs, Ancyromonadida, Malawimonadida, and Meta-

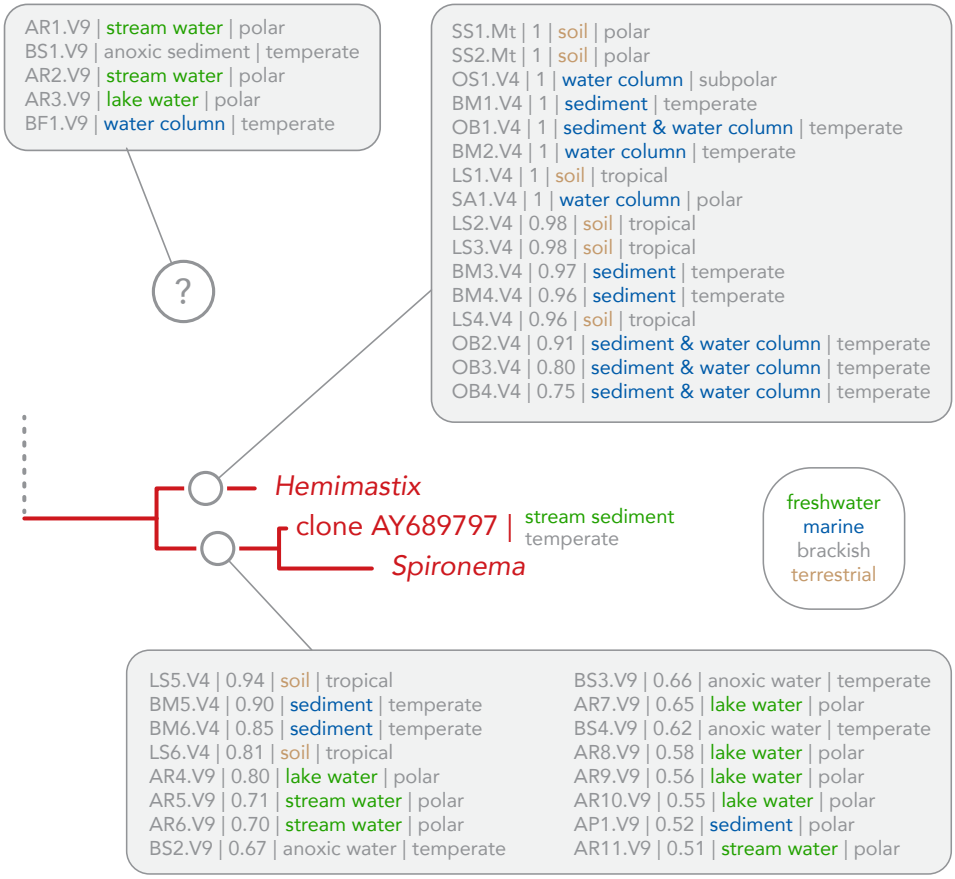

Figure 4.2: Environmental sequencing reads assigned to Hemimastigophora. For each read, the pplacer likelihood-to-weight-ratio (LWR), habitat, and environmental zone are reported. Reads with LWR > 0.5 are assigned to a branch. Five assigned sequences (circle with '?') were of uncertain placement within Hemimastigophora (i.e. LWR for any single branch within the clade was < 0.5, but the sum of all LWR=1). See Fig. C.3 for full reference tree. Table C.1 gives additional information on individual reads, including sample codes.

monada. The position of metamonads was unstable, mirroring conflicts amongst recent analyses (Brown et al. 2013, Cavalier-Smith et al. 2014).

Spironema and *Hemimastix* formed a maximally supported Hemimastigophora clade that was phylogenetically isolated. The 104-taxa analysis placed Hemimastigophora amongst the deepest branches within Diaphoretickes, sister to a clade of Sar, *Telonema*, haptophytes and centrohelids, though with equivocal support (ultrafast bootstrap approximation (UFboot) = 83%; Figures 4.4 and C.4). In the 61-taxa analysis Hemimastigophora again grouped with Diaphoretickes (bootstrap support (BS) = 100% (Posterior Mean Site Frequency method – PMSF); UFboot = 93%; Bayes-

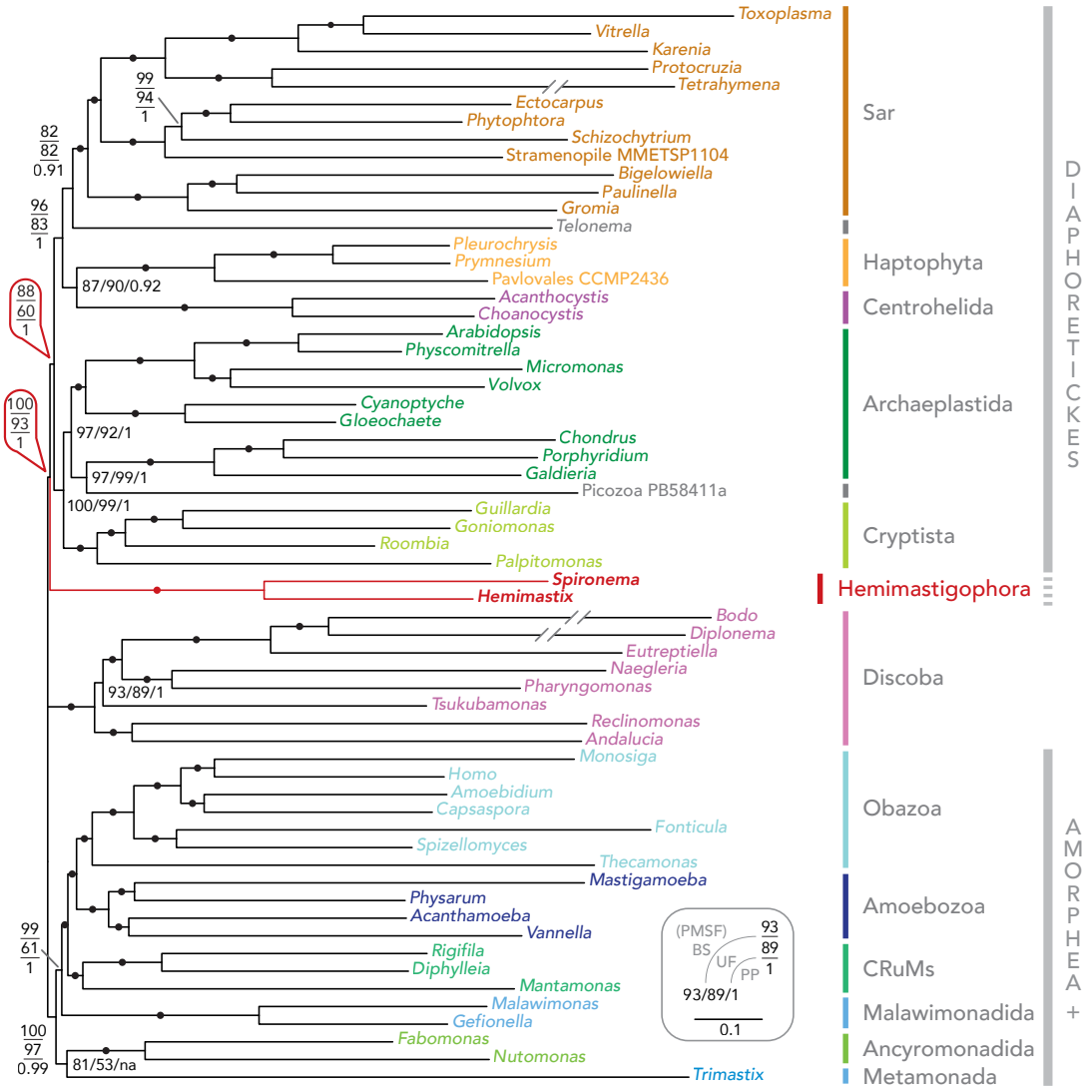

Figure 4.3: Phylogenetic placement of Hemimastigophora within eukaryotes. Unrooted phylogeny inferred from 351 genes and 61 taxa, using Maximum Likelihood under the LG+C60+F+Γ model. The numbers on branches show, in order: PMSF bootstrap percentages (BS; 200 true bootstrap replicates), ultrafast (UF) bootstrap approximation percentages (1000 replicates), and Bayesian posterior probabilities (PP; under the CAT+GTR model). Filled circles denote maximum support with all methods (i.e. 100/100/1). The three longest branches (leading to *Bodo, Diplonema and Tetrahymena*) are shown reduced by 1/3. Scale bar denotes 0.1 expected substitutions/site.

ian Posterior Probability (pp) = 1), but actually branched sister to all (other) Diaphoretickes, which formed a clade (BS = 88%; UFboot = 60%; pp = 1; Fig. 4.3).

To further explore the position of Hemimastigophora, we analysed several derivatives of the 61-taxa dataset that excluded potential sources of phylogenetic inaccuracy. Analyses that i) excluded the three taxa identified as outlier long-branches (58-nLB), or ii) excluded the three data-poorest taxa (site coverage < 30%; 58-nDP), or iii) recoded the amino acid data into four categories (61-SR4), all supported the same topology as the original 61-taxa analysis, that is Hemimastigophora outside of and sister to Diaphoretickes (Figures 4.3; C.5, C.6, C.7). Removing fast-evolving sites, however, did not systematically favour this tree over a topology where Hemimastigophora is sister to a Sar + *Telonema* + Haptophyta + Centrohelida clade (as in the 104-taxa analysis; Fig. C.8). Thus, while most analyses place Hemimastigophora as branching outside (other) Diaphoretickes, the alternative position where hemimastigotes fall one node inside Diaphoretickes remains credible (Fig. 4.4).

All previous proposals for the phylogenetic/systematic placement of Hemimastigophora were based on morphology alone. The sub-membranous thecal plates between the two flagellar rows suggested an affinity with euglenids, which have a pellicle (Foissner et al. 1988, Foissner & Foissner 1993). Later, affinities were proposed with completely different taxa presenting pellicular or thecal structures, namely alveolates (Cavalier-Smith 1998), or apusomonads and ancyromonads (Cavalier-Smith 2000). A placement within Rhizaria was also suggested, based on flagellum and extrusome substructure (Cavalier-Smith et al. 2008). None of these proposals is supported by our phylogenies, since Hemimastigophora is always distantly related to euglenids (Euglenozoa: Discoba), apusomonads and ancyromonads (both in Amorphea +), and Sar (which contains Alveolata and Rhizaria).

Instead, the extremely deep phylogenetic position of Hemimastigophora, most likely at the base of Diaphoretickes, implies that they represent a novel, suprakingdom-level lineage. This identifies hemimastigotes as a crucial group to include in descriptions of the tree of eukaryote life, and in most studies of the evolution of eukaryotic cells. This is especially important when inferring the history of eukaryotic innovations, or the nature of the last eukaryotic common ancestor (LECA), from the across-supergroup distributions of particular genes, genome characteristics, or cellular features (Speijer et al. 2015, de Mendoza et al. 2013, Fukasawa et al. 2017, Sebé-Pedrós et al. 2014, Barlow et al. 2018). Hemimastigotes may be equally important in the immensely challenging task of placing the root of the eukaryote tree. The root is usually inferred to lie somewhere between the largest eukaryote clans,

approximately positions a–c in Fig. 4.4 (He et al. 2014, Katz et al. 2012, Derelle & Lang 2012, Derelle et al. 2015), with position a (i.e. between Amorphea, and Diaphoretickes plus Discoba) currently most favoured (Derelle & Lang 2012, Derelle et al. 2015). Hemimastigophora appears to lie close to all of these positions (on the unrooted tree; see Fig. 4.4), and could be our only known representative of one of the most ancient divisions amongst extant eukaryotes. Accordingly, we searched the single-cell transcriptomes for genes that could have arisen during the divergences between supergroups (Table C.3; Fig. 4.4). Unexpectedly, we found in hemimastigotes several genes that are not known from Diaphoretickes, including myosin II (previously known from Amorphea, plus one subgroup of Discoba; Sebé-Pedrós et al. 2014, Richards & Cavalier-Smith 2005), and the Golgi protein GCP16/Golgin A7 (previously Amorphea-specific Barlow et al. 2018). The presence of such genes in hemimastigotes either pushes back/supports their likely origins before LECA or, more controversially, could be due to the root of eukaryotes being further from the base of Amorphea than generally supposed (Derelle et al. 2015, i.e. Amorphea and Hemimastigophora on the same side of the root; upper variant of 'position c' in Fig. 4.4). Yet, another hemimastigote myosin, for example, was previously unknown outside the Sar group (Fig. 4.4). Thus, irrespective of the final position of the root, this survey demonstrates how the antiquity of gene origins tends to be underestimated until all major lineages are considered. This bias can result in underestimation of the gene content of ancient eukaryotes and thus overestimation of the simplicity of their cell biology. Clearly, examining hemimastigote genomes, and ultimately cell biology, will be valuable for better understanding eukaryote evolution at the deepest levels.

This study is the first in which single-cell transcriptomics has unveiled a new deep-branching eukaryote lineage. Single-cell transcriptomics and genomics (Kolisko et al. 2014, Yoon et al. 2011, Gawryluk et al. 2016) bypass the 'culture bottleneck', and thus provide a rapid path to deeper taxon sampling, even when species from a group of interest are eventually cultivated. This is particularly valuable for phylogenomics, where inaccuracy due to poor taxon sampling is a perpetual concern (Keeling et al. 2014). For this application single-cell transcriptomics outperforms single-cell genomics because of better coverage of housekeeping genes (compare this study to e.g. Yoon et al. 2011, Gawryluk et al. 2016). Information on multiple related species is also valuable for ensuring data fidelity (e.g. detecting contaminants, gene transfers, etc; see methods). Single-cell techniques are especially promising for heterotrophic protozoa, which likely represent most 'undiscovered' major lineages, but

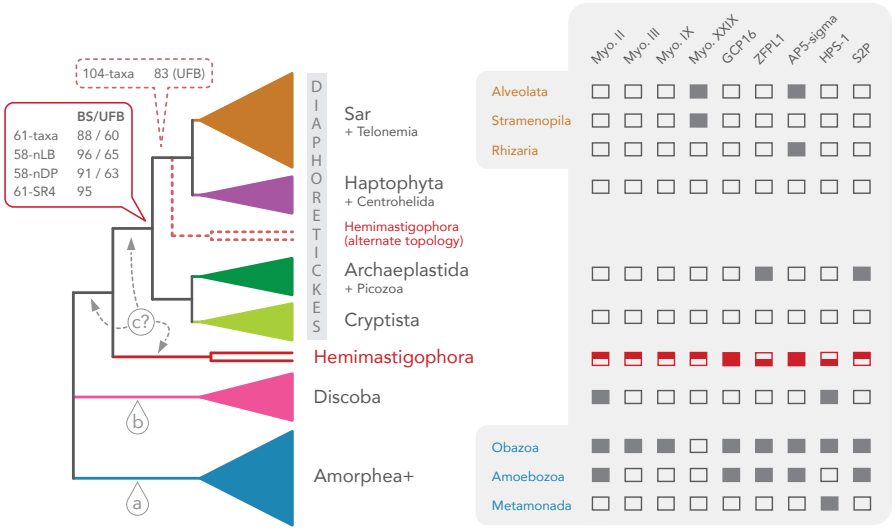

Figure 4.4: Summary of phylogenomic analyses and distribution of select genes across eukaryotes. Left panel shows inferred phylogenetic positions of Hemimastigophora. Solid box details support for Hemimastigophora as a deep branch relative to 'Diaphoretickes' supergroups in various analyses (BS: PMSF bootstrap support, except 61-SR4, where the GTR+R6+F model was used; UFB: Ultrafast bootstrap approximation support). Dashed box shows support for the alternative topology (Hemimastigophora a deep branch within Diaphoretickes) in the 104-taxa analysis. Stepwise fast-site removal analyses (61-SFSR; not displayed) equivocated between these alternatives (see Fig. C.8). Labels 'a', 'b', and 'c' show possible positions of the eukaryote root (see text), with the likely placement of Hemimastigophora implying several variants of 'position c'. Right panel maps known distributions of selected genes with proposed deep origins among living eukaryotes that were detected in hemimastigote transcriptomes (upper filled – in *Spironema*; lower filled – in *Hemimastix*); see Table C.3 for details.

where establishing cultures with suitable prey or hosts can be challenging (Kolisko et al. 2014, Gawryluk et al. 2016, Caron et al. 2016, Krabberød et al. 2017b).

In this first molecular phylogenetic investigation of Hemimastigophora, we show that they are a previously unrecognised supergroup of eukaryotes. Their phylogenetic distinctiveness is comparable to the whole animal plus fungi clade (Opisthokonta), or the assemblage containing all land plants and primary algae (Archaeplastida). We expect the discovery or recognition of other important lineages will greatly accelerate due to similar applications of single-cell methods.

4.3.1 Description of *Hemimastix kukwesjijk* Eglit and Simpson, n. sp. (ICZN)

Diagnosis: *Hemimastix* species, 16.5–20.5 μm long, with 17–19 flagella per row.

Type material: The name-bearing type (hapantotype) is an SEM stub mounting osmium-fixed sputter-coated material of strain BW2H, including trophic and dividing cells, deposited with the American Natural History Museum, New York, as AMNH _IZC 00267132. This material also contains prey *Spumella* sp. (Stramenopiles) and uncharacterised prokaryotes, both explicitly excluded from the hapantotype.

Type locality: Bluff Wilderness Trail, Nova Scotia, Canada (44.6610154 N, 63.767-4669 W); soil from mixed-species woodland.

Etymology: *kukwesjijk* (pronounced: Ku - Ga - Wes - Jij - K). 'Kukwes-' (Mi'kmaq): a rapacious, hairy ogre from the traditions of the Mi'kmaq First Nation of Nova Scotia. '-jijk' is diminutive (and pluralizing). The name 'little ogres' reflects the species' predatory and hairy nature. The use of Mi'kmaq language and tradition acknowledges the region where the species was isolated.

Gene sequence: The partial SSU rRNA gene sequence of strain BW2H is GenBank Accession MF682191.

Comments: Cells are larger and have several more flagella than *Hemimastix amphikineta*, the only previously described species (14x7 μm cell body; 12 flagella per row, Foissner et al. 1988).

4.4 Acknowledgements

The authors thank Ping Li and Patricia Scallion (Dalhousie University) for assistance with electron microscopy, Marlena Dlutek (Dalhousie University) for Illumina sequencing, Stefan Geisen (Wageningen University) for providing parsed metatranscriptomic data, Frédéric Mahé (CIRAD, Montpellier) for access to and parsing much of the V4 data, Matthew Brown (Mississippi State) for the seed phylogenomic dataset, Arnau Sebé-Pedrós (Weizmann Institute of Science) for the seed myosin alignments, Martin Kolisko (Institute of Parasitology, Czech Academy of Sciences) for data handling scripts, Bui Quang Minh (University of Vienna) for substantial help with phylogenomic analyses and troubleshooting in IQ-TREE, and Roger Lewis (Nova Scotia Museum) and Bernie Francis for advice on Mi'kmaq tradition and language. This work was supported by CIFAR, NSERC grant 298366-2014 to A.G.B.S., and NSERC grant 2016-016792 to A.J.R.

Chapter 5

Conclusion

5.1 Current State of Euglenid Research

5.1.1 Exploring the Biodiversity of Phagotrophic Euglenids

In Chapter 2 (Ploeotids) I have used a single-cell SSU rDNA approach combined with some culturing effort to sample the phylogenetic diversity of 'ploeotids', an assemblage of phagotrophs that, as it turns out, makes up much of the base of the euglenid tree. Chapter 3 used the same approach to sample Spirocuta, the large clade that includes both phototrophic and primary osmotrophic euglenids. Spirocuta are particularly morphologically diverse, are often highly metabolic, and express a variety of different gliding patterns, which is reflected in the phylogenies. The improved phylogenetic framework of euglenids showed that multiple morphologically defined genera are likely not monophyletic, and should in fact be considered to be split into separate genera (e.g. *Heteronema, Dinema, Anisonema, Ploeotia*). Three new genera *Olkasia, Liburna*, and *Hemiolia* were established in the cases were there was a clear disagreement between a traditional genus and the phylogenetic analyses (supported by scanning electron microscopy studies in the case of *Olkasia*). While the SSU rDNA trees with improved taxon sampling—as expected—failed to resolve a tree of euglenids, work in Chapter 2 and 3 substantially increased molecular sequences available for phagotrophic euglenids, helped resolve some standing taxonomic questions, and provided a base to inform sampling for multigene-analyses.

