Molecular Phylogenetics and Evolution 159 (2021) 107088

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Multigene phylogenetics of euglenids based on single-cell transcriptomics of diverse phagotrophs

G. Lax [a,b], M. Kolisko [c], Y. Eglit [a], W.J. Lee [d], N. Yubuki [e,g], A. Karnkowska [f], B.S. Leander [g], G. Burger [h], P.J. Keeling [b], A.G.B. Simpson [a,*]

[a] Department of Biology, and Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Canada
[b] Department of Botany, University of British Columbia, Vancouver, Canada[1]
[c] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic
[d] Department of Environment and Energy Engineering, Kyungnam University, Changwon, Republic of Korea
[e] Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Saclay, Orsay, France
[f] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Poland
[g] Department of Zoology, University of British Columbia, Vancouver, Canada
[h] Robert-Cedergren Centre for Bioinformatics and Genomics, Biochemistry Department, Université de Montréal, Montréal, Canada

ARTICLE INFO

Keywords:
Cell motility
Euglenozoa
Phylogenomics
Protozoa
Spirocuta
Symbiontida

ABSTRACT

Euglenids are a well-known group of single-celled eukaryotes, with phototrophic, osmotrophic and phagotrophic members. Phagotrophs represent most of the phylogenetic diversity of euglenids, and gave rise to the phototrophs and osmotrophs, but their evolutionary relationships are poorly understood. Symbiontids, in contrast, are anaerobes that are alternatively inferred to be derived euglenids, or a separate euglenozoan group. Most phylogenetic studies of euglenids have examined the SSU rDNA only, which is often highly divergent. Also, many phagotrophic euglenids (and symbiontids) are uncultured, restricting collection of other molecular data. We generated transcriptome data for 28 taxa, mostly using a single-cell approach, and conducted the first multigene phylogenetic analyses of euglenids to include phagotrophs and symbiontids. Euglenids are recovered as monophyletic, with symbiontids forming an independent branch within Euglenozoa. Spirocuta, the clade of flexible euglenids that contains both the phototrophs (Euglenophyceae) and osmotrophs (Aphagea), is robustly resolved, with the ploeotid *Olkasia* as its sister group, forming the new taxon Olkaspira. Ploeotids are paraphyletic, although Ploeotiidae (represented by *Ploeotia* spp.), *Lentomonas*, and *Keelungia* form a robust clade (new taxon Alistosa). Petalomonadida branches robustly as sister to other euglenids in outgroup-rooted analyses. Within Spirocuta, Euglenophyceae is a robust clade that includes *Rapaza*, and Anisonemia is a well-supported monophyletic group containing Anisonemidae (*Anisonema* and *Dinema* spp.), '*Heteronema II*' (represented by *H. vittatum*), and a clade of *Neometanema* plus Aphagea. Among 'peranemid' phagotrophs, *Chasmostoma* branches with included *Urceolus*, and *Peranema* with the undescribed '*Jenningsia II*', while other relationships are weakly supported and consequently the closest sister group to Euglenophyceae remains unresolved. Our results are inconsistent with recent inferences that *Entosiphon* is the evolutionarily pivotal sister either to other euglenids, or to Spirocuta. At least three transitions between posterior and anterior flagellar gliding occurred in euglenids, with the phylogenetic positions and directions of those transitions remaining ambiguous.

1. Introduction

Euglenida (Discoba; Euglenozoa) is a major and diverse group of unicellular microbial eukaryotes that inhabit freshwater, soil, and marine environments. Euglenids show a variety of trophic modes, including phototrophy, osmotrophy, and phagotrophy (Leander et al., 2017), with the phototrophic clade (Euglenophyceae) arising through a secondary endosymbiosis involving a phagotrophic euglenid host and a pyramimonadalean green alga (Jackson et al., 2018; Turmel et al., 2009). The main morphological apomorphy is the euglenid pellicle, consisting

* Corresponding author at: Dalhousie University, PO Box 15000, Halifax, NS B3H 4R2, Canada.
E-mail address: Alastair.simpson@dal.ca (A.G.B. Simpson).
[1] Present address.

https://doi.org/10.1016/j.ympev.2021.107088
Received 20 August 2020; Received in revised form 24 January 2021; Accepted 26 January 2021
Available online 2 February 2021
1055-7903/

of 4–100+ connected proteinaceous strips that run longitudinally or spirally beneath the cell membrane (Leander et al., 2017). The strips slide actively against each other in many species, enabling cells to change shape, with many undergoing squirming or peristalsis-like deformations called 'euglenoid motion' or 'metaboly' (Leander et al., 2001b, 2017). Euglenids exhibit a vast diversity in morphology, ranging from rigid cells with few but often elaborately shaped pellicle strips (e.g. some petalomonads), to large hyper-flexible cells with dozens of similar strips (e.g. many phototrophic euglenids). Almost all euglenids have one or two flagella. Most phagotrophic euglenids are poor swimmers, however, and use their flagella mainly in a gliding motility, though the particular flagella on which the cells glide are not the same across taxa (Leander, 2004; Leander et al., 2017; see below).

Phototrophic euglenids and primary osmotrophs (Aphagea) each represent constrained clades with reasonably- to well-resolved internal relationships (Karnkowska et al., 2015; Busse and Preisfeld, 2003; Preisfeld et al., 2001). By contrast, phagotrophs are a sprawling, paraphyletic assemblage (Lax et al., 2019;; Paerschke et al., 2017; Cavalier-Smith, 2016). Consequently, resolving deep-level euglenid phylogeny and evolution is fundamentally a question of understanding the relationships among phagotrophs. To date, almost every molecular phylogeny encompassing all euglenids has examined only the small subunit ribosomal RNA gene (SSU rDNA), but this marker still had very poor sampling of phagotrophic euglenids for most of the molecular era. Recent cultivation and (especially) cultivation-independent approaches have substantially increased the number and taxonomic breadth of phagotrophs in the SSU rDNA dataset (e.g. Lax et al., 2019; Cavalier-Smith et al., 2016; Schoenle et al., 2019; Lax and Simpson, 2020). Nonetheless, the SSU rDNA is divergent in many euglenids (Busse and Preisfeld, 2003; Busse et al., 2003; Łukomska-Kowalczyk et al., 2016; Cavalier-Smith et al., 2016), constraining its value as a phylogenetic marker. A couple of studies instead used Hsp90 sequences to infer euglenid phylogeny, but have included very few (<7) species of phagotrophs (Breglia and Leander, 2007; Cavalier-Smith et al., 2016).

Single-gene phylogenies do resolve some deeper-level relationships among euglenids, such as supporting the monophyly of Spirocuta, also known as Helicales (Lee and 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 and Farmer, 2001) and includes taxa with 14 or more pellicle strips, many of which are highly flexible. It encompasses both Euglenophyceae and Aphagea as well as various phagotrophs (Lax and Simpson, 2020; Leander, 2004; Leander et al., 2007, 2017). Most spirocute phagotrophs belong in one of two general morphological categories: anisonemids and peranemids (Leander et al., 2017). Anisonemids glide on their posterior flagellum only and can be moderately flexible (*Anisonema* and *Dinema*), whereas peranemids glide with the anterior flagellum in contact with the surface (the posterior flagellum is often not even emergent) and are usually highly metabolic (e.g., *Peranema, Jenningsia*, and *Urceolus*). Most relationships among phagotrophs within Spirocuta are poorly resolved, as is the sister group to phototrophic euglenids (Lax and Simpson, 2020).

The 'deeper-branching' euglenids outside Spirocuta are all phagotrophs, and can also be divided into two morphological categories: ploeotids and petalomonads. Ploeotids are rigid cells with a few pellicle strips (usually 10), which glide on their posterior flagellum (*Ploeotia, Serpenomonas, Entosiphon, Keelungia, Decastava, Lentomonas, Olkasia, Liburna*, and *Hemiolia*; Lax et al., 2019). They form a paraphyletic group in SSU rDNA phylogenies (Lax et al., 2019; Cavalier-Smith, 2016; Paerschke et al., 2017; Schoenle et al., 2019), though the relationships among ploeotid genera vary substantially across analyses, with the only robust grouping being a *Ploeotia* + *Serpenomonas* clade (Ploeotiidae; Lax et al., 2019, Lax and Simpson, 2020). Petalomonads (*Petalomonas, Notosolenus, Scytomonas, Sphenomonas* and several others) instead form a clade in SSU rDNA phylogenies (Lee and Simpson, 2014a; Cavalier-Smith, 2016; Lax and Simpson, 2020). They are rigid cells with 4–10 pellicle strips (or equivalents), and glide on their anterior flagellum (Leander et al., 2017). Morphological analyses and early molecular phylogenies placed petalomonads as the deepest branch(es) within euglenids (Leander et al., 2001a; Breglia and Leander, 2007; Montegut-Felkner and Triemer, 1997; Müllner et al., 2001; Preisfeld et al., 2000, 2001), but more recent single-gene phylogenies usually place them among ploeotids, generally with weak statistical support (Lax et al., 2019; Lax and Simpson, 2013; Chan et al., 2013; Schoenle et al., 2019; Cavalier-Smith, 2016; Cavalier-Smith et al., 2016; Lax and Simpson, 2020).

