Towards a molecular phylogeny of the Euglenozoa: analyses of ribosomal DNA sequences and deduced secondary structure elements

Thesis submitted in partial fulfilment
of the requirements of the degree
Doctor rer. nat. of the
Faculty for Mathematics and Natural Sciences
Chair in Biology and Didactics of Biology – Zoology
Bergische Universität Wuppertal, Germany

submitted by
Stefan Pärschke
Wuppertal
December 2015

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20160128-144301-5
[http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20160128-144301-5]

First reviewer:	Prof. Dr. A. Preisfeld
Second reviewer:	Prof. Dr. H. Wägele

Contents

Contents ..... I-IV

1 Abstract 1

2 Introduction 3

2.1 Characteristics of Euglenozoa 3

2.2 Biogeography of Euglenozoa 6

2.3 Euglenozoa in eukaryote systematics 8

2.4 Phylogenetic relationships of Euglenozoa 9

2.5 Euglenozoan ribosomal DNA 12

2.6 Scope of this thesis 13

3 Material 15

3.1 Organisms 15

3.1.1 Strains of euglenid flagellates 15

3.1.2 Strains of diplonemids 15

3.1.3 Strains of *Escherichia coli* 15

3.2 Media 16

3.2.1 Media for euglenid flagellates 16

3.2.2 Media for diplonemids 16

3.2.3 Media for *E. coli* 16

3.3 DNA samples 16

3.4 Buffers and solutions 17

3.5 Stock solutions 17

3.6 Oligonucleotides 18

3.7 Suppliers of reagents 19

3.8 Suppliers of enzymes 20

3.9 Suppliers of kits 20

3.10 Suppliers of standards 20
3.11 Suppliers of consumables 21
3.12 Suppliers of laboratory equipment 21
3.13 Databases and online tools 22
3.14 Computer software 23

4 Methods 24

4.1 Isolation of DNA and RNA 24
4.1.1 Isolation of total DNA 24
4.1.2 Isolation of plasmid DNA 24
4.1.3 Isolation of total RNA 25
4.2 Polymerase chain reactions (PCRs) 25
4.2.1 Standard PCR 26
4.2.2 Colony PCR 26
4.2.3 Reverse-Transcription-PCR 27
4.3 Agarose gel electrophoresis 27
4.4 Purification of PCR products 27
4.5 Cloning 28
4.6 Preparation of vector DNA 28
4.7 Quantification of vector DNA 29
4.8 Sequencing 29
4.9 Assembly of ribosomal DNA sequences 29
4.10 Alignments of sequence data 30
4.10.1 SSU rDNA sequences 30
4.10.2 LSU rDNA sequences 33
4.11 Datasets 35
4.12 Computer analyses 35
4.12.1 Statistical tests 35

4.12.2 Model testing 36
4.12.3 Maximum likelihood analyses 36
4.12.4 Bayesian inferences 36
4.12.5 Phylogenetic networks 37
4.12.6 Spectral analyses 37
4.13 Secondary structure analyses 37

5 Results and Discussion 38

5.1 Phylogenetic analyses of SSU rDNA sequences 38
5.1.1 Preliminary dataset 38
5.1.2 Marine versus freshwater Euglenozoa 43
5.1.3 Combined dataset 53
5.2 SSU rDNA nucleotide sequence analyses 62
5.2.1 Base composition 62
5.2.2 Identity matrix 65
5.2.3 Secondary structure 66
5.2.4 SSU rDNA variable regions 72
5.3 Phylogenetic analyses of LSU rDNA sequences 79
5.3.1 Broad dataset IV 79
5.3.2 Strict dataset V 83
5.4 LSU rDNA nucleotide sequence analyses 88
5.4.1 Base composition 88
5.4.2 LSU rDNA length comparison 90
5.4.3 Secondary structure 92
5.5 Phylogenetic analyses of ribosomal operon sequences 99

6 Conclusion 102

6.1 Methodological approach 102

6.2 Phylogenetic inferences 104
6.2.1 Long-branching *Entosiphon* 104
6.2.2 Most recent common ancestor of Euglenozoa 106
6.2.3 Major group relationships of Euglenozoa 107
6.2.4 Phagotrophic euglenids are polyphyletic 109
6.3 Taxonomic implications 109
6.4 Future prospects 112

7 References 113

8 Appendix 128

List of Abbreviations 157

Acknowledgements 159

1 Abstract

To shed light on the disputed molecular phylogeny of Euglenozoa, SSU rDNA sequences of an uncultured *Peranema* sp. (wild) and cultured *Ploeotia edaphica* (CCAP 1265/2) were isolated and a database was created containing a great many euglenozoan SSU rDNA sequences. Additionally, new LSU rDNA sequences were isolated from diplonemid *Rhynchopus euleeides* (ATCC 50226) and *Diplonema ambulator* (ATCC 50223) as well as of phagotrophic euglenids *Entosiphon sulcatum* (CCAP 1220/1B), *Notosolenus ostium* (wild), *Peranema trichophorum* (CCAP 1260/1B), *Petalomonas cantuscygni* (CCAP 1259/1), *Ploeotia costata* (CCAP 1265/1), and primary osmotrophic euglenids *Astasia curvata* (SAG 1204-5b), *Astasia torta* (SAG 217.80) and *Rhabdomonas costata* (SAG 1271-1) by the application of specifically designed primers for primer walking through unknown parts of this understudied gene region.

As a new approach, recently published SSU and LSU rRNA secondary structure data of *Saccharomyces cerevisiae* (Petrov et al. 2013 and 2014) was utilized as a blueprint for alignment of nucleotide sequences and deduction of secondary structure elements. Several datasets were formed to investigate phylogenetic inferences of SSU rDNA sequences with an equilibrated taxon sampling, separate marine and freshwater lineages and a combined set comprising more than 150 euglenozoan SSU rDNA sequences. Further examinations included two datasets derived from new LSU rDNA sequences as well as a concatenated dataset comprising genetic information of the ribosomal operon for the first time concerning euglenids, diplonemids and kinetoplastids. To address the adherent problem of weakness in statistical support regarding Euglenida found in prior studies, built datasets were additionally examined by phylogenetic network and spectral analyses. These analyses were also used to verify phylogenetic signals of identified monophyla and to test their tree compatibility. Finally, deduced secondary structures were utilized to pinpoint boundaries of coding and spacer regions as a precondition for examination of variable regions of SSU rDNA, SSU and LSU rDNA sequence lengths and corresponding base composition, identity matrices, ITS sequences and their insertion sites in LSU rDNA as well as unique nucleotide substitutions in the search for group specificities among Euglenozoa.

As a result, important findings concerning the phylogeny of major euglenozoan groups have been found, i.e. (1) Euglenida were not monophyletic, for Petalomonadida represented the deepest branch of Euglenozoa and the taxon Euglenida possessed no phylogenetic signal,

(2) Diplonemida were not the sister group of Kinetoplastida, phylogenetic and secondary structure analyses strongly inferred a relation of Diplonemida with Petalomonadida and Ploeotiida rather than with Kinetoplastida, (3) the existence of a euglenid crown group was confirmed by phylogenetic as well as spectral analyses and according to the eponymous autapomorphy, a helically pellicle pattern, prior taxon denominations were converted into Helicales *taxon nov.* PAERSCHKE & PREISFELD 2015, (4) the denominations 'phagotrophic euglenids' and 'Heteronematina' *sensu* Adl et al. (2012) describe a polyphyletic assemblage of euglenids and should be discarded.

The present work provides a basis for further examinations of euglenozoan LSU rDNA sequences and thus represents a precursor for future studies concerning the ribosomal operon of Euglenozoa.

2 Introduction

Since the Euglenozoa are not recorded in an all-embracing nomenclature, this taxon has an ambiregnal status in phylogenetic systematics. Diverse nutritional modes of euglenozoans and especially of euglenids might be the main reason for controversial views concerning the phylogeny of this protist group. In zoological systematics most heterotrophic forms were recognized as protozoa (e.g. Wehner & Gehring 2013) and have been regulated in the International Zoological Code of Nomenclature (ICZN) while phototrophic euglenids were regarded as algae (e.g. Kadereit et al. 2014) and have been listed in the International Code of Nomenclature for algae, fungi and plants (ICN). As both zoological and botanical descriptions exist in classifying taxonomy, mostly zoological names of taxa have been used in the present work whenever possible for reasons of comprehension and clarity.

2.1 Characteristics of Euglenozoa

Euglenozoa represent a large group of microbial eukaryotes comprising probably far more than 1,000 described species of euglenids plus an unknown number of diplonemids and kinetoplastids (Leander et al. 2007). Protozoan species numbers vary largely: about 8,000 species have been described and over 36,000 estimated according to Mora et al. (2011), earlier studies noted 30,000 described species and estimates even exceeded 250,000 (Fenchel 1993, May 1988). Nonetheless, the diversity of Euglenozoa found in marine environments surpassed earlier expectations (López-García et al. 2006, Moreira & López-García 2002, Zuendorf et al. 2006). Most euglenozoans do not contain calcium carbonate or other body parts that could outlast longer periods of time, therefore no fossil record is available; nonetheless euglenid cells have been identified in Triassic amber which are at least 220 million years old (Schönborn et al. 1999).

Nutrition

Most euglenozoans are free-living heterotrophic flagellates which play an important role in the food webs of Earth's watery environments (Fig. 1.1). Phagotrophic euglenozoans prey on bacteria and/or other eukaryotes (Boenigk & Arndt 2002, Lara et al. 2009), while some diplonemids and kinetoplastid bodonids are parasites of crustaceans and fishes, i.e. *Rhynchopus* and *Ichthyobodo* (von der Heyden et al. 2004). The kinetoplastid group of

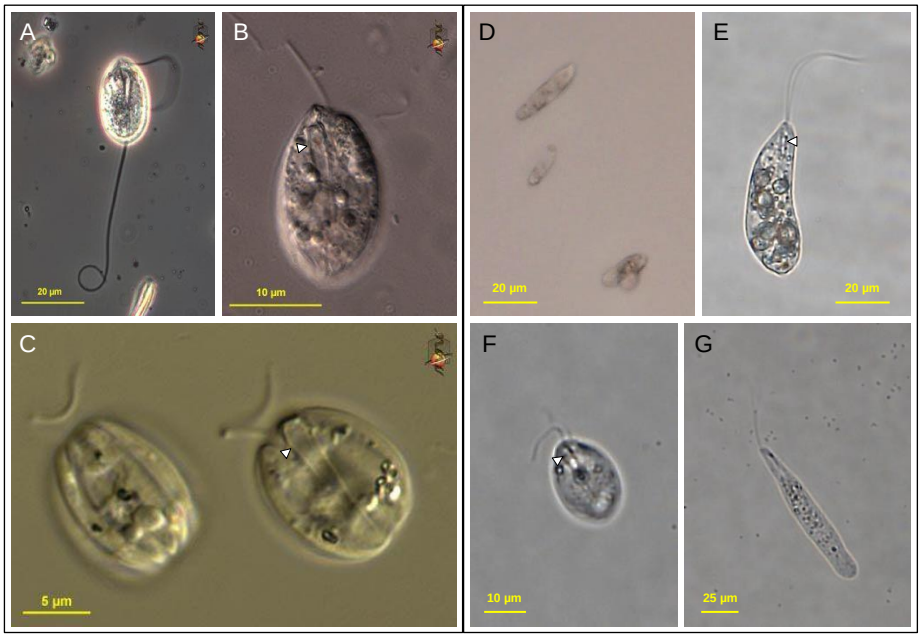


Fig. 1.1: Light-microscopic pictures of different euglenozoans. Numbers above yellow scale bars depict length in µm. Pictures A, B, C, E and F show phagotrophic euglenids. White arrowheads in B, C, E and F highlight the feeding apparatus. **A**: *Anisonema acinus*. **B**: *Entosiphon sulcatum*. **C**: two cells of *Ploeotia corrugata* visible in different focus depths, featuring longitudinally arranged pellicle strips (left) and feeding apparatus (right cell). **D**: cells of the diplonemid *Rhynchopus euleeides* (ATCC 50226). **E**: *Peranema* sp. (wild). **F**: *Ploeotia edaphica* (CCAP 1265/2). **G**: the primary osmotrophic euglenid *Distigma sennii* (SAG 222.80). Source of pictures A, B, C: micro*scope (http://www.pinkava.asu.edu/starcentral/microscope), © Angelika Preisfeld and David J. Patterson.

trypanosomatids is of large medical importance, it includes human parasites which are transmitted by insects, e.g. *Trypanosoma* can cause the African sleeping sickness, Chagas' disease (*T. brucei* and *T. cruzi*) as well as animal diseases like Nagana, organisms of genus *Leishmania* provoke several forms of Leishmaniasis, some of which are lethal (Walker et al. 2011). In 1907 the French physician Alphonse Laveran received the Nobel Prize in recognition for his work, for he identified protozoa as causative organisms of Malaria and Trypanosomiasis. The oldest record of Chagas' disease has been found in DNA samples of *T. cruzi* isolated from naturally desiccated human mummies, members of the Chonchurro culture, who lived in 7,000 B.C. (Aufderheide et al. 2004).

Three different modes of nutrition are known in Euglenida: most colorless forms are heterotrophic, i.e. phagotrophic euglenids prey on small eukaryotes and/or bacteria utilizing specialized ingestion apparatus, osmotrophic euglenids ingest dissolved nutrients by pinocytosis, and phototrophic "green" euglenids contain chloroplasts which enable photosynthesis besides pinocytotic uptake of nutrients. Chloroplasts of phototrophic euglenids incorporate chlorophyll a and b, are enfolded by three membranes and have been obtained by secondary endosymbiosis (Gibbs 1978).

Apomorphies

Traditionally, Euglenida have been characterized morphologically by ultrastructural features of their cells, e.g. pellicle, paraxonemal rods in flagella, and if present, paramylon or the feeding apparatus. The typical ‘euglenid’ pellicle can be rigid or flexible, longitudinally or helically arranged (Fig. 1.2), and consists of plasma membrane, proteinaceous strips, subjacent microtubules and cisternae of the endoplasmatic reticulum (Sommer 1965).

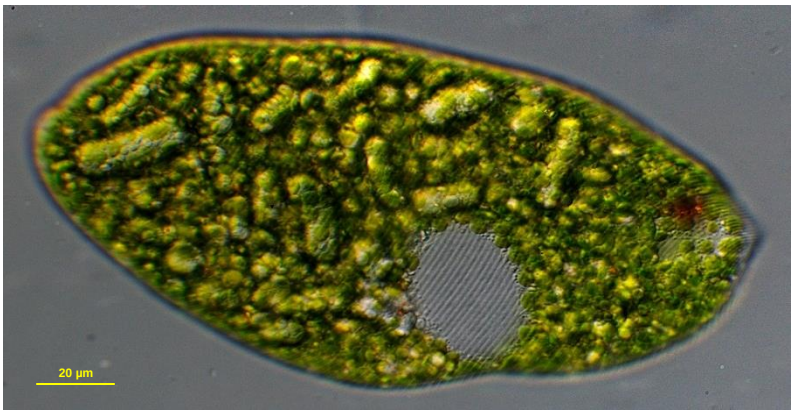


Fig. 1.2: Light-microscopic picture of a green-colored phototrophic *Euglena* sp. illustrating the helical pellicle pattern with proteinaceous strips (in grey) and the red-colored stigma, a carotenoid containing organelle which enables light-perception in combination with a crystalline structure on the basis of the dorsal flagellum (not visible). The scale bar depicts 20 µm. This picture was available under the Creative Commons License (CC0 1.0 Universal Public Domain Dedication, © David Shykind, 2012).

Euglenids with flexible pellicle are capable of metabolic movement that looks like squirming, compressing and stretching of the cell, and which is sometimes referred to as 'euglenid (or euglenoid) movement’ though not all euglenids are capable of metaboly (Leander et al. 2001). However, diplonemids are also capable of metaboly, at least in a certain life stage (Roy et al. 2007). Similarities of ultrastructural characteristics, i.e. in the flagellar apparatus, its paraxonemal rods, the ventral ingestion apparatus and a Golgi-associated contractile vacuole, early led to a postulated relationship of Euglenida and Kinetoplastida (Kivic & Walne 1984). The feeding apparatus of phagotrophic euglenids have been classified in four types constituted by complexity of presumably homologous morphological substructures (Belhadri et al. 1992, Triemer & Farmer 1991) and thus confirmed earlier findings. But a close relation of *Diplonema ambulator* to Kinetoplastida was hypothesized later by Montegut-Felkner & Triemer (1994 and 1996) based on similar morphological traits. A densely packed mitochondrial DNA which has been named kinetoplast (Meyer 1968), was made the eponymous apomorphy of kinetoplastids though a kinetoplast-like DNA inclusion had been

found also in *Petalomonas cantuscygni*, a presumably primordial euglenid (Breglia et al. 2007, Leander et al. 2001).

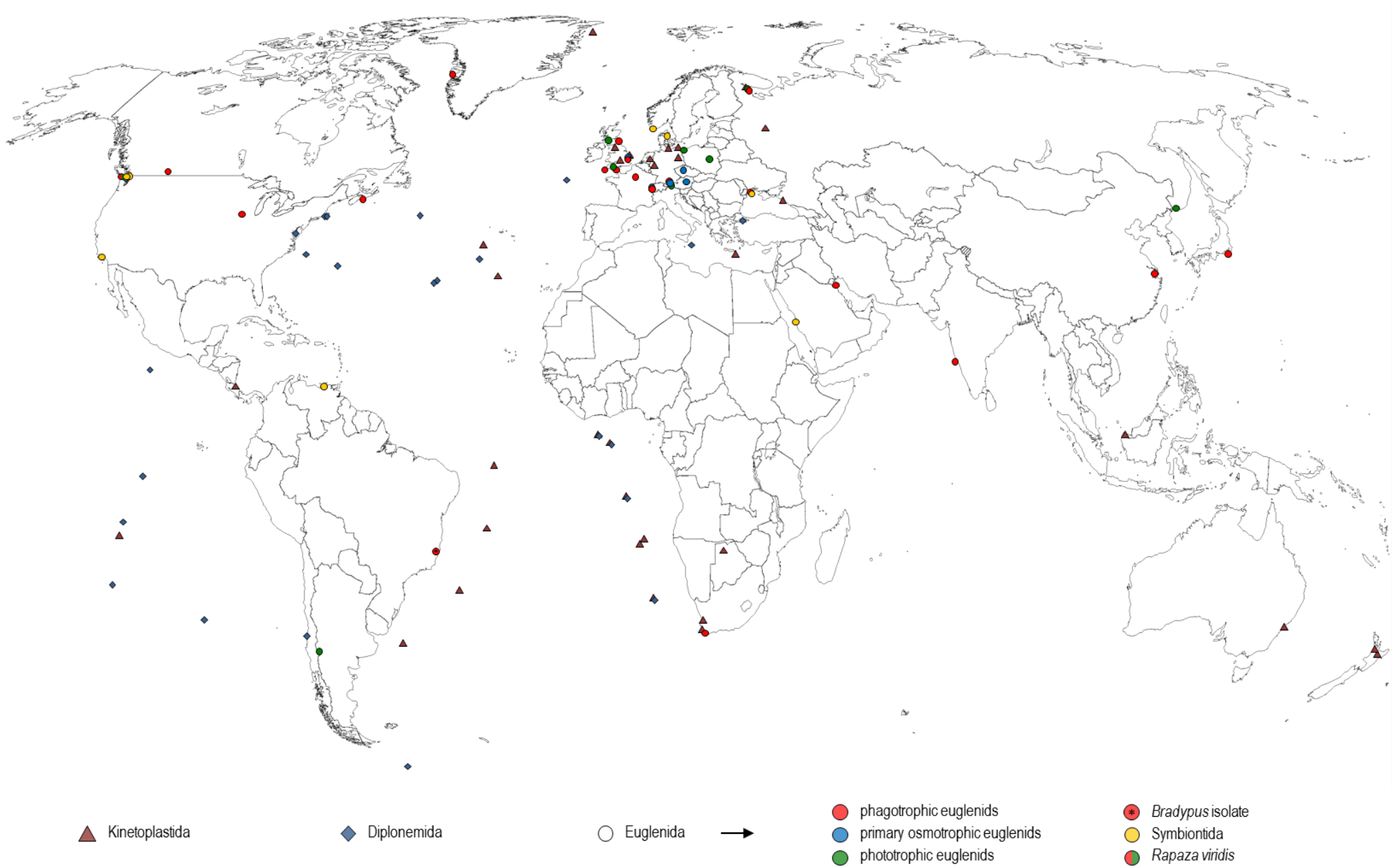
It has been shown that morphological characteristics contain ambiguities and thus are not applicable beyond question to pinpoint phylogenetic relationships of euglenozoan groups. This also applies to many molecular studies concerning gene evolution among Euglenozoa (see 2.4). Paraxonemal rods (Cachon et al. 1988, Talke & Preisfeld 2002, Walker et al. 2011), 'base J', i.e. β-D-glucosyl-hydroxymethyluracil (Dooijes et al. 2000, Gommers-Ampt et al. 1993), and trans-splicing (Frantz et al. 2000, Sturm et al. 2001, Tessier et al. 1991) represent genuine apomorphic features of Euglenozoa.