5.1.2 Large-Scale Euglenid Evolution

Understanding the relationship among euglenids is an ongoing task, and so far has been constricted to inferences by morphology, or the SSU rDNA gene which is often divergent in euglenids and poorly resolves their phylogeny (Paerschke et al. 2017, Lax & Simpson 2013, Busse et al. 2003). To further our understanding of euglenid evolution, multigene-analyses are clearly needed. A single-cell transcriptomics approach was used to generate multigene-data, and cell selection for this was heavily informed by the phylogenetic diversity identified with the SSU rDNA phylogenies. Several phylogenomic analyses show that phagotrophic species form the base and much of the 'crown' of euglenids, and fall into six distinct clades. Spirocuta are robustly recovered as monophyletic, while ploeotids are paraphyletic. One ploeotid, *Olkasia*, falls sister to Spirocuta with high support. Three ploeotid groups form a well supported clade (Ploeotiidae, *Lentomonas, Keelungia*), whereas the positions of *Liburna* and *Entosiphon* are still unclear. Spirocuta can be subdivided into highly supported clade 'Anisonemids plus', and a poorly supported, but potentially monophyletic clade 'Peranemids'. The basal position of petalomonads to all other euglenids is well supported, and is corroborated by some past morphological studies of the group. Multigene analyses that include two closely related symbiontids suggest that they are not derived euglenids, but an otherwise independent branch of Euglenozoa.

5.1.3 Hemimastigophora

Lastly, Chapter 4 explored Hemimastigophora, a long-known group of enigmatic protists which was never molecularly characterised. A single-cell transcriptomics approach enabled generation of data that populated most sites of a large phylogenomic dataset, and hemimastigotes were placed among eukaryotes: They do not appear to be specifically related to any other known supergroup, but instead form their own independent branch among eukaryotes. This work can be considered a useful example of how single-cell transcriptomics can be used in generating molecular data from unknown or unsampled protists and placing them within the tree of eukaryotes, even if they are extremely phylogenetically distinct from any previously studied forms.

5.2 Wider Impact on Euglenid/Eukaryote Research

5.2.1 Impact on Our Understanding of Euglenids

Phagotrophic euglenids have been known since at least the mid 19th century, yet have not been characterised molecularly to the same degree as their phototrophic cousins. This thesis aimed to provide a more complete picture of euglenid diversity by using SSU rDNA gene phylogenies. In fact, when compared to the most comprehensive recent sampling (Schoenle et al. 2019), I increased the number of phagotrophic euglenids from roughly 30 to 141 sequences, almost a five-fold increase. I have successfully used molecular single-cell methods like multiple-displacement amplification and emerging methods like single-cell transcriptomics to approach this subject, combined with a strong focus on capturing high-quality imagery of isolated cells. This link between molecular sequence(s) and images proved to be crucial, as it provides us the means of, for example, testing taxonomic assignments among phagotrophic euglenids. There are several examples of inconsistency between morphologically defined genera and phylogenetic clades in euglenids, and an ever increasing SSU rDNA dataset with associated morphological data is what is needed to resolve these issues. I hope that this research provides the beginnings of a phylogenetic framework that can—among many other questions—be used to test taxonomic assignments, evolutionary hypotheses (e.g. pellicle strip inheritance patterns), and how the secondary plastid in phototrophic euglenids was acquired. Lastly, understanding euglenid evolution is relevant to understanding the evolution of symbiontids, and is crucial to understanding the evolution of Euglenozoa in general.

5.2.2 Remaining Issues in Euglenids

So what is left to do in biodiversity and evolution research on phagotrophic euglenids? Perhaps unsurprisingly, many mysteries remain; for example, for multiple genera there are still no molecular data available at all (e.g. *Dolium, Dylakosoma*). Similarly, many morphospecies in already sampled genera remain unsequenced: *Dinema, Jenningsia, Urceolus*, and *Heteronema* currently all appear para- or polyphyletic, but with only two to five species sequenced for each genus, it is unclear to what extent this is due to an unresolved tree (i.e. missing sequences), or a reflection of their true phylogenetic placement. It is crucial to have sequences of the type species available, as only then can well-informed taxonomic decisions on which taxon subgroup (in the case of para- or polyphyletic taxa) will keep the

original genus name, be made. In a similar vein, multigene data for euglenids can still be much improved. For one, not all known genera are sampled in the analyses (e.g. *Teloprocta* and *Hemiolia* are missing). Furthermore, some of the transcriptomes included are not as high-quality as others, which is especially true for the two symbiontids included. Nevertheless, the transcriptomic data produced as part of this thesis will be critical in investigating cell biological, ecological, and evolutionary questions among euglenids. This might include exploring the molecular basis of motility across euglenids, the distribution of metabolic pathways, or understanding genes and proteins involved in pellicle structure and formation. The distribution and ecology of euglenids are barely known, and the SSU rDNA dataset will enable more accurate placing of any future environmental sequences among euglenids. This in turn can provide a starting point for asking more in-depth questions about their ecological roles in benthic ecosystems.

5.3 Are Single-Cell Approaches the Future?

A large focus of my thesis work has been on the adoption of existing molecular single-cell methods. This was largely borne out of necessity, as phagotrophic euglenids tend to be hard to culture, and these approaches provided a viable workaround. Single-cell approaches can be powerful—and they will only keep getting more powerful as technology progresses. For the foreseeable future, a transcriptomic approach will be key, as it will get easier to sequence cells. Additionally assembly of eukaryotic transcriptomes is already much more straightforward than assembling single-cell genomes (see Keeling & del Campo 2017). Still, single-cell approaches need to be used the right way, which depends on the questions asked. One could isolate cells en masse with an automated approach and generate hundreds of single-cell transcriptomes. This approach would certainly generate a lot of data and could be extremely useful to answer certain ecophysiological questions (see Liu et al. 2017), but misses the chance to capture and investigate morphological diversity. Protists are inherently morphologically diverse, and this is often reflected in their phylogenetic diversity. Even with a completely manual approach as was used in this thesis, many taxa can be sequenced quickly, all while generating high-quality imagery. For investigating phylogenetic relationships among and within protist groups, I consider the manual approach more useful.

I firmly believe—in the long run—cultures are equally, if not more important than single-cell approaches. Long-term cultures of protists provide a chance to study

organisms in-depth, including growth patterns, prey preference, and ultrastructure, among many other questions. This is crucial to understand their basic biology, and we cannot generate this data with any currently existing single-cell methodology. Investigating single-cells and investigating cultures essentially fulfill two different, but equally important, roles: The former is a quick way to sample diversity and to generate a framework, whereas the latter is slow, but instrumental in understanding the biology of an organism.

Appendix A

Ploeotids Represent Much of the Phylogenetic Diversity of Euglenids

Table A.1: Organism codes with assigned taxa used, GenBank accession codes and euglenid-biased SSU primer sequences and combinations applied in this study.

Electronic supplement: Deposited on DalSpace

Table A.2: Morphological measurements of isolated single cells.

Electronic supplement: Deposited on DalSpace

Table A.3: Sequence similarities of *Olkasia*, *Serpenomonas*, *Lentomonas*, and *Entosiphon*.

Electronic supplement: Deposited on DalSpace

Table A.4: Support values for euglenid and ploeotid groups in all analyses.

Electronic supplement: Deposited on DalSpace

Table A.5: Sampling sites of cultures and samples from this study.

Electronic supplement: Deposited on DalSpace

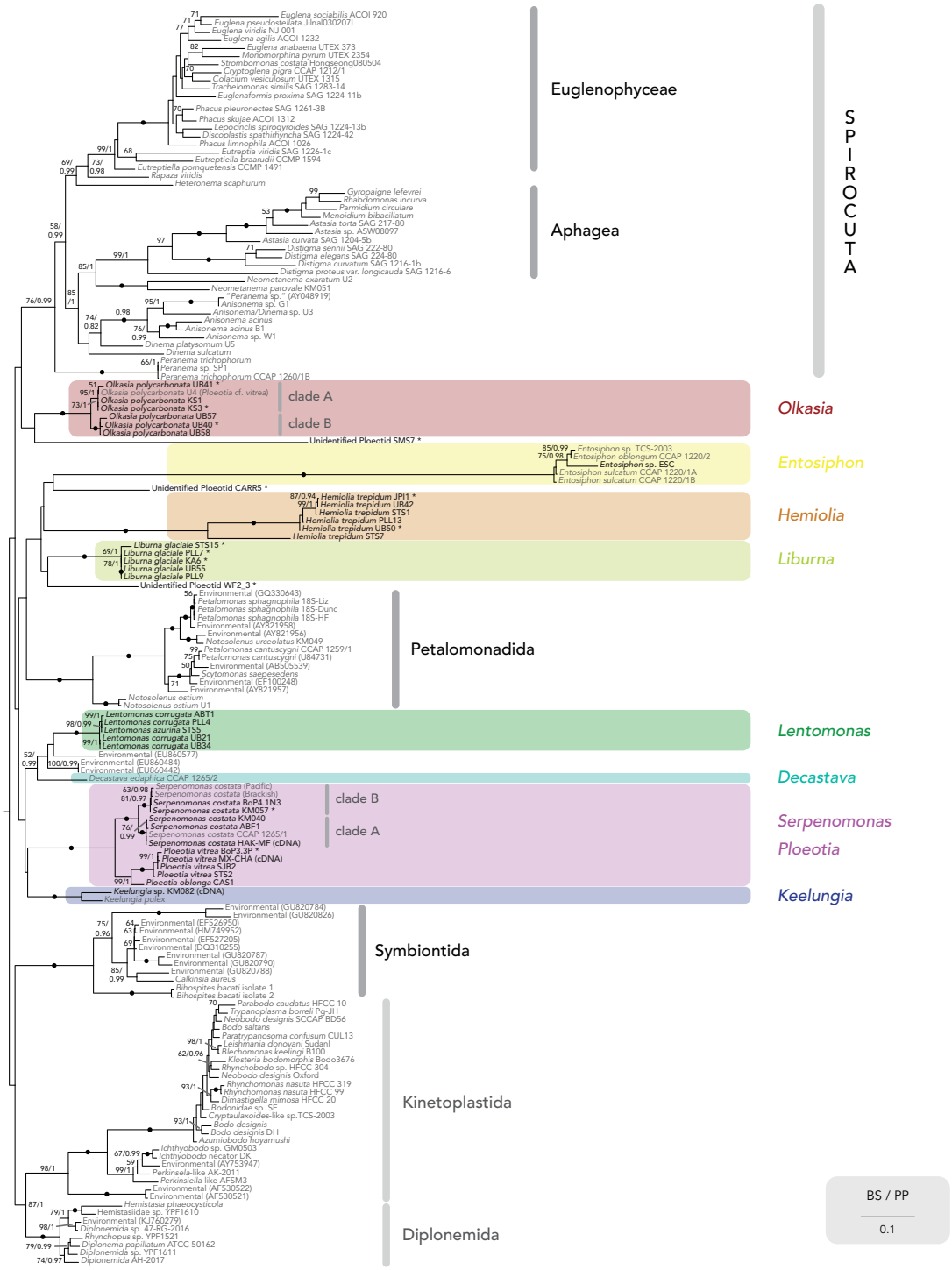

Figure A.1: Maximum Likelihood phylogeny of the SSU-rRNA gene of euglenids, including *Entosiphon*. Bootstrap values were derived under the GTR+Γ model, as were posterior probabilities. Major groups of ploeotids are shown, with sequences acquired in this study bolded, an asterisk (*) on taxa denoting partial sequences. This phylogeny is outgroup-rooted on diplonemids and kinetoplastids. Maximum bootstrap support (100%) and posterior probability (pp of 1) is shown with a black filled circle, values below 50% and 0.9 pp not shown. This is the same analysis as summarized in Figure 2.8

Appendix B

Towards a Resolved Tree of Euglenids: A Single-cell and Phylogenomics Approach

Table B.1: Explanation of sample site codes, with location, coordinates and material sampled.

Electronic supplement: Deposited on DalSpace

Table B.2: SSU rDNA primer sequences used in this study, as well as primer combinations used for individual cells.

Electronic supplement: Deposited on DalSpace

Table B.3: Sequencing information, assembly strategies, BUSCO-scores and transrate statistics for single-cell and bulk-transcriptome assemblies, and single-cell genomes.

Electronic supplement: Deposited on DalSpace

Table B.4: Overview over taxon sampling for multigene-analyses, with 20-gene availability for all taxa and sources for external assemblies.

Electronic supplement: Deposited on DalSpace

Table B.5: Isolated euglenid cells, with identified morphotype, associated clades/groups, basic measurements (length, width, length of flagella) and locomotion information.

Electronic supplement: Deposited on DalSpace

Table B.6: Support values for important bipartitions in SSU rDNA and multigene-phylogenies, including values used in Supplementary Fig. B.12.

Electronic supplement: Deposited on DalSpace

Videos B.7: Videos of representatives of phagotrophic euglenid groups/clades, as outlined in Figures 3.1, 3.2, 3.3, 3.4, and 3.5.

Electronic supplement: Deposited on DalSpace

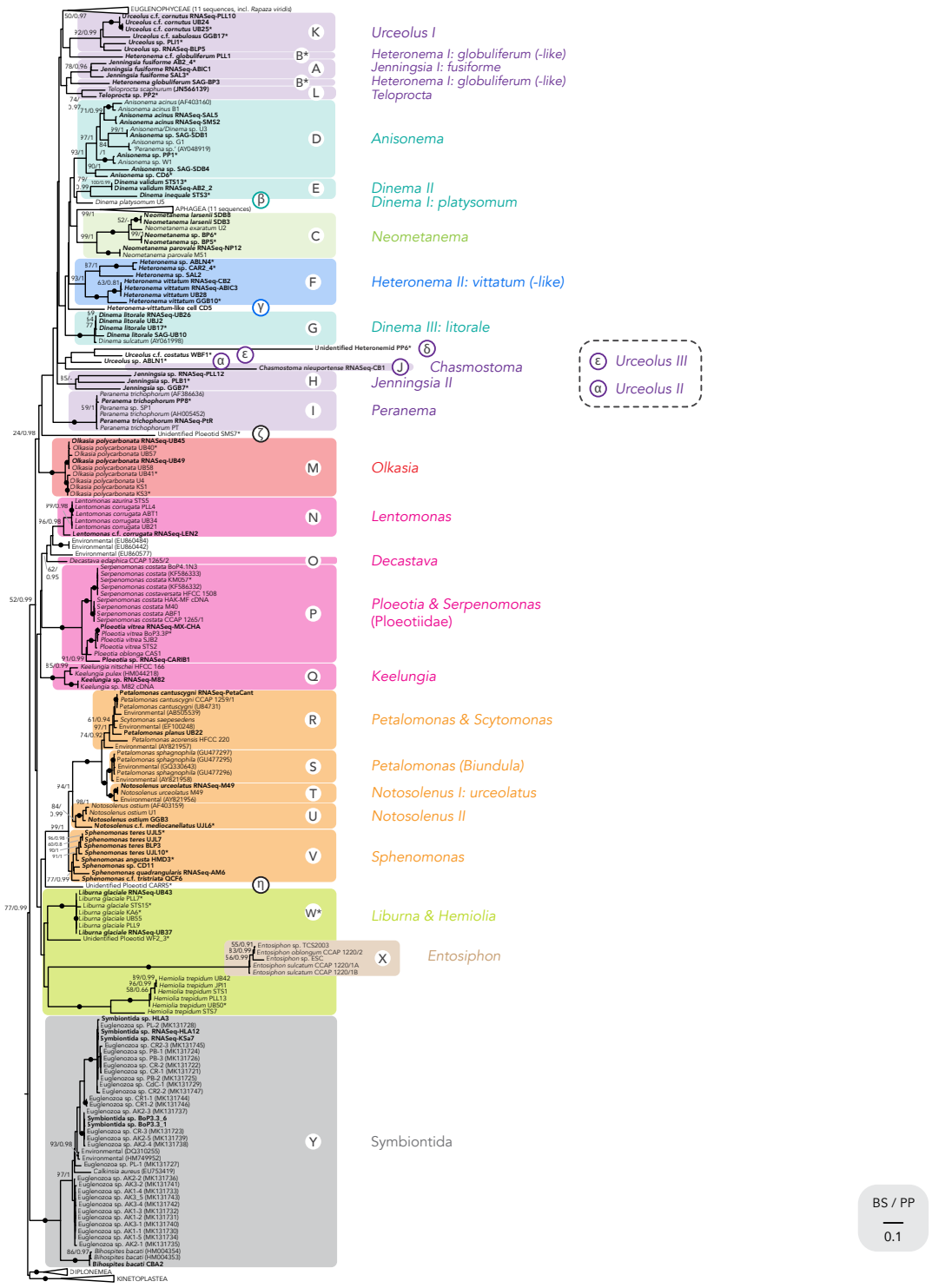

Figure B.1: Maximum Likelihood phylogeny of the SSU rDNA gene of Euglenozoa (including five long-branching *Entosiphon* sequences, Unidentified Heteronemid PP6, and *Chasmostoma nieuportense* CB1), estimated under the GTR+Γ model with 1000 bootstrap replicates ('SSU-LB'). Posterior probabilities were derived from the same model. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) and posterior probabilities (pp of 1) are denoted by a filled circle. Support values below 50% and 0.9 pp are not shown.