Finally, symbiontids (Symbiontida/Postgaardea, including *Postgaardi*, *Calkinsia* and *Bihospites*) are enigmatic flagellates that inhabit low-oxygen saline environments (Yubuki et al., 2009; Breglia et al., 2010; Edgcomb et al., 2010; Yubuki and Leander, 2018). They host various episymbiotic bacteria (Edgcomb et al., 2010), in some cases including magnetotactic Deltaproteobacteria that enable the consortium to orient to the geomagnetic field (Monteil et al., 2019). Symbiontids are clearly euglenozoans, yet lack the defining morphological synapomorphies of euglenids, kinetoplastids, or diplonemids (Yubuki and Leander, 2018; Yubuki et al., 2009, 2013; Simpson et al., 1997), although *Bihospites* has S-shaped cell-surface folds reminiscent of the euglenid pellicle (Yubuki et al., 2013; Breglia et al., 2010). Most SSU rDNA phylogenies place symbiontids either within euglenids, or as sister to them, usually with poor support either way (e.g. Breglia et al., 2010; Cavalier-Smith, 2016; Lax and Simpson, 2013, 2020). These conflicting results have led researchers to treat symbiontids either as derived euglenids (e.g. Breglia et al., 2010; Lee and Simpson, 2014a), or as a distinct euglenozoan group (e.g. Simpson et al., 1997; Cavalier-Smith, 2016).

In an effort to increase taxon sampling among phagotrophic euglenids, especially spirocutes, Lax and Simpson (2020) analysed ~70 SSU rDNA sequences obtained with single-cell approaches, and also divided phagotrophic euglenids (and symbiontids) into 34 groups of convenience, almost all of which were clades with moderate support or better, or single sequences (with these groups given letter codes A through Z, and α through θ, that are also used in the present study). Here, we generated transcriptomes from 28 euglenids that collectively represent most of these groups. The majority were obtained from uncultivated species via single-cell methods, giving a much broader sampling than possible from cultivated strains alone. We used this data for the first multi-gene phylogenetic analyses of euglenids encompassing phagotrophs. Our analyses address several important topics, including various relationships within Spirocuta, the branching order at deeper levels of the euglenid tree, and the placement of symbiontids.

2. Methods

2.1. Single-cell transcriptomics

Most samples were prepared as reported by Larsen and Patterson (1990): Briefly, sediment/detritus was collected from several marine and freshwater sites in Eastern and Western Canada, and Curaçao (Supplementary Table 1). In small trays, the samples were spread to a height of 1–2 cm, and a tissue paper (Kimwipe), then 50 × 20 mm coverslips, were placed on top. The samples were incubated for 12–72 h at room temperature, under ambient sunlight (following a day-night cycle). Coverslips were then examined on an inverted microscope, bottom facing up (Zeiss Axiovert 200 M under 1000× total magnification or Leica DM IL LED under 400× or 630×). Following imaging with a Zeiss AxioCam M5 or Sony NEX6, respectively, cells of interest were isolated using a drawn-out glass pipette, and washed several times in 1–5 μl of sterile seawater, tap water, or distilled water, with aspiration and washing carried out as gently as possible to reduce stress on cells. For *Lentomonas* c.f. *corrugata* LEN2, symbiontid KSa7 and symbiontid HLA12, cells were instead isolated from crude liquid-media enrichments of the natural samples, with a small volume of this fluid transferred to a clean coverslip for the observation, pipette-isolation and washing steps described above.

Similar isolation and washing procedures were also used to harvest cells from three cultures, namely *Peranema trichophorum* strain PtR (Carolina Biological Supplies; 1 cell), *Neometanema parovale* strain KM051 (see Lee and Simpson, 2014b; 2 cells, sequenced separately and co-assembled), and *Ploeotia* sp. strain CARIB1 (4 cells), in order to separate euglenids from their eukaryotic prey (and contaminants). The new culture *Ploeotia* sp. CARIB1 was sourced from subtidal marine sediment from Curaçao (Supplementary Table 1), isolated by drawn-out glass pipette, and grown with *Phaeodactylum tricornutum* in f/2 medium, with transfers every few weeks. Its identity as an (undescribed) *Ploeotia* species was confirmed by light microscopy observations and SSU rRNA gene phylogenies (Lax and Simpson, 2020).

Cells 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, Supplementary Table 2). 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 sample. After library generation with Nextera XT, samples were sequenced on Illumina HiSeq or MiSeq systems, using 2 × 250 bp paired-end sequencing (see Supplementary Table 2 for multiplexing information).

2.2. Transcriptomes from mass-harvested cultures

Cultures of *Ploeotia vitrea* strain MX-CHA (Lax et al., 2019) and *Notosolenus urceolatus* KM049 (Lee and Simpson, 2014a) were mass-cultured in 150 mm Petri dishes. For cell harvesting, most of the medium was discarded, and the dishes were 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. *Keelungia* sp. strain KM082 (Lax et al., 2019) was grown in a 25 cm^2 culture flask as normal, treated by brief vigorous shaking to dislodge cells, then ~10 ml of the culture was centrifuged to concentrate cells. After removing supernatant, RNA was harvested from the remaining ~2 ml of fluid using a NucleoSpin RNA XS column (Macherey-Nagel/Takara) and eluted into 5 μl. In all three cases, purity and quantity of the RNA was assessed with a NanoDrop spectrophotometer (ThermoFisher). The purified RNA was then used as input for the SmartSeq2 protocol and sequencing, following procedures described above.

Cultures of *Petalomonas cantuscygni* strain CCAP 1259/1 and an unnamed *Entosiphon sulcatum* strain (isolated by Mark Farmer, University of Georgia, from a contaminated *Peranema trichophorum* culture from Carolina Biological Supplies, item #131838) were maintained as described previously (Roy et al., 2007). Cells were collected by centrifuging and RNA was extracted with a homemade TRIzol substitute (Rodríguez-Ezpeleta et al., 2009). Residual DNA was removed 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 2 × 250 bp and 2 × 150 bp paired-end reads, respectively. For both organisms reads from both runs were adapter- and quality-trimmed, then co-assembled as described below.

Cells of *Rapaza viridis* (Yamaguchi et al., 2012; same strain as the now-defunct ATCC PRA-360) were collected by centrifugation, and RNA 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 2 × 150 bp paired-end reads. The assembly was done independently of the other data reported here (see below) with Trinity under default parameters (Haas et al., 2013) and translated to protein sequences with TransDecoder. The transcriptome and its analysis will be reported fully elsewhere (Yubuki et al., unpublished).

2.3. Assembly of transcriptomic data

Raw reads from the single-cell transcriptomes, as well as the transcriptomes derived from cultures (except *R. viridis*), were corrected using rcorrector version 1.0.4 (Song and Florea, 2015), quality- and adapter-trimmed with Trimmomatic version 0.39 with default parameters (Bolger et al., 2014), or Cutadapt (Martin, 2011; *E. sulcatum* and *P. cantuscygni* only), and assembled with rnaSPAdes version 3.13.1, under default parameters (Bushmanova et al., 2019). In some cases—due to assembly errors in rnaSPAdes (Supplementary Table 2)—corrected and trimmed reads were subsequently re-assembled with Trinity version 2.4.0. 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). Our single-cell transcriptomes varied considerably in terms of BUSCO scores and phylogenomic-pipeline coverage (Supplementary Tables 2 and 3, see below), with the lowest-quality assemblies—*Heteronema vittatum* CB2 and *Ploeotia* sp. CARIB1—recovering 17 BUSCOs each (complete, duplicate, and fragmented) and 15% and 26.2% of sites (respectively) of a previously published 351-gene phylogenomic dataset (Brown et al., 2018). By contrast, the highest-quality single-cell assembly (*Peranema trichophorum* PtR) included 144 BUSCOs and 57.3% of the phylogenomic marker sites from Brown et al. (2018). All metrics, BUSCO scores, assembly strategies and multiplexing information for individual samples are given in Supplementary Table 2.

2.4. Dataset construction

All new transcriptome assemblies were added to the 104-taxa, 351-gene eukaryote-wide dataset from Lax et al. (2018) 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 (Supplementary Table 3). All analyses were performed on inferred amino acid translations. After addition of new taxa, 351 single-gene alignments were generated with MAFFT L-INS-i version 7.0 (Katoh and Standley, 2013), trimmed with BMGE version 1.0 (-m BLOSUM30 -h 0.5 -g 0.2; Criscuolo and 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 inspected for obvious contaminant, paralogous, long-branching, or otherwise aberrant sequences, which were removed from the dataset. After re-aligning and re-trimming, single-gene phylogenies were re-inferred and inspected a second time.

A final dataset was constructed by filtering all 351 genes by taxon-completeness (threshold of ≥70% taxa present), then choosing the genes whose trimmed alignment lengths exceeded 250 amino acids (aa; 20 inferred proteins retained). This was done with the aim of having sufficient phylogenetic signal in the final single-gene phylogenies (see below), to reliably detect paralogs and other aberrant sequences. To attempt to recover additional orthologs that our pipeline had overlooked, we searched for each gene that was absent from the 20 gene dataset using blastx against the relevant transcriptome, with *Euglena gracilis, Eutreptiella gymnastica* or *Trypanosoma grayi* as query sequences, 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, as before.

The final 'base' dataset contained 20 genes and retained 6289 aa sites after masking. It included 25 phagotrophic euglenids, 4 euglenophyceans including *Rapaza viridis*, 1 aphagean (*Rhabdomonas costata*), 2 symbiontids, 8 kinetoplastids, 6 diplonemids, and 9 additional discobid taxa to act as outgroups to Euglenozoa, for a total of 55 taxa. A full list of data used and their sources is shown in Supplementary Table 3.

2.5. Phylogenetic 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 Posterior Mean-Site Frequency model (PMSF; Wang et al., 2017) in IQ-TREE, we additionally ran 200 non-parametric bootstraps under LG + C60 + F + Γ ('PMSF-BS'). A Bayesian analysis was performed in PhyloBayes version 4.1 (Lartillot et al., 2009) under the CAT + GTR model, with 4 parallel chains run for 30,000 generations and burnin = 300 (maxdiff = 0.226; convergence of all four chains was confirmed by comparing their separately computed consensus trees and posterior probabilities).