2.2 Biogeography of Euglenozoa

Being ubiquitously dispersed on Earth, Euglenozoa can be found in almost every watery environment. Free-living forms have been isolated from manifold marine, freshwater and brackish habitats, from temperate, humid cold and hot, as well as from extreme environments, e.g. cold-seeps or anoxic sediments. In the past 14 years, many studies examined environmental dispersal of Euglenozoa or discovered new species, and thus tremendously amended available SSU rDNA sequences (Fig. 1.3 and Tab. 1.1). Some Euglenozoa have been reported from rather extraordinary places: free-living kinetoplastids were isolated from tables in meat-cutting plants and from butterhead lettuce (Vaerewijck et al. 2008 and 2011), phagotrophic euglenids were found on pack ice in the Antarctic (Garrison & Buck 1989) and in furs of Brazilian sloths of genus *Bradypus* (Suutari et al. 2010), which exemplifies that the development of ecosystems needs a certain amount of deceleration.

Tab. 1.1: List of studies which have contributed most euglenozoan SSU rDNA sequences used in Fig. 1.3. Studies are sorted by publication year and alphabetically by first author's name.

No.	Year of publication	Author(s)
1	2001	López-García et al.
2	2001	Stonik & Selina
3	2003	López-García et al.
4	2003	Stoeck & Epstein
5	2003	Stoeck et al.
6	2004	Rat'kova et al.
7	2004	Von der Heyden et al.
8	2005	Šlapeta et al.
9	2006	Behnke et al.
10	2006	López-García et al.
11	2006	Scheckenbach et al.
12	2006	Tikhonenkov et al.
13	2006	Zuendorf et al.
14	2007	Countway et al.
15	2007	Stoeck et al.
16	2008	Chen et al.
17	2009	Lara et al.
18	2009	Marande et al.
19	2009	Saburova et al.
20	2009	Yubuki et al.
21	2010	Breglia et al.
22	2010	Jebaraj et al.
23	2010	Sauvadet et al.
24	2010	Scheckenbach et al.
25	2010	Suutari et al.
26	2010	Takishita et al.
27	2011	Lara et al.
28	2011	Wylezich & Jürgens
29	2012	Orsi et al.
30	2012	Salani et al.
31	2012	Thomas et al.
32	2012	Yamaguchi et al.
33	2013	Breglia et al.
34	2013	Chan et al.
35	2013	Lax & Simpson
36	2014	Lee & Simpson (2)
38	2014	Wang et al.


Kinetoplastida
Diplonemida
Euglenida
phagotrophic euglenids
primary osmotrophic euglenids
phototrophic euglenids
Bradypus isolate
Symbiontida
Rapaza viridis

2.3 Euglenozoa in eukaryote systematics

Within the super-group Excavata, the recently erected Discoba (SIMPSON in Hampl et al. 2009) form a clade that is indeed well supported by molecular data, although no withstanding morphological synapomorphies have been identified (Hampl et al. 2009, Walker et al. 2011). The Discoba embrace the usually well supported Euglenozoa (CAVALIER-SMITH 1981) SIMPSON 1997, the Heterolobosea (PAGE & BLANTON 1985) and the Jakobida (CAVALIER-SMITH 1993) ADL et al. 2005.

As close relatives of Euglenozoa, the Heterolobosea are defined as protists with two life phases, one of which with eruptive pseudopodia, the other with flagella (Page and Blanton 1985, Patterson 1999). Most heteroloboseans display discoidal mitochondrial cristae, therefore Discicristata CAVALIER-SMITH 1998 have been erected, comprising Euglenozoa and Heterolobosea with discoidal mitochondrial cristae as synapomorphy, although in some taxa these seem to be secondarily altered, e.g. some heterolobosean genera like *Psalteriomonas* own hydrogenosomes (de Graaf et al. 2009) and some diplonemids own longitudinally arranged, rather lamellar cristae (Marande et al. 2005, Roy et al. 2007). Discoidal mitochondrial cristae are also found outside of euglenozoans and heteroloboseans, in fact the excavate genus *Malawimonas* O'KELLY & NERAD 1999 also presents discoidal cristae (O'Kelly & Nerad 1999). The recently described excavate *Tsukubamonas globosa* YABUKI et al. 2011 possesses morphological characters which relate it to Heterolobosea, but molecular analysis often deviates (Brown et al. 2012, Harding et al. 2013, Park et al. 2012, Yabuki et al. 2011). However, some recent phylogenies did not confirm the existence of Discicristata as a monophyletic group (Cavalier-Smith 2002, Simpson et al. 2006), whereas others did (Baldauf et al. 2000, Cavalier-Smith 2003, Lara et al. 2006). Close relatives of Discicristata are the Jakobida (CAVALIER-SMITH 1993) ADL et al. 2005, which differ widely in mitochondrial features. While the genus *Jakoba* PATTERSON 1990 owns flat cristae (Simpson & Patterson 2001), the Histionidae FLAVIN & NERAD 1993 have tubular cristae, while affiliates of *Andalucia* LARA et al. 2006 bear tubular cristae or lack cristae at all (Lara et al. 2006).

◄ **Fig. 1.3**: World map of euglenozoan diversity illustrating isolation sites of studies mentioned in Tab. 1.1, major euglenozoan groups and subordinate taxa are color-coded as depicted in the legend.

Following these findings, Adl et al. (2012) recognized the Discoba as a robustly supported clade composed of Euglenozoa, Heterolobosea, Jakobida, and *Tsukubamonas*. Since no synapomorphies have been identified so far, the Discoba are described as a clade that stems from the most recent common ancestor of *Jakoba libera*, *Andalucia godoyi*, *Euglena gracilis*, and *Naegleria gruberi*. A recently published study ranked Euglenozoa into synthetically erected orders, classes, subphyla, phyla and infrakingdoms (Ruggiero et al. 2015), but it included no natural clades nor reflected the evolution of higher taxa (e.g. Euglenozoa are simultaneously an infrakingdom as well as a phylum), therefore such an artificial classification was not considered in this work.

2.4 Phylogenetic relationships of Euglenozoa

Small subunit ribosomal DNA (SSU rDNA)-based phylogenies have been widely used to infer relationships even among eukaryotic super-groups. Although most of the identified major clades are well supported, backup for some major as well as subordinate clades depends heavily on outgroup choice and taxon sampling, besides choice of gene(s) or methodologies while tackling the difficulties arising in the history of gene evolution (Parfrey et al. 2006). This also applies to the usually well supported Euglenozoa (CAVALIER-SMITH 1981) SIMPSON 1997. Former studies considered the Euglenozoa to consist of Diplonemida (CAVALIER-SMITH 1993) SIMPSON 1997, Kinetoplastida HONIGBERG 1963, Euglenida (BÜTSCHLI 1884) SIMPSON 1997 and *Postgaardi* FENCHEL et al. 1995, assorted among other characteristics by paraxonemal rods in the flagella featuring different structures in protein complexity (Adl et al. 2005, Kivic and Walne 1984, Simpson 1997, Roy et al. 2007, Walne & Dawson 1993). However, phylogenetic positions of major groups and particularly phagotrophic euglenids could not be resolved properly (Fig. 1.4). Yubuki et al. (2009) erected a further clade of euglenozoans, when they formally described *Calkinsia aureus*, a euglenid-like cell with rod-shaped epibiotic bacteria, and established the Symbiontida. In SSU rDNA phylogenies, *Calkinsia aureus* formed a clade together with other microbial eukaryotes isolated earlier from sub- and anoxic marine habitats (Behnke et al. 2006; Stoeck et al. 2003; Zuendorf et al. 2006). Adl et al. (2012) treated them as a major clade within the Euglenozoa, but also noted that symbiontids are probably derived phagotrophic euglenids, consisting of the genera *Calkinsia*, *Bihospites* and *Postgaardi*, which was recently confirmed by a morphological study (Yubuki et al. 2013).

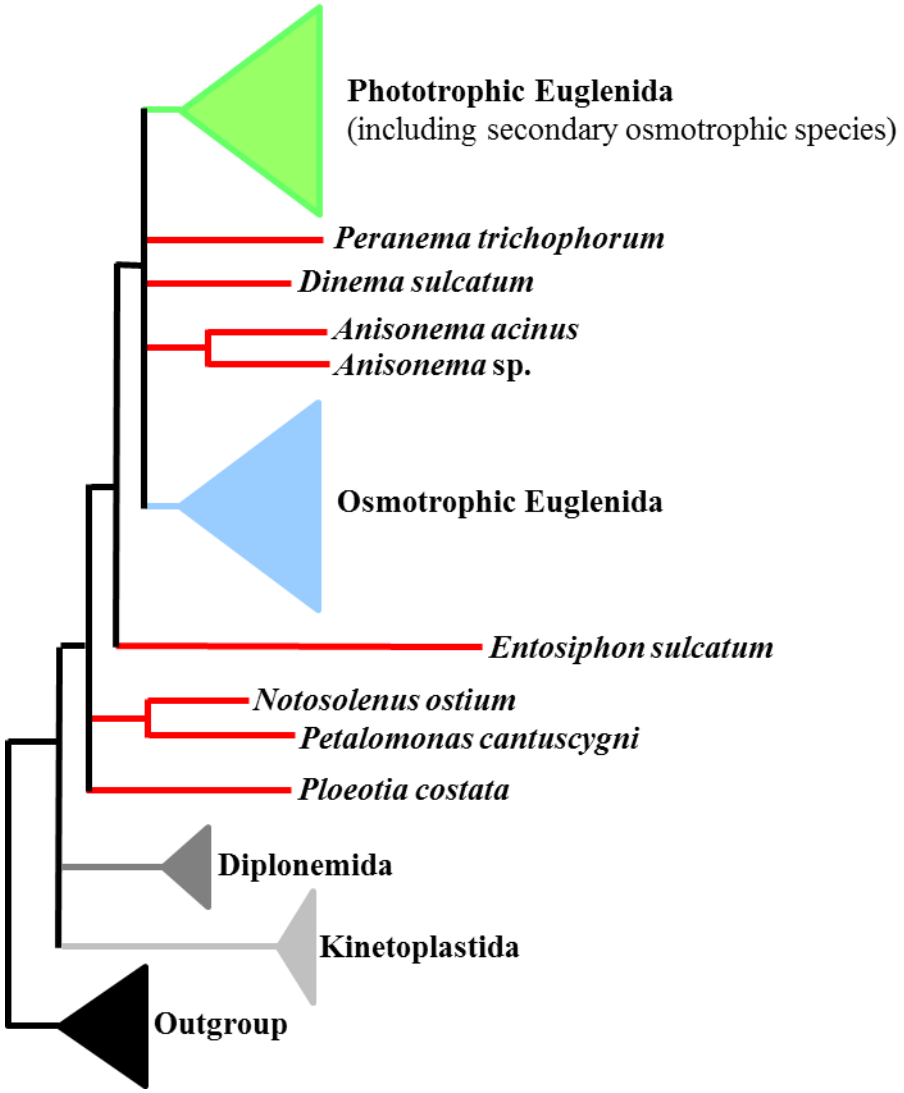


Fig. 1.4: Schematic phylogeny of Euglenozoa demonstrating unclear positions of diplonemids and kinetoplastids as well as phagotrophic euglenids which branch in multiple furcations (modified from Busse 2003).

Nowadays, Euglenozoa comprise the major clades Euglenida, Diplonemida and Kinetoplastida, with inner affiliations still controversial, and the limited taxon sampling especially of phagotrophic euglenids hinders molecular analysis (Adl et al. 2012; Triemer and Farmer 2007; Walker et al. 2011). Though recent discoveries of new euglenid species greatly amended available SSU rDNA sequence data (Tab. 1.1), phylogenetic studies still lead to incongruous tree topologies, probably due to differing methodological approaches or choice of diverse outgroups (Tab. 1.2). Furthermore, SSU rDNA based analyses of Euglenozoa almost traditionally suffered from weakness in statistical support regarding major group relationships (Busse and Preisfeld 2002a, 2003b, 2003c, Busse et al. 2003, Cavalier-Smith 2004, Moreira et al. 2004, Preisfeld et al. 2001, von der Heyden et al. 2004), recent analyses recovered paraphyletic Euglenida (Breglia et al. 2010, Chan et al. 2013, Yubuki et al. 2009), and even when other major euglenozoan clades have been included into the outgroup, monophyly of Euglenida lacked support (Lax and Simpson 2013).

Studies which investigated protein-based phylogenies utilized inappropriately poor taxon samplings concerning the Euglenozoa, in most cases of which the highly derived phototroph *Euglena gracilis* was the only taxon representing Euglenida (e.g. Simpson & Roger 2004, Simpson et al. 2004). So-called multigene studies may have contained information from many gene sequences, but included barely more than five taxa of all major euglenozoan groups into data analysis (e.g. Burki et al. 2007, Derelle & Lang 2012, He et al. 2014, Simpson et al. 2002 and 2008, Yoon et al. 2008). The sum of aforementioned circumstances would be a plausible reason for the tessellated status of euglenozoan molecular phylogeny (Tab. 1.2).

Tab. 1.2: List of most SSU rDNA based phylogenetic studies of euglenozoans including phagotrophic euglenids and their implications for euglenozoan phylogeny. When known, relevant parameters of analyses are given, i.e. applied methods, models chosen and tested, number of taxa (K = kinetoplastids, D = diplonemids, E = euglenids, S= symbiontids), and outgroups selected. Studies are sorted chronologically by publication year, then alphabetically by first author's name.* = symbiontids therein mislabeled as diplonemids. Studies with reliable results concerning euglenozoan phylogeny are highlighted in green, for explanation see text.

Study	Year	Method	ML-Model	Model tested	No. of positions	No. of taxa	K	D	E	S	outgroup	euglenozoan phylogeny
Montegut-Felkner & Triemer	1997	MP/NJ	multiple		964	8	2	-	4	-	Eukaryota	-
Maslov et al.	1999	ML	?	?	1,349	17	9	2	4	-	Eukaryota	E (K+D)
Preisfeld et al.	2000	ML/MP/NJ	multiple		?	13	2	-	9	-	K	-
Moreira et al.	2001	ML/MP/NJ	?	?	1,236	13	3	3	3	-	Eukaryota	K (D+E)
		ML	?	?	1,023	40	3	3	34	-	K + D	-
Müllner et al.	2001	NJ	multiple		1,036	35	2	-	33	-	K	-
Preisfeld et al.	2001	ML/MP/NJ	multiple		984	44	5	2	29	-	Eukaryota	E (K+D)
Busse & Preisfeld	2002	ML/MP	GTR+G+I	+	1,119	35	7	5	12	-	Eukaryota	D (K+E)
Busse & Preisfeld	2003a	ML	SYM+G+I	+	1,141	40	4	2	30	-	Eukaryota	E (K+D)
Busse & Preisfeld	2003b	ML/MP/NJ	GTR+G+I	+	1,117	74	5	5	52	-	Eukaryota	K (D+E)
Busse et al.	2003	ML/BI/MP	GTR+G+I	+	1,137	36	5	5	20	-	Eukaryota	E (E+[K+D])
Cavalier-Smith	2003	ML	K80+G+I	?	1,338	98	3	2	4	-	Eukaryota	E (K+[D+E])
Marin et al.	2003	ML/MP/NJ	TrN+G+I	+	1,454	64	11	2	51	-	K	-
Moreira et al.	2004	ML/BI/MP/NJ	GTR+G+I	+	1,150	24	4	4	4	-	Eukaryota	E (K+D)
Von der Heyden et al.	2004	NJ	GTR+G+I	no	1,233	145	47	12	68	-	Excavata	K (D+E)
		ME	GTR+G+I	no	1,233	80	-	12	68	-	D	-
Stoeck et al.	2006	ME	GTR+G+I	+	670	27	9	1	5	4*	Eukaryota	E (K+[D+S])
Behnke et al.	2006	ME	GTR+G+I	+	924	?	?	?	?	7	Eukaryota	E (K+[D+S])
Zuendorf et al.	2006	ML	TrN+G	+	711	52	7	4	3	6	Archaea	S (E*+[K+D])
Lara et al.	2009	ML	GTR+G+I	no	825	44	9	25	4	6	E	-
Yubuki et al.	2009	ML/BI	GTR+G+I	+	760	35	6	6	13	8	*Andalucia*	E(D[E(K+S)])
Breglia et al.	2010	ML/BI	GTR+G+I	no	760	37	6	6	13	10	*Andalucia*	E (E+S+D+K)
Kim et al.	2010	ME/MP	GTR+G+I	no	1,068	33	6	3	20	2	Jakobida	K (D+(S+E)
Yamaguchi et al.	2012	ML/BI	TIM1ef+G	+	805	39	7	3	27	2	K + D	-
Breglia et al.	2013	ML/BI	GTR+G+I	+	636	39	7	3	27	2	K + D	-
Chan et al.	2013	ML/MP/NJ	GTR+G	+	1,950	49	6	7	31	2	Jakobida (+K?)	K (S+[D+E])
Lax & Simpson	2013	ML/BI	GTR+G	no	1,161	80	9	6	49	5	Discoba (+K+D)	K (D+E)
Lee & Simpson	2014a	ML/BI	GTR+G	no	1,216	52	-	-	47	5	none (S?)	-
Lee & Simpson	2014b	ML/BI	GTR+G	no	1,293	53	-	-	48	5	none	-
this work	2015	ML/BI/network	GTR+G+I	+	1,194	85	8	8	33	8	Excavata	see Results
this work	2015	ML/BI/network	GTR+G+I	+	1,214	250	37	29	87	6	Eukaryota	see Results
this work	2015	ML/BI/network	GTR+G+I	+	1,158	199	33	22	30	19	Eukaryota	see Results
this work	2015	ML/BI/network	GTR+G+I	+	1,224	178	14	-	74	-	Eukaryota	see Results