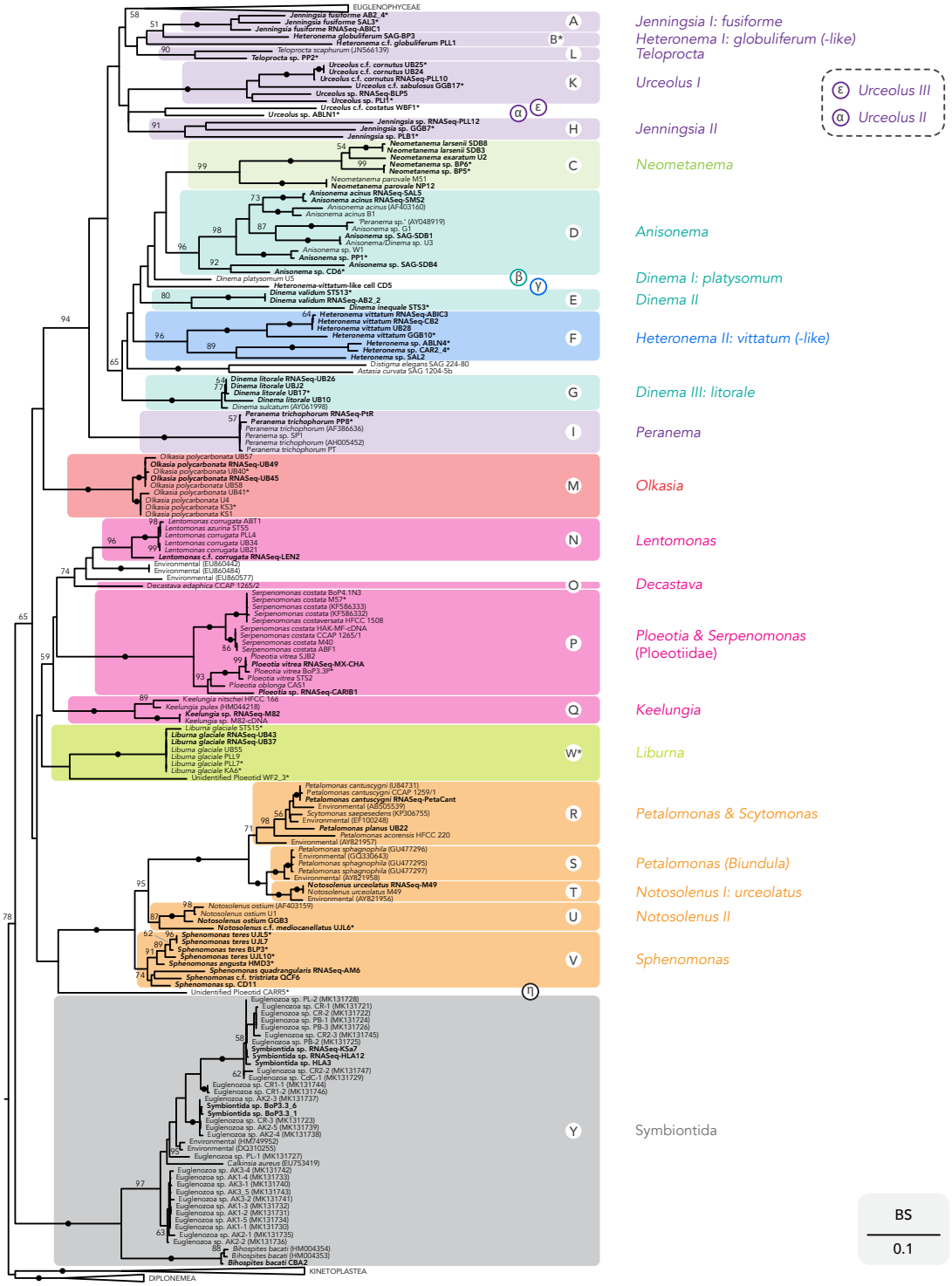

Figure B.2: Maximum Likelihood phylogeny of the SSU rDNA gene of Euglenozoa with long-branching taxa removed ('SSU-noLB'). Estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) are denoted by a filled circle, while support values below 50% are not shown.

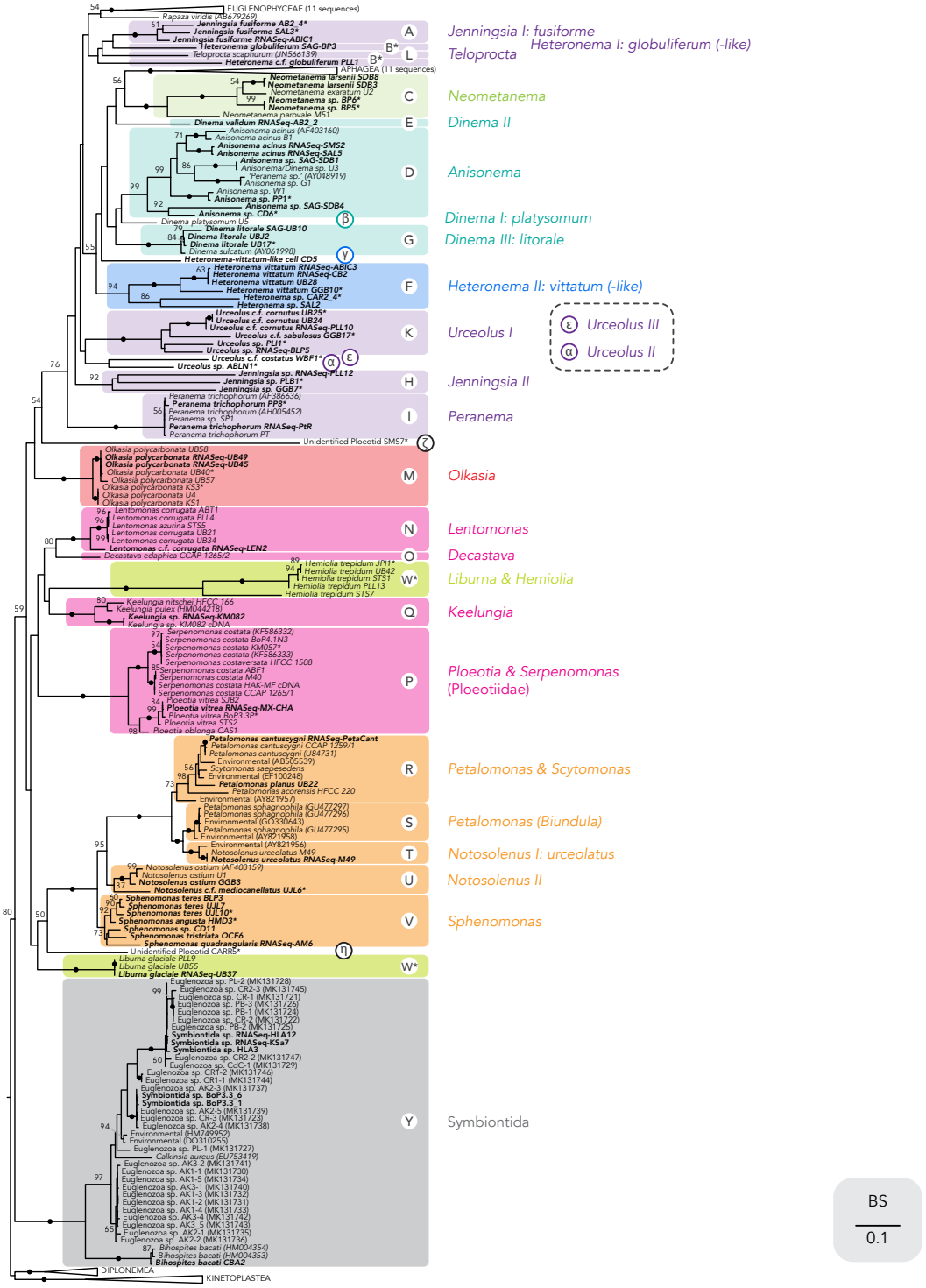

Figure B.3: Maximum Likelihood phylogeny of the SSU rDNA gene of Euglenozoa, based on 'SSU-base' (Fig. 3.1), with short sequences under 1000 bp removed ('SSU-noShort'). Estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) are denoted by a filled circle, while support values below 50% are not shown.

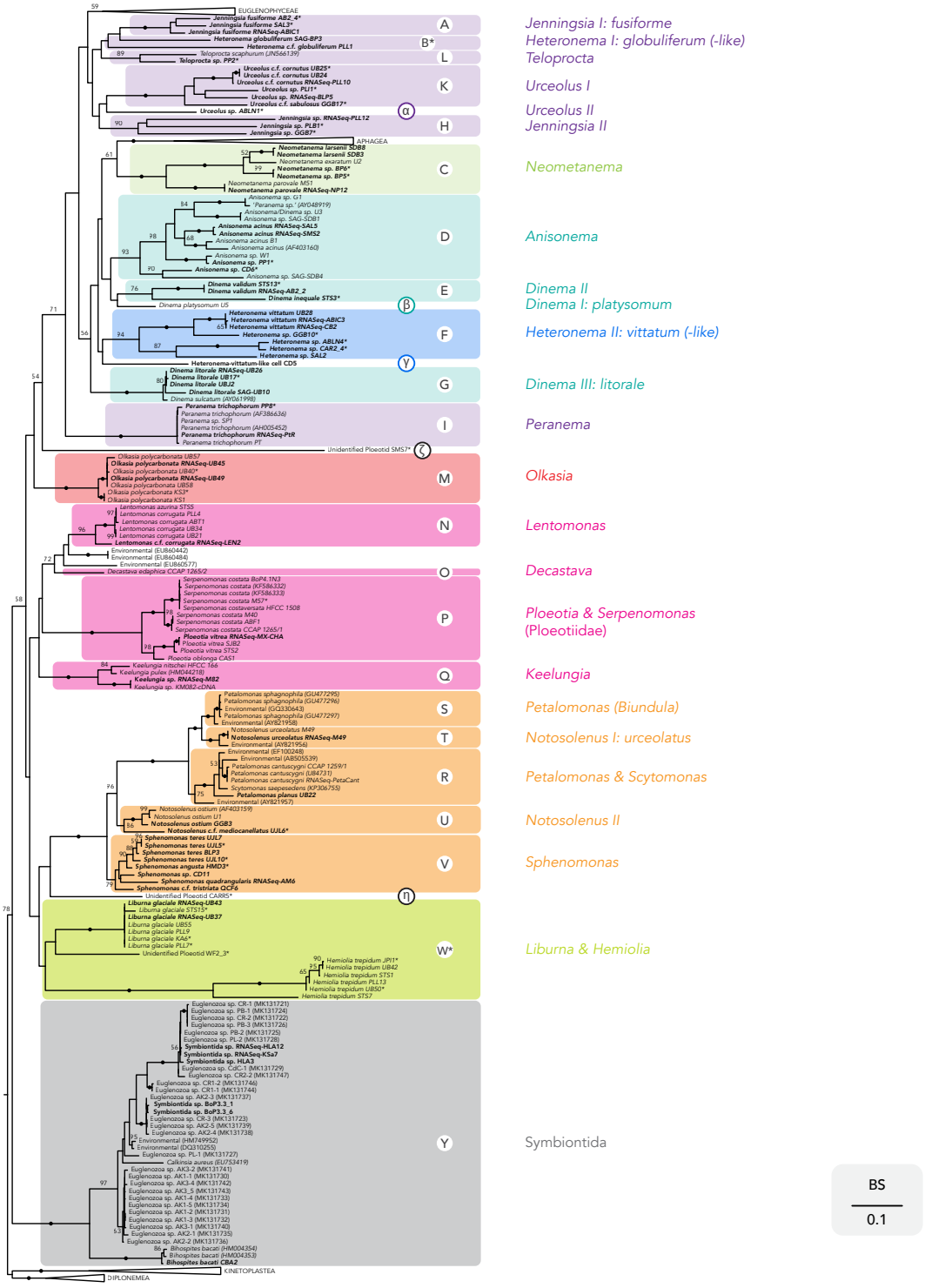

Figure B.4: Maximum Likelihood phylogeny of the SSU rDNA gene of Euglenozoa, based on 'SSU-base' (Fig. 3.1), but rogue taxa identified by RogueNaRok removed ('SSU-noRogues'). Estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) are denoted by a filled circle, while support values below 50% are not shown.

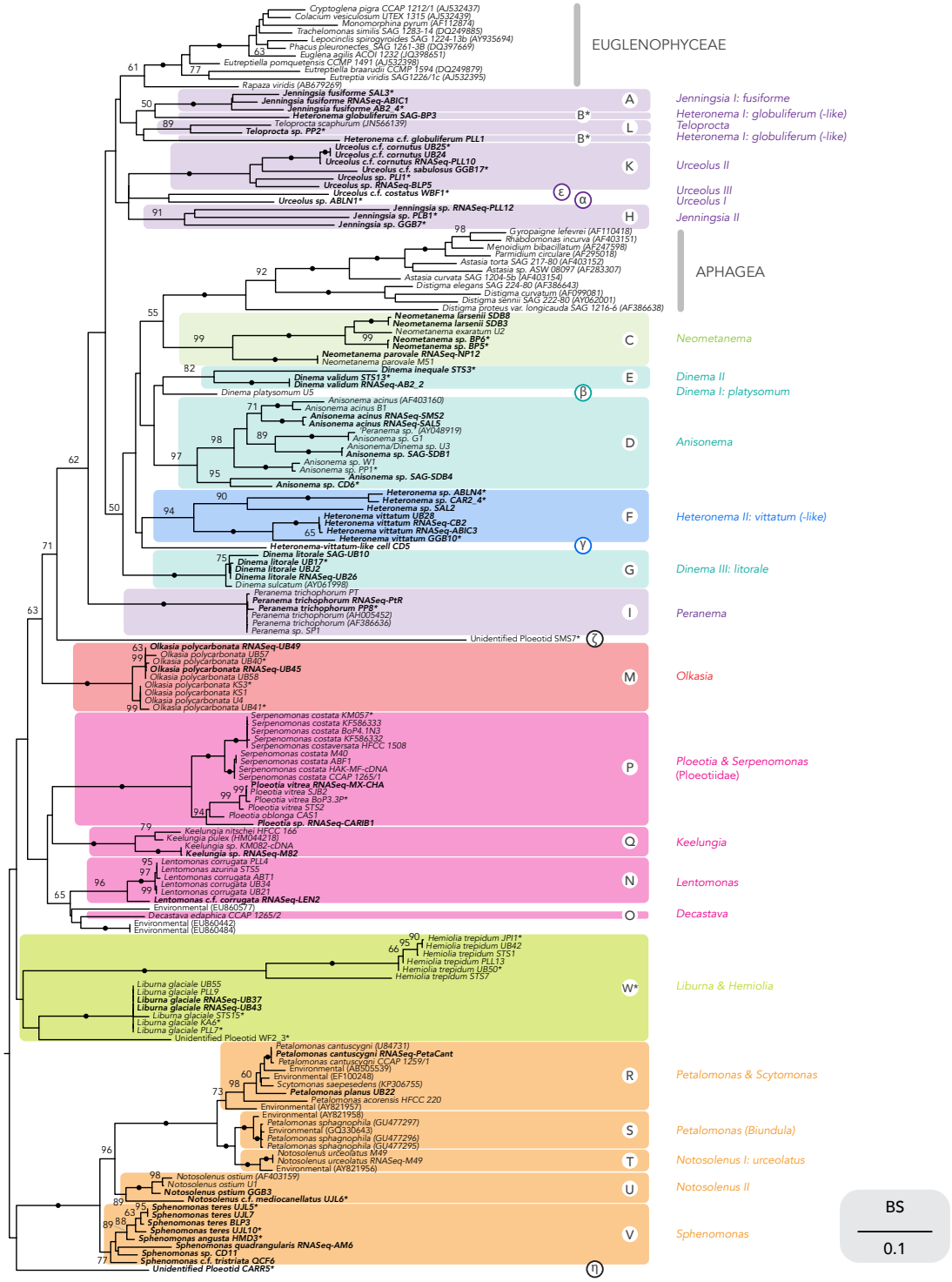

Figure B.5: Maximum Likelihood phylogeny of the SSU rDNA gene of euglenids ('SSU-Euglenids'). Estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) are denoted by a filled circle, while support values below 50% are not shown.

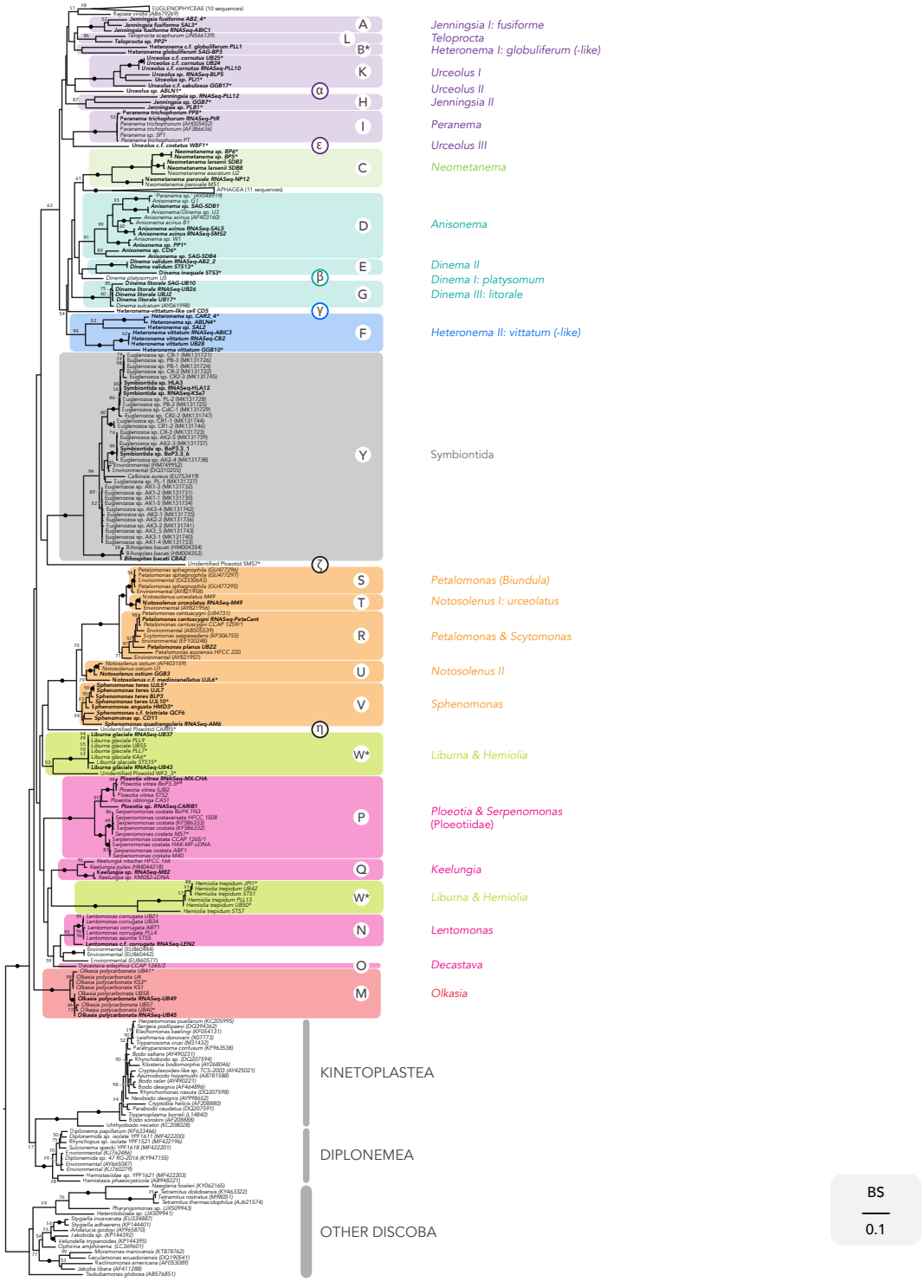

Figure B.6: Maximum Likelihood phylogeny of the SSU rDNA gene of Discoba ('SSU-Discoba'). Estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel sequences from this study are bolded and nodes receiving maximum support for bootstraps (100%) are denoted by a filled circle, while support values below 50% are not shown.