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 (the taxa removed had significantly longer branches than others), resulting in the 51-taxa 'noLB' dataset. (ii) RogueNaRok (RNR algorithm and strict consensus settings; Aberer et al., 2013) identified the following as rogue taxa in the 'base' analysis: 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 symbiontids to generate the 30-taxa 'Euglenida-only' dataset, plus two subvariants that additionally excluded the long-branching petalomonad *P. cantuscygni* (29 taxa; 'Euglenida-noPcant'), or excluded all three petalomonads (27 taxa; 'Euglenida-noPetalo'). Phylogenetic trees for these additional datasets were estimated under the LG + C60 + F + Γ model with 1000 UFB replicates. (v) 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 (using a custom script). At each step, a LG + C20 + F + Γ phylogeny with 1000 UFB replicates was inferred.

3. Results

The 20-gene, 55-taxa analysis included new transcriptomes from 28 euglenids and symbiontids, 22 of which were examined using single-cell methods, with most (19) of these being uncultivated species. Morphological measurements and movement patterns of isolated cells can be found in Supplementary Table 4. Most of the labelled groups in the SSU rDNA phylogenies from our companion publication (Lax and Simpson, 2020) had at least one representative taxon in the multigene analyses (Supplementary Table 6). For ease of reference, the colour codes and group lettering used here are the same as in Lax and Simpson (2020): Individual lineages of phagotrophic euglenids (at approximately the 'genus level'), plus symbiontids, are represented by Latin or Greek letters, while colours denote larger potential clades.

3.1. Deep phylogeny of euglenids and the position of Symbiontida

Our primary analyses that included outgroups to euglenids all returned broadly similar results (Figs. 1 and 2, Supplementary Figs. 1-3). Euglenozoa is highly-to-maximally supported (100% UFB, 96% PMSF-BS, 1 pp), and divides into two main subclades, one being Euglenida (95–99% UFB, 93% PMSF-BS, 1 pp) and the other including both diplonemids and kinetoplastids. Symbiontida always branches in the latter clade, with moderate to high support (92–99% UFB, 84% PMSF-BS, 1 pp). Symbiontids are usually sister to a diplonemid-kinetoplastid clade (Glycomonada), though Glycomonada monophyly to the exclusion of Symbiontida is often weakly supported (78–91% UFB; 53% PMSF-BS, 0.95 pp), and in some steps of the 'FSR-removal' analysis, symbiontids instead fall sister to Kinetoplastea (with 84% UFB support, or less, Supplementary Figure 5).

Spirocuta forms a major clade within euglenids, and is always fully supported. Ploeotids form 3 or 4 clades at the base of Spirocuta. *Olkasia polycarbonata* falls sister to Spirocuta with high support (95–99% UFB, 92% PMSF-BS, 0.99 pp; though see below), forming the grouping we refer to as Olkaspira (see below). *Entosiphon sulcatum* (sometimes with *Liburna glaciale* – see below) branches as sister to Olkaspira, but with low support (45–71% UFB, 69% PMSF-BS). *Ploeotia (Ploeotia vitrea, Ploeotia* sp.), *Keelungia* sp., and *Lentomonas* c.f. *corrugata* form a well-supported clade that we refer to as Alistosa (94–100% UFB, 97% PMSF-BS, 1 pp; see below). The final ploeotid group, *Liburna,* is unstably positioned relative to *Entosiphon* and Alistosa (Figs. 1, 2). At the base of euglenids, *Petalomonas cantuscygni, Notosolenus urceolatus*, and *Sphenomonas quadrangularis* form a maximally supported Petalomonadida clade (Fig. 1). Support for the basal position of petalomonads remained maximal when long-branching taxa were removed (limiting petalomonads to *N. urceolatus* and *S. quadrangularis*; Fig. 2A), or rogue taxa were omitted (Fig. 2B), or when the most distant outgroups were excluded (Fig. 2C). The basal position of petalomonads was also strongly supported through the FSR-analysis (lowest 97% UFB with 52% of the sites retained, Supplementary Figure 5).

Analysis of the unrooted 'Euglenida-only' dataset resulted in a poorly supported deep topology that was incongruent with the rooted analyses, with a weakly supported bipartition dividing Spirocuta and petalomonads from all ploeotids (55% UFB; Supplementary Figure 4A), and with *Olkasia* branching sister to *Liburna* within a clade of ploeotids. Support for this aberrant topology was even lower when the longest-branching petalomonad *Petalomonas cantuscygni* was excluded (Supplementary Figure 4A). When all three petalomonads were excluded, the grouping of *Olkasia* with Spirocuta was restored, with moderate statistical support (86% UFB; Supplementary Figure 4B). This is consistent with the aberrant topology representing phylogenetic error induced by the divergent nature of the petalomonad clade, which appears with a very long basal branch in the absence of outgroups.

3.2. Relationships within Spirocuta

Spirocuta divides into an often poorly supported Euglenophyceae + peranemids grouping (52–66% UFB, 67% PMSF-BS, 0.99 pp), and a highly supported grouping that corresponds well to the taxon Anisonemia Cavalier-Smith, 2016 (99–100% UFB, 99% PMSF-BS, 1 pp). Euglenophyceae (including *Rapaza viridis*) is maximally supported. Peranemids form a clade, but with negligible support due primarily to the poorly-resolved position of *Jenningsia fusiforme* (*Jenningsia I*/clade A). The remaining peranemids form a clade with low to moderate support (78–85% UFB, 83% PMSF-BS, 0.71 pp). Within this, *Peranema* (I) and *Jenningsia II* (H) form a strongly supported clade (99–100% UFB, 100% PMSF-BS, 1 pp), while *Chasmostoma* (J) forms a highly supported clade with *Urceolus* (*Urceolus* I/clade K) in all outgroup-rooted analyses (97–99% UFB, 94% PMSF-BS, 1 pp), though with the caveat that *Urceolus II* and *Urceolus III* (groups α and ε in Lax and Simpson, 2020) are not represented in our multigene analysis (see discussion).

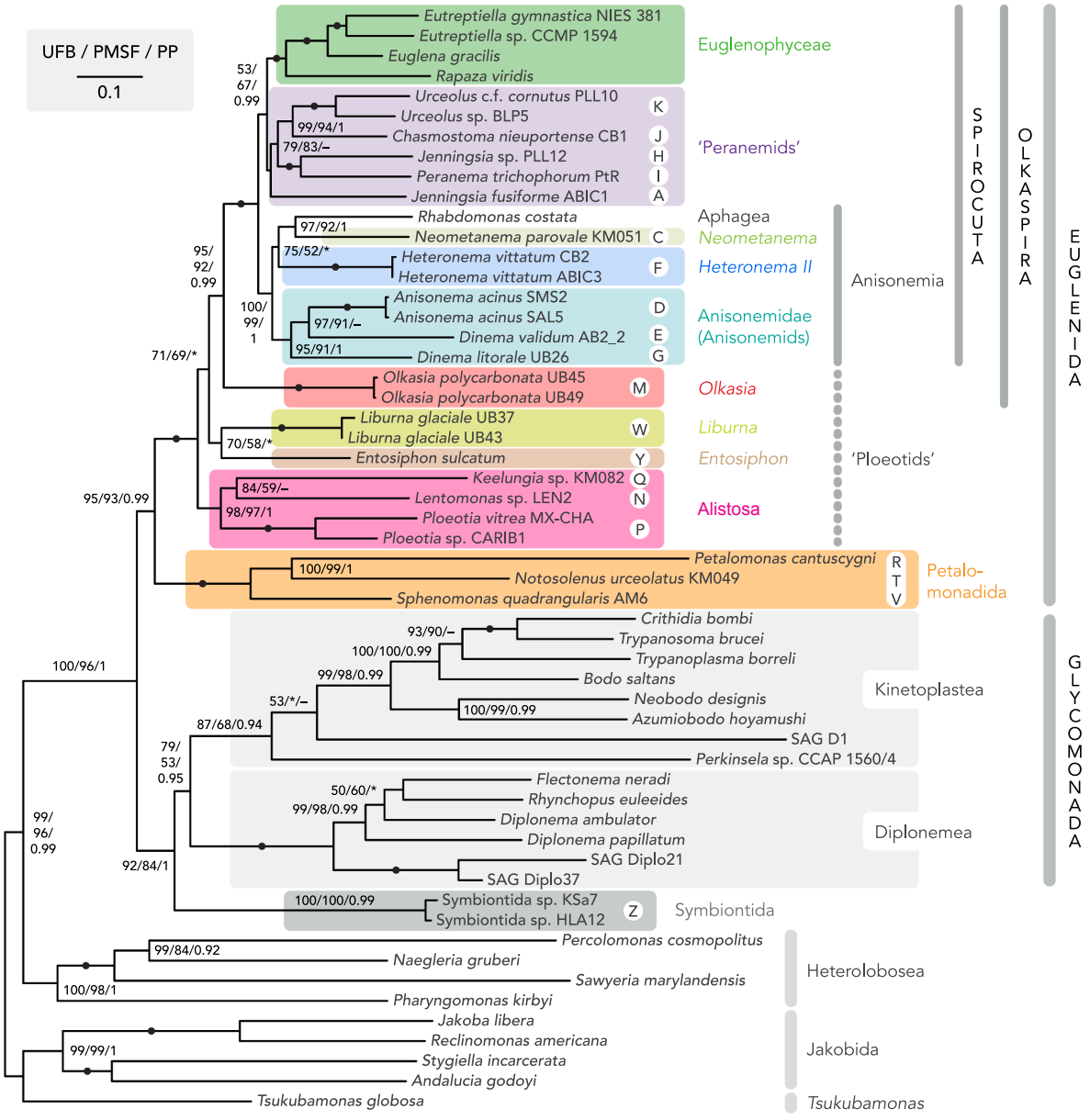

Fig. 1. Phylogeny of Discoba inferred from 20 genes and estimated in the maximum likelihood framework under the LG + C60 + F + Γ model, with robustness assessed with 1000 Ultrafast Bootstrap replicates (UFB), 200 'true' bootstrap replicates (PMSF), and Bayesian Posterior Probabilities (pp, under CAT + GTR model). Major clades of euglenids are shown in coloured/shaded 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 (not shown, or identified with '–'). Asterisk (*) indicates a clade that was not recovered in the Bayesian consensus tree.