2.5 Euglenozoan ribosomal DNA

The typical eukaryotic ribosome comprises four nascent rRNA species, i.e. 18S, 5.8S, 28S and 5S, as well as more than 70 associated proteins (Torres-Machorro et al. 2010). In most species, equivalent 18S, 5.8S and 28S ribosomal DNA is present in chromosomes as multiple copies of tandemly organized repeats, so-called cistrons. These cistrons are intermitted by spacer regions between rDNA regions, the internal transcribed spacers (ITSs), and between cistrons by the intergenic spacer (IGS). It has been shown that at least kinetoplastids and euglenids differ from other protozoa in many aspects regarding nucleotide sequence length, gene copy number, localization and organization of ribosomal genes. Due to insertions,

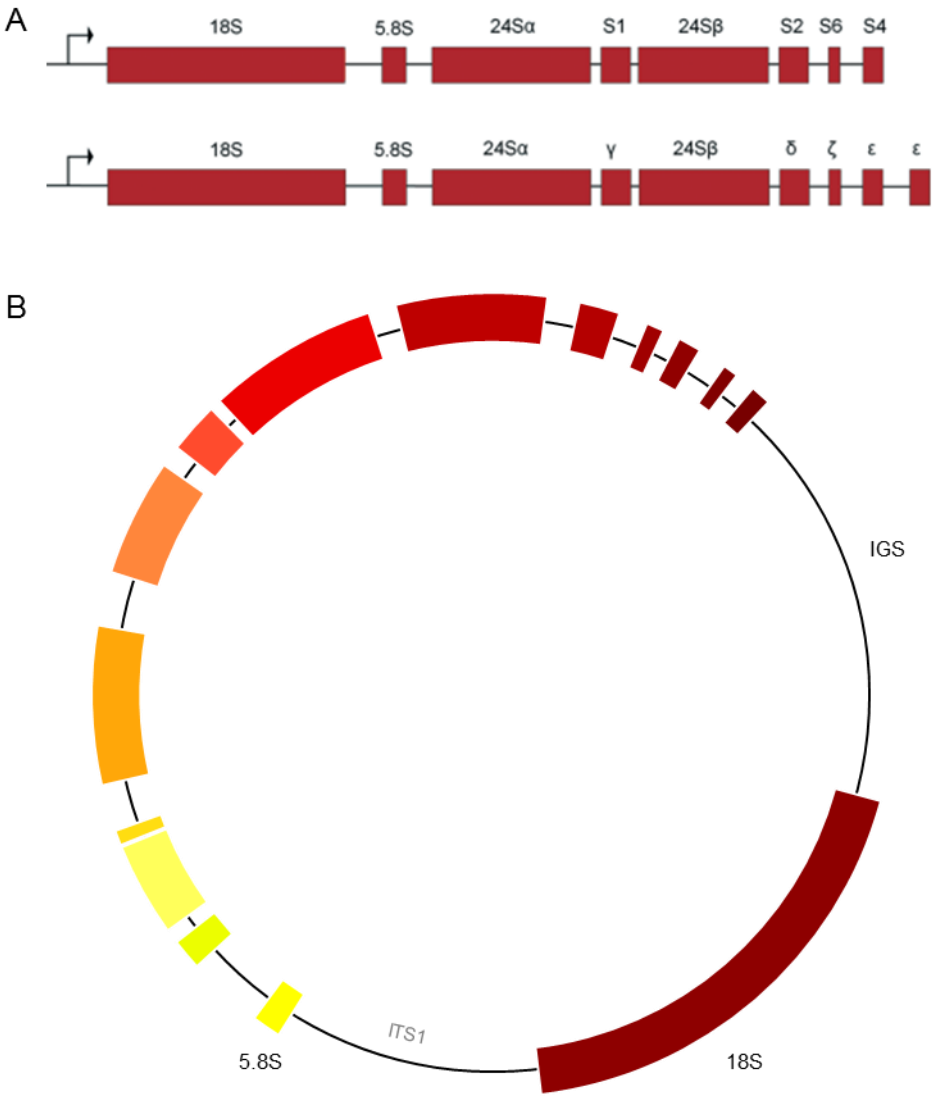


Fig. 1.5: Fragmentation of nuclear ribosomal genes in kinetoplastids and the phototrophic euglenid *Euglena gracilis*. **A**: Linear organized rDNA cistrons of Trypanosoma cruzi which 28S rDNA is fragmented into 6 elements, and Leishmania major, which 28S rDNA is fragmented into 7 components (modified after Torres-Machorro et al. 2010). **B**: Circular organized extrachromosomal ribosomal RNA gene of *Euglena gracilis* (modified after Greenwood et al. 2001).

primary osmotrophic euglenids of genus *Distigma* exhibited the largest SSU rDNA size ever measured (Busse & Preisfeld 2002b). While trypanosomatids contain 56 to 166 chromosomal copy numbers of rDNA, the phototrophic euglenid *Euglena gracilis* comprises only four chromosomal, but 800-4000 extrachromosomal copies (El-Sayed et al. 2005, Ravel-Chapuis 1988). 28S rDNA of trypanosomatids is fragmented into six or seven components (Fig. 1.5), but that of *Euglena gracilis* exhibits the highest fragmentation known today, it contains 13 intermittent ITSs (Schnare & Gray 1990, Schnare et al. 1990, Spencer et al. 1987, Torres-Machorro et al. 2010). As found in the rather derived heterolobosean *Naegleria gruberi*, extrachromosomal rDNA copies of *Euglena gracilis* are organized as a circular plasmid which contains 18S, 5.8S and 28S rDNAs that are separated by an additional intergenic spacer region (Clark & Cross 1987, Greenwood et al. 2001, Maruyama & Nozaki 2007).

Nowadays, mostly SSU rDNA has been used to infer phylogenetic relationships of Euglenozoa, e.g. the Kinetoplastida represent a so-called ribogroup according to Adl et al. (2012), characterized by phylogenetic inferences of ribosomal genes. Partial LSU rDNA sequences have been employed exclusively to examine phylogenetic relationships of phototrophic euglenids (Brosnan et al. 2003, Ciugulea et al. 2008, Kim et al. 2013, Triemer et al. 2006).

2.6 Scope of this thesis

To shed light on the molecular phylogeny of phagotrophic euglenids, SSU rDNA sequences of uncultured *Peranema* sp. and cultured *Ploeotia edaphica* (CCAP 1265/2) have been isolated. At the time preparing this work, number of available SSU rDNA sequences was overwhelmingly high which allowed the creation of a database containing more than 200 euglenozoan SSU rDNA sequences.

As a new (renewed) approach, recently published secondary structure data of *Saccharomyces cerevisiae* (Petrov et al. 2014) was utilized as a blueprint for recognition of homologous positions and alignment of sequences. Furthermore, a double-strategy was applied to address well-known problems regarding the phylogeny of Euglenida. Firstly, marine and freshwater lineages of the Euglenozoa were investigated separately, to test their phylogenetic implications. Euglenida, like Kinetoplastida, include free-living flagellates, which occupy marine as well as freshwater habitats, unlike free-living Diplonemida, which have been

isolated exclusively from marine environment. This diversity is dispersed throughout all Euglenida, it is found in phagotrophic lineages, e.g. in Petalomonadida and in Euglenea (BÜTSCHLI 1884) BUSSE & PREISFELD 2002 (i.e. phototrophic euglenids incapable of phagotrophy), which comprise predominantly marine Eutreptiales (LEEDALE 1967) MARIN & MELKONIAN 2003, and mostly freshwater Euglenales (LEEDALE 1967) MARIN & MELKONIAN 2003. Secondly, aforementioned datasets were combined to allow more thorough phylogenetic analyses with an even greater taxon sampling (Tab. 3.3).

Additionally, new LSU rDNA sequences of diplonemids as well as phagotrophic, osmotrophic and phototrophic euglenids were obtained by application of specifically designed primers for primer walking through mostly unknown parts of this understudied gene region. Only LSU rDNA sequences of three trypanosomatids and a diplonemid were available at the time preparing this thesis, so the addition of two new diplonemids and ten new euglenid LSU rDNA sequences would allow tentative steps into LSU rDNA-based phylogenetic examinations of Euglenozoa (Tab. 3.4). SSU and LSU rDNA datasets were merged to a new dataset, which for the first time contained genetic information of the ribosomal operon concerning euglenids, diplonemids and kinetoplastids, and was ultimately utilized for a phylogenetic study of euglenozoan ribosomal operon data.

To address the adherent problem of weakness in phylogenetic signal concerning the Euglenida, built datasets were examined by phylogenetic network analyses, which were used to investigate phylogenies in a tree-unlike manner, for networks graphically present compatible as well as incompatible and ambiguous phylogenetic signals hidden in SSU and LSU rDNA data. Additionally, phylogenetic tree and network topologies were tested by spectral analyses to verify phylogenetic signals and check for tree compatibility of identified monophyla.

Finally, secondary structures were deduced from obtained SSU and LSU rDNA sequences to pinpoint genetic boundaries as a prerequisite for identification of potential clade specificities by comparison of coding and spacer sequence lengths, base composition, variable regions of SSU rDNA, identity matrices, ITS sequences and unique nucleotide substitutions. In the end, these data were used to hypothesize phylogenetic relationships among major euglenozoan groups inferred from evolutionary traits of their ribosomal gene sequences.

3 Material

3.1 Organisms

3.1.1 Strains of euglenid flagellates

Astasia pertyi (CCAP 1204/3)

Distigma sennii (CCAP 1216/4)

Entosiphon sp (CCAP 1220/2)

Entosiphon sulcatum (CCAP 1220/1B)

Peranema sp (environmental isolate)

Petalomonas cantuscygni (CCAP 1259/1)

Ploeotia edaphica (CCAP 1265/2)

Strains from the Culture Collection of Algae and Protozoa (CCAP, United Kingdom) were received via ASSEMBLE Grant Agreement No. 227799, which was provided to the author by the European Community. *Peranema* sp. was found by Marisa Bartling in a culture of *Amoeba proteus*, which had been ordered from Lebendkulturen.de for teaching purposes. It was identified and isolated by the author, then cultivated according to 3.2.1.

3.1.2 Strains of diplonemids

Diplonema ambulator (ATCC 50223)

Rhynchopus euleeides (ATCC 50226)

The two diplonemid strains were purchased from the American Type Culture Collection (ATCC, United States of America) including medium.

3.1.3 Strains of *Escherichia coli*

E. coli DH5α obtained from Sabine Stratmann-Lettner

E. coli TOP10™	component of TOPO Cloning™ kit purchased from Life Technologies GmbH, Darmstadt, Germany

3.2 Media

3.2.1 Media for euglenid flagellates

Primary osmotrophic euglenids were cultivated in sterilized Volvic™ mineral water. Strains of freshwater phagotrophic euglenids (*Entosiphon*, *Peranema* and *Ploeotia*) were cultivated in sterilized Volvic™ mineral water and fed with baker's yeast once a week. *Petalomonas cantuscygni* was shortly cultivated in artificial sea salt medium, which has been obtained from Nadja Dabbagh.

3.2.2 Media for diplonemids

After thawing, diplonemid strains were cultivated in the included sea salt medium (ATCC medium 1728, enriched *Isonema* medium) with provided bacteria as prey organisms.

3.2.3 Media for *E. coli*

LB medium	25 g/l lysogeny broth in diH_2O
LB agar plates	15 g/l agar added to LB medium
SOB medium	26.64 g/l super optimal broth in diH_2O
SOC medium	20 mM glucose added to SOB medium

3.3 DNA samples

Samples of total DNA were obtained from Dr. Ingo Busse and Prof. Dr. Angelika Preisfeld, which have been isolated earlier from euglenid flagellates. *Anisonema acinus* and *Notosolenus ostium* were environmental samples and both collected by A. Preisfeld in Western Australia. Some DNA samples were extracted years ago from cultured strains which

have been obtained from CCAP, i.e. *Peranema trichophorum* (CCAP 1260/1B), *Ploeotia costata* (CCAP 1265/1), *Menoidium* sp (CCAP 1247/6), or from SAG, i.e. *Astasia curvata* (SAG 1204-5b), *Astasia torta* (SAG 217.80), *Rhabdomonas costata* (SAG 1271-1), *Eutreptia viridis* (SAG 1226-1c). DNA isolates from *Eutreptiella braarudii* (CCMP 1594) and *Eutreptiella pomquetensis* (CCMP 1491) were provided by Nadja Dabbagh. A culture of *Euglena gracilis* was allocated to the author by Dr. Renate Radek (Institute of Biology, Freie Universität Berlin).

3.4 Buffers and solutions

DNA loading dye (6x)	1 µl per 5 µl purified PCR product
DNA Stain *Clear* G	3 µl in 100 ml agarose gel (applied after boiling)
DreamTaq Green buffer (10x)	
TAE buffer (1x)	20 ml/l TAE stock solution in diH_2O
X-gal	40 µg/ml solubilized in Dimethylformamide

3.5 Stock solutions

Ampicillin stock solution	100 mg/ml in ddH_2O working concentration: 100 µg/ml
TAE buffer stock solution (50x)	2 M Tris 5.71 % (v/v) acetic acid 50 mM EDTA pH 8.3 - 8.5

3.6 Oligonucleotides

Oligonucleotides for the amplification of ribosomal DNA sequences were purchased from Eurofins Genomics, Ebersberg, Germany. Lyophilized oligonucleotides were solubilized in deionized sterile water (dsH_2O) to stock solutions with a concentration of 100 µmol according to manufacturer's manual.

Tab. 3.1: Oligonucleotides used as primers in PCR experiments sorted by gene, then alphabetically. Names, orientation (Or.; F = forward; R = reverse), length in nucleotides, sequences and references/sources are shown.

Name	Or.	Length	Sequence	Reference / Source
SSU				
AP 2	F	20 nt	5' – AAT CTG GTT GAT CCT GCC AG – 3'	Preisfeld et al. (2000)
AP 5	F	20 nt	5' – CAA CTG GAG GGC AAG TCT GG – 3'	Busse & Preisfeld (2002)
AP 6	R	21 nt	5' – GTT GAG TCA AAT TAA GCC GCA – 3'	Busse & Preisfeld (2002)
AP 8	R	25 nt	5' – TCA CCT ACA GCW ACC TTG TTA CGA C – 3'	Busse & Preisfeld (2002)
AP 10	R	20 nt	5' – CCA GAC TTG CCC TCC AGT TG – 3'	Dr. Busse
ITS 1	F	21 nt	5' – TGC GGC TTA ATT TGA CTC AAC – 3'	Dr. Busse
LSU				
AP 40	R	24 nt	5' – GTT GAT CCT GCC AGT AGT CAT ATG – 3'	*this work*
ITS 2	R	20 nt	5' – TCC TCC ACT GAG TGA TAT GC – 3'	Dr. Busse
LSU 2	R	25 nt	5' – TCA CGC TAC TTG TTC GCT ATC GGT C – 3'	Dr. Busse
LSU 4	R	21 nt	5' – ACT CCT TGG TCC GTG TTT CAA – 3'	Dr. Busse
LSU 6	R	18 nt	5' – AGT GAT ATG CTT AAG TCC – 3'	Dr. Busse
LSU 8	R	21 nt	5' – CTT GAT GAA ATG CTT TAA TCC – 3'	Dr. Busse
LSU 10	R	22 nt	5' – AGC TAT CCT GAG GGA AAC TTC G – 3'	Dr. Busse
LSU 11	F	19 nt	5' – ACC CGC TGA ACT TAA GCA T – 3'	Dr. Busse
LSU 12	R	21 nt	5' – GCT ACT CCA ACC AAG ATC TGC – 3'	Dr. Busse
LSU 13	F	19 nt	5' – ACG CCC TGG ATT AAA GCA T – 3'	Dr. Busse
LSU 14	R	21 nt	5' – GTC ATA GTT ACT CCC GCC GTT – 3'	*this work*
LSU 15	F	21 nt	5' – TTG AAA CAC GGA CCA AGG AGT – 3'	Dr. Busse
LSU 16	R	21 nt	5' – GTC TAA ACC CAG CTC ACG TTC – 3'	*this work*
LSU 17	F	21 nt	5' – GTC GTA ACA AGG TTG CTG TAG – 3'	*this work*
LSU 19	F	21 nt	5' – ATC GAA CCA CCT AGT AGC TGG – 3'	*this work*
LSU 21	F	22 nt	5' – GAA TGT GTA ACA ACT CAC CTG C – 3'	*this work*
LSU 23	F	19 nt	5' – TGA CTT CTG CCC AGT GCT C – 3'	*this work*
LSU 25	F	21 nt	5' – ATC CTT CGA TGT CGG CTC TTC – 3'	*this work*
LSU 27	F	19 nt	5' – TTA TGG CCG GTT CCT ACG G – 3'	*this work*

3.7 Suppliers of reagents

α-D(+)-Glucose monohydrate	Carl Roth GmbH, Karlsruhe, Germany
Acetic acid p.a.	Carl Roth GmbH, Karlsruhe, Germany
Agar-agar	AppliChem GmbH, Darmstadt, Germany
Ampicillin	Carl Roth GmbH, Karlsruhe, Germany
Dimethylformamide	Carl Roth GmbH, Karlsruhe, Germany
DNA Stain *Clear* G	Serva GmbH, Heidelberg, Germany
dNTPs	Fisher Scientific GmbH, Schwerte, Germany
DreamTaq Green buffer (10x)	Fisher Scientific GmbH, Schwerte, Germany
EDTA	Carl Roth GmbH, Karlsruhe, Germany
Ethanol	Carl Roth GmbH, Karlsruhe, Germany
Fluid nitrogen	kindly provided by Soner Öner-Sieben and Tim Kreutzer from the Institute of Botany and by Andreas Siebert from the Dept. of NMR spectroscopy
Glycerol	Carl Roth GmbH, Karlsruhe, Germany
LB Broth	Becton Dickinson, Heidelberg, Germany
peqGold Universal agarose	PeqLab, Erlangen, Germany
SOB Broth	Carl Roth GmbH, Karlsruhe, Germany
Tris	Carl Roth GmbH, Karlsruhe, Germany
Ultrapure™ agarose	Life Technologies GmbH, Darmstadt, Germany
X-gal	Carl Roth GmbH, Karlsruhe, Germany

3.8 Suppliers of enzymes

DreamTaq DNA polymerase (5 U/µl)	Fisher Scientific GmbH, Schwerte, Germany
T4 DNA ligase (1 U/µl)	Life Technologies GmbH, Darmstadt, Germany

3.9 Suppliers of kits

DNeasy® Plant Mini kit	Qiagen GmbH, Hilden, Germany
E.Z.N.A.™ Plasmid Miniprep II kit	Omega Bio-Tek Inc., Norcross, USA
innuPrep Double Pure kit	Analytik Jena AG, Jena, Germany
My-budget DNA Mini kit	Bio-Budget Technologies GmbH, Krefeld, Germany
OneStep RT-PCR kit	Qiagen GmbH, Hilden, Germany
RNeasy® Plant Mini kit	Qiagen GmbH, Hilden, Germany
TA Cloning™ kit (with pCR®2.1 vector)	Life Technologies GmbH, Darmstadt, Germany
TOPO Cloning™ kit (with pCR®2.1 vector)	Life Technologies GmbH, Darmstadt, Germany

3.10 Suppliers of standards

GeneRuler™ DNA Ladder Mix	Fisher Scientific GmbH, Schwerte, Germany
GeneRuler™ Low Range DNA Ladder	Fisher Scientific GmbH, Schwerte, Germany

3.11 Suppliers of consumables

Biosphere® filter tips	Sarstedt GmbH, Nürnbrecht, Germany
Microscope slides and cover slips	Carl Roth GmbH, Karlsruhe, Germany
Parafilm® M	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Pipette tips	Sarstedt GmbH, Nürnbrecht, Germany
Reaction tubes (0.2 – 2 ml)	Sarstedt GmbH, Nürnbrecht, Germany

Other consumables not listed here have been purchased from Carl Roth GmbH, Karlsruhe, Germany and from Sarstedt GmbH, Nürnbrecht, Germany.