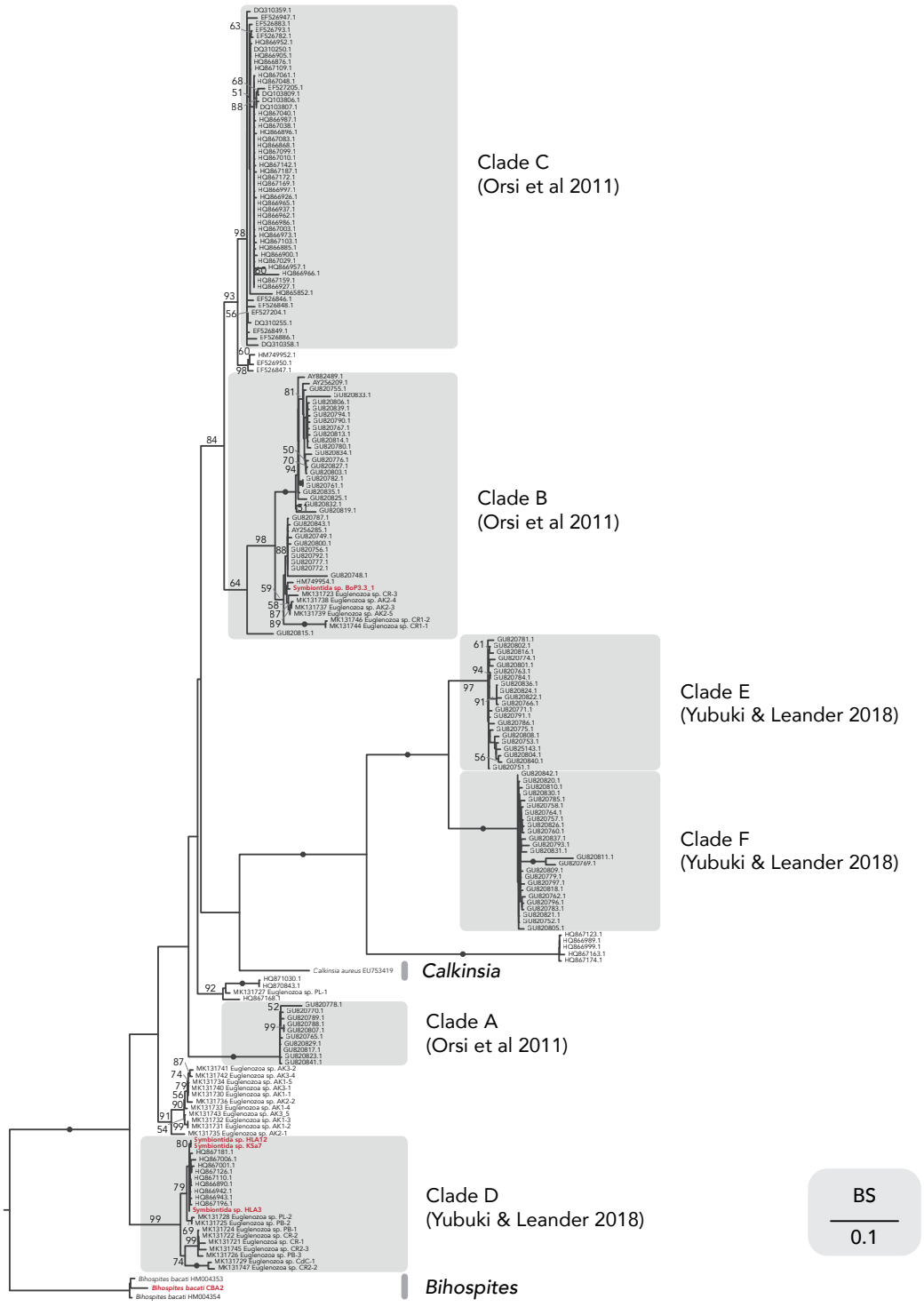

Figure B.7: Maximum Likelihood phylogeny of the SSU rDNA of Symbiontids ('SSU-Symbiontids'), estimated under the GTR+Γ model with 1000 bootstrap replicates. Novel symbiontid sequences are in red and bolded, and environmental clades established by Orsi et al. (2011) and Yubuki & Leander (2018) are marked. This phylogeny is rooted based on Yubuki & Leander (2018). Nodes receiving maximum support (100%) are denoted by a filled circle, whereas support values below 50% are not shown.

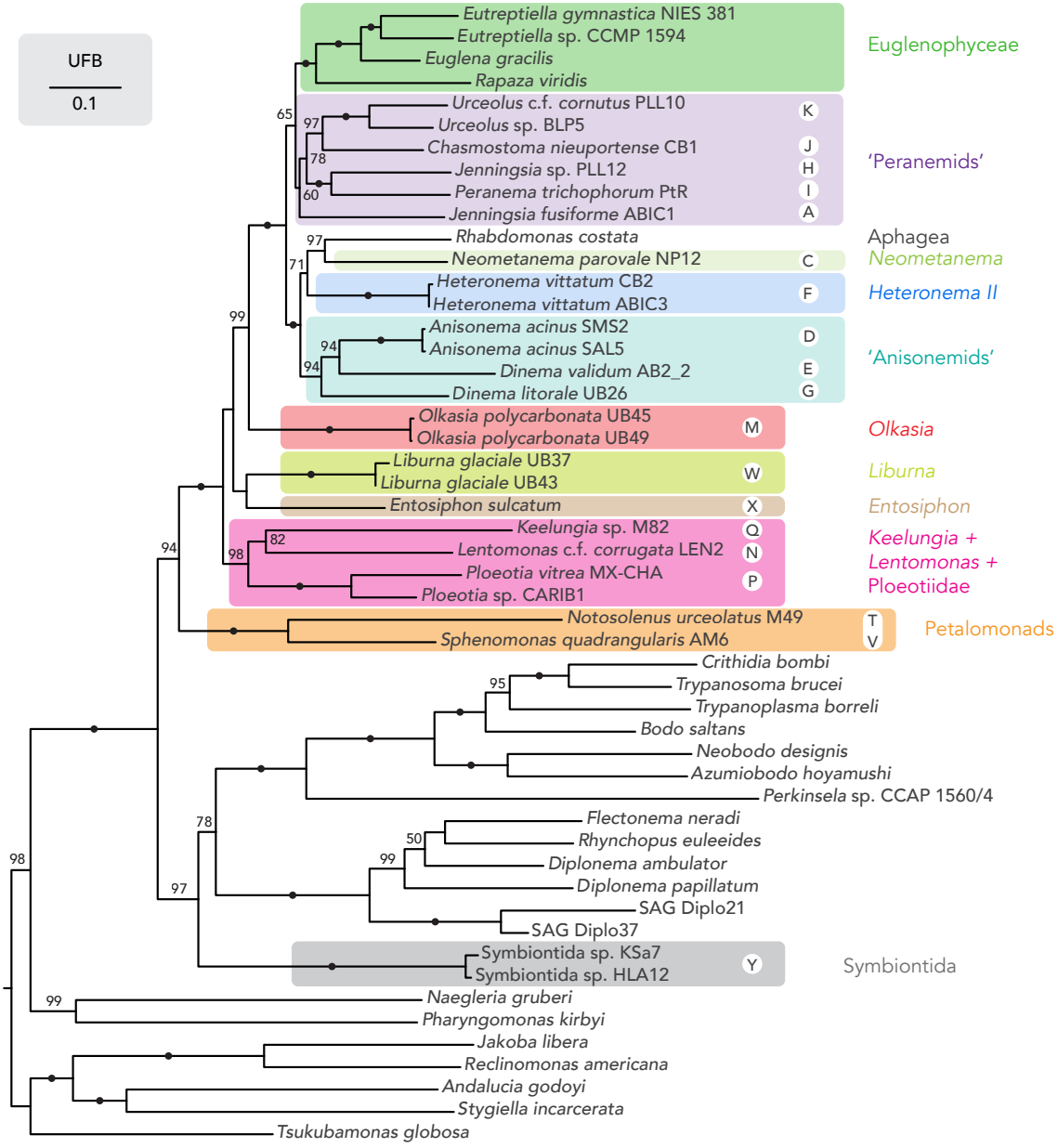

Figure B.8: 20-gene phylogeny of Discoba with long-branching taxa removed (*Petalomonas cantuscygni, Percolomonas cosmopolitus*, SAG D1, and *Sawyeria marylandensis*, 'noLB'). Estimated in the Maximum Likelihood framework under the LG+C60+F+Γ model, with robustness assessed by 1000 Ultrafast Bootstrap replicates (UFB). Major clades of euglenids are shown in coloured boxes. Nodes receiving maximum support for both bootstrapping methods (100%) are denoted by a filled circle, and support values below 50% are omitted.

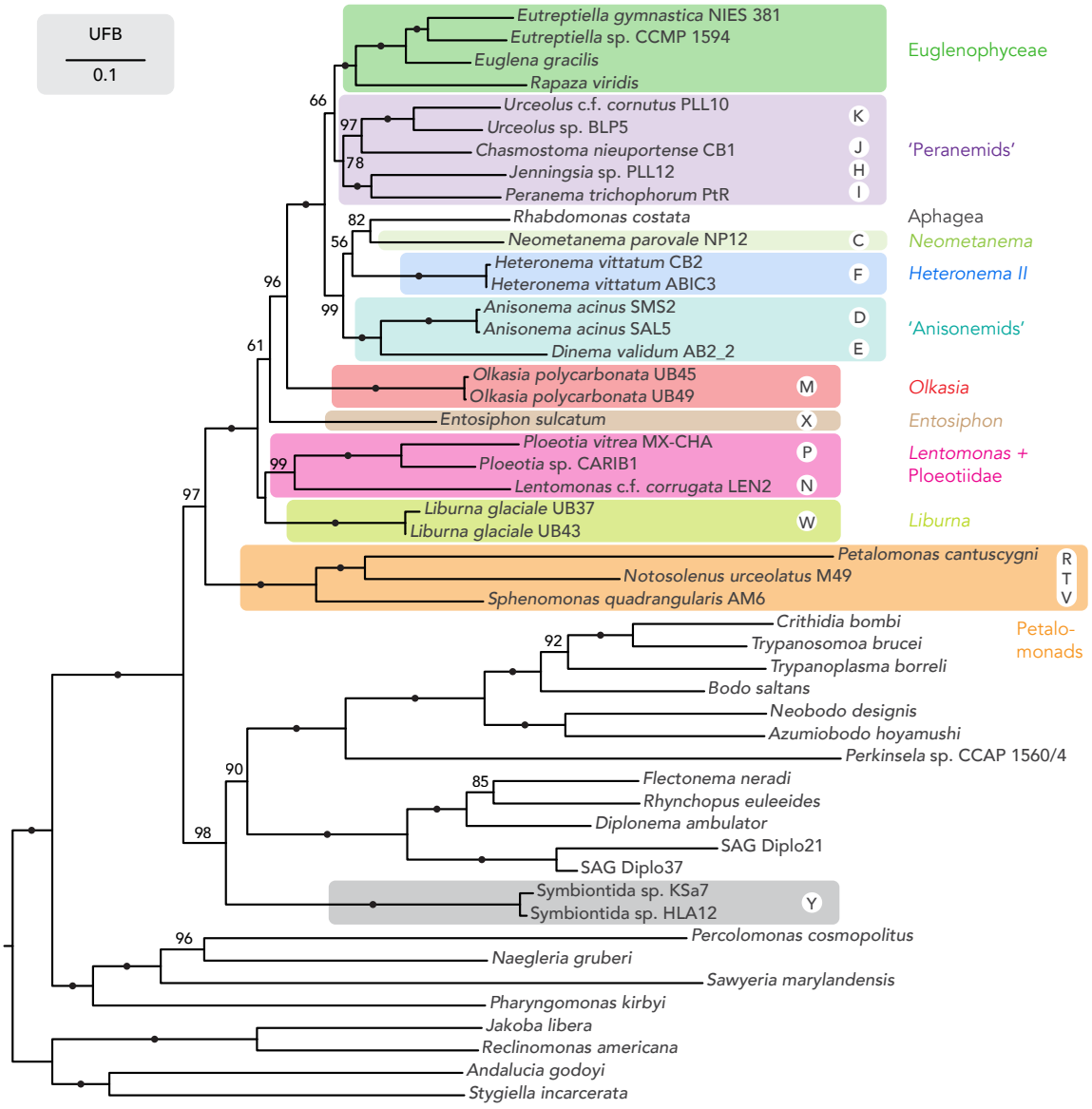

Figure B.9: 20-gene phylogeny of Discoba with rogue taxa removed ('noRogues', identified by RogueNaRok: SAG D1, *Diplonema papillatum, Tsukubamonas globosa*, *Keelungia* sp. M82, *Jenningsia fusiforme* ABIC1, and *Dinema litorale* UB26). Estimated in the Maximum Likelihood framework under the LG+C60+F+Γ model, with robustness assessed by 1000 Ultrafast Bootstrap replicates (UFB). Major clades of euglenids are shown in coloured boxes. Nodes receiving maximum support for both bootstrapping methods (100%) are denoted by a filled circle, and support values below 50% are omitted.

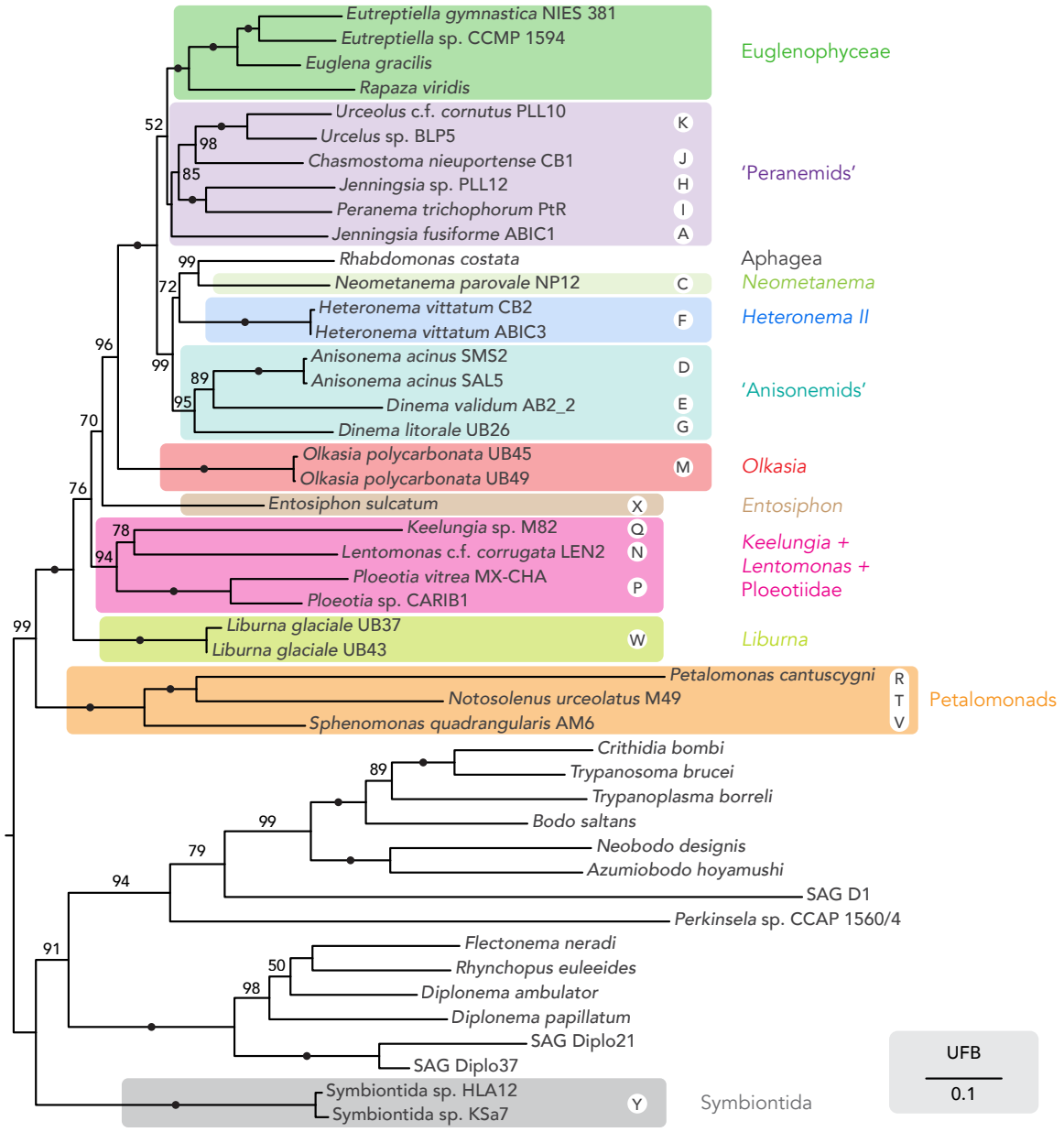

Figure B.10: 20-gene phylogeny of Euglenozoa ('Euglenozoa-only'), estimated in the Maximum Likelihood framework under the LG+C60+F+Γ model, with robustness assessed by 1000 Ultrafast Bootstrap replicates (UFB). Major clades of euglenids are shown in coloured boxes. Nodes receiving maximum support for both bootstrapping methods (100%) are denoted by a filled circle, and support values below 50% are omitted.

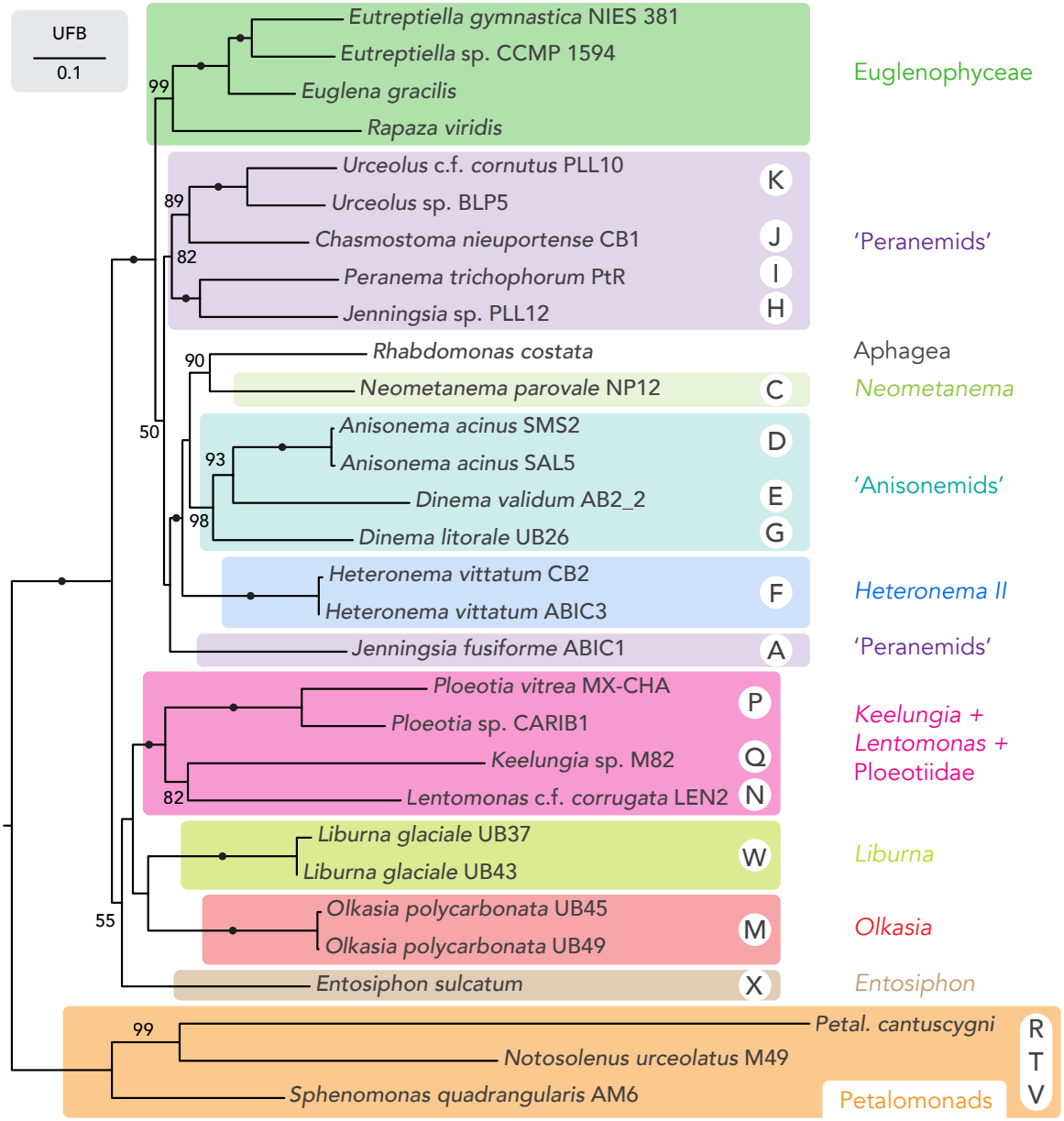

Figure B.11: 20-gene phylogeny of Euglenids ('Euglenida-only'), estimated in the Maximum Likelihood framework under the LG+C60+F+Γ model, with robustness assessed by 1000 Ultrafast Bootstrap replicates (UFB). Major clades of euglenids are shown in coloured boxes. Nodes receiving maximum support for both bootstrapping methods (100%) are denoted by a filled circle, and support values below 50% are omitted.