Anisonemia is composed of three main groups: a highly supported *Anisonema* + *Dinema* clade ('anisonemids' *sensu stricto*, corresponding well to Anisonemidae *sensu* Cavalier-Smith, 2016; 94–100% UFB, 91% PMSF-BS, 1 pp), a usually well-supported Aphagea + *Neometanema* clade (i.e. *Rhabdomonas costata* with *Neometanema parovale*; 82–99% UFB, 92% PMSF-BS, 1 pp), and *Heteronema vittatum* (i.e. *Heteronema II*/clade F). *Dinema* is recovered as paraphyletic, with *Dinema validum (Dinema II*/clade E) branching closer to *Anisonema acinus*/clade D (89–100% UFB, 91% PMSF, but 0.53 pp) than *Dinema litorale* (*Dinema III*/clade G). *Dinema I* (group β; Lax and Simpson, 2020) is not represented in our analyses, since we lack transcriptomic data.

3.3. Data availability

Raw reads of single-cell transcriptomes are deposited under NCBI SRA accession PRJNA624171, and bulk RNA-Seq reads of *Entosiphon sulcatum* and *Petalomonas cantuscygni* under NCBI accession PRJNA663625. Alignment files and tree files (for the concatenated dataset, as well as single genes before and after removal of problematic sequences), videos, and assemblies of single-cell transcriptomes are deposited on Datadryad, accessible under https://doi.org/10.5061/dryad.tht76hdx7.

4. Discussion

4.1. Large-scale euglenid evolution

The vast majority of euglenid cultures are of phototrophs (Leander et al., 2017), and the few cultures of phagotrophs available for study today represent just a small portion of their known diversity—approximately 10 nominal genera, with about half represented by just

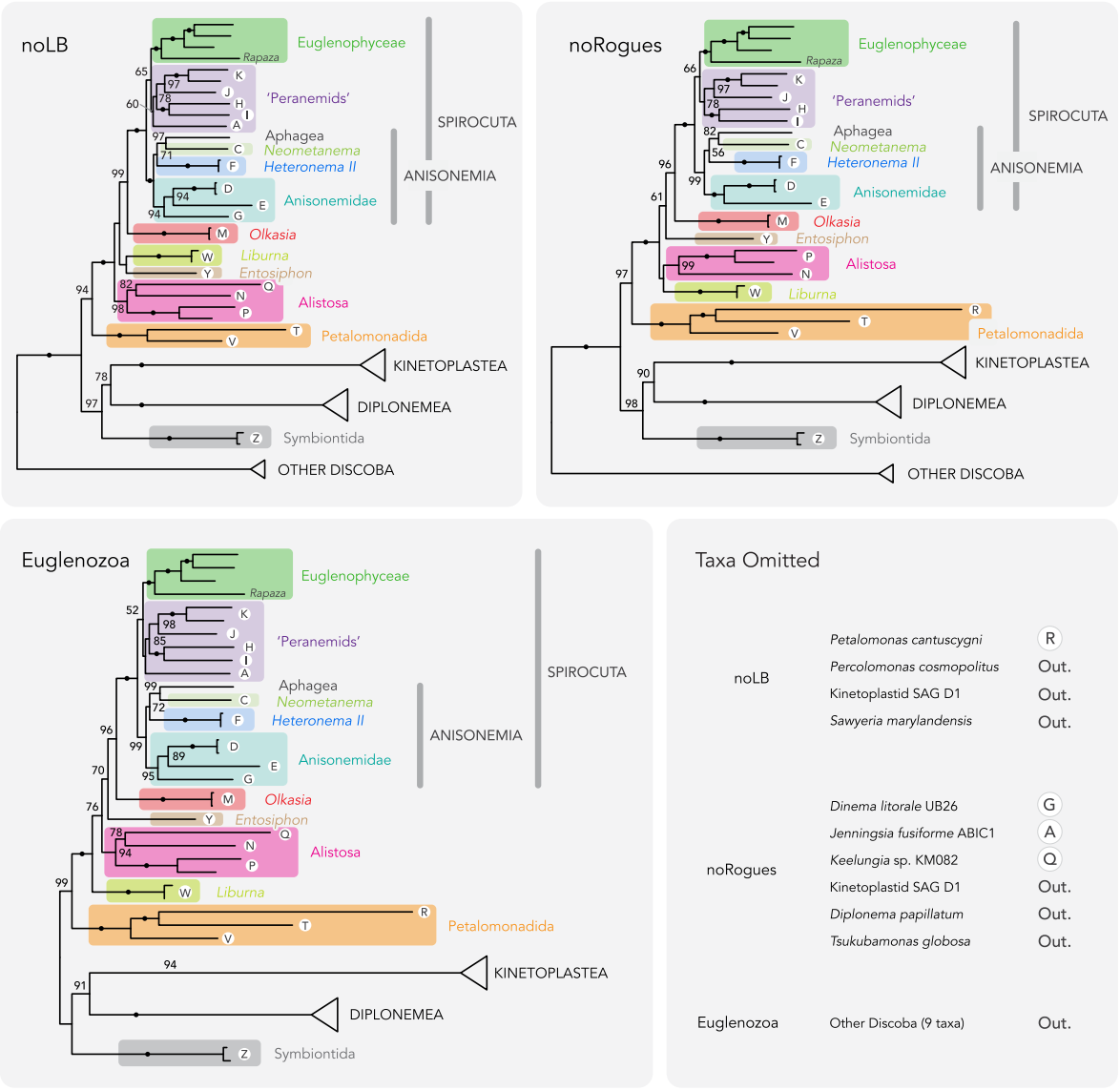

Fig. 2. Comparison of three phylogenomic analyses of euglenids, each inferred from 20 genes and estimated in the maximum likelihood framework under the LG + C60 + F + Γ model, with robustness assessed with 1000 Ultrafast Bootstrap replicates (UFB). **noLB** is the base dataset with four long-branching taxa removed. **noRogues** is the base dataset with six rogue taxa removed, following identification by RogueNaRok (see Methods). **Euglenozoa** is the base dataset with the outgroup of nine discobid taxa removed. In all panels major clades of euglenids are shown in coloured/shaded boxes. Nodes receiving maximum bootstrap support (100%) are denoted by filled circles, while support values below 50% are omitted. The taxa excluded from each analysis are listed in the bottom right panel.

one extant strain (e.g. Lee and Simpson, 2014a, b; Cavalier-Smith et al., 2016). This scarcity, amongst other factors, has historically skewed taxon sampling in SSU rDNA phylogenies heavily towards phototrophic euglenids (e.g. Müllner et al., 2001), a bias only partially overcome by the recent broad application of single-cell SSU rDNA sequencing (Lax et al., 2019; Lax and Simpson, 2020). Meanwhile, multigene phylogenetic analyses of euglenids do not include phagotrophs at all (e.g. Karnkowska et al., 2015), and the occasional inclusion of a single phagotroph, *Peranema*, in broader phylogenomic analyses of eukaryotes (Cavalier-Smith et al., 2014) cannot reveal relationships among phagotrophs. Using a single-cell approach we generated transcriptomes from 19 uncultured phagotrophic euglenids (and symbiontids). Supplemented by transcriptome data from some cultures, this transforms the taxon sampling for markers other than SSU rDNA.

4.2. *Ploeotid clades, Entosiphon and the sister to Spirocuta*

As in SSU rDNA analyses, we recover a Spirocuta clade that includes phototrophs, primary osmotrophs and various phagotrophs (Figs. 1 and 2). Ploeotids are not recovered as a clade, but various genera fall in several groups at the base of Spirocuta, as in SSU rDNA phylogenies (Lax et al., 2019; Lax and Simpson, 2020). The multigene approach offers improved phylogenetic resolution over SSU rDNA studies: *Olkasia* is sister to Spirocuta with robust support, and a clade of several genera is well-supported, uniting Ploeotiidae, *Keelungia,* and *Lentomonas*.

The Spirocuta plus *Olkasia* clade we recover is consistent with some SSU rDNA analyses that place *Olkasia* as the closest described sister to Spirocuta, with weak statistical support (either alone or with the poorly characterised and unidentified ploeotid cell 'SMS7'; Cavalier-Smith, 2016; Lax et al., 2019; Lax and Simpson, 2020). The pellicle strips of *Olkasia* are similar to those typical of Spirocuta in being S-shaped with considerable overhangs (Lax et al., 2019), which may reflect their common ancestry. *Olkasia* also has a chisel-shaped feeding apparatus which is common among phagotrophic spirocutes, and contrasts with the hook-shaped feeding apparatus of most other ploeotids (Lee, 2012; Lax et al., 2019). This last feature needs additional scrutiny, since the

molecularly-sampled ploeotids with hook-shaped feeding apparatuses all belong to the proposed clade Alistosa (Cavalier-Smith et al., 2016; Chan et al., 2013; Lee, 2008; Larsen and Patterson, 1990; Lax et al., 2019; see below), and the phylogenetic positions of other ploeotids with chisel-shaped feeding apparatuses are currently unknown (e.g. *Ploeotia pseudanisonema*, *Ploeotia plumosa*; Larsen and Patterson, 1990; Lee, 2012). Still, the *Olkasia* + Spirocuta clade is now robust and useful enough to merit taxonomic recognition, in our opinion. We propose 'Olkaspira' as a new unranked name for the smallest clade containing *Olkasia* and a representative spirocute (see below).