3.12 Suppliers of laboratory equipment

Autoclav Systec VX-120	Systec GmbH, Wettenberg, Germany
Centrifuge 5424 (rotor F45-24-11)	Eppendorf AG, Hamburg, Germany
Centrifuge 5804 (swing-bucket-rotor A-4-44)	Eppendorf AG, Hamburg, Germany
Electrophoresis power supply EV 231	Consort n.v., Turnhout, Belgium
Electrophoresis apparatus	Consort n.v., Turnhout, Belgium
Freezer HERA Ultra-low temperature	Fisher Scientific GmbH, Schwerte, Germany
Gel documentation UV-System	Intas Science Imaging Instruments, Göttingen, Germany
Magnetic stirrer MR Hei-Standard	Heidolph Instruments GmbH, Schwabach, Germany
Microliter pipettes	VWR International GmbH, Langenfeld, Germany
Microliter tray cell Label Guard™	Implen GmbH, München, Germany

Microscope BA300	Motic GmbH, Wetzlar, Germany
Microscope Fluorescence Lifetime Imaging	Keyence GmbH, Neu-Isenburg, Germany
Mini centrifuge MCF-2360	Laboratory & Medical Supplies Inc., Tokyo, Japan
PCR Mastercycler® gradient	Eppendorf AG, Hamburg, Germany
PCR Mastercycler® personal	Eppendorf AG, Hamburg, Germany
pH meter HI 223	Hanna Instruments GmbH, Kehl, Germany
Photometer GeneSys 10 uv	Fisher Scientific GmbH, Schwerte, Germany
Rotator SB3	VWR International GmbH, Langenfeld, Germany
Thermoblock TB2	Analytik Jena AG, Jena, Germany
Thermomixer TS1	Analytik Jena AG, Jena, Germany
Vortexer VV3	VWR International GmbH, Langenfeld, Germany

3.13 Databases and online tools

Nucleotide database (National Center for Biotechnology Information, USA):
http://www.ncbi.nlm.nih.gov/nucleotide/

Nucleotide BLAST® – Basic Local Alignment Search Tool (National Center for Biotechnology Information, USA):
http://www.blast.ncbi.nlm.nih.gov/

SILVA ribosomal RNA database (Max Planck Institute for Marine Microbiology, Germany):
http://www.arb-silva.de/

Ribosomal RNA secondary structure database (Center for Ribosomal Origins and Evolution, USA):

http://www.apollo.chemistry.gatech.edu/ribosomegallery/index.html

Group I intron sequence and structure database (Wuhan University, China):
http://www.rna.whu.edu.cn/gissd/

PubMed database (National Library of Medicine and National Institutes of Health, USA):
http://www.ncbi.nlm.nih.gov/pubmed/

3.14 Computer software

GeneDoc	Multiple Sequence Alignment Editor and Shading Utility, Version 2.7.000 by K.B. Nicholas and H.B. Nicholas *formerly* http://www.psc.edu/biomed/genedoc [*link outdated or moved elsewhere*, author's comment]
Geneious 7	Version 7.1.4 purchased from Biomatters Inc., New Zealand http://www.geneious.com
MEGA 5	Molecular Evolution Genetics Analysis, Version 5.2.2 by K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar http://www.megasoftware.net
SAMS	Splits Analysis Methods, Version 1.4.3 beta by C. Mayer, S. Meid and W. Wägele https://www.zfmk.de/en/research/research-centres-and-groups/sams
SplitsTree 4	Version 4.13.1 by D. Huson and D. Bryant with contributions from M. Franz, M. Jette, T. Kloepper and M. Schröder http://www.splitstree.org

4 Methods

4.1 Isolation of DNA and RNA

Prior to preparation of DNA or RNA, a 2 ml aliquot of a culture of flagellates was centrifuged (1 min at 14,000 rpm) and the supernatant removed carefully.

4.1.1 Isolation of total DNA

Total DNA was isolated from cultures of flagellates applying different methods:

(1) Pelleted cells were quick-frozen in liquid nitrogen, broken up mechanically with a sterilized pestle, and then total DNA was isolated using the DNeasy® Plant Mini kit (Qiagen) according to manufacturer's protocol.
(2) Pelleted cells were broken up chemically with Proteinase K, thus following standard procedure for preparation of DNA from a cell culture using the My-budget DNA Mini kit (Bio-Budget).

As a modification of both methods, elution buffers were pre-warmed to 65 °C and elution of total DNA utilized in two centrifugation steps, a first eluate was recovered using 50 µl of elution buffer, and then a separate eluate with 100 µl of elution buffer using the same column. Eluted DNA was quantified by agarose gel electrophoresis, then used as template for standard PCR experiments and/or stored at -20 °C.

4.1.2 Isolation of plasmid DNA

Preparation of plasmid DNA from pelleted flagellate cells was done with the E.Z.N.A.™ Plasmid Miniprep II kit corresponding to operation manual with slightly modified elution steps. The elution buffer was pre-warmed to 65 °C and elution was carried out in two steps with the same column to maximize DNA yields. Eluted plasmid DNA was quantified by agarose gel electrophoresis, then used as template for standard PCR experiments and/or stored at -20 °C.

4.1.3 Isolation of total RNA

For the isolation of total RNA from flagellate cells the RNeasy® Plant Mini kit (Qiagen) was used accordant to manufacturer's reference. Quantity of eluted RNA was measured by agarose gel electrophoresis, then directly inserted as template in RT-PCR experiments and/or stored at -20 °C.

4.2 Polymerase chain reactions (PCRs)

For successful cloning (4.5) and sequencing (4.8) methods, polymerase chain reaction-(PCR-) amplified ribosomal DNA fragments should not exceed a size of ~1,500 bp, therefore primer pairs were designed and chosen accordingly (Fig. 4.1).

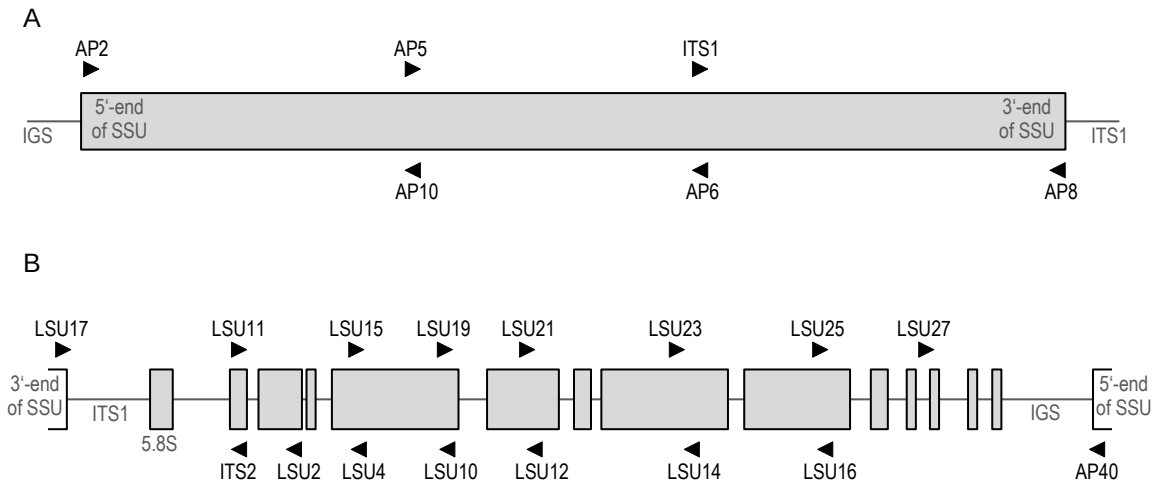


Fig. 4.1: Primers used in this work and their corresponding positions within the linearized ribosomal RNA gene of *Euglena gracilis* (after Greenwood et al. 2001). Forward primers shown above, reverse primers below gene, both marked by black arrowheads; gene regions, e.g. ITS1 and 5.8 S regions are shown in grey. **A**: Small subunit rDNA specific primers. **B**: Large subunit rDNA specific primers. Structural regions of the LSU rRNA gene are depicted in grey, i.e. LSU rDNA species 1 to 14 *sensu* Schnare et al. (1990).

Prior to the amplification of ribosomal RNA gene fragments by standard PCR, gradient PCRs with the same components were performed in a gradient thermocycler (Eppendorf) to test for optimal annealing temperatures of specific primer pairs.

4.2.1 Standard PCR

Standard PCR experiments for the amplification of ribosomal DNA fragments were executed in a thermocycler (Eppendorf) using DreamTaq Green buffer (Fisher Scientific), which contains a colorant component that made application of a loading dye in agarose gel electrophoresis obsolete. Annealing temperatures and duration of denaturation, annealing and elongation cycle steps varied depending on choice of primer pairs (Tab. 4.1).

Tab. 4.1: Standard PCR components and cycler conditions

Component	Amount
DreamTaq Green buffer 10 x	2.5 µl
dNTPs	1.0 µl
DreamTaq DNA polymerase	0.5 µl
Primer forward	2.0 µl
Primer reverse	2.0 µl
template DNA	1 – 3 µl
sterile deionized water	*ad* 25 µl

Cycler conditions			
Phase	T	t	
initial denaturation	94 °C	3 min	
denaturation	94 °C	30-60 sec	30 x
annealing	51-55 °C	30-90 sec	
elongation	72 °C	1-3 min	
final elongation	72 °C	10 min	
storage	4 °C	∞	

4.2.2 Colony PCR

Colony PCRs were used to test presence and size of insert in the vector of transformants in a bacterial clone culture. In colony PCR experiments, 1 µl of *E. coli* cell suspension was used directly as template for each PCR reaction. The M13-forward and -reverse regions, which flanked the insertion site within the pCR®2.1 vector, were used as primer locations for amplification of the inserted fragment (Tab. 4.2).

Tab. 4.2: Colony PCR components and cycler conditions

Component	Amount
DreamTaq Green buffer 10 x	2.5 µl
dNTPs	1.0 µl
DreamTaq DNA polymerase	0.5 µl
Primer M13 forward	1.0 µl
Primer M13 reverse	1.0 µl
template DNA	1.0 µl
sterile deionized water	*ad* 20 µl

Cycler conditions			
Phase	T	t	
initial denaturation	95 °C	3 min	
denaturation	94 °C	30 sec	25 x
annealing	52 °C	30 sec	
elongation	72 °C	1 min	
final elongation	72 °C	10 min	
storage	4 °C	∞	

4.2.3 Reverse-Transcription-PCR

In Reverse-Transcription-PCR experiments (RT-PCR), RNA was reversely transcribed into cDNA and a primer-specific region of the ribosomal RNA was amplified with the OneStep RT-PCR kit (Qiagen) executing the manufacturer's protocol carefully. Pipetting was performed with sterile RNase-free Biosphere® filter tips. RT-PCR products were examined in agarose gel electrophoresis and stored at -20 °C.

4.3 Agarose gel electrophoresis

Agarose gel electrophoresis is a standard method for the separation of negatively charged DNA fragments through an agarose gel matrix within an electric field. Migration speed depends on voltage of the electric field, agarose concentration and conformation of DNA. In a given concentration, longer DNA molecules move slower through the matrix than smaller ones, and higher agarose concentrations are preferred for separation of smaller DNA molecules. All experiments were performed in a horizontally arranged gel electrophoresis apparatus with 1x TAE buffer and a constant voltage of 100 V. Depending on expected fragment proportions, standards were used to determine DNA fragment size (3.10). Stain Clear G (3.4) was added to boiled agarose as a dyeing component for the visualization of DNA gel bands in UV light after electrophoresis.

For analyses of standard PCR, colony PCR and RT-PCR products, gel electrophoresis was performed using 1.5 % agarose (w/v) in TAE buffer with Stain Clear G, without loading dye (see 4.2.1), and 5 µl of each sample were applied directly to the gel. When standard PCR amplified multiple products of different size in a sample, 2 % Ultrapure™ agarose (w/v) was used for a preparative gel, and then desired bands were carefully extracted after gel electrophoresis. For quantification of purified PCR products, 5 µl of each sample were mixed with 1 µl of (6x) loading dye (3.4) and then applied to the gel. Each gel run was documented photographically.

4.4 Purification of PCR products

Prior to ligation, standard PCR products were purified using the InnuPrep Double Pure kit (Analytik Jena) according to manufacturer's protocol to remove primers, dNTPs and other

possible contaminants. PCR products from preparative gels were purified following the “Mini elute” protocol. Purified samples were used for cloning or stored at -20 °C.

4.5 Cloning

Ligation of purified PCR products and pCR®2.1 vector was done with the TA Cloning™ kit (Invitrogen) following manufacturer’s protocol. Purified samples, vector, buffer and T4 DNA ligase were mixed in diH_2O and reaction tubes were stored overnight at 14 °C.

Ligated vector samples were transferred into chemically competent *E. coli* cells (DH5α, see 3.1.3) according to One Shot™ chemical transformation manual of the TA Cloning™ kit. After heat shock treatment, bacterial cells were regenerated in 500 µl SOC medium for 1 h. Transformants were plated on LB agar dishes containing Ampicillin and X-gal for blue-white screening and incubated at 37 °C for 10 to 12 hours.

White transformants were picked from agar plates and cultivated in reaction vials containing 4 ml LB medium with Ampicillin and stored in an incubator at 37 °C for 10 to 12 hours. After incubation, the reaction vials were stored in a refrigerator to stop bacterial growth. 1.5 ml of each transformant culture were mixed carefully with sterile glycerol in a screw cap ampoule and stored in a clone library at -80 °C. The clone library was documented according to the Gentechnikgesetz § 6 (GenTG).

Unsuccessful cloning procedures were repeated once, in case of a second failure, cloning was performed with the TOPO Cloning™ kit (Invitrogen) according to manufacturer’s protocol using TOP10™ *E. coli* cells of higher chemical competence. Products and ingredients from unsuccessful cloning experiment were discarded at last.

4.6 Preparation of vector DNA

Prior to isolation of vector DNA from transformant *E. coli* cells, a 2 ml aliquot of a bacterial transformant culture was centrifuged (14,000 rpm, 1 min) and the supernatant discarded. Then vector DNA was isolated using the E.Z.N.A.™ Plasmid Miniprep II kit (Omega) according to manufacturer’s manual, elution was executed in two steps, each with 50 µl elution buffer,

which was pre-warmed to 65 °C. Extracted vector DNA was quantified photometrically (4.7) and then stored at -20 °C.

4.7 Quantification of vector DNA

Isolated vector DNA was quantified using a photometer with a Label Guard™ tray cell (Implen). The blank value was calibrated with 3 µl of pure elution buffer from the E.Z.N.A.™ Plasmid Miniprep II kit, then each DNA sample was measured three times. The arithmetic mean of measured values was taken for calculation of DNA quantity in ng/µl per sample. According to the sample submission guide of Eurofins Genomics, successful sequencing required a final DNA concentration of 50-100 ng/µl, therefore higher concentrations of vector DNA were diluted correspondingly.

4.8 Sequencing

Isolated vector DNA samples containing inserts were sequenced by Eurofins Genomics (Ebersberg, Germany), each sample twice, with standard M13 primers (forward and reverse) respectively.

4.9 Assembly of ribosomal DNA sequences

Obtained nucleotide sequence data was quality checked with sequencing reports, then 5'- and 3'-end M13 primer sequences and vector nucleotides were discarded. Insert nucleotide sequences were identified as ribosomal DNA via BLAST search (Altschul et al. 1990). When the search result was positive, forward and reverse nucleotide sequences of each sample were manually assembled to larger rDNA fragments using the software MEGA5 (version 5.2.2, Tamura et al. 2011). Assembly of larger overlapping sequence fragments followed the same manner, until nearly complete and/or complete rDNA sequences were obtained. In case of a negative BLAST search result, PCR experiments were performed again using an alternative primer pair (see Fig. 4.1).

4.10 Alignments of sequence data

4.10.1 SSU rDNA sequences

A preliminary mask file was created using the multiple sequence alignment editor software GeneDoc (version 2.7, Nicholas et al. 1997) combining SSU rDNA secondary structure information of *Euglena gracilis* (Schnare and Gray 1990) and osmotrophic euglenids (Busse and Preisfeld 2002b), for these comprise the largest SSU rDNA sequences known to date, with a slightly modified helix numbering after Wuyts et al. (2001). Then SSU rDNA sequences with at least 1,500 bp length available from GenBank representing the Discoba were chosen carefully to build a consequently equilibrated dataset, containing Euglenozoa, i.e. diplonemids, kinetoplastids, symbiontids (eight taxa each), and euglenid groups of different modes of nutrition, i.e. phototrophic and primary osmotrophic euglenids (eight taxa each), *Rapaza viridis* plus all phagotrophic euglenids available (sixteen taxa in the year 2013), as well as Heterolobosea (eight taxa), *Tsukubamonas globosa*, Jakobida (seven taxa) and twelve excavate taxa representing *Malawimonas*, Fornicata and Preaxostyla as outgroup. Missing nucleotides which could be assigned to homologous positions due to outgroup/ingroup comparison were replaced by 'N', other non-homologous (gapped) positions were omitted from the alignment and a preliminary, equilibrated dataset was generated (dataset 0 in Tab. 4.5).

A final mask file was shaped using GeneDoc consistent with recently published SSU rRNA secondary structure information of *Saccharomyces cerevisiae* (Petrov et al. 2014) as a blueprint for the identification of homologous positions and as a backbone for the alignment of nucleotide sequences. Therefore, data from the preliminary alignment was modified to fit revised secondary structure domains including helix numbering, and obtained SSU rDNA sequences from *Peranema* sp. and *Ploeotia edaphica* were added to the final mask file. Then most available discoban and several eukaryote SSU rDNA sequences with a length of at least 1,500 bp were downloaded from GenBank using Geneious software (version 7.1.4, Biomatters Inc.) to produce a SSU rDNA sequence database comprising over 300 nucleotide sequences (Tab. 3.3). These sequences were aligned to the secondary structure mask file by hand with GeneDoc for the identification of helices and homologous nucleotide positions. Based on this mask file, several SSU rDNA datasets of different taxon samplings were generated for phylogenetic analyses (see 4.11, Tab. 4.5).

Tab. 4.3: Sampling of euglenozoan and outgroup taxa used for the SSU rDNA alignment and related accession numbers, sorted into higher groups, then alphabetically. This work = accession pending.