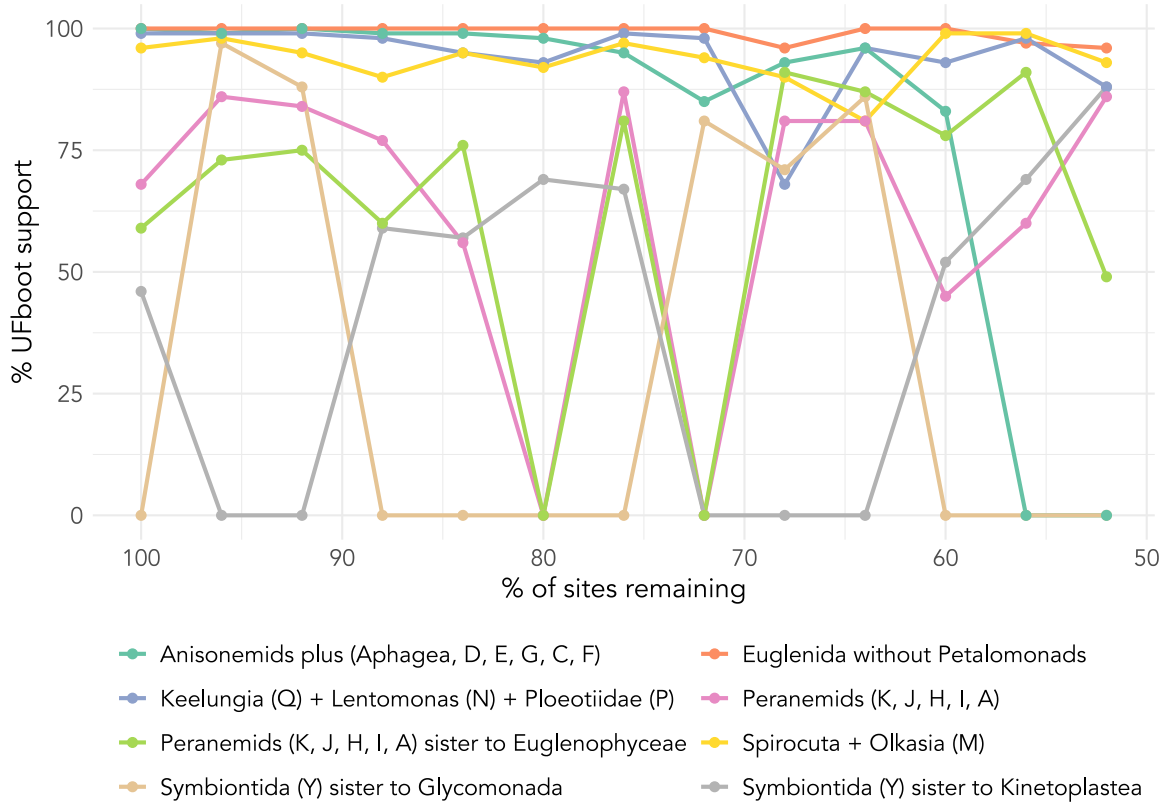

Figure B.12: Graph summarizing support for important bipartitions with progressive removal of fast-evolving sites from our 20-gene base dataset ('FSR'). Support values are ultrafast bootstraps (1000 replicates; UFboot) inferred under the LG+C20+F+Γ model. Raw values can be found in Supplementary Table B.6.

Appendix C

Hemimastigophora is a Novel Supra-kingdom-level Lineage of Eukaryotes

Table C.1: Full listing of environmental sequences attributable to Hemimastigophora, with habitat and location data.

Electronic supplement: Deposited on DalSpace

Table C.2: Taxa used in phylogenomic analyses, organized by major group, with gene- and site-coverage statistics, and sources of data identified.

Electronic supplement: Deposited on DalSpace

Table C.3: Genes of potential deep evolutionary significance in eukaryotes, searched for in the single-cell transcriptomes of *Spironema* and *Hemimastix*.

Electronic supplement: Deposited on DalSpace

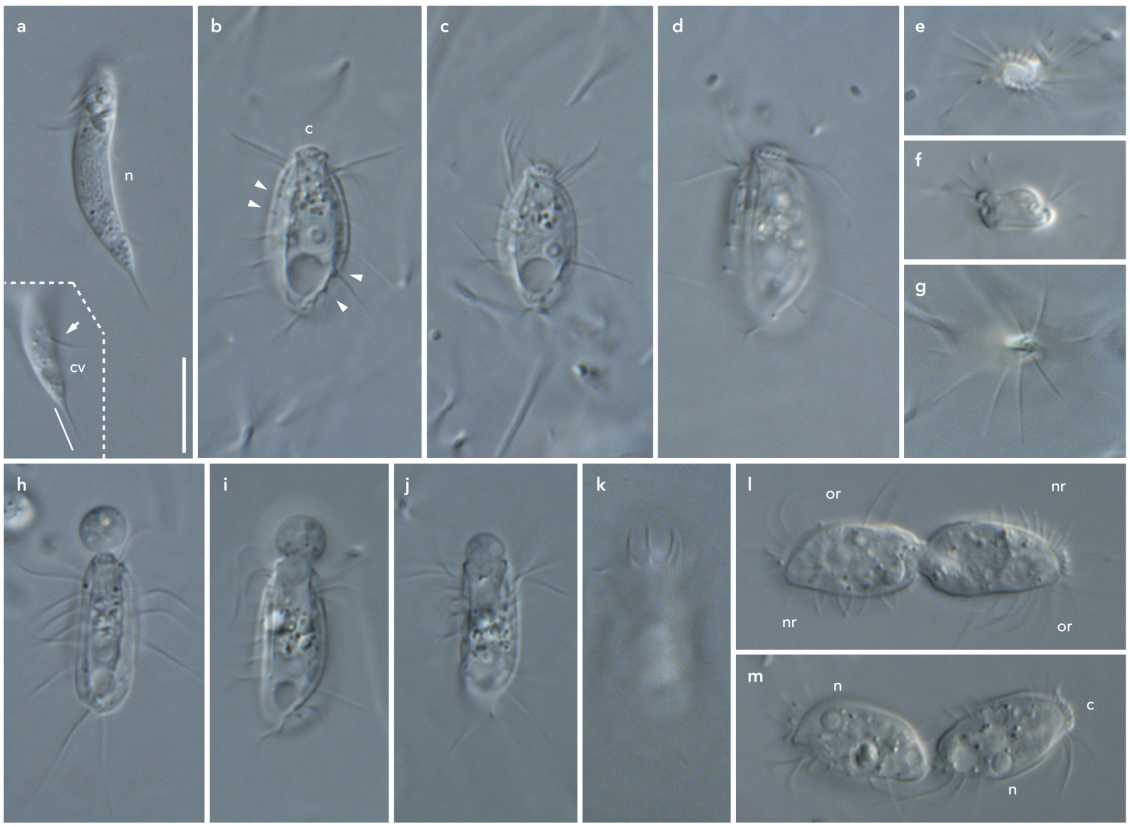

Figure C.1: *Spironema* c.f. *multiciliatum* (a) and *Hemimastix kukwesjijk* (b–m), DIC light micrographs of live cells. (a) Two views of a *S.* c.f. *multiciliatum* cell, with inset detailing the posterior end. Note nucleus (n), detail of one of the posterior flagella (inset: arrow) and small contractile vacuole (inset: cv), as well as posterior spike/tail (inset: line). (b–c) Optical sections through one *H. kukwesjijk* cell, detailing notches from which flagella emerge (arrowheads), a section through the capitulum (c) and a conspicuous contractile vacuole in the cell posterior (in b). (d) Surface view of one of the two thecal plates. (e–g) Optical cross sections of different cells showing the capitulum (e), mid-body region with rotationally symmetrical plate overlap (f) and the posterior (g) with radial arrangement of the posterior-most flagella. (h–j) Pseudoseries illustrating the feeding process, showing progression of prey ingestion stages. Note forming phagocytic vacuole (in i, asterisk) and widened capitulum (in j). (k) Same cell as in j, showing the anterior flagella curving forward to surround prey (seen especially in early feeding). (l, m) Dividing cells. showing the diagonal symmetry of short new rows (nr) and longer old rows (or) of flagella, as well as the daughter nuclei (n). Scale bar: 10 µm for all figures.

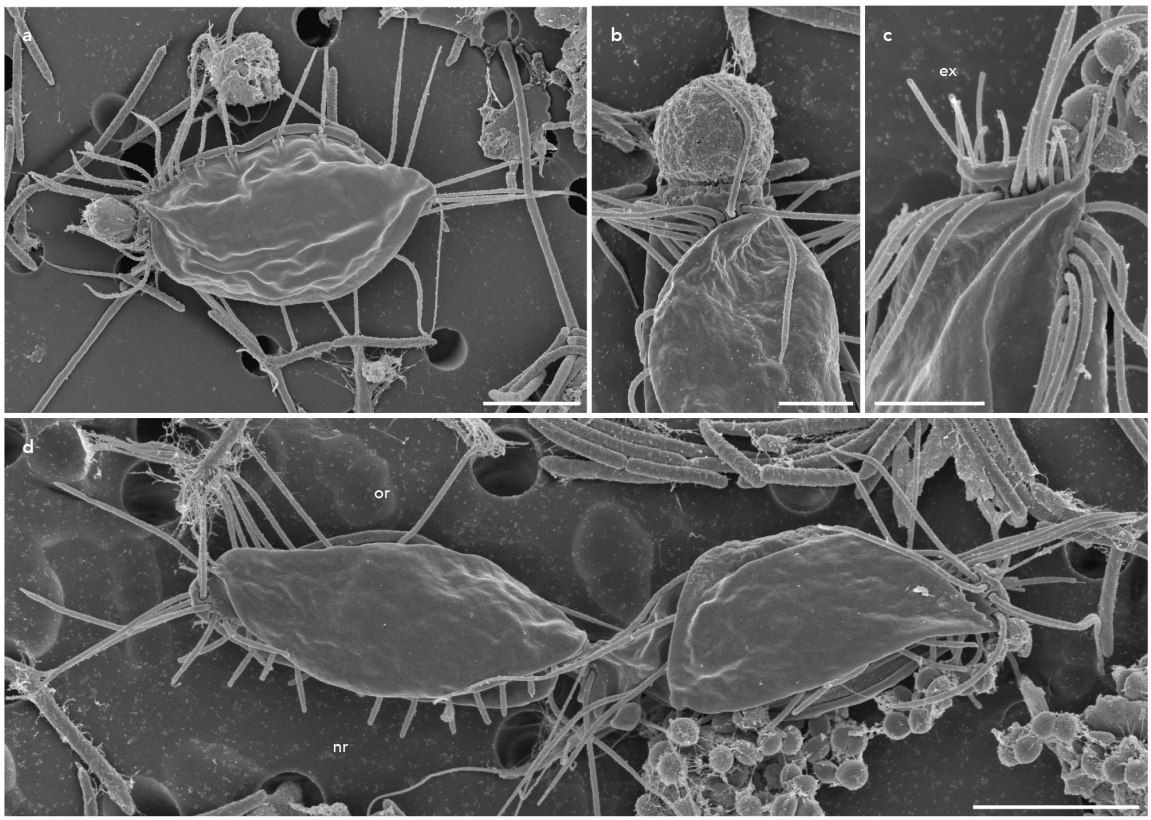

Figure C.2: SEM images of *Hemimastix kukwesjijk*. (a) Feeding cell, general view (anterior to left; note prey item attached to capitulum). (b) Close-up of anterior end showing ingestion in progress at the capitulum. (c) Discharged extrusomes (ex; triggered by the fixation process) along margin of the capitulum (compare to undischarged extrusomes in Fig. 4.1d). (d) Dividing cells, with the left-most cell clearly showing the old row of full-length flagella (or) and the new row with short flagella (nr). Scale bars: (a) 5 µm, (b-c) 2 µm, (d) 5 µm.

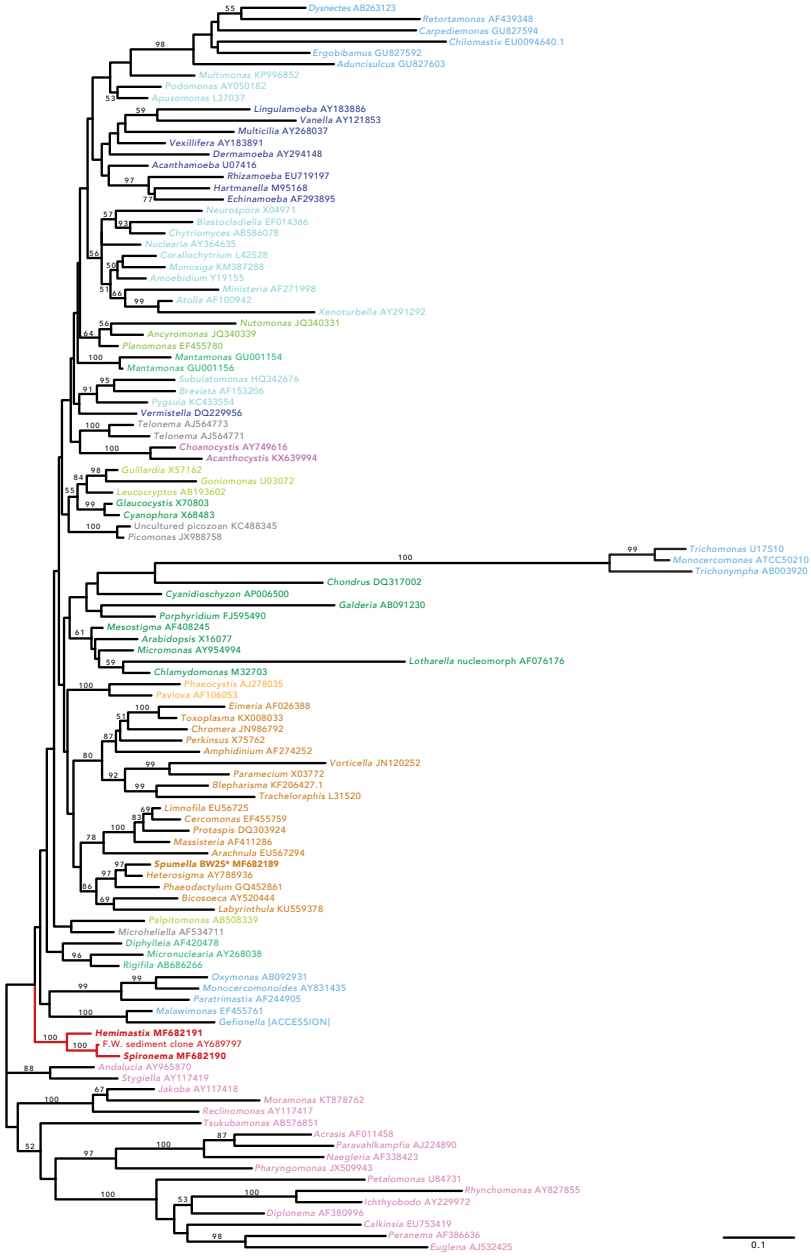

Figure C.3: SSU rDNA phylogeny of eukaryotes inferred from 111 taxa and 1252 sites under the GTR+Γ model in RAxML. Hemimastigophora, including *Hemimastix kukwesjijk* and *Spironema* c.f. *multiciliatum* from this study, are shown in red. Colours of other sequence names correspond to the same taxonomic groupings as in Fig. 4.3. The sequence of *Spumella* sp. strain BW2S, the prey for *H. kukwesjijk*, is included and marked with an asterisk. The numbers on branches show bootstrap percentages (1000 replicates; values below 50% not shown). Branches in gray are 1/2 their original length. This tree was the reference phylogeny for pplacer analyses shown in Fig. 4.2. Scale bar denotes 0.1 expected substitutions/site.

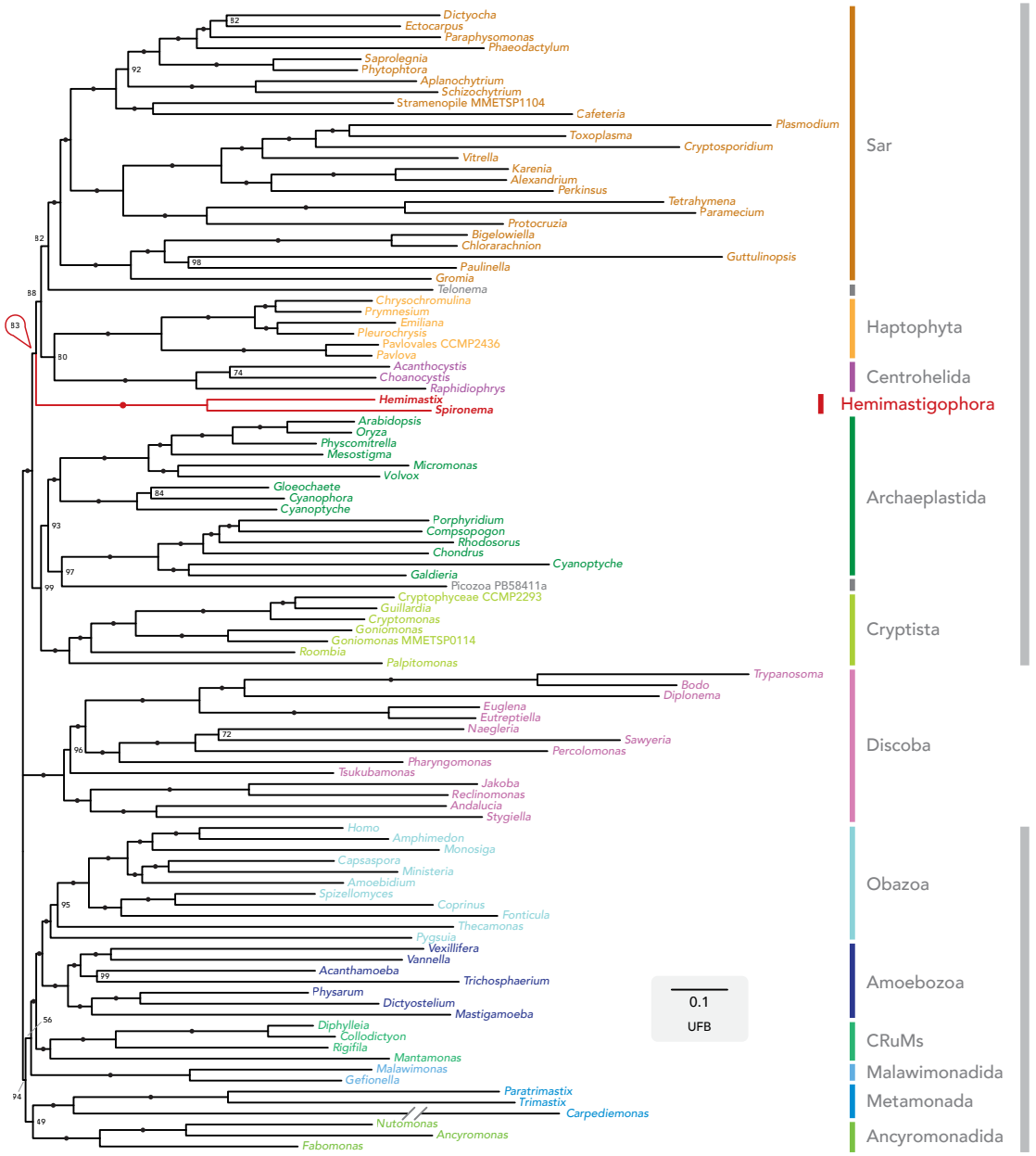

Figure C.4: Unrooted phylogeny of eukaryotes inferred from 351 genes, with 104 taxa, using maximum likelihood under the LG+C60+F+Γ model. The numbers on branches show ultrafast bootstrap approximation percentages (UFB), with filled circles denoting 100% support. The *Carpediemonas* branch is shown reduced by 1/3 of the original length for display purposes. Scale bar denotes 0.1 expected substitutions/site.