The grouping of *Keelungia, Lentomonas* and *Ploeotia* into a well-supported clade is consistent with a very poorly supported group of Ploeotiidae (*Ploeotia* and *Serpenomonas*), *Keelungia*, *Lentomonas* and *Decastava* in some SSU rDNA phylogenies (Lax et al., 2019; Lax and Simpson, 2020). We could not include *Decastava* in our multigene analyses, but SSU rDNA phylogenies have placed this taxon as sister to *Keelungia* (Cavalier-Smith, 2016; Paerschke et al., 2017; Schoenle et al., 2019), or as sister to *Lentomonas* once data from *Lentomonas* became available (Lax et al., 2019; Lax and Simpson, 2020), typically with moderate statistical support or better. These taxa all share a similar pellicle structure in that they have strips with bifurcations at their abutting joints, though this is also known in *Entosiphon* (Cavalier-Smith et al., 2016; Chan et al., 2013; Farmer and Triemer, 1994, 1988; Triemer, 1986). They also share a 'hook-like' appearance of the feeding apparatus under light microscopy (Lax et al., 2019), unlike the other ploeotids for which there are molecular data, although hook-like feeding apparatuses are also present in many morphospecies that are currently assigned to *Ploeotia*, but whose true phylogenetic positions are unknown (e.g. *Ploeotia adherens*, *Ploeotia discoides*, *Ploeotia plana*; Larsen and Patterson, 1990; Schroeckh et al., 2003; Lee, 2012). The taxa *Ploeotia*, *Serpenomonas*, *Keelungia*, *Lentomonas,* and *Decastava* were recently divided amongst three different subclass-to-class-ranked taxa that prove not to reflect phylogenetic relationships. *Ploeotia* was grouped with *Lentomonas* only (also equated with the order Ploeotiida), while *Serpenomonas* was in its own separate subclass, and *Keelungia* and *Decastava* were placed together with Petalomonadida in a different subclass (Cavalier-Smith, 2016). Rather than introduce a dramatically different concept for one or other of these taxa, we propose a new (unranked) name for the clade, Alistosa. This taxon is defined as the smallest clade containing *Ploeotia*, *Lentomonas*, and *Keelungia*, but is inferred to also include *Decastava*, as well as *Serpenomonas* (see below).

In recent studies, *Entosiphon* was considered either as the deepest-branching euglenid (Cavalier-Smith, 2016; Cavalier-Smith et al., 2016), or as sister to Spirocuta (Paerschke et al., 2017). These incompatible inferences were both based partly on SSU rDNA phylogenies, but supported in the first case by unavoidably taxon-sparse Hsp90 phylogenies as well as interpretations of feeding apparatus morphology (Cavalier-Smith et al., 2016), and in the second case by the inferred phylogenetic distributions of paramylon bodies and an SSU rRNA secondary structure (Paerschke et al., 2017). In our rooted analyses these important phylogenetic positions are instead taken by Petalomonadida and *Olkasia,* respectively. Instead, we generally recovered *Entosiphon* as a sister to the *Olkasia* + Spirocuta clade (Olkaspira), albeit with poor support. The extremely unstable position of *Entosiphon* in SSU rDNA phylogenies is likely due to their exceptionally divergent SSU rDNA sequences (Cavalier-Smith et al., 2016; Lax et al., 2019), whereas *Entosiphon* is not particularly long-branching in our multigene analyses, suggesting a different cause of the (more modest) phylogenetic uncertainty. Future research may test whether *Entosiphon*'s position is more robustly resolved with improved gene sampling and/or better taxon sampling for ploeotids in multigene datasets.

4.3. The deep phylogenetic position of petalomonads

Petalomonads were proposed by some researchers as the deepest branch or branches among euglenids on the basis of various morphological or molecular data, giving them a special importance in understanding euglenid evolution (Leander, 2004; Leander et al., 2007, 2001a; Leander and Farmer, 2001; Triemer and Farmer, 1991; Breglia et al., 2007). Recent SSU rDNA phylogenies give various—often poorly supported—positions for petalomonads, sometimes placing them close to the base of euglenids (e.g. Chan et al., 2013; Paerschke et al., 2017) but more often nested within ploeotids (e.g. Lax et al., 2019; Schoenle et al., 2019; Cavalier-Smith, 2016; Lax and Simpson, 2020). Our multigene analyses instead recover petalomonads as the deepest branch among euglenids, with strong statistical support. The petalomonad taxa appear to have relatively rapidly evolving nuclear protein-coding genes, raising the possibility that their deep-branching phylogenetic position might be a long-branch attraction (LBA) artefact. To try to limit the influence of LBA, we used site-heterogeneous mixture models (LG + C60 + F + Γ and CAT-GTR) and conducted several additional analyses, including exclusion of the longest-branching taxa (one being *Petalomonas cantuscygni)*, and progressive removal of rapidly-evolving sites. All of the additional analyses still recover the basal position of petalomonads with high-to-full support (Supplementary Table 5). While we strongly advocate further careful examinations of the phylogenetic placement of petalomonads, this robustness makes a basal position for the petalomonad clade the best-supported hypothesis at present for the deepest divergence within euglenids. Within petalomonads, we recover the same structure as in SSU rDNA phylogenies, with *Sphenomonas* sister to other petalomonads (Lax and Simpson, 2020).

A deep position for petalomonads implies a pivotal importance for understanding euglenid evolution. For example, based on the low number of pellicle strips in petalomonads, it was suggested that early euglenids had very few strips (e.g. four or eight), and through various evolutionary changes (including inferred strip-doubling events), they gave rise to forms with more pellicle strips. This later resulted in the emergence of highly metabolic taxa with dozens of pellicle strips (Leander et al., 2001a; Leander and Farmer, 2001; Leander, 2004; Leander et al., 2007). Conversely, some recent inferences about deep euglenid evolution are unsupported if petalomonads are basal. For example, Cavalier-Smith (2016) proposed that the last common ancestor of euglenids was a posterior glider and bore a feeding apparatus with well-developed rods, based on *Entosiphon* being sister to other euglenids, and petalomonads being derived from ploeotid ancestors. However, the ancestral states of both characters are equivocal if petalomonads are sister to other euglenids, since petalomonads are anterior gliders and apparently lack feeding apparatus rods (Lee and Simpson, 2014a; Mignot, 1966). Unfortunately, petalomonads are poorly characterised at present: There is one detailed transmission electron microscopy study published (*Notosolenus urceolatus*; Lee and Simpson, 2014a), but only cursory or fragmentary ultrastructural data for the several other investigated species (see Lee and Simpson, 2014a; also Cavalier-Smith et al., 2016, Paerschke et al., 2017) and the phylogenetically important taxon *Sphenomonas* has not been examined by electron microscopy at all. Prior to our study only *P. cantuscygni* has had protein-coding genes sequenced (e.g. Hsp90, EFL, glutamyl-tRNA synthase, triosephosphate isomerase; Lakey and Triemer, 2017; Markunas and Triemer, 2015; Breglia and Leander, 2007; Gile et al., 2009), plus a preliminary study of mitochondrial DNA architecture (Roy et al., 2007). More detailed examinations across the diversity of petalomonads may be crucial to better understanding early euglenid evolution.

4.4. Closest living relatives to Euglenophyceae

The expansive phototrophic clade (Euglenophyceae) arose following a secondary endosymbiotic event involving a phagotrophic euglenid and a pyramimonadalean green alga (Jackson et al., 2018; Turmel et al., 2009). Our analyses strongly place *Rapaza* in this clade as sister to other known euglenophyceans, as inferred previously from morphological examinations and SSU rDNA phylogenies (Yamaguchi et al., 2012; Cavalier-Smith, 2016). The nature of the phagotrophic host of

Euglenophyceae would be substantially illuminated by identifying their closest living relatives. *Teloprocta* (formerly *Heteronema proto parte*) is one candidate sister to Euglenophyceae, based on an unstable and usually weak affinity in various SSU rDNA analyses in which *Peranema* was the only other included peranemid (Cavalier-Smith, 2016). An earlier morphological analysis instead placed *Urceolus* as the sister to Euglenophyceae, based partly on a paraflagellar swelling and potential stigma similar to the eyespot of phototrophic euglenids (Leander et al., 2001a). By contrast, our multigene analysis places 'peranemids' as a whole as sister to Euglenophyceae, rather than any particular genus-level group. This 'peranemids' clade does include *Urceolus*, but together with *Chasmostoma,* two distinct clades of *Jenningsia,* and *Peranema* (see below). This topology is consistent with Euglenophyceae arising relatively early in spirocute evolution. Yet the position of *Jenningsia I* (clade A) is uncertain: In most analyses it branches sister to all other peranemids, but with no support and with its position (and thus peranemid monophyly) being poorly supported throughout the fast-site removal analysis (Supplementary Figure 5). As such, *Jenningsia I* remains a possible candidate for a closer relative of phototrophic euglenids than the other peranemids sampled here. Tantalisingly, recent taxon-broad SSU rDNA phylogenies sometimes recovered *Jenningsia I* (clade A), *Heteronema* (c.f.) *globuliferum* (*Heteronema I*/group B) and *Teloprocta* (clade L) as a clade that is sister to Euglenophyceae, though statistical support for this position, and for the A + B + L clade itself, was very weak at best (Lax and Simpson, 2020). Obtaining multigene data from these other peranemids should be a high priority in the search for the closest relatives of Euglenophyceae.