No.	Name/Taxon	Accession
	EUGLENIDA (including symbiontids)	
1	*Anisonema acinus*	AF403160
2	*Anisonema acinus*	KC990937
3	*Anisonema* sp.	KC990936
4	*Anisonema* sp.	KC990935
5	*Anisonema/Dinema* sp.	KC990932
6	*Anisonema ("Peranema")* sp.	AY048919
7	*Astasia curvata*	AJ532394
8	*Astasia curvata*	AY004245
9	*Astasia curvata*	AF403153
10	*Astasia curvata*	AF403154
11	*Astasia* sp.	AF283307
12	*Astasia torta*	AF403152
13	*Bihospites bacati*	HM004354
14	*Bihospites bacati*	HM004353
15	*Calkinsia aureus*	EU753419
16	*Colacium mucronatum*	AF326232
17	*Colacium* sp.	FJ719601
18	*Colacium vesiculosum*	AF081592
19	*Cryptoglena pigra*	JQ356764
20	*Cryptoglena skujae*	JQ356774
21	*Cryptoglena* sp.	JQ356779
22	*Cyclidiopsis acus*	JQ681752
23	*Dinema platysomum*	KC990934
24	*Dinema sulcatum*	AY061998
25	*Discoplastis* sp.	FJ719606
26	*Discoplastis spathirhyncha*	AJ532454
27	*Distigma curvatum*	AF099081
28	*Distigma gracile*	AY061997
29	*Distigma gracilis*	AF386637
30	*Distigma pringsheimii*	AF386639
31	*Distigma proteus*	AF106036
32	*Distigma sennii*	AF386644
33	*Entosiphon* sp.	AY425008
34	*Entosiphon sulcatum*	AF220826
35	*Entosiphon sulcatum*	AY061999
36	*Euglena* cf. *Mutabilis*	AY082988
37	*Euglena gracilis*	M12677
38	*Euglena gracilis* var. *bacillaris*	AY029409
39	*Euglena longa*	AF112871
40	*Euglena* sp.	AF112873
41	*Euglena stellata*	AF150936
42	*Euglena tripteris*	AF445459
43	*Eutreptia pertyi*	AF081589
44	*Eutreptia* sp.	AJ532396
45	*Eutreptia viridis*	AF157312
46	*Eutreptia viridis*	AJ532395
47	*Eutreptiella braarudii*	AJ532397
48	*Eutreptiella eupharyngea*	AJ532399
49	*Eutreptiella gymnastica*	FJ719618
50	*Eutreptiella pomquetensis*	AJ532398
51	*Eutreptiella* sp.	AF112875
52	*Eutreptiella* sp.	JQ337867
53	*Gyropaigne lefèvrei*	AF110418
54	*Heteronema scaphurum*	JN566139
55	*Hyalophacus ocellata*	AF445458
56	*Keelungia pulex*	HM044218
57	*Khawkinea quartana*	U84732
58	*Lepocinclis oxyuris*	HQ287920
59	*Lepocinclis spirogyroides*	FJ719619
60	*Menoidium bibacillatum*	AF247598
61	*Menoidium cultellus*	AF295019
62	*Menoidium gibbum*	AF247599
63	*Menoidium intermedium*	AF295022
64	*Menoidium obtusum*	AF403155
65	*Menoidium pellucidum*	AF403156
66	*Menoidium* sp.	AY083243
67	*Monomorphina aenigmatica*	AF190814
68	*Monomorphina megalopsis*	JN603844
69	*Monomorphina pyrum*	JN603861
70	*Neometanema* cf. *exaratum*	KC990931
71	*Neometanema parovale*	KJ690254
72	*Notosolenus ostium*	AF403159
73	*Notosolenus ostium*	KC990930
74	*Notosolenus urceolatus*	KJ778682
75	*Parmidium circulare*	AF295018
76	*Parmidium scutulum*	AF309633
77	*Peranema* sp.	this work
78	*Peranema trichophorum*	AF386636
79	*Peranema trichophorum*	AH005452
80	*Petalomonas cantuscygni*	AF386635
81	*Petalomonas cantuscygni*	U84731
82	*Petalomonas sphagnophila*	GU477295
83	*Petalomonas sphagnophila*	GU477296
84	*Petalomonas sphagnophila*	GU477297
85	*Phacus oscillans*	AF181968
86	*Phacus pusillus*	AF190815
87	*Ploeotia* cf. *vitrea*	KC990933
88	*Ploeotia costata*	AF525486
89	*Ploeotia costata*	KF586332
90	*Ploeotia costata*	KF586333
91	*Ploeotia edaphica*	this work
92	*Rapaza viridis*	AB679269
93	*Rhabdomonas costata*	AF295021
94	*Rhabdomonas incurva*	AF247601
95	*Rhabdomonas intermedia*	AF295020
96	*Rhabdomonas spiralis*	AF247600
97	*Strombomonas acuminata*	FJ719639
98	*Strombomonas verrucosa*	AF445461
99	*Trachelomonas grandis*	this work
100	*Trachelomonas hispida*	AF445462
101	*Trachelomonas volvocinopsis*	AY015004
102	uncultured clone Blacksea cl 50	HM749952
103	uncultured clone Blacksea cl 51	HM749953
104	uncultured clone Blacksea cl 52	HM749954
105	uncultured clone CH1 S1 57	AY821956
106	uncultured clone CH1 S2 16	AY821957
107	uncultured clone CH1 S2 19	AY821958
108	uncultured clone D2P04B10	EF100248
109	uncultured clone D3P06F06	EF100316
110	uncultured clone FV23 1E10	DQ310358
111	uncultured clone FV23 2D3C4	DQ310255
112	uncultured clone FV36 2E04	DQ310359
113	uncultured clone M4 18D10	DQ103806
114	uncultured clone M4 18E09	DQ103807
115	uncultured clone M4 18H08	DQ103809
116	uncultured clone NA1 1G12	EF526883
117	uncultured clone NA1 3E11	EF526849
118	uncultured clone NA1 4B5	EF526782
119	uncultured clone NA1 4H11	EF526793
120	uncultured clone NA2 3B2	EF526848
121	uncultured clone NA2 3D8	EF526846
122	uncultured clone NA2 3E9	EF526847
123	uncultured clone PR3 3E 63	GQ330643

Tab. 4.3: continued.

No.	Name/Taxon	Accession
124	uncultured clone SA2 3B11	EF526950
	DIPLONEMIDA	
125	*Diplonema ambulator*	AY425009
126	*Diplonema ambulator*	AF380996
127	*Diplonema ambulator*	this work
128	*Diplonema papillatum*	AF119811
129	*Diplonema* sp.	AF119812
130	*Diplonema* sp.	AY425010
131	*Diplonema* sp.	AY425011
132	*Diplonema* sp.	AY425012
133	*Rhynchopus euleeides*	this work
134	*Rhynchopus* sp.	AY425014
135	*Rhynchopus* sp.	AY490209
136	*Rhynchopus* sp.	AY490210
137	*Rhynchopus* sp.	AY490211
138	uncultured clone CCW85	AY180037
139	uncultured clone DH148-EKB1	AF290080
140	uncultured clone LC22 5EP 17	DQ504321
141	uncultured clone LC22 5EP 18	DQ504322
142	uncultured clone LC22 5EP 19	DQ504323
143	uncultured clone LC22 5EP 32	DQ504349
144	uncultured clone LC23 5EP 5	DQ504350
145	uncultured clone Ma121 D1 12	EU635674
146	uncultured clone Ma131 1A46	FJ032684
147	uncultured clone PRTBE7274	HM799985
148	uncultured clone PRTBE7330	HM800011
149	uncultured clone PRTBE7353	HM799846
150	uncultured clone PRTBE7392	HM799859
151	uncultured clone PRTBE7426	HM800050
152	uncultured clone PRTBE7438	HM800057
153	uncultured clone PRTBE7445	HM800063
154	uncultured clone PRTBE7455	HM799887
155	uncultured clone PRTBE7509	HM799914
156	uncultured clone PRTBE7517	HM800094
157	uncultured clone PRTBE7533	HM799925
158	uncultured clone RM2-SGM31	AB505539
159	uncultured clone RM2-SGM32	AB505540
160	uncultured clone SCM15C6	AY665087
	KINETOPLASTIDA	
161	*Angomonas deanei*	HM593011
162	*Bodo edax*	AY028451
163	*Bodo rostratus*	AY425017
164	*Bodo saltans*	AF208889
165	*Crithidia dedva*	JN624299
166	*Crithidia fasciculata*	Y00055
167	*Cruzella marina*	AF208878
168	*Cryptobia bullockii*	AF080224
169	*Cryptobia helicis*	AF208880
170	*Cryptobia salmositica*	AF080225
171	*Dimastigella mimosa*	DQ207576
172	*Dimastigella trypaniformis*	AY028447
173	*Endotrypanum* sp.	EU021240
174	*Herpetomonas nabiculae*	JN624300
175	*Herpetomonas* sp.	JQ359724
176	*Ichthyobodo necator*	AY028448
177	*Ichthyobodo necator*	KC208028
178	*Leishmania major*	FR796423
179	*Leptomonas collosoma*	JN582046
180	*Leptomonas mirabilis*	JQ359729
181	*Neobodo designis*	AY753616
182	*Neobodo saliens*	DQ207589
183	*Parabodo caudatus*	JF754435
184	*Parabodo nitrophilus*	AF208886
185	*Perkinsela*-like sp. AK-2011	JN202437
186	*Perkinsiella*-like sp. AFSM3	AY163355
187	*Perkinsiella*-like sp. PLO-DE4A	HQ132931
188	*Phanerobia pelophila*	AY425020
189	*Procryptobia sorokini*	AY425018
190	*Rhynchomonas nasuta*	DQ465526
191	*Sergeia podlipaevi*	DQ394362
192	*Strigomonas culicis*	ATMH010127
193	*Strigomonas galati*	HM593010
194	*Trypanosoma brucei*	M12676
195	*Trypanosoma rangeli*	KJ742907
196	uncultured clone AND31	AY965872
197	uncultured clone AT1-3	AF530519
198	uncultured clone AT4-103	AF530522
199	uncultured clone AT4-56	AF530520
200	uncultured clone AT5-25	AF530518
201	uncultured clone AT5-48	AF530517
202	uncultured clone AT5-9	AF530516
203	uncultured clone Discovery	JN542579
204	uncultured clone FV18-8TS	AY963571
205	uncultured clone Kryos IF A3	JN542577
206	uncultured clone L7.7	AY753946
207	uncultured clone LC103 5EP 19	DQ504351
208	uncultured clone Urania B B5	JN542569
209	uncultured clone ZJ2007	JQ928406
210	*Wallaceina* sp.	JN582045
	OUTGROUP (DISCOBA)	
211	*Acrasis helenhemmesae*	GU437219
212	*Acrasis rosea*	HM114342
213	*Allovahlkampfia spelaea*	EU696948
214	*Andalucia godoyi*	AY965865
215	*Andalucia godoyi*	AY965870
216	*Andalucia incarcerata*	AY117419
217	*Andalucia incarcerata*	EU334887
218	*Euplaesiobystra hypersalinica*	FJ222604
219	*Harpagon descissus*	JN606337
220	*Harpagon schusteri*	JN606339
221	*Heteramoeba clara*	AF439350
222	*Heterolobosea* sp.	HQ898857
223	*Heterolobosea* sp.	JX509941
224	*Heterolobosea* sp.	JX441981
225	*Heterolobosea* sp.	JX509942
226	*Heterolobosea* sp.	FN668558
227	*Heterolobosea* sp.	HQ898858
228	*Jakoba libera*	AF411288
229	*Jakoba libera*	AY117418
230	*Macropharyngomonas halophila*	AF011465
231	*Monopylocystis visvesvarai*	AF011463
232	*Naegleria clarki*	DQ768725
233	*Naegleria gruberi*	NC_01018
234	*Neovahlkampfia damariscottae*	AJ224891
235	*Paravahlkampfia* sp.	FJ169185
236	*Percolomonas cosmopolitus*	AF519443
237	*Pharyngomonas* sp.	JX509943
238	*Plaesiobystra hypersalinica*	AF011459
239	*Pleurostomum flabellatum*	DQ979962
240	*Psalteriomonas lanterna*	X94430
241	*Psalteriomonas magna*	JN606351
242	*Pseudoharpagon pertyi*	JN606356
243	*Reclinomonas americana*	AF053089
244	*Reclinomonas americana*	AY117417
245	*Sawyeria marylandensis*	AF439351
246	*Seculamonas ecuadoriensis*	DQ190541

Tab. 4.3: continued.

No.	Name/Taxon	Accession
247	*Selenaion koniopes*	JX025226
248	*Stachyamoeba* sp.	AF011461
249	*Stephanopogon minuta*	AB365646
250	*Stygamoeba regulata*	JF694285
251	*Tetramitus thermacidophilus*	AJ621575
252	*Tetramitus thorntoni*	X93085
253	*Tsukubamonas globosa*	AB576851
254	*Tulamoeba peronaphora*	FJ222603
255	uncultured clone Blacksea cl 54	HM749956
256	uncultured clone Blacksea cl 55	HM749957
257	uncultured clone cLA12C05	EU446381
258	uncultured clone CN207St155	HM581638
259	uncultured clone CN207St70	HM581633
260	uncultured clone	EU368037
261	uncultured clone FV23 CilE10	DQ310279
262	uncultured clone M2 18G04	DQ103829
263	uncultured clone MA1 2H5L	EF527199
264	uncultured clone NKS105	JX296584
265	uncultured clone NKS177	JX296588
266	uncultured clone NKS188	JX296589
267	uncultured clone NKS82	JX296587
268	uncultured clone NKS96	JX296586
269	uncultured clone NKS97	JX296585
270	uncultured clone SA1 1D05	EF526978
271	uncultured clone WIM43	AM114803
272	*Vrihiamoeba italica*	AB513360
	OUTGROUP (DISTANT)	
273	*Acanthamoeba castellanii*	U07401
274	*Amoeba leningradensis*	AJ314605
275	*Amoeba proteus*	AJ314604
276	*Ancyromonas sigmoides*	AF174363
277	*Apusomonas* sp.	AY752987
278	*Besnoitia besnoiti*	AY833646
279	*Bigelowiella natans*	DQ158857
280	*Blastocystis* sp.	KF447163
281	*Breviata anathema*	AF153206
282	*Byssochlamys spectabilis*	AB023946
283	*Capsaspora owczarzaki*	AF436886
284	*Carpediemonas membranifera*	AY117416
285	*Carpediemonas*-like sp.	AF439347
286	*Cercozoa* sp.	FJ824128
287	*Chaos nobile*	AJ314606
288	*Chilomonas paramecium*	L28811
289	*Chlamydomonas pulsatilla*	DQ009748
290	*Chlorarachnion reptans*	U03275
291	*Chrysowaernella hieroglyphica*	HQ710556
292	*Chytriomyces hyalinus*	DQ536487
293	*Ciliophrys infusionum*	AB846665
294	*Cryptomonas paramecium*	NC_015331
295	*Cyanidioschyzon merolae*	AB158483

No.	Name/Taxon	Accession
296	*Cyrtohymena shii*	JQ513386
297	*Dictyostelium discoideum*	X00134
298	*Dinenympha exilis*	AB092924
299	*Dysnectes brevis*	AB263123
300	*Eimeria tenella*	AF026388
301	*Entamoeba histolytica*	AB426549
302	*Ergobibamus cyprinoides*	GU827592
303	*Galdieria sulphuraria*	KB454502
304	*Glomus mosseae*	NG_01717
305	*Goniomonas avonlea*	JQ434475
306	*Goniomonas* sp.	AY360454
307	*Guillardia theta*	X57162
308	*Gymnodinium sanguineum*	U41085
309	*Heterosigma akashiwo*	DQ191681
310	*Hicanonectes teleskopos*	FJ628363
311	*Kipferlia bialata*	GU827604
312	*Levinella prolifera*	AM180956
313	*Malawimonas jakobiformis*	AY117420
314	*Malawimonas jakobiformis*	EF455761
315	*Marchantia polymorpha*	X75521
316	*Micractinium reisseri*	AB506070
317	*Monosiga brevicollis*	AF100940
318	*Monosiga ovata*	AF271999
319	*Nannochloropsis gaditana*	KF040086
320	*Nephroselmis olivacea*	FN562436
321	*Oxymonas* sp.	AB326383
322	*Palmaria palmata*	X53500
323	*Palpitomonas bilix*	AB508339
324	*Paramecium tetraurelia*	X03772
325	*Paramicrosporidium vannellae*	JQ796368
326	*Pessonella sp.*	EU273458
327	*Physarum polycephalum*	X13160
328	*Picomonas judraskeda*	JX988758
329	*Planomonas micra*	EF455780
330	*Plasmodium falciparum*	AL844501
331	*Plasmodium ovale*	KF018657
332	*Prorocentrum rhathymum*	HF565181
333	*Proteromonas lacertae*	U37108
334	*Pterocystis tropica*	AY749603
335	*Pyramimonas tetrarhynchus*	FN562441
336	*Pyrsonympha grandis*	AB092942
337	*Salpingoeca infusionum*	AF100941
338	*Sarcocystis alceslatrans*	KF831276
339	*Tetraselmis cordiformis*	HE610130
340	*Theileria annulata*	KF429793
341	*Trichoplax* sp.	AY652578
342	*Trimastix pyriformis*	AF244904
343	uncultured clone BBSR 323	U52356
344	uncultured clone DH141 3A30N	FJ032651
345	uncultured clone DH22 2A36	FJ032657

4.10.2 LSU rDNA sequences

Obtained partial and complete LSU rDNA sequences were integrated into a new mask file, which was built manually using GeneDoc based on recently published LSU rRNA secondary structure data of *Saccharomyces cerevisiae*, adopting domains and helix numbering from

Petrov et al. (2013). LSU rDNA sequences from some phototrophic euglenids and all available diplonemids and kinetoplastids were added to the alignment (Tab. 4.4). Hence no LSU rDNA sequences from jakobid or other excavates were available (in 2014), only one heterolobosean and many other sequences from various higher protozoan groups (and two metazoan taxa) were used as outgroups for alignment and formation of LSU rDNA datasets.

Tab. 4.4: Sampling of euglenozoan and outgroup taxa used for LSU rDNA sequence alignment and corresponding accession numbers sorted by higher groups, then alphabetically. This work = accession pending.