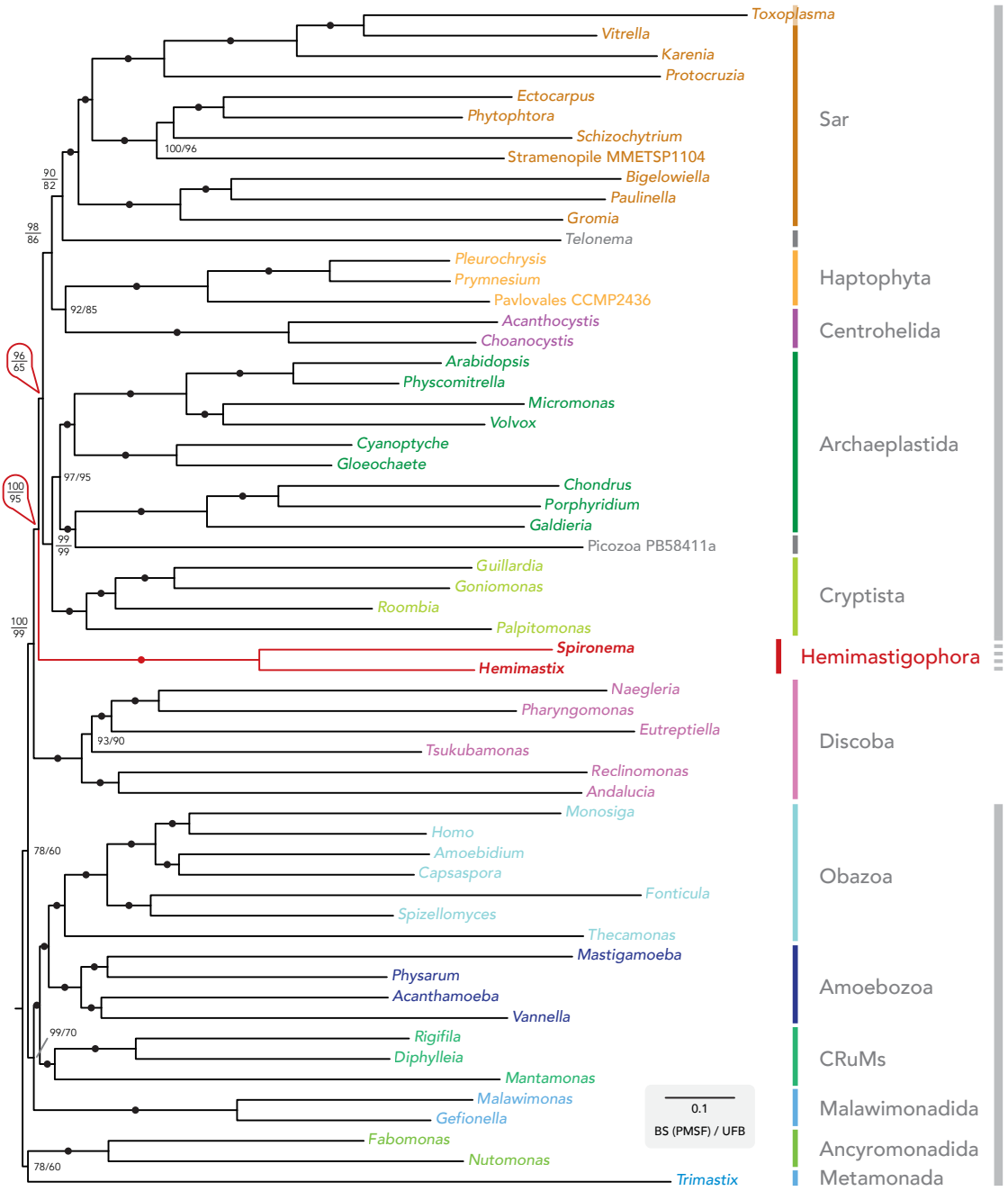

Figure C.5: Unrooted phylogeny of eukaryotes from the '58 taxa, no longest branches' (58-nLB) dataset of 351 genes, inferred using Maximum Likelihood under the LG+C60+F+Γ model. The numbers on branches show PMSF bootstrap percentages (BS PMSF; 200 true bootstrap replicates), then ultrafast bootstrap approximation percentages (UFB; 1000 replicates). Filled circles denote 100% support with both methods. Scale bar denotes 0.1 expected substitutions/site.

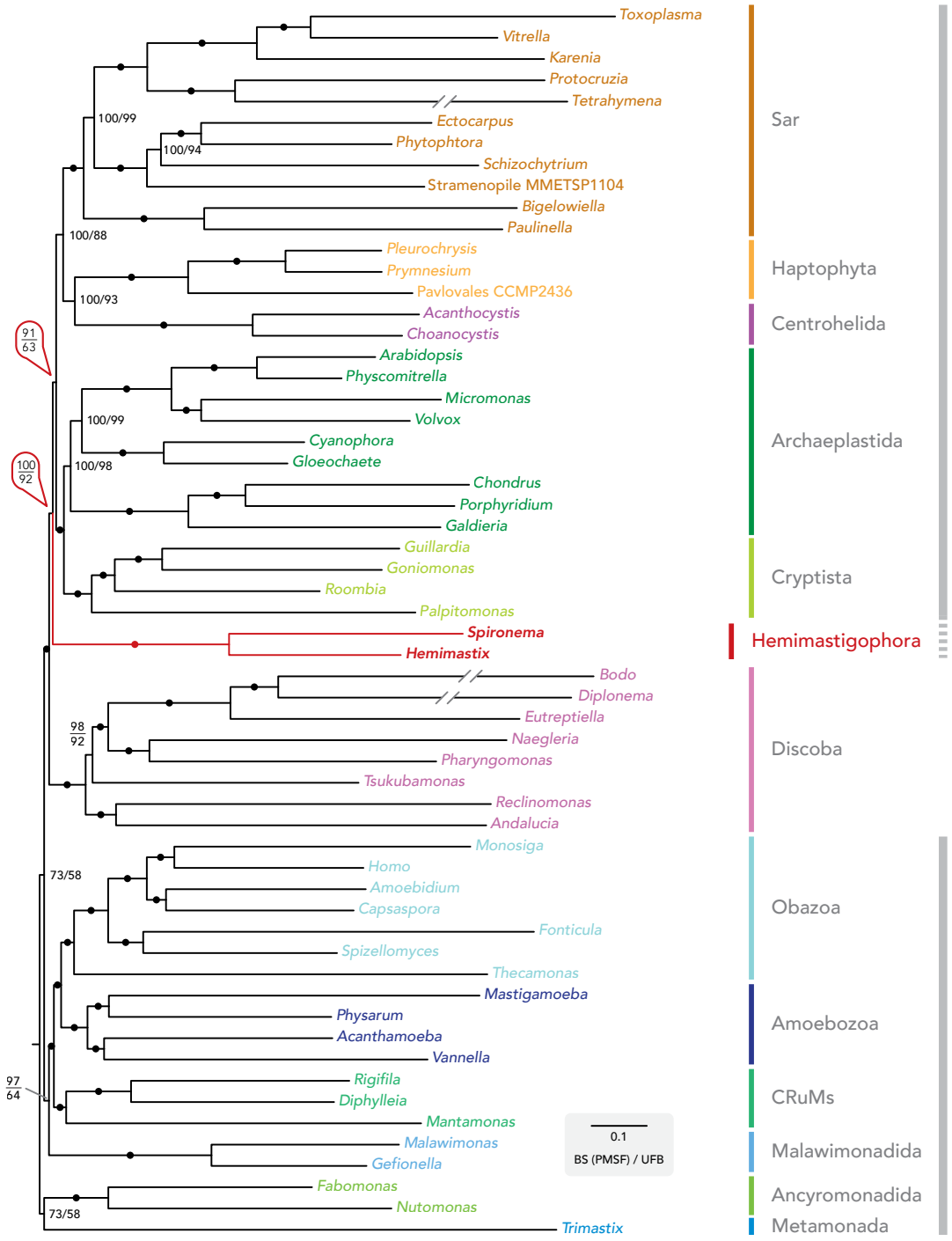

Figure C.6: Unrooted phylogeny of eukaryotes from the '58 taxa, no data-poor taxa' (58-nDP) dataset of 351 genes, inferred using Maximum Likelihood under the LG+C60+F+Γ model. The numbers on branches show PMSF bootstrap percentages (BS PMSF; 100 true bootstrap replicates), then ultrafast bootstrap approximation percentages (UFB; 1000 replicates). Filled circles denote 100% support with both methods. The branches leading to *Bodo, Diplonema* and *Tetrahymena* are shown reduced by 1/3. Scale bar denotes 0.1 expected substitutions/site.

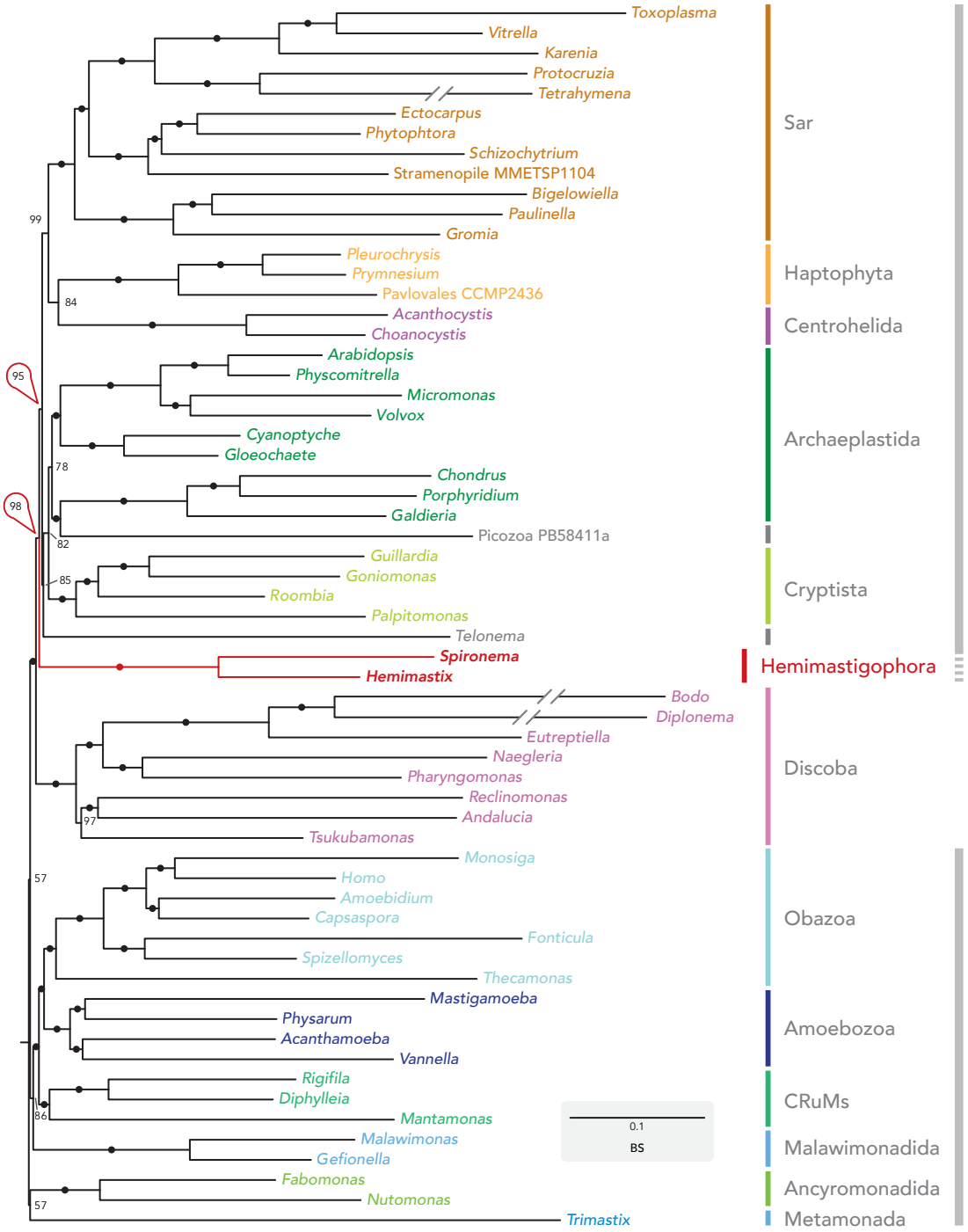

Figure C.7: Unrooted phylogeny of eukaryotes from the '61-SR4' dataset of 61 taxa and 351 genes, with amino acids recoded as four states, inferred using Maximum Likelihood under the GTR+R6+F model. The numbers on branches show bootstrap percentages (BS; 500 true bootstrap replicates). Filled circles represent 100% support. The branches leading to *Bodo, Diplonema* and *Tetrahymena* are shown reduced by 1/3. Scale bar denotes 0.1 expected substitutions/site.

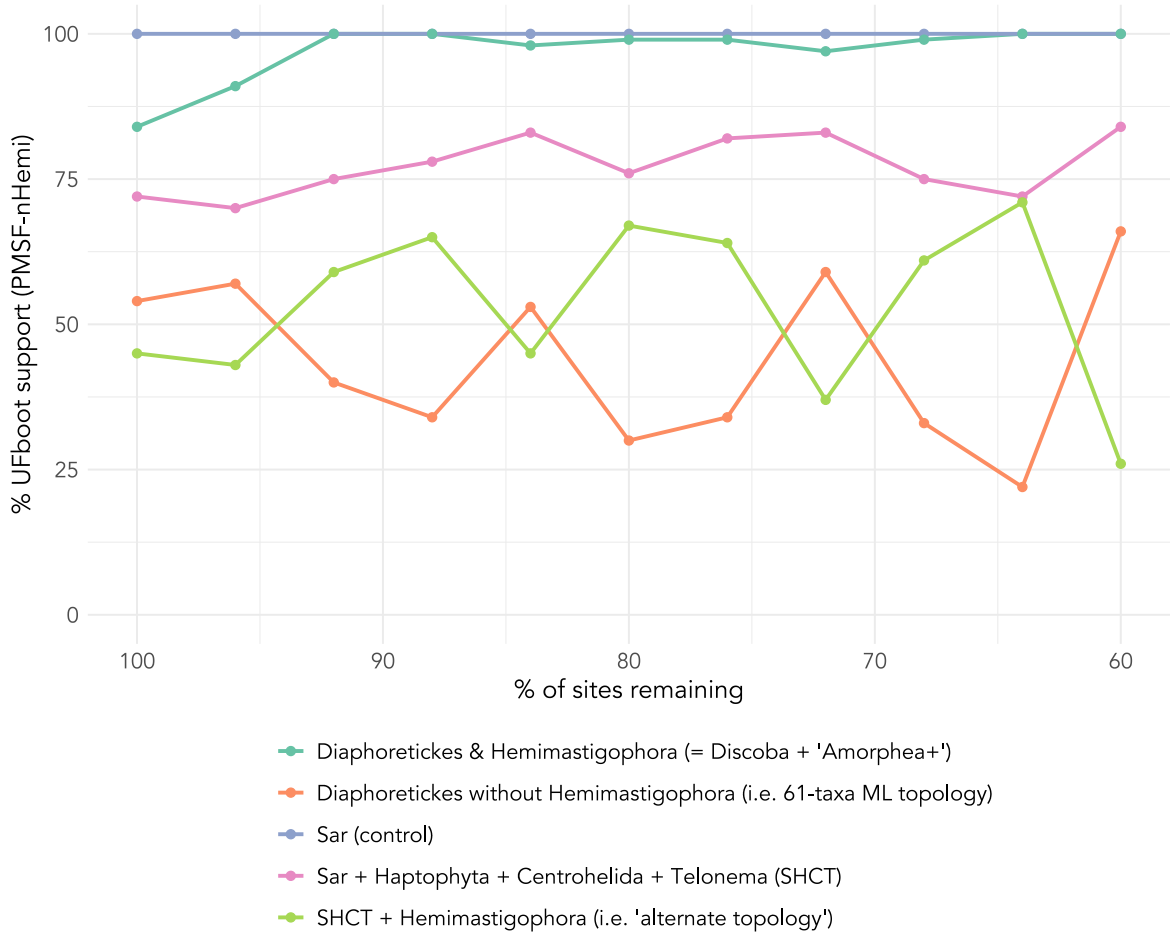

Figure C.8: Summary of support for several important bipartitions with the sequential removal of the fastest evolving sites from the 61 taxa, 351 gene dataset (61-SFSR analysis). The support values are UFboot percentages (1000 replicates) inferred using Maximum Likelihood under the LG+C60+F+Γ derived PMSF model using a guide tree pruned of hemimastigotes (i.e. PMSF-nHEMI – see Methods); these values are not directly comparable to those from the other illustrated analyses.

Bibliography

Aberer, A.J., Krompass, D. & Stamatakis, A., 2013. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. *Systematic Biology*, 62(1):162–166.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana i Molera, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella i Cortes, G., Youssef, N., Zlatogursky, V. & Zhang, Q., 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*, 66(1):4–119.

Ahrendt, S.R., Quandt, C.A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B., Cheng, J.F., Woyke, T., Pelin, A., Henrissat, B., Reynolds, N.K., Benny, G.L., Smith, M.E., James, T.Y. & Grigoriev, I.V., 2018. Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology*, 3(12):1417–1428.

Al-Qassab, S., Lee, W.J., Murray, S., Simpson, A.G.B. & Patterson, D.J., 2002. Flagellates from stromatolites and surrounding sediments in Shark Bay, Western Australia. *Acta Protozoologica*, 41:91–144.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.

Amann, R. & Fuchs, B.M., 2008. Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nature Reviews Microbiology*, 6(5):339–348.

Amann, R.I., Krumholz, L. & Stahl, D.A., 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *Journal of Bacteriology*, 172(2):762–770.

Balzano, S., Corre, E., Decelle, J., Sierra, R., Wincker, P., Da Silva, C., Poulain, J., Pawlowski, J. & Not, F., 2015. Transcriptome analyses to investigate symbiotic relationships between marine protists. *Frontiers in Microbiology*, 6(98):1–14.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. & Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5):455–477.

Barlow, L.D., Nývltová, E., Aguilar, M., Tachezy, J. & Dacks, J.B., 2018. A sophisticated, differentiated Golgi in the ancestor of eukaryotes. *BMC Biology*, 16:1–15.

Bennett, M.S. & Triemer, R.E., 2012. A new method for obtaining nuclear gene sequences from field samples and taxonomic revisions of the photosynthetic euglenoids *Lepocinclis (Euglena) helicoideus* and *Lepocinclis (Phacus) horridus (Euglenophyta)*. *Journal of Phycology*, 48(1):254–260.

Berger, S.A. & Stamatakis, A., 2011. Aligning short reads to reference alignments and trees. *Bioinformatics*, 27(15):2068–2075.

Bhattacharya, D., Price, D.C., Yoon, H.S., Yang, E.C., Poulton, N.J., Andersen, R.A. & Das, S.P., 2012. Single cell genome analysis supports a link between phagotrophy and primary plastid endosymbiosis. *Scientific Reports*, 2(356):1–8.

Bicudo, C.E.d.M. & Menezes, M., 2016. Phylogeny and Classification of Euglenophyceae: A Brief Review. *Frontiers in Ecology and Evolution*, 4(98):429–15.

Biomarks Consortium, 2011. BioMarKs Data Portal.

Boenigk, J. & Arndt, H., 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. *A Van Leeuw J Microb*, 81:465–480.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Breckenridge, D.G., Watanabe, Y.i., Greenwood, S.J., Gray, M.W. & Schnare, M.N., 1999. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. *Proceedings of the National Academy of Sciences*, 96(3):852–856.

Breglia, S.A. & Leander, B.S., 2007. Phylogeny of Phagotrophic Euglenids (Euglenozoa) as Inferred from Hsp90 Gene Sequences. *Journal of Eukaryotic Microbiology*, 54(1):86–92.