4.5. Other relationships within Spirocuta

Despite the overall poor resolution of SSU rDNA phylogenies of euglenids, two deep-level relationships within Spirocuta that have usually been supported, are (1) a clade of *Anisonema*, various *Dinema* lineages, *Neometanema* and Aphagea, and (2) *Peranema* as the sister group to all other spirocutes (e.g. Busse et al., 2003; Lee and Simpson, 2014; Cavalier-Smith, 2016; Lax et al., 2019). The first is strongly confirmed by our multigene phylogenies, including the addition of *Heteronema II* (clade F, e.g. *H. vittatum)*, for which sequence data have only recently been obtained (Lax and Simpson, 2020). This monophyletic group corresponds as closely as possible with current molecular sampling to the taxon Anisonemia (Cavalier-Smith, 2016). One unresolved issue is whether *Heteronema vittatum* can remain assigned to *Heteronema*, or will require a new genus, which is tangled in the complex taxonomic status of *Heteronema* (see Larsen and Patterson, 1990; Cavalier-Smith, 2016; Lax and Simpson, 2020). Within Anisonemia, the generally robust clade of *Neometanema* and Aphagea in multigene analyses reinforces most SSU rDNA phylogenies, which also show this grouping (though sometimes with equivocal statistical support; Lax and Simpson, 2013, 2020; Lee and Simpson, 2014a,b; Cavalier-Smith, 2016). This confirms the importance of *Neometanema* for understanding the evolution of primary osmotrophic euglenids (Lee and Simpson, 2014a) and supports the concept of a taxon to include them both (Natomonadida; Cavalier-Smith, 2016).

By contrast, a basal position for *Peranema* within Spirocuta was not supported by our multigene analyses. Instead *Peranema* appears robustly as sister to *Jenningsia II* (clade H), and most probably forms a clade with some/all other 'peranemids', and perhaps Euglenophyceae, to the exclusion of Anisonemia (see above). Until recently, the SSU rDNA database had a very poor sampling of peranemids, with taxa assigned to *Jenningsia*, *Urceolus* and *Chasmostoma* missing entirely. Interestingly, new SSU rDNA phylogenies that include most/all of these taxa do not support the basal position of *Peranema* within Spirocuta, or do so with only weak support (Lax and Simpson, 2020). This is consistent with the previous stronger support being an artefact due to poor taxon sampling, albeit some of the new SSU rDNA sequences are relatively divergent, and this may also depress support. A close relationship between *Peranema* and *Jenningsia II* was not strongly contradicted by these SSU rDNA phylogenies, since the position of *Jenningsia II* was also poorly resolved (Lax and Simpson, 2020). In short, an isolated position for *Peranema* at the base of Spirocuta appears unlikely at present, and should not be assumed when inferring euglenid evolutionary history, or for systematics.

While *Urceolus* is one of the most distinctive and well-known phagotrophic euglenid genera, its biodiversity and phylogenetic affinities are poorly understood. The first SSU rDNA sequences of *Urceolus* were reported recently, and placed *Urceolus* in an unresolved position within Spirocuta (Lax and Simpson, 2020). Further, the monophyly of *Urceolus* was not demonstrated, with sequences forming an (unsupported) clade in some analyses but up to three separate lineages otherwise (*Urceolus I–III*; Lax and Simpson, 2020). *Chasmostoma* is a lesser-known and rarely observed taxon with a single described species (Lee et al., 1999). The SSU rDNA sequence extracted from the *Chasmostoma* CB1 transcriptome is very divergent, but branched without support with some *Urceolus* (II and III) in a maximum likelihood analysis (Lax and Simpson, 2020). Our multigene phylogenies instead recover a relatively robust clade of *Chasmostoma* and *Urceolus,* represented in the dataset by *Urceolus I*. Both *Urceolus* and *Chasmostoma* are highly metabolic and glide on their single emergent anterior flagellum, with at least half of the flagellum in direct contact with the substrate, although this also describes *Jenningsia* (Lee et al., 1999; Lax and Simpson, 2020). *Urceolus* however, is distinguished by a flared collar that projects laterally at the anterior end of the cell beyond the opening of the flagellar canal (Larsen and Patterson, 1990; Leander et al., 2001a). *Chasmostoma* instead possesses a distinctive flagellar cavity that extends anterior to the canal (Lee et al., 1999). The flagellum normally extends from the narrow apical opening of the cavity, but can retract and curl up within it (Lee et al., 1999). Since the only two phagotrophic spirocutes with conspicuous anterior pellicle-supported structures appear to be specifically related, we suggest that the collar of *Urceolus*, and the walls of the flagellar cavity of *Chasmostoma* are homologous. It is possible that the cavity of a *Chasmostoma*-like common ancestor flared outwards to form the collar in an ancestor of *Urceolus* (also see Lax and Simpson, 2020). It is also possible that the common ancestor had a *Urceolus*-like flared collar, which constricted to form the flagellar cavity in *Chasmostoma*. With respect to this second scenario, current data does not reject the possibility that *Urceolus* is a paraphyletic group that gave rise to *Chasmostoma*. Expanding multigene sampling to encompass the *Urceolus II* and *Urceolus III* lineages would be valuable for testing the monophyly of *Urceolus*.

4.6. Are symbiontids an independent branch of Euglenozoa?

Some symbiontid morphospecies have been known for several decades (Lackey, 1960; Fenchel et al., 1995), but molecular (SSU rDNA) data only became available relatively recently (initially for *Calkinsia aureus* and *Bihospites bacati*; Yubuki et al., 2009; Breglia et al., 2010), and is currently dominated by environmental sequences (Monteil et al., 2019; Yubuki and Leander, 2018; Orsi et al., 2011). SSU rDNA phylogenies place symbiontids as a monophyletic group of uncertain affinity within Euglenozoa, recovered variously as sister to kinetoplastids (e.g. Yubuki et al., 2009), sister to euglenids (e.g. Monteil et al., 2019; Schoenle et al., 2019; Cavalier-Smith, 2016; Lax and Simpson, 2020), sister to diplonemids and kinetoplastids (Cavalier-Smith, 2016), a branch within euglenids (e.g. Paerschke et al., 2017; Lax and Simpson, 2013, 2020), or with unresolved affinities (Breglia et al., 2010). As a result, there are competing views on whether symbiontids are derived euglenids (that presumably lost the euglenid pellicle; e.g. Breglia et al., 2010), or a separate clade of Euglenozoa (e.g. Cavalier-Smith, 2016), with these hypotheses based partly on morphological considerations, in addition to SSU rDNA phylogenies.

Our study is the first to include multigene data for symbiontids. In our analyses, symbiontids never fall within euglenids, or even sister to them, but are most closely related to diplonemids and kinetoplastids.

While our study argues against symbiontids being derived euglenids (with high statistical support for euglenid monophyly throughout), their exact phylogenetic placement remains incompletely resolved: In most phylogenetic analyses they are sister to a diplonemid + kinetoplastid clade (Glycomonada) often with high support, yet this latter clade is unstable across the FSR-analysis, with a weak symbiontid + kinetoplastid clade often recovered instead (Supplementary Figure 5). The quality of our symbiontid single-cell transcriptomes is comparatively low (Supplementary Tables 2 and 3), and the resulting low coverage of the alignment may play a role in their uncertain placement. Furthermore, the two examined cells (HLA12 and KSa7) span only a small part of the known phylogenetic diversity of symbiontids, as both fall within 'Symbiontida clade d' sensu Yubuki and Leander (2018) in SSU rDNA analyses (Lax and Simpson, 2020). Therefore, higher quality multigene data from a broader sampling across symbiontid diversity is needed to more precisely place this group. As no symbiontids have been cultured so far, additional high-quality single-cell transcriptomes, or perhaps single-cell genomes (see Monteil et al., 2019), will likely be key.

4.7. Motility across euglenids

Euglenids employ several mechanisms of locomotion. Euglenophyceans tend to swim, either with a single or multiple flagella, whereas phagotrophs are overwhelmingly surface-associated and glide on their flagella (Leander, 2004; Leander et al., 2017; see Supplementary material, videos A1-A4). This gliding smoothly transports cells across surfaces independently of flagellar beating activity. The underlying molecular mechanisms are not known, but a study of *Peranema* localised the gliding machinery along the adhering portion of the driving flagellum (anterior in this species), and recorded speeds around 30 μm/sec (Saito et al., 2003). Some phagotrophic euglenids glide on their anterior flagellum (e.g. *Petalomonas, Notosolenus, Jenningsia, Urceolus, Peranema*; Supplementary material, videos A1, A2), whereas others glide on their posterior flagellum (e.g. *Anisonema, Dinema, Ploeotia, Lentomonas*; Supplementary material, videos A3, A4). This system and its diversity is fascinating because the gliding speeds can greatly exceed the maximum linear speeds of individual molecular motors and because flagellar gliding mechanisms localised to anterior and posterior flagella would have opposite polarities relative to the flagellum cytoskeleton. *Neometanema* instead performs 'skidding', a form of swimming where the posterior flagellum maintains contact with the surface while the beating of the anterior flagellum propels the cell forward (Lee and Simpson, 2014b).

Anterior gliding is present in two large assemblages of phagotrophs, since anterior-gliding petalomonads branch at the base of the euglenid tree, while peranemids lie within Spirocuta (Figs. 1, 2 and Supplementary Figs. 1-4). Their flagellar movement patterns differ, with petalomonads moving only the absolute distal tip of the flagellum (mostly up-and-down) with the remainder staying attached to the substrate (Supplementary material, video A1). Different peranemids, by contrast, keep varying lengths of their anterior flagellum in contact with the surface, with some taxa beating as much as the distal 1/2 of their flagellum, and with a more sinuous motion of the free portion (Supplementary material, video A2). In both groups there are some taxa in which the posterior flagellum is emergent and in contact with the substrate, but if so it is shorter than the anterior flagellum as a rule (especially in petalomonads), and we provisionally assume it is not involved in driving the cell forwards.