No.	Name/Taxon	Accession
	EUGLENIDA	
1	*Astasia curvata*	this work
2	*Astasia torta*	this work
3	*Colacium mucronatum*	EF999906
4	*Cryptoglena skujae*	this work
5	*Cyclidiopsis acus*	this work
6	*Entosiphon sulcatum*	this work
7	*Euglena gracilis*	X53361
8	*Euglena longa*	AY130223
9	*Eutreptia viridis*	DQ140108
10	*Eutreptiella braarudii*	EU624026
11	*Eutreptiella pomquetensis*	EU624012
12	*Lepocinclis oxyuris*	HQ287919
13	*Monomorphina pyrum*	AY130238
14	*Notosolenus ostium*	this work
15	*Peranema trichophorum*	this work
16	*Petalomonas cantuscygni*	this work
17	*Phacus helikoides*	HQ287923
18	*Ploeotia costata*	this work
19	*Rhabdomonas costata*	this work
20	*Trachelomonas lefèvrei*	AY359949
	DIPLONEMIDA	
21	*Diplonema ambulator*	this work
22	*Rhynchopus euleeides*	this work
23	*Diplonema papillatum*	KF633467
24	Uncultured clone Ma131 1A46	FJ032685
	KINETOPLASTIDA	
25	*Bodo caudatus*	AY028450
26	*Bodo saltans*	AF208890
27	*Bodo saltans*	AY028452
28	*Bodo saltans*	FJ176704
29	*Crithidia fasciculata*	Y00055
30	*Dimastigella mimosa*	FJ176708
31	*Dimastigella trypaniformis*	AY028447
32	*Leishmania major*	FR796423
33	*Neobodo saliens*	FJ176711
34	*Trypanosoma brucei*	NC_008409
35	*Trypanosoma rangeli*	KJ742907
	OUTGROUP (PROTOZOA)	
36	*Acanthamoeba castellani*	GU001160
37	*Ancyromonas sigmoides*	AY752988
38	*Apusomonas* sp.	AY752987
39	*Besnoitia besnoiti*	AY833646
40	*Bigelowiella natans*	DQ158857
41	*Breviata anathema*	GU001164
42	*Capsaspora owczarzaki*	AY724688
43	*Cercozoa* sp.	GQ144690
44	*Chytriomyces hyalinus*	DQ536499
45	*Ciliophrys infusionum*	AB846664
46	*Codosiga gracilis*	EU011935
47	*Cryptomonas paramecium*	CP002174
48	*Cyanidioschyzon merolae*	AB158483
49	*Eimeria tenella*	AF026388
50	*Entamoeba histolytica*	X65163
51	*Galdieria sulphuraria*	KB454502
52	*Glomus mossae*	NG_027652
53	*Goniomonas avonlea*	JQ434476
54	*Goniomonas* sp.	AY752989
55	*Guillardia theta*	AJ010592
56	*Laminaria digitata*	AF331153
57	*Mallomonas asmundae*	AF409122
58	*Micractinium reisseri*	AB506070
59	*Monosiga ovata*	AF271999
60	*Monosiga* sp.	EU011940
61	*Naegleria gruberi*	AB298288
62	*Nannochloropsis gaditana*	AZIL01002195
63	*Nephroselmis olivacea*	HE61046
64	*Palpitomonas bilix*	AB508340
65	*Paramecium tetraurelia*	EU828456
66	*Picomonas judraskeda*	JX988758
67	*Planomonas micra*	GU001169
68	*Plasmodium falciparum*	AL844501
69	*Pyramimonas tetrarhynchus*	HE610152
70	*Saccharomyces cerevisiae*	NR_132218
71	*Salpingoeca amphoridium*	EU011942
72	*Salpingoeca infusionum*	AY026380
73	*Tetrahymena pyriformis*	X54004
74	*Tetrahymena thermophila*	JN547815
75	*Tetraselmis cordiformis*	HE610130
76	Uncultured clone DH141 3A30N	FJ032652
77	Uncultured clone DH22 2A36	FJ032658
	OUTGROUP (METAZOA)	
78	*Levinella prolifera*	JQ272292
79	*Trichoplax* sp.	AY652583

4.11 Datasets

Gapped nucleotide positions were discarded from the alignments, except for secondary structure relevant positions in which only some taxa showed missing nucleotides, those were replaced by 'N', and then several SSU rDNA and LSU rDNA datasets were formed. Derived SSU rDNA and LSU rDNA datasets were concatenated to operon datasets for phylogenetic analyses (Tab. 4.5). Dataset 0 was built as a preliminary dataset with an equilibrated number of taxa representing subordinate groups, while datasets I to III comprised continuously increased taxon samplings. LSU rDNA-based broad dataset IV was built to contain as many nucleotides as possible; modified criteria for dataset formation were applied after Castresana (2000) and the strict LSU rDNA dataset V formed (see 5.3.2). Finally, dataset VI included concatenated SSU and LSU rDNA data.

Tab. 4.5: Datasets built for phylogenetic analyses. Examined genes, size of each dataset and the number of taxa including euglenozoan subgroups therein are shown (S: symbiontids; D: diplonemids; K: kinetoplastids). Operon = concatenated SSU rDNA and LSU rDNA data.

No.	Gene(s) in dataset	Size of dataset	No. of taxa in dataset	Euglenea	Aphagea	phagotrophic euglenids	S	D	K
0	SSU rDNA	1,194	85 / 82	8	8	17 / 14	8	8	8
I	SSU rDNA	1,158	199	10	-	20	19	22	33
II	SSU rDNA	1,224	178 / 175	30	26	18 / 15	-	-	14
III	SSU rDNA	1,030	241 / 238	39	25	33 / 30	9	23	51
IV	LSU rDNA	2,406	44 / 43	4	3	5 / 4	-	4	4
V	LSU rDNA	862	25 / 24	4	-	5 / 4	-	4	6
VI	Operon	3,741	56 / 55	7	3	5 / 4	-	4	4

4.12 Computer analyses

4.12.1 Statistical tests

Several statistical analyses were performed using different tests which are implemented in the software MEGA5. For instance, homogeneity of substitution patterns between sequences was measured with disparity index tests (Kumar and Gadagkar 2001), p-values were calculated using Monte Carlo computations with 500 replicates for dataset 0 and with 1,000 replicates for all other datasets. Evolutionary divergence estimates within and between groups were calculated employing the Maximum Composite Likelihood model (Tamura and Kumar 2002), initially with 500 bootstrap replicates for all datasets.

4.12.2 Model testing

Prior to phylogenetic analyses all datasets were tested for the best-fit model of evolution. Preliminary dataset 0 was tested using the software jModelTest (version 0.1.1, Posada 2008) with default parameters. For practical reasons all other datasets were tested with the Modeltest program (Nei and Kumar 2000) implemented in MEGA5 using all sites in an automatic tree search and branch swap filter set to very strong. The best-fit model of evolution was selected for each dataset according to lowest scores of the three criteria Bayesian information criterion (BIC), the corrected Akaike information criterion (AICc) and the negative Maximum likelihood value (-lnL). In case of conflicting scores, i.e. when scores of one parameter were equal for two different models, then the model was chosen which complied with at least two out of three criteria.

4.12.3 Maximum likelihood analyses

All phylogenetic reconstructions inferring the Maximum likelihood method were performed with MEGA5 conducting 1,000 bootstrap replications using all sites of each dataset. The Nearest-Neighbor-Interchange heuristic method was applied with default initial tree options (automatic, BioNJ) and the branch swap filter was set to very strong. Maximum likelihood analyses of dataset 0 and datasets II to VIII were performed including long-branching *Entosiphon* sequences and then re-iterated without *Entosiphon* sequences.

4.12.4 Bayesian inferences

For the determination of posterior probabilities, Bayesian inferences were calculated for each dataset using the plug-in MrBayes (Huelsenbeck and Ronquist 2001) implemented in Geneious (Kearse et al. 2012), applying unheated Markov chain Monte Carlo sampling with unconstrained branch lengths, including 2,000,000 steps with a sub-sampling frequency of 2,000 and a burn-in of 10 % for the preliminary dataset 0. Posterior probability estimation for all other datasets involved 5,000,000 computation steps, a sub-sampling frequency of 5,000 steps and a 10 % burn-in, which means that 900 sample trees were selected out of 4,500,000 computations.

4.12.5 Phylogenetic networks

Network analyses were performed with the software SplitsTree (Huson 1998; version 4.13.1 by Huson and Bryant 2013) using default parameters for calculations of neighbor-net (Bryant and Moulton 2004) phylogenetic networks.

4.12.6 Spectral analyses

Phylogenetic signal and spectral analyses were conducted with the software SAMS (version 1.4.3 beta by Mayer, Meid and Wägele 2012) applying pairwise sequence comparisons using splits search mode, the maximum number of splits set to 100 and gap mode ‘missing’.

4.13 Secondary structure analyses

Properties of the secondary structure were examined using the software GeneDoc. The alignment of SSU rDNA sequences included the defined helix numbering from Petrov et al. (2014) based on secondary structure information of *Saccharomyces cerevisiae*. This enabled an identification of the exact boundaries of SSU rDNA ‘coding’ and variable regions and thereby a comparative examination of nucleotide composition, GC-values, length variation and other parameters.

The alignment of LSU rDNA sequences based on secondary structure information of *Saccharomyces cerevisiae* and helix numbering which was adopted from Petrov et al. (2013). This permitted a determination of exact boundaries of the internal transcribed spacer (ITS) regions which allowed a comparative survey of nucleotide sequence composition, group specific length variation of ITSs and LSU rDNA domains as well as other parameters of this rather understudied gene region.

5 Results and Discussion

In this chapter, results of phylogenetic and sequence analyses are presented and for a better understanding, evolutionary implications and other aspects are discussed as well as compared with results from other studies and the literature. For a final discussion of results and concluding remarks see chapter 6 Conclusion.

5.1 Phylogenetic analyses of SSU rDNA sequences

5.1.1 Preliminary dataset

In Maximum likelihood (ML) and Bayesian inference (BI) analyses of dataset 0, the euglenid *Entosiphon* clade arose as deepest and longest branch of the Euglenozoa and thereby, together with the next branching clade, the very good supported monophyletic Petalomonadida, rendered the Euglenida paraphyletic (not shown). But the *Entosiphon* clade exhibited a disproportionate long branch, which made the position as the deepest branching euglenozoan questionable and raised the assumption, that *Entosiphon* caused a long-branch attraction artefact in the analyses. A disparity index test showed no clade specific homogeneities in substitution patterns between *Entosiphon* and other euglenozoan sequences (see xlsx-file 'DItestSSU' in folder 'Supplement' on the CD). In order to gain an understanding for the deep and long branching of the *Entosiphon* group, the evolutionary divergence between and within relevant euglenozoan groups were estimated. In comparison of divergence estimates between groups *Entosiphon* revealed the highest divergence estimates between all groups

Tab. 5.1: Evolutionary divergence over sequence pairs between selected groups of dataset 0. Evolutionary divergence estimates are shown in the lower left, corresponding standard deviations to the upper right. Phagotrophs are phagotrophic euglenids excluding Petalomonadida and *Entosiphon*. Estimates for *Entosiphon* sequences in bold.

No.	Name of group	1	2	3	4	5	6	7	8	9	10	11
1	Euglenea		0.043	0.036	0.032	0.038	0.024	0.040	0.104	0.047	0.048	0.047
2	Aphagea	0.502		0.047	0.044	0.057	0.037	0.046	0.103	0.063	0.058	0.060
3	Symbiontida	0.372	0.534		0.029	0.039	0.031	0.038	0.094	0.045	0.044	0.046
4	Diplonemida	0.339	0.487	0.293		0.031	0.024	0.034	0.094	0.042	0.034	0.038
5	Kinetoplastida	0.428	0.619	0.396	0.309		0.034	0.046	0.104	0.050	0.042	0.044
6	phagotrophs	0.310	0.486	0.368	0.298	0.415		0.034	0.088	0.041	0.039	0.039
7	Petalomonadida	0.446	0.556	0.386	0.340	0.468	0.408		0.091	0.047	0.043	0.043
8	*Entosiphon*	**0.964**	**1.023**	**0.856**	**0.859**	**0.989**	**0.875**	**0.881**		0.109	0.108	0.106
9	Heterolobosea	0.572	0.734	0.530	0.498	0.580	0.537	0.566	**1.058**		0.030	0.034
10	Jakobida	0.502	0.647	0.458	0.349	0.456	0.459	0.454	**0.991**	0.413		0.025
11	distant outgroup	0.575	0.730	0.538	0.476	0.538	0.526	0.529	**1.033**	0.494	0.334	

(Tab. 5.1). Although *Entosiphon* displayed the lowest divergence estimates within groups, omission of *Entosiphon* sequences from selected phagotrophic clades resulted in a distinctive decrease of divergence estimates, regardless of complexity or taxonomic hierarchy of groups investigated (Tab. 5.2). These findings affirmed the presumption that *Entosiphon* caused a long-branch attraction artefact in the first analyses. As a consequence, the *Entosiphon* sequences were discarded from dataset 0 and phylogenetic analyses reiterated without them.

Tab. 5.2: Evolutionary divergence over sequence pairs within groups of dataset 0. Divergence estimates (d) and corresponding standard deviations (Δ) are shown for euglenozoan subgroups and higher ranking taxa of the ingroup and outgroups.

Taxon	d	Δ
Euglenea	0.1624	0.0145
Aphagea	0.2803	0.0235
Euglenid crown clade	0.3650	0.0266
Symbiontida	0.0811	0.0073
Kinetoplastida	0.1427	0.0145
Diplonemida	0.0591	0.0068
Euglenida	0.4848	0.0348
Euglenida excl. *Entosiphon*	0.3941	0.0266
phagotrophic Euglenida	0.5042	0.0390
phagotrophs excl. Petalomonadida	0.5369	0.0461
phagotrophs excl. *Entosiphon*	0.3355	0.0251
phagotrophs excl. Petalomonadida and *Entosiphon*	0.3005	0.0222
Petalomonadida	0.1442	0.0138
Entosiphon clade	0.0394	0.0071
Euglenozoa	0.4364	0.0299
Euglenozoa excl. *Entosiphon*	0.3821	0.0260
Heterolobosea	0.3690	0.0277
Discicristata (Heterolobosea + Euglenozoa)	0.4713	0.0319
Jakobida	0.1396	0.0133
Discoba (Jakobida + Discicristata)	0.4744	0.0323
distant outgroup	0.3532	0.0244

As a result of reiterated ML and BI analyses without *Entosiphon* sequences, the monophyly of Euglenozoa received maximal support (Fig. 5.1). Monophyletic Petalomonadida recovered maximal support as the deepest branch of the Euglenozoa, thus confirming the paraphyly of Euglenida. Other taxa maintained their positions in the tree topology. Aphagea, Diplonemida, Euglenea, Kinetoplastida and Symbiontida formed very good supported monophyla within the Euglenozoa. Diplonemida and Kinetoplastida were resolved as poorly supported sister group, and the phagotrophic *Keelungia pulex* was found to be the sister taxon of Symbiontida, though weakly supported. Both of these putative sister clades situated between early branching Petalomonadida and *Ploeotia costata*, but received no statistical support, depicting

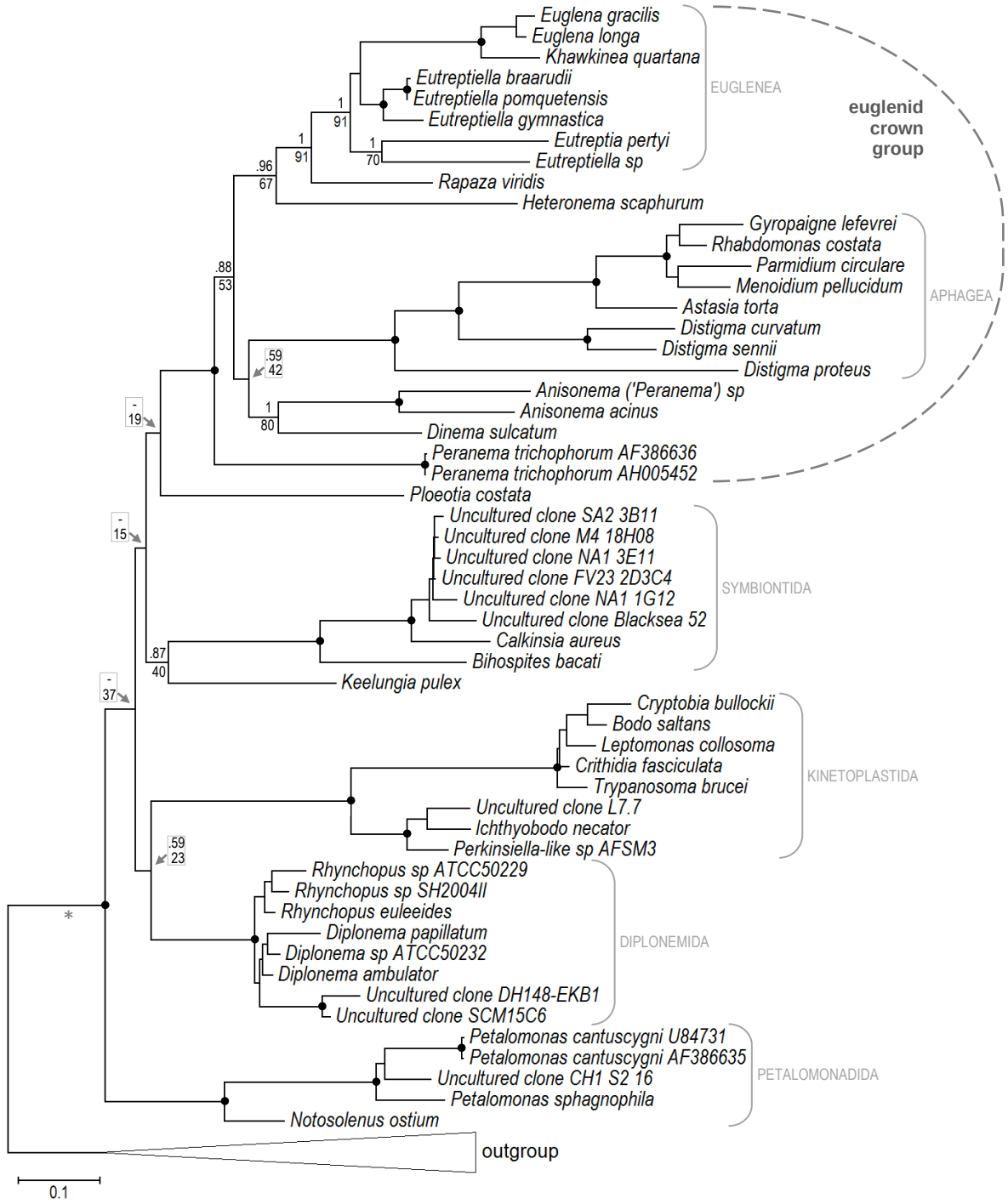


Fig. 5.1: Consensus tree of reiterated Maximum likelihood (ML) and Bayesian inference (BI) analyses of preliminary dataset 0 comprising 82 taxa without long-branching *Entosiphon* sequences (initial *Entosiphon* branching point is marked by a grey asterisk). Congruent BI posterior probability estimates were mapped onto the ML tree (GTR+Γ+I, -lnL = 39422.78, gamma shape = 0.760, p-invar = 0.172) and are shown above, bootstrap values below relevant nodes. Black circles highlight congruent nodes with very high statistical support, i.e. bootstrap values of at least 98 and posterior probabilities of 1.00, respectively. Discrepancies to the Bayesian tree are hyphenated. Values for residual nodes within monophyletic subclades were ignored for reasons of clarity. The scale bar represents 10 % divergence.

a rather weak ‘backbone’ or ‘stem’ for the Euglenozoa. Discrepancies between ML and BI trees were a result from a change of branching points of *Ploeotia costata* and the above-mentioned sister clades. *Peranema trichophorum* sequences were found to be sister to all

remaining euglenids within a maximal supported euglenid crown group. A robustly supported clade consisting of *Dinema sulcatum* and *Anisonema* sequences appeared as sister group of the Aphagea. The phagotrophic euglenid *Heteronema scaphurum* recovered ambiguous support as sister taxon of a robustly supported clade containing *Rapaza viridis* and Euglenea.

Monophyly of the Euglenozoa was strongly supported in all analyses, concurrent with earlier results based on SSU rDNA without Symbiontida (e.g. Busse et al. 2003, Maslov et al. 1999, Preisfeld et al. 2001) and recent studies including the Symbiontida (Behnke et al. 2006, Breglia et al. 2010). In addition, the recovered monophyly of major euglenozoan groups corroborated results from previous studies which discovered monophyly of Diplonemida (Busse & Preisfeld 2002a, Moreira et al. 2001), Kinetoplastida (Doležel et al. 2000, Simpson et al. 2002), Symbiontida (Yubuki et al. 2009, Zuendorf et al. 2006) and of major euglenid groups as Aphagea (Busse & Preisfeld 2002b, Marin et al. 2003), Euglenea (Busse & Preisfeld 2003b, Linton et al. 2000) and Petalomonadida (Šlapeta et al. 2005, von der Heyden et al. 2004).