Breglia, S.A., Yubuki, N., Hoppenrath, M. & Leander, B.S., 2010. Ultrastructure and molecular phylogenetic position of a novel euglenozoan with extrusive episymbiotic bacteria: *Bihospites bacati* n. gen. et sp. (Symbiontida). *BMC Microbiology*, 10(1):145.

Breglia, S.A., Yubuki, N. & Leander, B.S., 2013. Ultrastructure and Molecular Phylogenetic Position of *Heteronema scaphurum*: A Eukaryovorous Euglenid with a Cytoproct. *Journal of Eukaryotic Microbiology*, 60(2):107–120.

Brown, J.W., Walker, J.F. & Smith, S.A., 2017. Phyx: phylogenetic tools for unix. *Bioinformatics*, 33(12):1886–1888.

Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.I., Hashimoto, T., Simpson, A.G.B. & Roger, A.J., 2018. Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biology and Evolution*, 10(2):427–433.

Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G.B. & Roger, A.J., 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proceedings of the Royal Society Series B: Biological Sciences*, 280(1769):20131755–20131755.

Burki, F., 2014. The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harbor Perspectives in Biology*, 6(5):a016147–a016147.

Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P. & Keeling, P.J., 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings Biological sciences / The Royal Society*, 283(1823):20152802–10.

Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A.D., 2018. rnaSPAdes: a de novotranscriptome assembler and its application to RNA-Seq data. *bioRxiv*, 32(7):1009–1015.

Busse, I., Patterson, D.J. & Preisfeld, A., 2003. Phylogeny of Phagotrophic Euglenids (Euglenozoa): a Molecular Approach Based on Culture Material and Environmental Samples. *Journal of Phycology*, 39:828–836.

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

Busse, I. & Preisfeld, A., 2003b. Discovery of a Group I Intron in the SSU rDNA of *Ploeotia costata* (Euglenozoa). *Protist*, 154(1):57–69.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.

Caron, D.A., Alexander, H., Allen, A.E., Archibald, J.M., Armbrust, E.V., Bachy, C., Bell, C.J., Bharti, A., Dyhrman, S.T., Guida, S.M., Heidelberg, K.B., Kaye, J.Z., Metzner, J., Smith, S.R. & Worden, A.Z., 2016. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews Microbiology*, 15(1):6–20.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4):540–552.

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

Cavalier-Smith, T., *The Flagellates. The Systematics Association Special Volume Series 59*, chap. Flagellate Megaevolution: The Basis for Eukaryote Diversification, (pages 361–390) (Taylor & Francis, London/New York 2000).

Cavalier-Smith, T., 2016. Higher classification and phylogeny of Euglenozoa. *European Journal of Protistology*, 56:250–276.

Cavalier-Smith, T., 2017. Euglenoid pellicle morphogenesis and evolution in light of comparative ultrastructure and trypanosomatid biology: Semi-conservative microtubule/strip duplication, strip shaping and transformation. *European Journal of Protistology*, 61:137–179.

Cavalier-Smith, T., Chao, E.E., Snell, E.A., Berney, C., Fiore-Donno, A.M. & Lewis, R., 2014. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Molecular Phylogenetics and Evolution*, 81:71–85.

Cavalier-Smith, T., Chao, E.E. & Vickerman, K., 2016. New phagotrophic euglenoid species (new genus *Decastava*; *Scytomonas saepesedens*; *Entosiphon oblongum*), Hsp90 introns, and putative euglenoid Hsp90 pre-mRNA insertional editing. *European Journal of Protistology*, 56:147–170.

Cavalier-Smith, T., Lewis, R., Chao, E.E., Oates, B. & Bass, D., 2008. Morphology and Phylogeny of *Sainouron acronematica* sp. n. and the Ultrastructural Unity of Cercozoa. *Protist*, 159(4):591–620.

Chan, Y.F., Chiang, K.P., Chang, J., Moestrup, Ø. & Chung, C.C., 2015. Strains of the morphospecies *Ploeotia costata* (Euglenozoa) isolated from the western North Pacific (Taiwan) reveal substantial genetic differences. *Journal of Eukaryotic Microbiology*, 62(3):318–326.

Chan, Y.F., Moestrup, Ø. & Chang, J., 2013. On *Keelungia pulex* nov. gen. et nov. sp., a heterotrophic euglenoid flagellate that lacks pellicular plates (Euglenophyceae, Euglenida). *European Journal of Protistology*, 49(1):15–31.

Chantangsi, C. & Leander, B.S., 2010. An SSU rDNA barcoding approach to the diversity of marine interstitial cercozoans, including descriptions of four novel genera and nine novel species. *International Journal of Systematic and Evolutionary Microbiology*, 60(8):1962–1977.

Clingenpeel, S., Clum, A., Schwientek, P., Rinke, C. & Woyke, T., 2015. Reconstructing each cell's genome within complex microbial communities—dream or reality? *Frontiers in Microbiology*, 5:1–6.

Criscuolo, A. & Gribaldo, S., 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10:210.

Davis, W.J., Amses, K.R., Benny, G.L., Carter-House, D., Chang, Y., Grigoriev, I., Smith, M.E., Spatafora, J.W., Stajich, J.E. & James, T.Y., 2019. Genome-scale phylogenetics reveals a monophyletic Zoopagales (Zoopagomycota, Fungi). *Molecular Phylogenetics and Evolution*, 133:152–163.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejčić, M., Torruella, G., Domazet-Lošo, T. & Ruiz-Trillo, I., 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National Academy of Sciences*, 110(50):E4858–E4866.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans Coordinators, Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P. & Karsenti, E., 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348(6237):1–12.

Dean, F.B., Nelson, J.R., Giesler, T.L. & Lasken, R.S., 2001. Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome Research*, 11(6):1095–1099.

DeLong, E.F., Wickham, G.S. & Pace, N.R., 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science*, 243(4896):1360–1363.

Derelle, R. & Lang, B.F., 2012. Rooting the Eukaryotic Tree with Mitochondrial and Bacterial Proteins. *Molecular Biology and Evolution*, 29(4):1277–1289.

Derelle, R., Torruella, G., Klimeš, V., Brinkmann, H., Kim, E., Vlček, Č., Lang, B.F. & Eliáš, M., 2015. Bacterial proteins pinpoint a single eukaryotic root. *Proceedings of the National Academy of Sciences*, 112(7):E693–699.

Dietrich, D. & Arndt, H., 2004. Benthic heterotrophic flagellates in an Antarctic melt water stream. *Hydrobiologia*, 529(1):59–70.

Diez, B., Pedrós-Alió, C., Marsh, T.L. & Massana, R., 2001. Application of Denaturing Gradient Gel Electrophoresis (DGGE) To Study the Diversity of Marine Picoeukaryotic Assemblages and Comparison of DGGE with Other Molecular Techniques. *Applied and Environmental Microbiology*, 67(7):2942–2951.

Dujardin, F., *Histoire Naturelle des Zoophytes: Infusoires* (Librairie Encyclopédique de Roret, Paris 1841).

Ebenezer, T.E., Zoltner, M., Burrell, A., Nenarokova, A., Vanclová, A.M.G.N., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankissoor, N.N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A.P., Devos, D.P., Lukeš, J., Lebert, M., Vaughan, S., Hampl, V., Carrington, M., Ginger, M.L., Dacks, J.B., Kelly, S. & Field, M.C., 2019. Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology*, 17(11):1–23.

Eddy, S.R., 2011. Accelerated Profile HMM Searches. *PLoS Computational Biology*, 7(10):e1002195–16.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797.

Edgcomb, V.P., Breglia, S.A., Yubuki, N., Beaudoin, D., Patterson, D.J., Leander, B.S. & Bernhard, J.M., 2010. Identity of epibiotic bacteria on symbiontid euglenozoans in O_2-depleted marine sediments: evidence for symbiont and host co-evolution. *ISME Journal*, 5(2):231–243.

Ekebom, J., Patterson, D.J. & Vørs, N., 1995. Heterotrophic Flagellates from Coral Reef Sediments (Great Barrier Reef, Australia). *Archiv für Protistenkunde*, 146(3-4):251–272.

Esson, H.J. & Leander, B.S., 2006. A model for the morphogenesis of strip reduction patterns in phototrophic euglenids: evidence for heterochrony in pellicle evolution. *Evolution & Development*, 8(4):378–388.

Esson, H.J. & Leander, B.S., 2008. Novel pellicle surface patterns on *Euglena obtusa* (Euglenophyta) from the marine benthic environment: Implications for pellicle development and evolution. *Journal of Phycology*, 44(1):132–141.

Farmer, M.A. & Triemer, R.E., 1988. A Redescription of the Genus *Ploeoti*a Duj. (Euglenophyceae). *Taxon*, 37(2):319–325.

Farmer, M.A. & Triemer, R.E., 1994. An Ultrastructural Study of *Lentomonas applanatum* (Preisig) N. G. (Euglenida). *Journal of Eukaryotic Microbiology*, 41(2):112–119.

Fenchel, T., Bernard, C., Esteban, G.F., Finlay, B.J., Hansen, P.J. & Iversen, N., 1995. Microbial diversity and activity in a Danish fjord with anoxic deep water. *Ophelia*, 43(1):45–100.

Feng, J.M., Jiang, C.Q., Sun, Z.Y., Hua, C.J., Wen, J.F., Miao, W. & Xiong, J., 2020. Single-cell transcriptome sequencing of rumen ciliates provides insight into their molecular adaptations to the anaerobic and carbohydrate-rich rumen microenvironment. *Molecular Phylogenetics and Evolution*, 143:106687.

Fiore-Donno, A.M., Tice, A.K. & Brown, M.W., 2018. A Non-Flagellated Member of the Myxogastria and Expansion of the Echinosteliida. *Journal of Eukaryotic Microbiology*, 94:1–7.

Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., Bowler, C., Lukeš, J. & Horák, A., 2016. Extreme Diversity of Diplonemid Eukaryotes in the Ocean. *Current Biology*, 26(22):3060–3065.

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. *Journal of Eukaryotic Microbiology*, 40(4):422–438.

Foissner, W., Blatterer, H. & Foissner, I., 1988. The Hemimastigophora (*Hemimastix amphikineta* nov. gen., nov. spec.), a new protistan phylum from Gondwanian soils. *European Journal of Protistology*, 23(4):361–383.

Foissner, W. & Foissner, I., *An Illustrated Guide to the Protozoa*, chap. Hemimastigophora, (pages 1185–1186) (Society of Protozoologists / Allen Press, Lawrence, Kansas 2002), 2nd ed.

Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152.

Fukasawa, Y., Oda, T., Tomii, K. & Imai, K., 2017. Origin and Evolutionary Alteration of the Mitochondrial Import System in Eukaryotic Lineages. *Molecular Biology and Evolution*, 34(7):1574–1586.

Gawad, C., Koh, W. & Quake, S.R., 2016. Single-cell genome sequencing: current state of the science. *Scientific Reports*, 17(3):175–188.

Gawryluk, R.M.R., del Campo, J., Okamoto, N., Strassert, J.F.H., Lukeš, J., Richards, T.A., Worden, A.Z., Santoro, A.E. & Keeling, P.J., 2016. Morphological Identification and Single-Cell Genomics of Marine Diplonemids. *Current Biology*, 26(22):3053–3059.

Gawryluk, R.M.R., Tikhonenkov, D.V., Hehenberger, E., Husnik, F., Mylnikov, A.P. & Keeling, P.J., 2019. Non-photosynthetic predators are sister to red algae. *Nature*, 572:240–243.

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M. & Urich, T., 2015. Metatranscriptomic census of active protists in soils. *ISME Journal*, 9(10):2178–2190.

Goddard, M.R. & Burt, A., 1999. Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences*, 96(24):13880–13885.

Gómez, F., Moreira, D., Benzerara, K. & López-García, P., 2011. *Solenicola setigera* is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environmental Microbiology*, 13(1):193–202.

Gouy, M., Guindon, S. & Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, 27(2):221–224.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–652.

Guminska, N., Płecha, M., Zakryś, B. & Milanowski, R., 2018. Order of removal of conventional and nonconventional introns from nuclear transcripts of *Euglena gracilis*. *PLoS Genetics*, 14(10):e1007761–16.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N. & Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8):1494–1512.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M. & Troedsson, C., 2014. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE*, 9(2):e87624–11.

Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G.B. & Roger, A.J., 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proceedings of the National Academy of Sciences*, 106(10):3859–3864.

Haugen, P., Simon, D.M. & Bhattacharya, D., 2005. The natural history of group I introns. *Trends in Genetics*, 21(2):111–119.

He, D., Fiz-Palacios, O., Fu, C.J., Fehling, J., Tsai, C.C. & Baldauf, S.L., 2014. An Alternative Root for the Eukaryote Tree of Life. *Current Biology*, 24(4):465–470.

Hedlund, B.P., Dodsworth, J.A., Murugapiran, S.K., Rinke, C. & Woyke, T., 2014. Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter". *Extremophiles*, 5:865–875.

Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P. & Keeling, P.J., 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):2043–2050.

Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, W., 1995. A nuclear gene of eubacterial origin in Euglena gracilis reflects cryptic endosymbioses during protist evolution. *Proceedings of the National Academy of Sciences*, 92(20):9122–9126.

Huang, L., Ma, F., Chapman, A., Lu, S. & Xie, X.S., 2015. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annual Review of Genomics and Human Genetics*, 16(1):79–102.

Huber-Pestalozzi, G., *Das Phytoplankton des Süßwassers: Euglenophyceen* (Schweizerbart'sche Verlagsbuchhandlung, Stuttgart 1955).

Huse, S.M., Welch, D.B.M., Voorhis, A., Shipunova, A., Morrison, H.G., Eren, A.M. & Sogin, M.L., 2014. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics*, 15(1):41.

Huys, G.R. & Raes, J., 2018. Go with the flow or solitary confinement: a look inside the single-cell toolbox for isolation of rare and uncultured microbes. *Current Opinion in Microbiology*, 44:1–8.

Irwin, N.A.T., Tikhonenkov, D.V., Hehenberger, E., Mylnikov, A.P., Burki, F. & Keeling, P.J., 2019. Phylogenomics supports the monophyly of the Cercozoa. *Molecular Phylogenetics and Evolution*, 130:416–423.

Jackowiak, P., Nowacka, M., Strozycki, P.M. & Figlerowicz, M., 2011. RNA degradome—its biogenesis and functions. *Nucleic Acids Research*, 39(17):7361–7370.

Jackson, C., Knoll, A.H., Chan, C.X. & Verbruggen, H., 2018. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Scientific Reports*, 8:1523.

Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcântara, D.M.C., Roger, A.J., Shadwick, L.L., Smirnov, A., Kudryavtsev, A., Lahr, D.J.G. & Brown, M.W., 2017. Between a Pod and a Hard Test: The Deep Evolution of Amoebae. *Molecular Biology and Evolution*, (pages 1–13).

Karnkowska, A., Bennett, M.S., Watza, D., Kim, J.I., Zakryś, B. & Triemer, R.E., 2015. Phylogenetic Relationships and Morphological Character Evolution of Photosynthetic Euglenids (Excavata) Inferred from Taxon-rich Analyses of Five Genes. *Journal of Eukaryotic Microbiology*, 62(3):362–373.

Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.

Katz, L.A., Grant, J.R., Parfrey, L.W. & Burleigh, J.G., 2012. Turning the Crown Upside Down: Gene Tree Parsimony Roots the Eukaryotic Tree of Life. *Systematic Biology*, 61(4):653–660.

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. *Bioinformatics*, 28(12):1647–1649.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D.H., Marchetti, A., McManus, G.B., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A. & Worden, A.Z., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology*, 12(6):e1001889.

Keeling, P.J. & del Campo, J., 2017. Marine Protists Are Not Just Big Bacteria. *Current Biology*, 27(11):R541–R549.

Kim, E., Park, J.S., Simpson, A.G.B., Matsunaga, S., Watanabe, M., Murakami, A., Sommerfeld, K., Onodera, N.T. & Archibald, J.M., 2010. Complex array of endobionts in *Petalomonas sphagnophila*, a large heterotrophic euglenid protist from Sphagnum-dominated peatlands. *ISME Journal*, 4(9):1108–1120.

Klebs, G., *Flagellatenstudien* (Akademische Verlags-Gesellschaft, Leipzig 1893).

Kolisko, M., Boscaro, V., Burki, F., Lynn, D.H. & Keeling, P.J., 2014. Single-cell transcriptomics for microbial eukaryotes. *Current Biology*, 24(22):R1081–R1082.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. & Teichmann, S.A., 2015. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4):610–620.

Krabberød, A.K., Bjorbækmo, M.F.M., Shalchian-Tabrizi, K. & Logares, R., 2017a. Exploring the oceanic microeukaryotic interactome with metaomics approaches. *Aquatic Microbial Ecology*, 79(1):1–12.

Krabberød, A.K., Bråte, J., Dolven, J.K., Ose, R.F., Klaveness, D., Kristensen, T., Bjørklund, K.R. & Shalchian-Tabrizi, K., 2011. Radiolaria Divided into Polycystina and Spasmaria in Combined 18S and 28S rDNA Phylogeny. *PLoS ONE*, 6(8):e23526–10.

Krabberød, A.K., Orr, R.J.S., Bråte, J., Kristensen, T., Bjørklund, K.R. & Shalchian-Tabrizi, K., 2017b. Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria. *Molecular Biology and Evolution*, 34(7):1557–1573.

Lackey, J.B., 1960. *Calkinsia aureus* gen. et sp. nov:, a New Marine Euglenid. *Transactions of the American Microscopical Society*, 79(1):105–107.

Lahr, D.J.G., Grant, J., Nguyen, T., Lin, J.H. & Katz, L.A., 2011. Comprehensive Phylogenetic Reconstruction of Amoebozoa Based on Concatenated Analyses of SSU-rDNA and Actin Genes. *PLoS ONE*, 6(7):e22780–17.

Lahr, D.J.G., Kosakyan, A., Lara, E., Mitchell, E.A.D., Morais, L., Porfirio-Sousa, A.L., Ribeiro, G.M., Tice, A.K., Pánek, T., Kang, S. & Brown, M.W., 2019. Phylogenomics and Morphological Reconstruction of Arcellinida Testate Amoebae Highlight Diversity of Microbial Eukaryotes in the Neoproterozoic. *Current Biology*, 29(6):991–1001.

Larsen, J., 1987. Algal studies of the Danish Wadden Sea. IV. A taxonomic study of the interstitial euglenoid flagellates. *Nordic Journal of Botany*, 7(5):589–607.

Larsen, J. & Patterson, D.J., 1990. Some flagellates (Protista) from tropical marine sediments. *Journal of Natural History*, 24(4):801–937.

Larsen, J. & Patterson, D.J., The diversity of heterotrophic euglenids. In: D.J. Patterson & J. Larsen (Editors), *The Biology of Free-living Heterotrophic Flagellates*, (pages 205–217) (Clarendon Press, Oxford 1991).

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278.

Lartillot, N., Lepage, T. & Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, 25(17):2286–2288.

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):1095–1109.

Lasken, R.S. & McLean, J.S., 2014. Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews Genetics*, 15:577–584.

Lax, G., Eglit, Y., Eme, L., Bertrand, E.M., Roger, A.J. & Simpson, A.G.B., 2018. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature*, 564(7736):410–414.

Lax, G., Lee, W.J., Eglit, Y. & Simpson, A., 2019. Ploeotids represent much of the phylogenetic diversity of euglenids. *Protist*, 170(2):233–257.

Lax, G. & Simpson, A.G.B., 2013. Combining Molecular Data with Classical Morphology for Uncultured Phagotrophic Euglenids (Excavata): A Single-Cell Approach. *Journal of Eukaryotic Microbiology*, 60(6):615–625.

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

Leander, B.S., Esson, H.J. & Breglia, S.A., 2007. Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays*, 29(10):987–1000.