Posterior gliders include the ploeotids close to the base of euglenids, and the anisonemids within the Anisonemia clade of Spirocuta. Gliding patterns in those two groups appear very similar, with the posterior flagellum fully in contact with the surface (or at least not beating laterally), while the anterior flagellum beats in front of the cell through its whole length. Other members of Anisonemia have different motilities. In *Heteronema vittatum* the posterior flagellum lies along the surface, but the proximal half or more of the long anterior flagellum (~2x cell length) also attaches to the surface while the distal portion 'sweep-beats' regularly (Supplementary material, video A4). It is inferred that *H. vittatum* is an anterior glider, or that both flagella contribute to forward motion (Lax and Simpson, 2020). *Neometanema* and Aphagea are known *primarily* as swimmers ('skidding' is functionally a form of swimming – see above), with this similarity being consistent with their close relationship in molecular phylogenies (Lax and Simpson, 2013; Lee and Simpson, 2014; Cavalier-Smith, 2016) and reflected in the taxon name Natomonadida (*nato-* to swim, Cavalier-Smith, 2016). However, some species of *Distigma,* which is basal within Aphagea (Busse et al., 2003), reportedly also exhibit flagellar gliding, with the anterior flagellum assumed to be driving (Jahn, 1946). The relative placements of anisonemids, the *Neometanema* + Aphagea clade, and *Heteronema vittatum* are poorly resolved. Irrespective, it is likely that Anisonemia includes one or more subgroups of anterior gliders.

Based on our phylogenies, it is parsimonious to infer that the posterior gliding mode was part of the ancestry of Spirocuta, since it is present across the paraphyletic ploeotid assemblage, and that anterior gliding in peranemids evolved (from posterior gliding) independently of petalomonads (see also Cavalier-Smith, 2016). If petalomonads are indeed the deepest branch within euglenids, then the ancestral form of gliding in euglenids cannot be inferred confidently (unlike in scenarios where *Entosiphon* is inferred to be basal, see Cavalier-Smith, 2016). At our current state of phylogenetic knowledge, it is also unclear whether the last common ancestor of spirocutes was an anterior glider or posterior glider, owing to uncertainty over whether all peranemids form one clade (see above), as well as the occurrence of anterior gliding in some Anisonemia. We prefer the hypothesis that the ancestral spirocute was a posterior glider, and that the examples of anterior gliding within Anisonemia (see above) represent one or two additional origins of this motility mode in euglenids, independently of peranemids (and petalomonads). Nonetheless, further phylogenetic analyses will be required to resolve this issue, ideally supported by more detailed live observations.

4.8. Culture vs. single-cell transcriptomes vs. SAGs

We mainly used a single-cell approach to gather multigene data. This approach enables a broad molecular sampling of euglenid diversity without the need to establish cultures. Most phagotrophic euglenid taxa appear to be relatively 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 (Lax et al., 2019;Lax and Simpson, 2020). Yet, the availability of cultures is also crucial to further our understanding of the biology of euglenids through ultrastructural studies and other detailed examinations.

The single-cell transcriptomics approach we employed (SmartSeq2; Picelli et al., 2014) has been used successfully to investigate the phylogenies of various eukaryote groups, including Amoebozoa (Kang et al., 2017), Holozoa (Hehenberger et al., 2017) and Retaria (Krabberød et al., 2017), as well as to determine the position of Hemimastigophora within eukaryotes (Lax et al., 2018). By contrast, the use of single-cell amplified genomes (SAGs) from eukaryotes in phylogenetics has a somewhat checkered track record. Many eukaryotic genomes are intrinsically difficult to sequence and assemble, due to their size and features such as repeats and long stretches of non-coding sequence (Keeling and 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 rDNA operon was analysed (e.g. Yoon et al., 2011; Ahrendt et al., 2018). Some of the SSU rDNA sequence data examined in our accompanying study did come from SAGs (cells BP3, SDB1, SDB4, UB10; Lax and Simpson, 2020), but the assemblies did not provide sufficient data for

multigene analysis. The 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 the 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-cell transcriptomics approaches tend to generate higher quality data at a lower price than single-cell genomics—at least when used for phylogenomic or multigene phylogenetic analyses, and on larger eukaryotic cells (Kolisko et al., 2014; Lax et al., 2018).

4.9. Does phylogenomics resolve a tree of Euglenids?

Previously published multigene phylogenies of euglenids or Euglenozoa only included phototrophic euglenids (e.g. Yazaki et al., 2017; Karnkowska et al., 2015; Simpson et al., 2006; Hampl et al., 2009), or—more recently—phototrophs and an osmotroph (Butenko et al., 2020). 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 multigene phylogenetic analysis of the group to include phagotrophic euglenids, and thereby include most of the range of euglenid diversity.

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* monophyly is not validated using SSU rDNA (Lax and Simpson, 2020), and we currently only have sampling for one clade of three from this genus. The availability of these taxa will allow further testing of important relationships within Spirocuta; for example, whether peranemids are monophyletic. Our analyses identify ploeotid taxa as making up most of the backbone of the euglenid tree, but still with overall low support for some of the deep branching order. This might partly result from missing taxa, as for example, both *Decastava* and *Hemiolia* are not sampled. 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 (CARR5, SMS7, WF2_3 in Lax et al., 2019), and pointed out that the taxon *Ploeotia* currently contains various nominal species that are morphologically quite different from *Ploeotia vitrea* (and *Ploeotia oblonga*) and for which there are no sequence data at all (see also above). 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.

To summarize, 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.

5. New taxa

Alistosa G. Lax and A.G.B. Simpson

Definition: The smallest crown clade containing *Ploeotia vitrea* Dujardin, 1841 (Ploeotiidae), *Keelungia pulex* Chan and Moestrup 2013, and *Lentomonas corrugata* (Larsen and Patterson) Cavalier-Smith, 2016, but not *Euglena gracilis* Klebs 1883 (Spirocuta; Euglenophyceae), *Olkasia polycarbonata* Lax et al., 2019 and *Entosiphon sulcatum* (Dujardin) Stein 1878. This is a minimum-crown-clade definition with external specifiers.

Etymology: From alistos (Greek adj. 'pickled'). 'PKLD' (pronounced PicKLeD) is an acronym for 'Ploeotiidae, *Keelungia*, *Lentomonas*, *Decastava*', the present inferred composition of the clade.

Reference phylogeny: Fig. 1, this paper. *Keelungia pulex* is closely related to *Keelungia* sp. KM082, and *Lentomonas corrugata* to *Lentomonas* c.f. *corrugata* LEN2, in SSU rDNA phylogenies (Lax et al., 2019; Lax and Simpson, 2020). See Lax et al. (2019, Fig. 7) for inferred composition of the clade.

Inferred Composition: *Ploeotia*, *Serpenomonas*, *Keelungia*, *Lentomonas*, *Decastava*.

Comments: This clade is inferred from molecular phylogenies. This is a zoological name above the level of the family category and as such falls outside the zoological (and botanical) codes of nomenclature.

Olkaspira G. Lax and A.G.B. Simpson

Definition: The smallest crown clade containing *Euglena gracilis* Klebs 1883 (Spirocuta; Euglenophyceae), and *Olkasia polycarbonata* Lax et al., 2019 but not *Entosiphon sulcatum* (Dujardin) Stein 1878 and *Ploeotia vitrea* Dujardin 1841 (Ploeotiidae). This is a minimum-crown-clade definition with external specifiers.

Etymology: A portmanteau of '*Olkasia*' and 'Spirocuta', representing the two primary subclades.

Reference phylogeny: Fig. 1, this paper. See Lax and Simpson (2020, Fig. 1) for inferred composition of Spirocuta.

Inferred Composition: Spirocuta (incl. Euglenophyceae, Aphagea, *Anisonema*, *Chasmostoma*, *Dinema*, *Heteronema*, *Jenningsia*, *Neometanema*, *Peranema*, *Teloprocta* and *Urceolus*), and *Olkasia*. Note that some genera within Spirocuta are likely not monophyletic with current compositions, notably *Heteronema* and *Dinema*.

Comments: This clade is inferred from molecular phylogenies. This is a zoological name above the level of the family category and as such falls outside the zoological (and botanical) codes of nomenclature.

Authors' contributions

GL and AGBS conceived and designed the project. GL isolated, identified and imaged single cells, generated single-cell transcriptome data, and conducted phylogenetic analyses. MK and PJK supported some of the single-cell transcriptome data generation. YE isolated cultures. WJL provided cultures. NY, AK, BL and GB provided some mass-culture transcriptome data. GL and AGBS wrote the manuscript with input from other authors.

Acknowledgments

The authors would like to thank Joannie Roy, Shona Teijeiro and Georgette Kiethega (Université de Montréal), and Drahomira Faktorova (University of South Bohemia) for culturing and preliminary molecular analyses of *Petalomonas cantuscygni* and *Entosiphon sulcatum*. The research was supported by NSERC grants 298366-2014 to AGBS, 2014-03994 to PJK, RGPIN-2014-05286 to GB, RGPIN- 2019-03986 to BSL, and by the Canadian Institute for Advanced Research (CIfAR). WJL was supported by a grant from the National Research Foundation of Korea (NRF-2018R1A2A3075567). GL gratefully acknowledges partial stipend support from the Lett Graduate Student Bursary (Department of Biology, Dalhousie University).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107088.

References

Aberer, A.J., Krompass, D., Stamatakis, A., 2013. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. Syst. Biol. 62, 162–166.