Contrasting juxtaposition

The relationships between the major euglenozoan lineages and their positions within the Euglenozoa were not fully resolved, because the putative sister clades Kinetoplastida/ Diplonemida and Symbiontida/*Keelungia pulex* hampered the 'backbone' of the tree by displaying weak node support. To test for perseverance in tree topology and further investigate the inter-relationships of these three monophyletic major groups within the Euglenozoa, additional analyses were performed to examine the branching effects of these lineages by contrasting juxtaposition. Therefore, ML and BI analyses were rerun after elimination of each group from the dataset (i.e. firstly without kinetoplastids, but with diplonemids and symbiontids included, then without diplonemids, but with kinetoplastids and symbiontids included, etc.) and after elimination of group pairs (i.e. only kinetoplastids, then only diplonemids, etc.), to observe the impact of each group and/or group pairs on the tree for reasons of comparison. All three groups were tested separately and in mixture of each other, which means that all possible combinations (2^3) were calculated once, and then reiterated excluding *Entosiphon* to test for a long-branch attraction effect. Intriguingly, the exclusion of long-branching *Entosiphon* had no effect on the tree topologies or the positions of major euglenozoan groups and a persisting tree topology was observed throughout additional analyses: the basal position of Petalomonadida as deepest branch within the Euglenozoa was confirmed in all analyses with maximal support, and the strongly supported crown group,

embracing *Peranema trichophorum*, *Dinema sulcatum*, *Anisonema acinus*, *Anisonema* sp., Aphagea, *Heteronema scaphurum*, *Rapaza viridis* and Euglenea, maintained its position as euglenid crown clade uniting euglenid flagellates owning a helical pellicle (termed Helicales

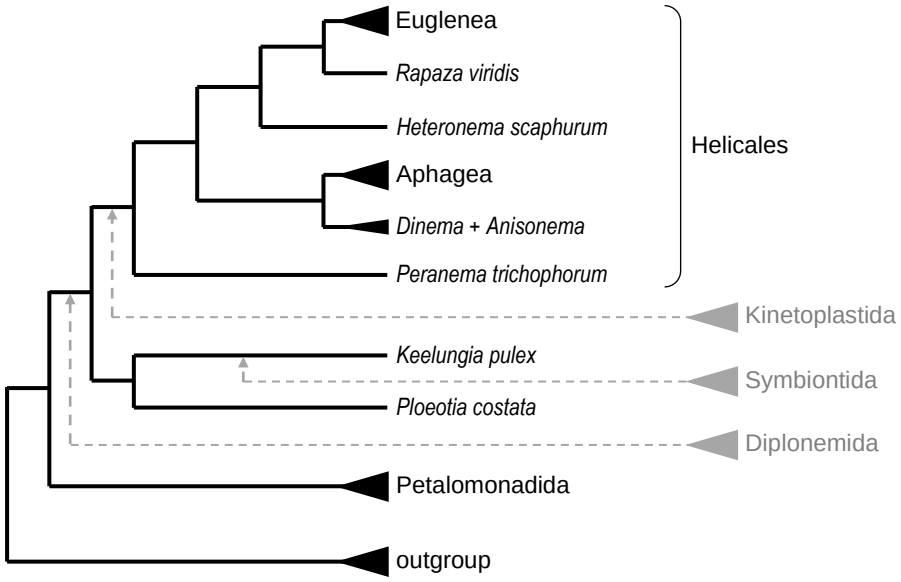


Fig. 5.2: Schematic phylogeny of the Euglenozoa obtained from additional ML analyses of dataset 0 showing the branching effects of major euglenozoan lineages examined by contrasting juxtaposition. The perseverative tree topology is depicted in black, subsidiary lineages in grey and corresponding branching points with dashed arrowheads. The euglenid crown group was named Helicales, for explanation see text.

in Fig. 5.2, and see next paragraph). Statistical support for this clade was strengthened by the exclusion of *Entosiphon* in the results of rerun analyses. The phagotrophic euglenids *Ploeotia costata* and *Keelungia pulex* changed positions between juxtaposed euglenozoan lineages, in 50 % as sister taxa, but with weak support. Comparison of the branching effects of major euglenozoan lineages revealed that segregated Kinetoplastida branched as unreliably supported sister group to the euglenid crown clade, that separated Diplonemida emerged between Petalomonadida and the *Ploeotia costata*/*Keelungia pulex* clade, although weakly supported, and that isolated Symbiontida appeared as poorly supported sister group to *Keelungia pulex*. Nonetheless, these results confirmed the paraphyly of the Euglenida.

Helicales – a new name for a new taxon?

The recurrently identified euglenid crown clade unites euglenid taxa which are morphologically characterized by a helical pellicle. In fact, this taxon is not new – it has been identified earlier, whether without being named (e.g. Busse & Preisfeld 2002a, Preisfeld et al. 2001) or with different names: it was termed ‘clade G’ in a study based on morphological

characters (Leander et al. 2001), as well as ‘clade H’ (Busse et al. 2003) and ‘HP grouping’ (Lee & Simpson 2014a) in studies based on SSU rDNA analyses. For reasons of clarity, a taxon should be termed properly, e.g. with a name describing the shared evolutionary characteristic (i.e. ideally the autapomorphy) of this certain group of organisms. Most descriptive terms corresponding to the denotation ‘helical’ already existed in zoology and were applied to gastropod taxa of different hierarchic level, but the term ‘Helicales’ was not found to be occupied. As a consequence, for reasons of intelligibility, the name ‘Helicales’ was used within this work as a descriptive denomination for a well-known monophyletic euglenid crown clade comprising euglenid flagellates that are morphologically characterized by a helical pellicle, but without classifying a taxonomic rank necessarily. Taxonomic implications of the Helicales are discussed in Chapter 6.3.

5.1.2 Marine versus freshwater Euglenozoa

Although monophyly of major lineages within the Euglenozoa and Euglenida was confirmed by analyses of the preliminary dataset, the weakness of node support values for internal branches remained an impediment for euglenozoan SSU rDNA phylogeny. Another approach was used to address this problem, for the Euglenozoa are known to differentiate in regard to their aquatic habitat. Some available sequences have been extracted from organisms that represent groups which inhabit marine environments exclusively, e.g. SSU rDNA sequences of Diplonemida and Symbiontida, while the representatives of other groups populate as well marine as freshwater environments, e.g. the Kinetoplastida, Petalomonadida, Euglenea (marine Eutreptiales and freshwater Euglenales) and the phagotrophic euglenids. The primary osmotrophic Aphagea live in freshwater biotopes without exception. Therefore marine and freshwater Euglenozoa were examined separately, each with a greatly increased taxon sampling (Tab. 4.5). Dataset I comprised only marine euglenozoans including sequences from phagotrophic euglenids not available before 2013, e.g. *Ploeotia* cf *vitrea*, *Neometanema* cf *exaratum*, *Dinema platysomum* and some new anisonemid sequences. Dataset II contained only freshwater Euglenozoa including SSU rDNA sequences from *Ploeotia edaphica* (CCAP 1265/2) and *Peranema* sp. which both have been isolated in the context of this work.

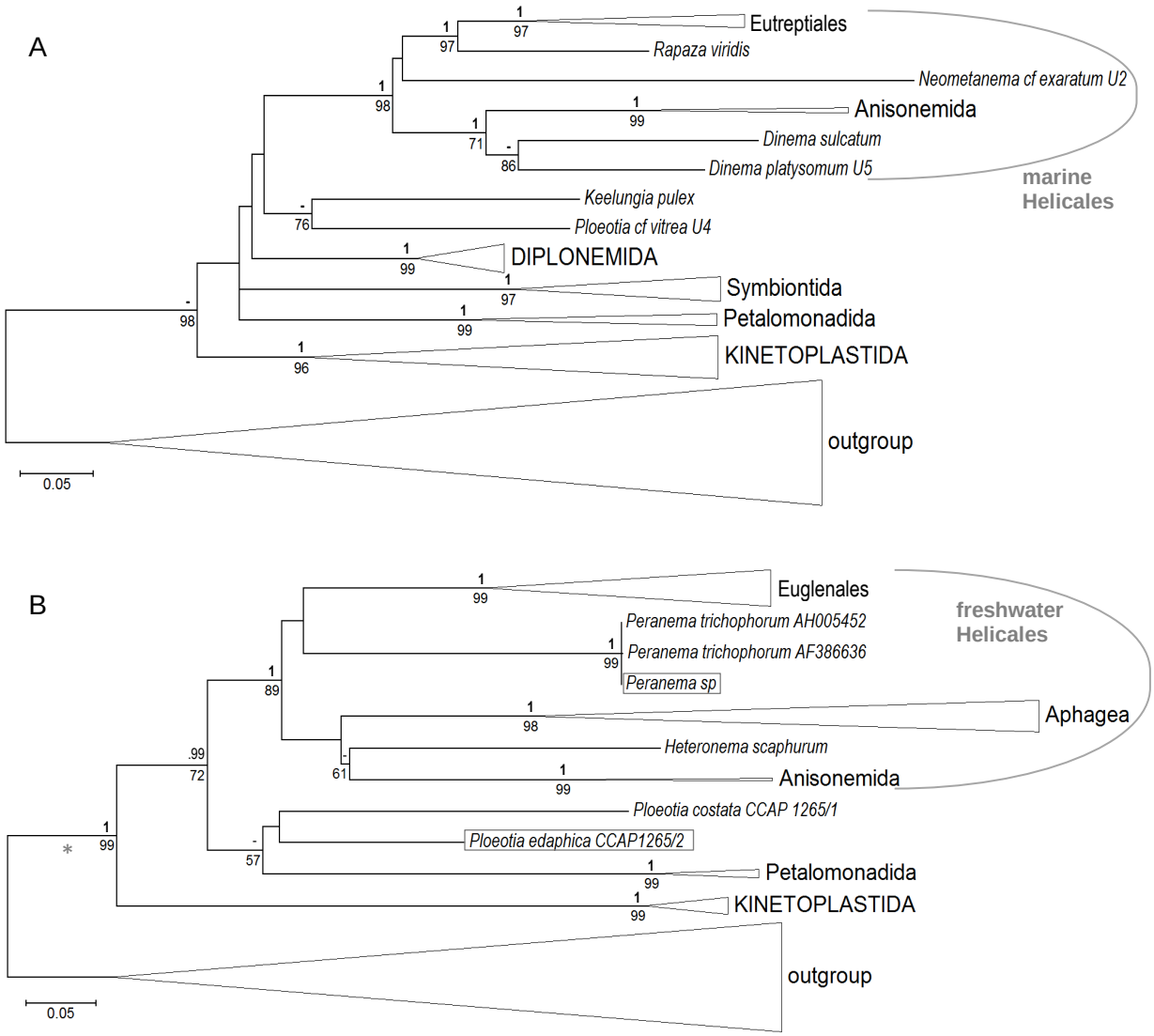


Fig. 5.3: Consensus trees of Maximum likelihood (ML) and Bayesian inference (BI) analyses of marine and freshwater datasets. Congruent posterior probabilities are mapped onto ML trees and shown above, bootstrap values below corresponding nodes; hyphens represent discrepant tree topologies. Scale bars depict 5 % sequence divergence. **A**: Consensus tree of dataset I comprising 104 taxa of marine Euglenozoa (GTR+Γ+I, -lnL = 59748.54, gamma shape = 0.758, p-invar = 0.086). **B**: Consensus tree of reiterated analyses of dataset II containing 85 freshwater euglenozoans excluding *Entosiphon* sequences (GTR+Γ+I, -lnL = 63863.98, gamma shape = 0.784, p-invar = 0.102). New isolated sequences are boxed. The branching point of *Entosiphon* sequences in initial analyses of dataset II is marked by a grey asterisk.

Marine euglenozoan trees

The ML analysis of marine dataset I retrieved well supported monophyletic Euglenozoa with marine kinetoplastids as strongly supported monophyletic sister group to all remaining marine euglenozoans (Fig. 5.3A). Interestingly, marine Petalomonadida appeared as deepest-branching euglenozoan lineage with maximum support in the BI tree. Diplonemida emerged as sister group to Kinetoplastida in the BI tree with maximum support, but branched between symbiontids and remaining euglenids in the ML tree. *Keelungia pulex* appeared as sister to *Ploeotia* cf *vitrea* with moderate support, but in the BI tree as very good supported sister to

the Symbiontida. The monophyly and backbone branches of marine euglenids recovered only poor statistical support, yet marine Petalomonadida, Symbiontida, Anisonemida and Eutreptiales formed very good supported monophyla within marine Euglenozoa. A crown clade representing marine euglenids with a helical pellicle, i.e. marine Helicales, emerged as strongly supported monophylum. Within this clade, *Dinema* sequences gained moderate support as sister group to very good supported monophyletic Anisonemida in the ML tree. In the BI tree *Dinema platysomum* was sister to a *Dinema sulcatum*/Anisonemida clade. The position of *Neometanema* cf *exaratum* as sister to a *Rapaza viridis*/Eutreptiales clade was weakly supported in the ML tree and thus remained doubtful within marine Helicales. The results from analyses of dataset I confirmed the monophyly of Euglenozoa, likewise the monophyly of marine Helicales and major lineages, but statistical support for internal branches remained weak and conflicts between the ML and BI tree topologies persisted even to the deepest branch of the Kinetoplastida.

Freshwater euglenozoan trees

Maximum likelihood analysis of dataset II recovered well-supported monophyletic freshwater Euglenozoa, *Entosiphon* branched basally as sister to all other euglenozoans, but as a result of reiterated analyses, freshwater Kinetoplastida represented the deepest-branching clade (Fig. 5.3B). Monophyly of euglenids gained moderate statistical support, and freshwater Petalomonadida, Aphagea, Anisonemida and Euglenales emerged as very good supported monophyla within the Euglenida. The new isolate *Ploeotia edaphica* and *P. costata* formed a clade though poorly supported, and both formed the sister to Petalomonadida. *Peranema* sp. appeared as sister to *Peranema trichophorum* sequences with very good support. A well-supported monophylum consisting of freshwater euglenids with a helical pellicle, i.e. the freshwater Helicales, arose as crown group of freshwater Euglenida. Contrary to previous findings, results obtained from analyses of this dataset revealed even fewer discrepancies between ML and BI trees: *Peranema* represented the deepest branch within freshwater Helicales, *Heteronema scaphurum* emerged as sister taxon of Euglenales, and *Ploeotia costata* was sister to a Petalomonadida/*Ploeotia edaphica* clade in the BI tree. However, monophyly of Euglenida was most likely an artificial effect, because Kinetoplastida represented the only non-euglenid ingroup in the taxon sampling of dataset II.

Although phylogenetic reconstruction of datasets I and II confirmed the monophyly of Euglenozoa, Helicales and phagotrophic euglenid clades, e.g. Petalomonadida or Anisonemida, the inter-relationships of most euglenozoan lineages remained unclear due to

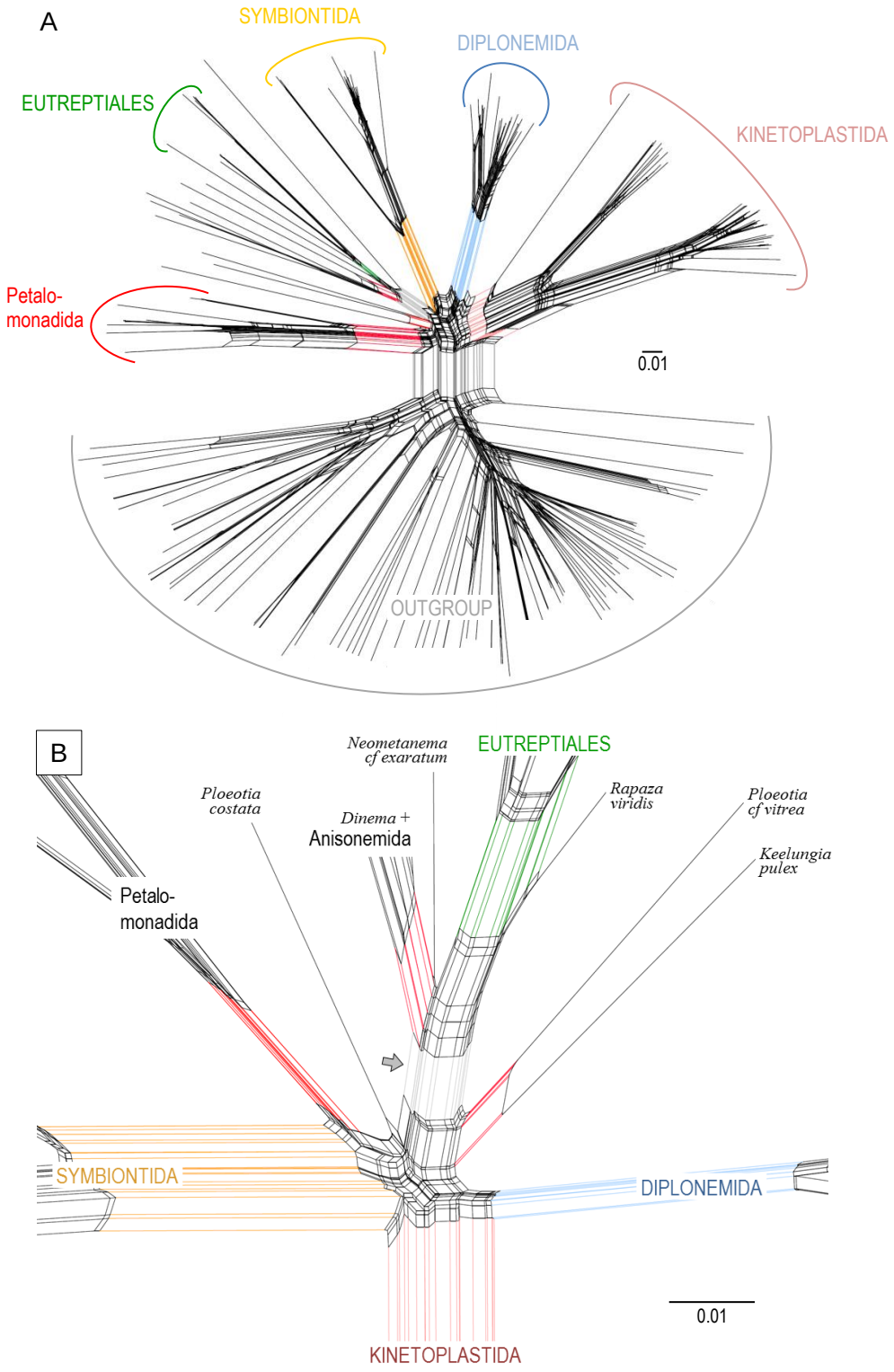


Fig. 5.4: Neighbor-net graph of dataset I comprising marine Euglenozoa and outgroup taxa. Network splits supporting monophyletic clades are colored. Scale bars represent 1 % sequence divergence. **A**: Splits graph overview displaying terminal splits of marine Euglenozoa. **B**: Detailed center view on network after exclusion of the outgroup taxa. Splits supporting the monophyly of marine Helicales are marked by a grey arrow.

lack of statistical node support or persistent incongruences between ML and BI tree topologies. Since tree-like reconstruction methods like Bayesian inference or maximum likelihood force the data into a bifurcate form, the application of phylogenetic networks is

more suitable to analyze weaknesses in phylogenies (Huson and Bryant 2006). Moreover, neighbor-net graphs are capable of visualizing compatible as well as incompatible splits of a phylogenetic network (Bryant and Moulton 2004), and when combined with spectral analysis, enable to better examine ambiguity in a dataset or even identify different types of long-branch attraction artefacts (Wägele and Mayer 2007). Both datasets were examined in phylogenetic network and spectral analyses to further investigate the abiding weakness of node support and the impact of long-branching *Entosiphon* sequences.