Leander, B.S. & Farmer, M.A., 2001. Comparative Morphology of the Euglenid Pellicle. II. Diversity of Strip Substructure. *Journal of Eukaryotic Microbiology*, 48(2):202–217.

Leander, B.S., Lax, G., Karnkowska, A. & Simpson, A.G.B., Euglenida. In: J.M. Archibald, C.H. Slamovits & A.G.B. Simpson (Editors), *Handbook of the Protists*, (pages 1047–1088) (Springer International Publishing, Cham 2017).

Leander, B.S., Triemer, R.E. & Farmer, M.A., 2001a. Character evolution in heterotrophic euglenids. *European Journal of Protistology*, 37(3):337–356.

Leander, B.S., Witek, R.P. & Farmer, M.A., 2001b. Trends in the evolution of the euglenid pellicle. *Evolution*, 55(11):2215–2235.

Lee, W.J., 2001. *Diversity and distribution of free-living benthic flagellates in Botany Bay, Australia*. Ph.D. thesis, University of Sydney, Sydney.

Lee, W.J., 2006. Heterotrophic Euglenids from Marine Sediments of Cape Tribulation, Tropical Australia. *Ocean Science*, 41(2):59–73.

Lee, W.J., 2008. Free-living heterotrophic euglenids from marine sediments of the Gippsland Basin, southeastern Australia. *Marine Biology Research*, 4:333–349.

Lee, W.J., 2012. Free-living benthic heterotrophic euglenids from Botany Bay, Australia. *Marine Biology Research*, 8(1):3–27.

Lee, W.J., Blackmore, R. & Patterson, D.J., 1999. Australian records of two lesser known genera of heterotrophic euglenids – *Chasmostoma* Massart, 1920 and *Jenningsia* Schaeffer, 1918. *Protistology*, 1(1):10–16.

Lee, W.J. & Patterson, D.J., 2000. Heterotrophic flagellates (Protista) from marine sediments of Botany Bay, Australia. *Journal of Natural History*, 34(4):483–562.

Lee, W.J. & Patterson, D.J., 2002. Abundance and Biomass of Heterotrophic Flagellates, and Factors Controlling Their Abundance and Distribution in Sediments of Botany Bay. *Microbial Ecology*, 43(4):467–481.

Lee, W.J. & Simpson, A.G.B., 2014a. Morphological and Molecular Characterisation of *Notosolenus urceolatus* Larsen and Patterson 1990, a Member of an Understudied Deep-branching Euglenid Group (Petalomonads). *Journal of Eukaryotic Microbiology*, 61(5):463–479.

Lee, W.J. & Simpson, A.G.B., 2014b. Ultrastructure and Molecular Phylogenetic Position of *Neometanema parovale* sp. nov. (*Neometanema* gen. nov.), a Marine Phagotrophic Euglenid with Skidding Motility. *Protist*, 165(4):452–472.

Lee, W.J., Simpson, A.G.B. & Patterson, D.J., 2005. Free-living heterotrophic flagellates from freshwater sites in Tasmania (Australia), a field survey. *Acta Protozoologica*, 44:321–350.

Leedale, G.F., *Euglenoid Flagellates* (Prentice-Hall, Englewood Cliffs, NJ 1967).

Lemmermann, E., Eugleninae. In: A. Pascher (Editor), *Die Süßwasser-Flora Deutschlands, Österreichs und der Schweiz*, (pages 115–174) (Verlag von Gustav Fischer, Jena 1913).

Linton, E.W., Karnkowska-Ishikawa, A., Kim, J.I., Shin, W., Bennett, M.S., Kwiatowski, J., Zakryś, B. & Triemer, R.E., 2010. Reconstructing Euglenoid Evolutionary Relationships using Three Genes: Nuclear SSU and LSU, and Chloroplast SSU rDNA Sequences and the Description of *Euglenaria* gen. nov. (Euglenophyta). *Protist*, 161(4):603–619.

Linton, E.W. & Triemer, R.E., 1999. Reconstruction of the feeding apparatus in *Ploeotia costata* (Euglenophyta) and its relationship to other euglenoid feeding apparatuses. *Journal of Phycology*, 35(2):313–324.

Liu, Z., Hu, S.K., Campbell, V., Tatters, A.O., Heidelberg, K.B. & Caron, D.A., 2017. Single-cell transcriptomics of small microbial eukaryotes: limitations and potential. *ISME Journal*, 11(5):1282–1285.

Łukomska-Kowalczyk, M., Karnkowska, A., Krupska, M., Milanowski, R. & Zakryś, B., 2016. DNA barcoding in autotrophic euglenids: evaluation of COI and 18s rDNA. *Journal of Phycology*, 52(6):951–960.

Lynn, D.H. & Pinheiro, M., 2009. A Survey of Polymerase Chain Reaction (PCR) Amplification Studies of Unicellular Protists Using Single-Cell PCR. *Journal of Eukaryotic Microbiology*, 56(5):406–412.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., Siemensmeyer, T., Trautmann, I., Romac, S., Berney, C., Kozlov, A., Mitchell, E.A.D., Seppey, C.V.W., Egge, E., Lentendu, G., Wirth, R., Trueba, G. & Dunthorn, M., 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Scientific Reports*, 1(4):0091.

Mangot, J.F., Logares, R., Sánchez, P., Latorre, F., Seeleuthner, Y., Mondy, S., Sieracki, M.E., Jaillon, O., Wincker, P., de Vargas, C. & Massana, R., 2017. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Scientific Reports*, 7:41498.

Marcy, Y., Ishoey, T., Lasken, R.S., Stockwell, T.B., Walenz, B.P., Halpern, A.L., Beeson, K.Y., Goldberg, S.M.D. & Quake, S.R., 2007a. Nanoliter Reactors Improve Multiple Displacement Amplification of Genomes from Single Cells. *PLoS Genetics*, 3(9):e155–7.

Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., Martin, H.G., Szeto, E., Platt, D., Hugenholtz, P., Relman, D.A. & Quake, S.R., 2007b. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences*, 104(29):11889–11894.

Marin, B., Palm, A., Klingberg, M. & Melkonian, M., 2003. Phylogeny and Taxonomic Revision of Plastid-Containing Euglenophytes based on SSU rDNA Sequence Comparisons and Synapomorphic Signatures in the SSU rRNA Secondary Structure. *Protist*, 154(1):99–145.

Marín, I., Aguilera, A., Reguera, B. & Abad, J.P., 2001. Preparation of DNA suitable for PCR amplification from fresh or fixed single dinoflagellate cells. *BioTechniques*, 30(1):88–93.

Marquardt, M., Vader, A., Stübner, E.I., Reigstad, M. & Gabrielsen, T.M., 2016. Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). *Applied and Environmental Microbiology*, 82(6):1868–1880.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*, 17(1):10–12.

Maruyama, S., Suzaki, T., Weber, A.P., Archibald, J.M. & Nozaki, H., 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evolutionary Biology*, 11(1):105.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. & Pedrós-Alió, C., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and Environmental Microbiology*, 70(6):3528–3534.

Matsen, F.A., Kodner, R.B. & Armbrust, E.V., 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538.

Milanowski, R., Karnkowska, A., Ishikawa, T. & Zakryś, B., 2014. Distribution of conventional and nonconventional introns in tubulin (α and β) genes of euglenids. *Molecular Biology and Evolution*, 31(3):584–593.

Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5):1188–1195.

Montegut-Felkner, A.E. & Triemer, R.E., 1997. Phylogenetic Relationships of Selected Euglenoid Genera Based on Morphological and Molecular Data. *Journal of Phycology*, 33:512–519.

Monteil, C.L., Vallenet, D., Menguy, N., Benzerara, K., Barbe, V., Fouteau, S., Cruaud, C., Floriani, M., Viollier, E., Adryanczyk, G., Leonhardt, N., Faivre, D., Pignol, D., López-García, P., Weld, R.J. & Lefevre, C.T., 2019. Ectosymbiotic bacteria at the origin of magnetoreception in a marine protist. *Nature Microbiology*, 434(8):1–11.

Moon-van der Staay, S.Y., De Wachter, R. & Vaulot, D., 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 409(6820):607–610.

Müllner, A.N., Angeler, D.G., Samuel, R., Linton, E. & Triemer, R.E., 2001. Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *International Journal of Systematic and Evolutionary Microbiology*, 51:783–791.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1):268–274.

Olsen, G.J., Lane, D.J., Giovannoni, S.J. & Pace, N.R., 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology*, 40:337–365.

Orsi, W., Edgcomb, V., Jeon, S., Leslin, C., Bunge, J., Taylor, G.T., Varela, R. & Epstein, S.S., 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. II. Habitat specialization. *ISME Journal*, 5:1357–1373.

Paerschke, S., Vollmer, A.H. & Preisfeld, A., 2017. Ultrastructural and immunocytochemical investigation of paramylon combined with new 18S rDNA-based secondary structure analysis clarifies phylogenetic affiliation of *Entosiphon sulcatum* (Euglenida: Euglenozoa). *Organisms Diversity & Evolution*, 59(3):1–12.

Patterson, D.J. & Simpson, A.G.B., 1996. Heterotrophic flagellates from coastal marine and hypersaline sediments in Western Australia. *European Journal of Protistology*, 32(4):423–448.

Pawlowski, J. & Burki, F., 2009. Untangling the Phylogeny of Amoeboid Protists. *Journal of Eukaryotic Microbiology*, 56(1):16–25.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G. & Sandberg, R., 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, 10(11):1096–1098.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S. & Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, 9(1):171–181.

Prakadan, S.M., Shalek, A.K. & Weitz, D.A., 2017. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nature Reviews Genetics*, 18(6):345–361.

Preisfeld, A., Berger, S., Busse, I., Liller, S. & Ruppel, H.G., 2000. Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data. *Journal of Phycology*, 36:220–226.

Preisfeld, A., Busse, I. & Klingberg, M., 2001. Phylogenetic position and interrelationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *International Journal of Systematic and Evolutionary Microbiology*, 51:751–758.

Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., Schroth, G.P. & Sandberg, R., 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, 30(8):777–782.

Rappé, M.S. & Giovannoni, S.J., 2003. The uncultured microbial majority. *Annual Review of Microbiology*, 57(1):369–394.

Richards, T.A. & Cavalier-Smith, T., 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature*, 436(7054):1113–1118.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R. & Woyke, T., 2014. Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols*, 9(5):1038–1048.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P. & Woyke, T., 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–437.

Rodríguez-Ezpeleta, N., Teijeiro, S., Forget, L., Burger, G. & Lang, B.F., Construction of cDNA Libraries: Focus on Protists and Fungi. In: J. Parkinson (Editor), *Methods in Molecular Biology: Expressed Sequence Tags (ESTs)*, (pages 33–47) (Humana Press, Totowa, NJ, Totowa, NJ 2009).

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3):539–542.

Roy, J., Faktorová, D., Lukeš, J. & Burger, G., 2007. Unusual Mitochondrial Genome Structures throughout the Euglenozoa. *Protist*, 158(3):385–396.

Roy, R.S., Price, D.C., Schliep, A., Cai, G., Korobeynikov, A., Yoon, H.S., Yang, E.C. & Bhattacharya, D., 2014. Single cell genome analysis of an uncultured heterotrophic stramenopile. *Scientific Reports*, 4(1):518–518.

Saito, A., Suetomo, Y., Arikawa, M., Omura, G., Khan, S.M.M.K., Kakuta, S., Suzaki, E., Kataoka, K. & Suzaki, T., 2003. Gliding Movement in *Peranema trichophorum* Is Powered by Flagellar Surface Motility. *Cell Motility and the Cytoskeleton*, 55:244–253.

Schnepf, E., Schlegel, I. & Hepperle, D., 2002. *Petalomonas sphagnophila* (Euglenophyta) and its endocytobiotic cyanobacteria: a unique form of symbiosis. *Phycologia*, 41(2):153–157.

Schoenle, A., Živaljić, S., Prausse, D., Voss, J., Jakobsen, K. & Arndt, H., 2019. New phagotrophic euglenids from deep sea and surface waters of the Atlantic Ocean (*Keelungia nitschei*, *Petalomonas acorensis*, *Ploeotia costaversata*). *European Journal of Protistology*, 69:102–116.

Schroeckh, S., Lee, W.J. & Patterson, D.J., 2003. Free-living heterotrophic euglenids from freshwater sites in mainland Australia. *Hydrobiologia*, 493:131–166.

Sebastián, C.R. & O'Ryan, C., 2001. Single-cell sequencing of dinoflagellate (Dinophyceae) nuclear ribosomal genes. *Molecular Ecology Notes*, 1(1):329–331.

Sebé-Pedrós, A., Grau-Bové, X., Richards, T.A. & Ruiz-Trillo, I., 2014. Evolution and Classification of Myosins, a Paneukaryotic Whole-Genome Approach. *Genome Biology and Evolution*, 6(2):290–305.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212.

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

Simpson, A.G.B. & Eglit, Y., Protist Diversification. In: *Encyclopedia of Evolutionary Biology*, (pages 344–360) (Elsevier Ltd. 2016).

Simpson, A.G.B., Inagaki, Y. & Roger, A.J., 2006. Comprehensive Multigene Phylogenies of Excavate Protists Reveal the Evolutionary Positions of "Primitive" Eukaryotes. *Molecular Biology and Evolution*, 23(3):615–625.

Simpson, A.G.B., Van Den Hoff, J., Bernard, C., Burton, H.R. & Patterson, D.J., 1997. The Ultrastructure and Systematic Position of the Euglenozoon *Postgaardi mariagerensis*, Fenchel et al. *Archiv für Protistenkunde*, 147(3):213–225.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M. & Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*, 26(8):1134–1144.

Song, L. & Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, 4(48):1–8.

Speijer, D., Lukeš, J. & Eliáš, M., 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *Proceedings of the National Academy of Sciences*, 112(29):8827–8834.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I. & Sermon, K., 2006. Optimization and evaluation of single-cell whole-genome multiple displacement amplification. *Human Mutation*, 27(5):496–503.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313.

Stein, F.v., *Der Organismus der Infusionsthiere* (Verlag von Wilhelm Engelmann, Leipzig 1878).

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

Strassert, J.F.H., Karnkowska, A., Hehenberger, E., del Campo, J., Kolisko, M., Okamoto, N., Burki, F., Janouškovec, J., Poirier, C., Leonard, G., Hallam, S.J., Richards, T.A., Worden, A.Z., Santoro, A.E. & Keeling, P.J., 2018. Single cell genomics of uncultured marine alveolates shows paraphyly of basal dinoflagellates. *ISME Journal*, 12(1):304–308.

Susko, E. & Roger, A.J., 2007. On Reduced Amino Acid Alphabets for Phylogenetic Inference. *Molecular Biology and Evolution*, 24(9):2139–2150.

Takano, Y. & Horiguchi, T., 2005. Acquiring Scanning Electron Microscopical, Light Microscopical and multiple gene sequence data from a single dinoflagellate cell. *Journal of Phycology*, 42(1):251–256.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., Lao, K. & Surani, M.A., 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382.

Tice, A.K., Shadwick, L.L., Fiore-Donno, A.M., Geisen, S., Kang, S., Schuler, G.A., Spiegel, F.W., Wilkinson, K.A., Bonkowski, M., Dumack, K., Lahr, D.J.G., Voelcker, E., Clauß, S., Zhang, J. & Brown, M.W., 2016. Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. *Biology Direct*, 11(69):1–21.

Triemer, R.E., 1986. Light and Electron Microscopic Description of a Colorless Euglenoid, *Serpenomonas costata* n. g., n. sp. *Journal of Protozoology*, 33(3):412–415.

Triemer, R.E. & Farmer, M.A., 1991. An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. *Protoplasma*, 164(1-3):91–104.

Triemer, R.E. & Fritz, L., 1987. Structure and Operation of the Feeding Apparatus in a Colorless Euglenoid, *Entosiphon sulcatum*. *Journal of Protozoology*, 34(1):39–47.

Triemer, R.E. & Fritz, L., 1988. Ultrastructural Features of Mitosis in *Ploeotia costata* (Heteronematales, Euglenophyta). *Journal of Phycology*, 24(4):514–519.

Turmel, M., Gagnon, M.C., O'Kelly, C.J., Otis, C. & Lemieux, C., 2008. The Chloroplast Genomes of the Green Algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* Shed New light on the Evolutionary History of Prasinophytes and the Origin of the Secondary Chloroplasts of Euglenids. *Molecular Biology and Evolution*, 26(3):631–648.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. & Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43.

von der Heyden, S., Chao, E.E., Vickerman, K. & Cavalier-Smith, T., 2005. Ribosomal RNA Phylogeny of Bodonid and Diplonemid Flagellates and the Evolution of Euglenozoa. *Journal of Eukaryotic Microbiology*, 51(4):402–416.

Wagner, M., Horn, M. & Daims, H., 2003. Fluorescence *in situ* hybridisation for the identification and characterisation of prokaryotes. *Current Opinion in Microbiology*, 6(3):302–309.

Wang, H.C., Minh, B.Q., Susko, E. & Roger, A.J., 2017. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Systematic Biology*, 67(2):216–235.

Woese, C.R., Kandler, O. & Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12):4576–4579.

Worden, A.Z., Dupont, C. & Allen, A.E., 2011. Genomes of uncultured eukaryotes: sorting FACS from fiction. *Genome Biology*, 12(6):117.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E. & Keeling, P.J., 2015. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*, 347(6223):1257594–1257594.

Woyke, T., Doud, D.F.R. & Schulz, F., 2017. The trajectory of microbial single-cell sequencing. *Nature Methods*, 14(11):1045–1054.

Yamaguchi, A., Yubuki, N. & Leander, B.S., 2012. Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evolutionary Biology*, 12(1):29.

Yazaki, E., Ishikawa, S.A., Kume, K., Kumagai, A., Kamaishi, T., Tanifuji, G., Hashimoto, T. & Inagaki, Y., 2017. Global Kinetoplastea phylogeny inferred from a large-scale multigene alignment including parasitic species for better understanding transitions from a free-living to a parasitic lifestyle. *Genes & Genetic Systems*, 92(1):35–42.

Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S. & Bhattacharya, D., 2011. Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists. *Science*, 332(6030):714–717.

Yubuki, N., Edgcomb, V.P., Bernhard, J.M. & Leander, B.S., 2009. Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. *BMC Microbiology*, 9(1):16.

Yubuki, N. & Leander, B.S., 2012. Reconciling the bizarre inheritance of microtubules in complex (euglenid) microeukaryotes. *Protoplasma*, 249(4):859–869.

Yubuki, N. & Leander, B.S., 2018. Diversity and Evolutionary History of the Symbiontida (Euglenozoa). *Frontiers in Ecology and Evolution*, 6:100.

Yubuki, N., Pánek, T., Yabuki, A., Čepička, I., Takishita, K., Inagaki, Y. & Leander, B.S., 2015. 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. *Journal of Eukaryotic Microbiology*, 62(4):532–542.

Yubuki, N., Simpson, A.G.B. & Leander, B.S., 2013. Reconstruction of the feeding apparatus in *Postgaardi mariagerensis* provides evidence for character evolution within the Symbiontida (Euglenozoa). *European Journal of Protistology*, 49(1):32–39.

Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W. & Church, G.M., 2006. Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology*, 24(6):680–686.

Zhao, S., Burki, F., Bråte, J., Keeling, P.J., Klaveness, D. & Shalchian-Tabrizi, K., 2012. *Collodictyon*—An Ancient Lineage in the Tree of Eukaryotes. *Molecular Biology and Evolution*, 29(6):1557–1568.

Zhou, Y., Lu, C., Wu, Q.J., Wang, Y., Sun, Z.T., Deng, J.C. & Zhang, Y., 2007. GISSD: Group I Intron Sequence and Structure Database. *Nucleic Acids Research*, 36:D31–D37.

Zolffel, M. & Skibbe, O., 1997. Rediscovery of the multiflagellated protist *Paramastix conifera* Skuja 1948 (Protista incertae sedis). *Nova Hedwigia*, 65:443–452.