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. Nat. Microbiol. 3 (12), 1417–1428.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Breglia, S.A., Leander, B.S., 2007. Phylogeny of Phagotrophic Euglenids (Euglenozoa) as Inferred from Hsp90 Gene Sequences. J. Eukaryot Microbiol. 54, 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 Microbiol. 10, 145.
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 Biol. Evol. 10, 427–433.
Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience 8:giz100.
Busse, I., Preisfeld, A., 2003. Application of spectral analysis to examine phylogenetic signal among euglenid SSU rDNA data sets (Euglenozoa). Org. Divers. Evol. 3, 1–12.
Busse, I., Patterson, D.J., Preisfeld, A., 2003. Phylogeny of Phagotrophic Euglenids (Euglenozoa): a Molecular Approach Based on Culture Material and Environmental Samples. J. Phycol. 39, 828–836.
Butenko, A., Opperdoes, F.R., Flegontova, O., Horák, A., Hampl, V., Keeling, P., Gawryluk, R.M.R., Tikhonenkov, D., Flegontov, P., Lukeš, J., 2020. Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and euglenids. BMC Biol. 18 (1) https://doi.org/10.1186/s12915-020-0754-1.
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. Mol. Phylogenet. Evol. 81, 71–85.
Cavalier-Smith, T., 2016. Higher classification and phylogeny of Euglenozoa. Europ. J. Protistol. 56, 250–276.
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. Europ. J. Protistol. 56, 147–170.
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). Europ. J. Protistol. 49 (1), 15–31.
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 Evol. Bio.l 10 (1), 210. https://doi.org/10.1186/1471-2148-10-210.
Dujardin, F., 1841. Histoire Naturelle des Zoophytes: Infusoires. Librairie Encyclopédique de Roret, Paris.
Ebenezer, T.E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A.M.G., 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 Biol. 17 (1) https://doi.org/10.1186/s12915-019-0626-8.
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 J. 5, 231–243.
Farmer, M.A., Triemer, R.E., 1988. In: Flagellar systems in the euglenoid flagellates. Elsevier, pp. 283–291.
Farmer, M.A., Triemer, R.E., 1994. An Ultrastructural Study of *Lentomonas applanatum* (Preisig). J. Eukaryot. Microbiol. 41, 112–119.
Fenchel, T., Bernard, C., Esteban, G., 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.
Gile, G.H., Faktorová, D., Castlejohn, C.A., Burger, G., Lang, B.F., Farmer, M.A., Lukeš, J., Keeling, P.J., 2009. Distribution and phylogeny of EFL and EF-1α in Euglenozoa suggest ancestral co-occurrence followed by differential loss. PLoS ONE 4, e5162–9.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B.o., 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. Nat. Protoc. 8 (8), 1494–1512.
Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G.B., Roger, A.J., 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc. Natl. Acad. Sci. 106 (10), 3859–3864.
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. Curr. Biol. 27 (13), 2043–2050.e6.
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. Sci Rep 8 (1). https://doi.org/10.1038/s41598-017-18805-w.
Jahn, T.L., 1946. The Euglenoid Flagellates. Q. Rev. Biol. 21 (3), 246–274.
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. Mol. Biol. Evol. 34, 2258–2270.
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. J. Eukaryot. Microbiol. 62 (3), 362–373.
Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30 (4), 772–780.
Keeling, P.J., del Campo, J., 2017. Marine Protists Are Not Just Big Bacteria. Curr Biol 27, R541–R549.
Kolisko, M., Boscaro, V., Burki, F., Lynn, D.H., Keeling, P.J., 2014. Single-cell transcriptomics for microbial eukaryotes. Curr. Biol. 24 (22), R1081–R1082.
Krabberød, A.K., Orr, R.J.S., Bråte, J., Kristensen, T., Bjørklund, K.R., Shalchian-Tabrizi, K., 2017. Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria. Mol. Biol. Evol. 34, 1557–1573.
Lackey, J.B., 1960. *Calkinsia aureus* gen. et sp. nov:, a new marine euglenid. Trans. Am. Microscopical Soc. 79 (1), 105. https://doi.org/10.2307/3223980.
Lakey, B., Triemer, R., 2017. The tetrapyrrole synthesis pathway as a model of horizontal gene transfer in euglenoids. J. Phycol. 53 (1), 198–217.
Larsen, J., Patterson, D.J., 1990. Some flagellates (Protista) from tropical marine sediments. J. Nat. Hist. 24 (4), 801–937.
Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.
Lax, G., Simpson, A.G.B., 2013. Combining Molecular Data with Classical Morphology for Uncultured Phagotrophic Euglenids (Excavata): A Single-Cell Approach. J. Eukaryot. Microbiol. 60 (6), 615–625.
Lax, G., Simpson, A.G.B., 2020. The Molecular Diversity of Phagotrophic Euglenids Examined Using Single-cell Methods. Protist 171 (5), 125757. https://doi.org/10.1016/j.protis.2020.125757.
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.
Leander, B.S., 2004. Did trypanosomatid parasites have photosynthetic ancestors? Trends Microbiol. 12 (6), 251–258.
Leander, B.S., Farmer, M.A., 2001. Comparative Morphology of the Euglenid Pellicle. II. Diversity of Strip Substructure. J. Eukaryot. Microbiol. 48, 202–217.
Leander, B.S., Triemer, R.E., Farmer, M.A., 2001a. Character evolution in heterotrophic euglenids. Europ. J. Protistol. 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.
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., Lax, G., Karnkowska, A., Simpson, A.G.B., 2017. Euglenida. In: Archibald, J.M., Slamovits, C.H., Simpson, A.G.B. (Eds.), Handbook of the Protists. Springer International Publishing, Cham, pp. 1047–1088.
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). J. Eukaryot. Microbiol. 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., 2008. Free-living heterotrophic euglenids from marine sediments of the Gippsland Basin, southeastern Australia. Mar. Biol. Res. 4, 333–349. https://doi.org/10.1080/17451000802040137.
Lee, W.J., 2012. Free-living heterotrophic euglenids from Botany Bay, Australia. Mar. Biol. Res. 8, 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, 10–16.
Łukomska-Kowalczyk, M., Karnkowska, A., Krupska, M., Milanowski, R., Zakryś, B., Mock, T., 2016. DNA barcoding in autotrophic euglenids: evaluation of COI and 18s rDNA. J. Phycol. 52 (6), 951–960.
Markunas, C.M., Triemer, R.E., 2015. Evolutionary history of the enzymes involved in the Calvin-Benson cycle in euglenids. J. Eukaryot. Microbiol. 63, 326–339.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 17, 10–12.
Maruyama, S., Suzaki, T., Weber, A.PM., Archibald, J.M., Nozaki, H., 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. BMC Evol. Biol. 11 (1) https://doi.org/10.1186/1471-2148-11-105.
Mignot, J.P., 1966. Structure et ultrastructure de quelques euglenomonadines. Protistologica 2, 51–117.
Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast Approximation for Phylogenetic Bootstrap. Mol. Biol. Evol. 30 (5), 1188–1195.
Montegut-Felkner, A.E., Triemer, R.E., 1997. Phylogenetic Relationships of Selected Euglenoid Genera Based on Morphological and Molecular Data. J. Phycol. 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. Nat. Microbiol. 4 (7), 1088–1095.
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. Int. J. Syst. Evol. Micr. 51, 783–791.

Nguyen, L.-T., Schmidt, H.A., Haeseler von, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Orsi, W., Edgcomb, V., Jeon, S., Leslin, C., Bunge, J., Taylor, G.T., Varela, R., Epstein, S., 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. II. Habitat specialization. ISME J. 5 (8), 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). Org. Divers. Evol. 17 (3), 509–520.

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. Nat. 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. Nat. Protoc. 9 (1), 171–181.

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. J. Phycol. 36 (1), 220–226.

Preisfeld, A., Busse, I., Klingberg, M., 2001. Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). Int. J. Syst. Evol. Micr. 51, 751–758.

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

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

Saito, A., Suetomo, Y., Arikawa, M., Omura, G., Mostafa Kamal Khan, S.M., Kakuta, S., Suzaki, E., Kataoka, K., Suzaki, T., 2003. Gliding movement in *Peranema trichophorum* is powered by flagellar surface motility. Cell Motil. Cytoskeleton 55 (4), 244–253.

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*). Europ. J. Protistol. 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. https://doi.org/10.1023/A:1025457801420.

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., 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. Arch. Protistenk. 147 (3-4), 213–225.

Simpson, A.G.B., Inagaki, Y., Roger, A.J., 2006. Comprehensive Multigene Phylogenies of Excavate Protists Reveal the Evolutionary Positions of "Primitive" Eukaryotes. Mol. Biol. Evol. 23, 615–625.

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

Song, L.i., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaSci 4 (1). https://doi.org/10.1186/s13742-015-0089-y.

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., 1986. Light and Electron Microscopic Description of a Colorless Euglenoid, *Serpenomonas costata* n. g., n. sp. J. Protozool. 33, 412–415.

Turmel, M., Gagnon, M.-C., O'Kelly, C.J., Otis, C., Lemieux, C., 2009. 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. Mol. Biol. Evol. 26, 631–648.

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. Syst. Biol. 67, 216–235.

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 Evol. Biol. 12 (1), 29. https://doi.org/10.1186/1471-2148-12-29.

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 Genet. Syst. 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., Leander, B.S., 2018. Diversity and Evolutionary History of the Symbiontida (Euglenozoa). Front. Ecol. Evol. 6, 100.

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). Europ. J. Protistol. 49 (1), 32–39.

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 Microbiol. 9 (1), 16. https://doi.org/10.1186/1471-2180-9-16.