Marine euglenozoan networks

Neighbor-net analysis of dataset I corroborated the monophyly of marine Euglenozoa and major euglenozoan groups, i.e. Kinetoplastida, Diplonemida, Symbiontida, Petalomonadida, Anisonemida and Eutreptiales (Fig. 5.4). Marine Euglenozoa displayed a remarkable radiation which was nearly as broad as that of all outgroup taxa together. Interestingly, no splits were found that supported a sister group relationship of any of the major euglenozoan groups. While Kinetoplastida represented the deepest branch in the ML tree, no ancestral euglenozoan lineage was identified in the network graph. Conflicting splits dominated between major euglenozoan taxa, and intriguingly, no splits supported the monophyly of marine Euglenida. Although closely related in the consensus tree, marine Petalomonadida and Symbiontida shared no common splits. *Keelungia pulex* branched near *Ploeotia* cf *vitrea*, but *Ploeotia costata* was positioned elsewhere. Monophyly of marine Anisonemida and the sister group

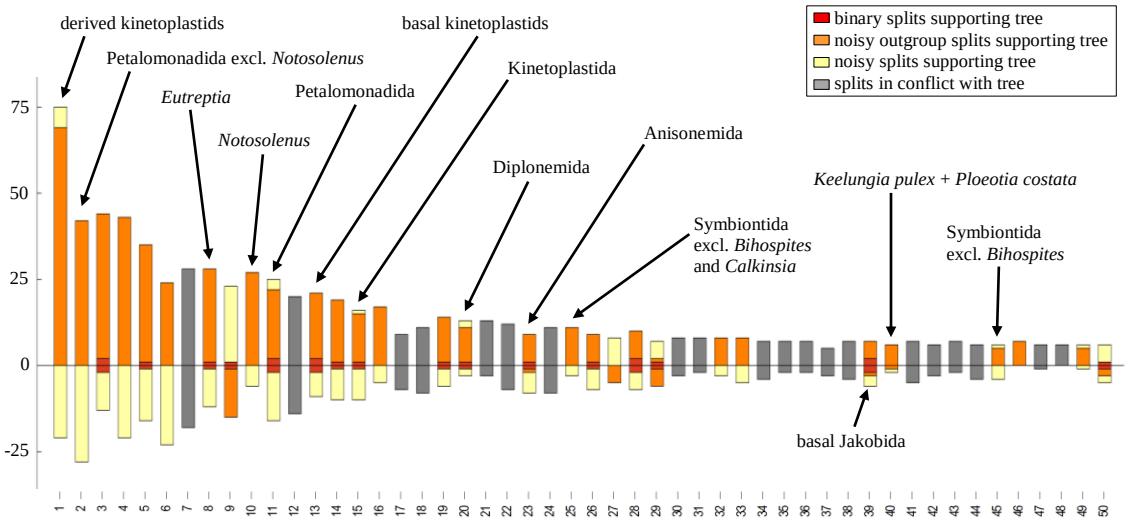


Fig. 5.5: Split support spectrum of the 50 best splits for data used in Fig. 5.4 A. Columns display the number of sequence positions (y-axis) supporting each partition of a specified split (above and below the x-axis) sorted by height. Quality of splits is color coded and depicted in a box to the upper right. Grey columns represent splits that are incompatible with a binary constructed topology. Splits of euglenozoan taxa are assigned by black arrows above, and splits of one jakobid outgroup taxon below the x-axis. All other splits belonged to the outgroup taxon Heterolobosea.

relationship with *Dinema* sequences were confirmed by common splits and *Rapaza viridis* emerged as sister taxon of monophyletic Eutreptiales. Remarkably also the monophyly of marine Helicales was corroborated by supporting splits (grey arrow in Fig. 5.4 B). The spectral analysis of dataset I revealed that 37 of the 50 best splits belonged to the outgroup taxon Heterolobosea which was represented by 33 nucleotide sequences. The majority of these splits (20 out of 37) stood in conflict with a binary constructed (tree) topology, albeit good support for split partitions (Fig. 5.5). Apparently, the heterolobosean splits rather interfered with euglenozoan splits, i.e. conspicuous signals of one outgroup taxon largely overlapped the phylogenetic signals of the ingroup, so the taxon sampling of dataset I needed an adjustment. As a consequence, all outgroup taxa were discarded from the dataset except for a small jakobid group comprising *Jakoba libera*, *Reclinomonas americana* and *Seculamonas ecuadoriensis*, which contained most compatible binary splits. As expected, the split support spectrum of modified dataset I showed a similar distribution of splits regarding marine Euglenozoa, but a higher number of supporting positions (column height), and most major euglenozoan taxa gained substantial split support (Fig. 5.6). Compatible splits clearly affirmed monophyly of marine Petalomonadida and Anisonemida as well as monophyly of Diplonemida, Symbiontida and marine Kinetoplastida. Interestingly, six compatible splits that belonged to internal branches of Kinetoplastida gained better split support than monophyletic

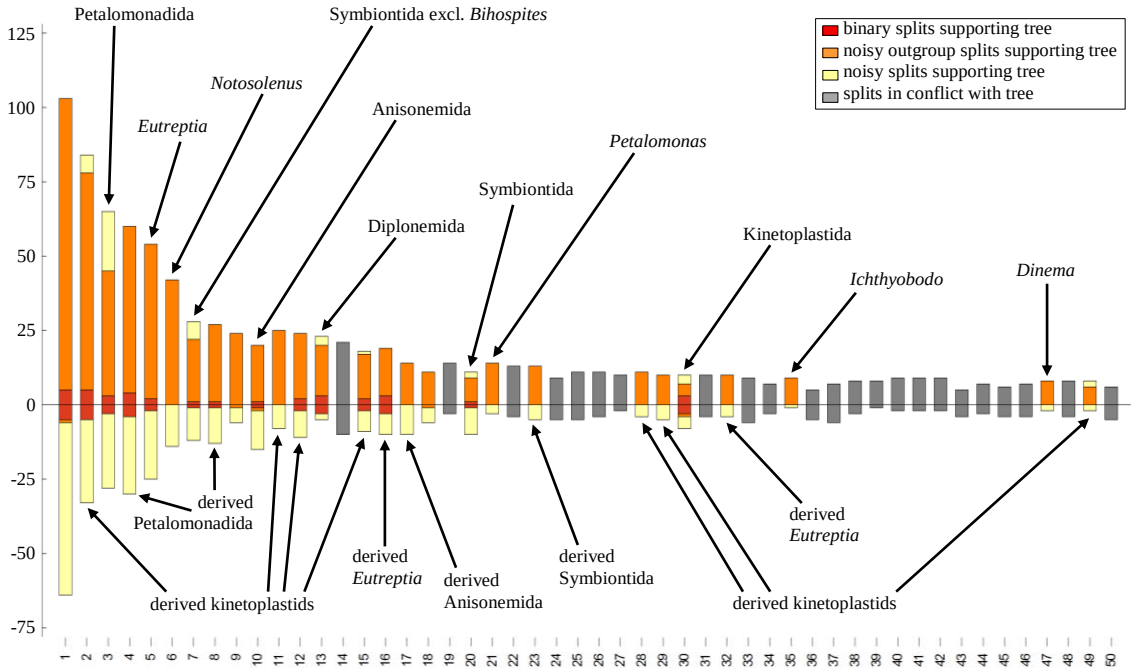


Fig. 5.6: Split support spectrum for modified dataset I illustrating the best 50 splits comprising marine Euglenozoa and the jakobid clade from Fig. 5.5 as outgroup. Euglenozoan taxa are marked by black arrows, compatible splits supporting monophyletic major groups and genera are depicted above columns, other internal splits below columns. Splits 1, 9, 14, 25 and 37 belonged to the outgroup. Grey columns represent incompatible or nonsense splits, e.g. *Ploeotia costata* + outgroup (split 37).

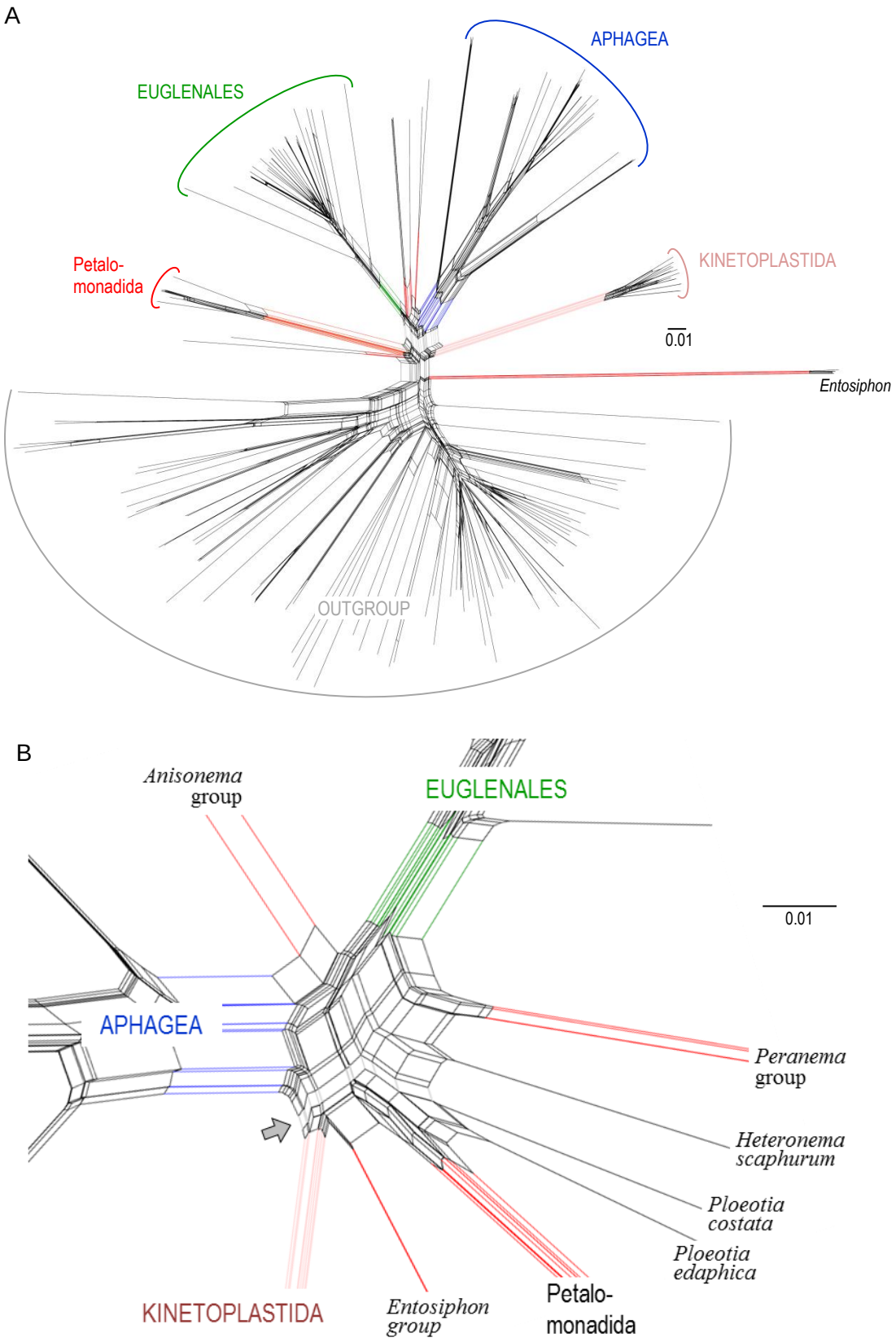


Fig. 5.7: Neighbor-net graph of dataset II comprising freshwater Euglenozoa and outgroup taxa. Network splits sustaining monophyletic clades are colored. Scale bars represent 1 % sequence divergence. **A**: Network overview including terminal splits. *Entosiphon* branched within the outgroup **B**: Detailed center view showing freshwater euglenozoans after exclusion of outgroup taxa. Splits separating marine Helicales from all other taxa are marked by a grey arrow.

Kinetoplastida, which endorsed the relatively high diversity of this group already shown in the network graph (Fig. 5.4). Split support also approved monophyly of several euglenid genera, i.e. *Eutreptia*, *Petalomonas*, *Dinema* and the kinetoplastid genus *Ichthyobodo*.

Freshwater euglenozoan networks

Monophyly of the Euglenozoa was not confirmed in the neighbor-net analysis of dataset II for *Entosiphon* diverged within the outgroup (Fig. 5.7). This was clearly due to its long-branch attraction effect, as observed in earlier results (Figs. 5.1 and 5.3 B), and in addition, Euglenozoa appeared to be monophyletic in reiterated network analysis after exclusion of *Entosiphon* sequences. Similar to findings from analysis of dataset I, no splits were found that supported a sister group relationship of any major euglenozoan groups in the network graph of the freshwater dataset. Interestingly, no splits supported the monophyly of freshwater

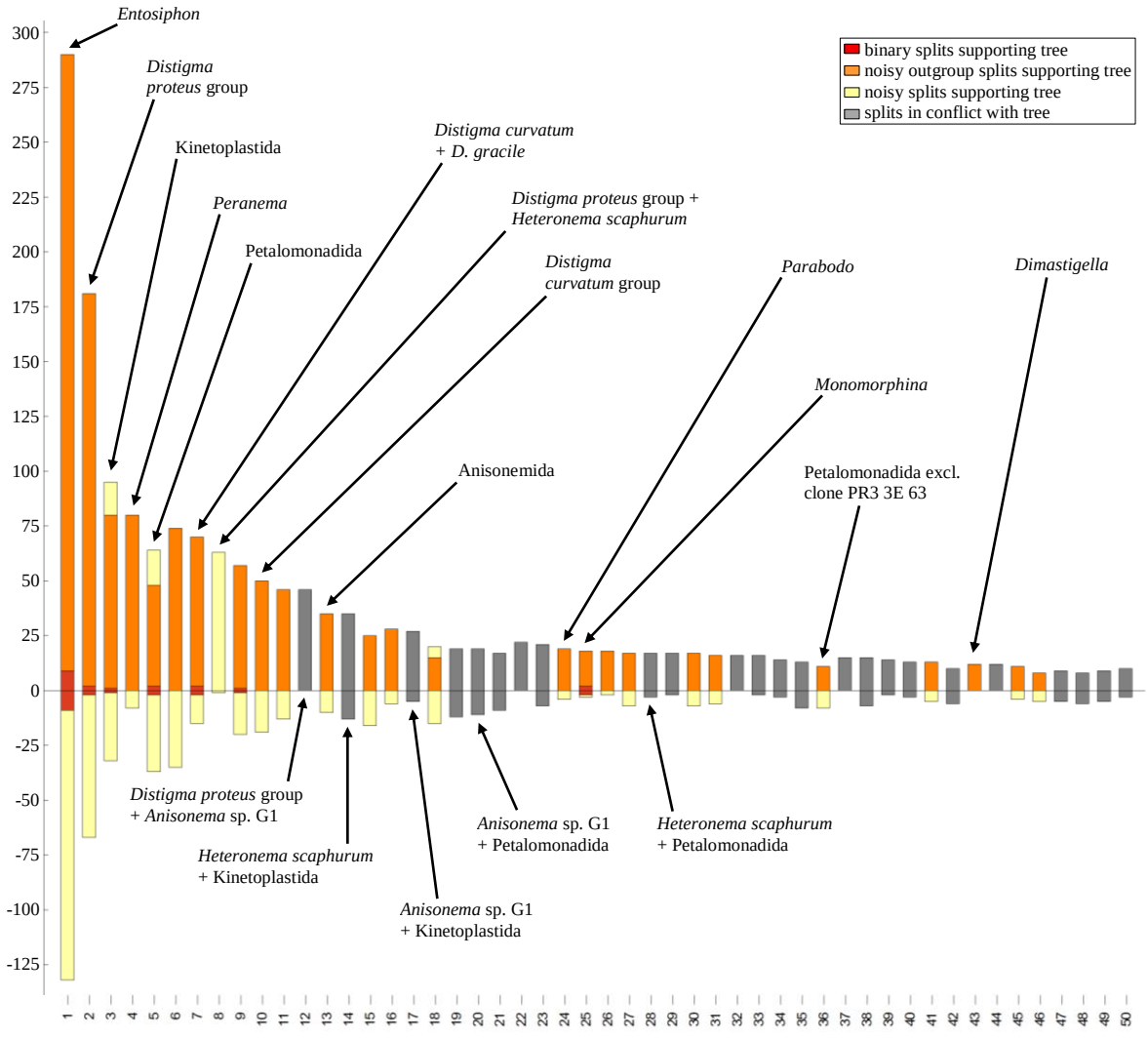


Fig. 5.8: Split support spectrum of the 50 best splits for dataset II. Compatible splits referring to euglenozoan taxa are marked by black arrows above, incompatible splits below columns. All unassigned splits were related to the outgroup Heterolobosea.

Euglenida. Kinetoplastida, Petalomonadida, Aphagea and Euglenales formed monophyletic major groups of freshwater euglenozoans. While *Entosiphon* appeared as the deepest branch in the initial ML tree and Kinetoplastida represented the deepest branch in the reiterated ML analysis, no ancestral euglenozoan lineage was identified in the network graph of dataset II. Nonetheless, freshwater kinetoplastids and non-helical phagotrophic euglenids (i.e. phagotrophic euglenids with longitudinally arranged pellicle strips) like Petalomonadida and *Ploeotia* were located at the basis of the network graph, whereas freshwater Helicales including Aphagea and Euglenales divided as a monophyletic crown clade (Fig. 5.7 A). Although plenty of potentially incompatible splits were determinant in the network graph even after exclusion of outgroup taxa, major groups of freshwater Euglenozoa remained monophyletic, and the monophyly of freshwater Helicales, i.e. Anisonemida, Aphagea, *Heteronema scaphurum*, *Peranema* and Euglenales, gained weak split support (grey arrow in Fig. 5.7 B). *Entosiphon* sequences were found in the vicinity of Kinetoplastida and *Ploeotia* sequences branched near Petalomonadida, but not as sister taxa.

Spectral analysis of dataset II uncovered an extraordinary large number of sequence positions providing support for compatible splits regarding *Entosiphon* (Fig. 5.8), but since *Entosiphon* branched outside of Euglenozoa in the network graph, it must be considered an outgroup taxon. The split support spectrum also revealed that 32 of the 50 best splits belonged to the outgroup taxon Heterolobosea and that most of these splits (19 out of 32) stood in conflict with a binary tree topology. As observed before in spectral analysis of dataset I, phylogenetic signals of the freshwater ingroup have been overlapped by signals of the highly diverse heterolobosean outgroup, though four of the five best splits belonged to Euglenozoa. As a result, outgroup taxa including long-branching *Entosiphon* sequences were removed and spectral analysis rerun. The split support spectrum of modified dataset II showed an overall increase in sequence positions providing support for compatible splits (Fig. 5.9). Inner branches of the Aphagea received the best split support, i.e. the deep-branching *Distigma proteus* clade and the *Distigma curvatum* group, but also the osmotrophic genus *Menoidium* and more derived Aphagea were found among the 50 best splits. Split support for major euglenozoan groups confirmed monophyly of freshwater Kinetoplastida, Petalomonadida, Anisonemida and Euglenales. Compatible splits also supported monophyly of some phototrophic (*Monomorphina*, *Strombomonas*) and kinetoplastid genera (*Parabodo*, *Dimastigella*) as well as that of phagotrophic *Peranema*. Interestingly, a compatible split supported an affiliation of phagotrophic *Heteronema scaphurum* with the deep-branching *Distigma proteus* clade of the Aphagea (split 7) and another split grouped both together with

Peranema (split 28), whereas a split combining the *Distigma proteus* clade with *Peranema* appeared to be incompatible (14). Incompatible splits represented nonsense combinations of taxa: most incompatible splits affiliated Kinetoplastida with a single phagotrophic euglenid taxon (i.e. splits 8, 11, 15, 23, 24, 27, 31, 33, 37, 39, 43 and 46) and others united a single euglenid taxon of the Helicales with a single taxon or group of non-helical euglenids (e.g. *Heteronema scaphurum* with Petalomonadida, split 17).

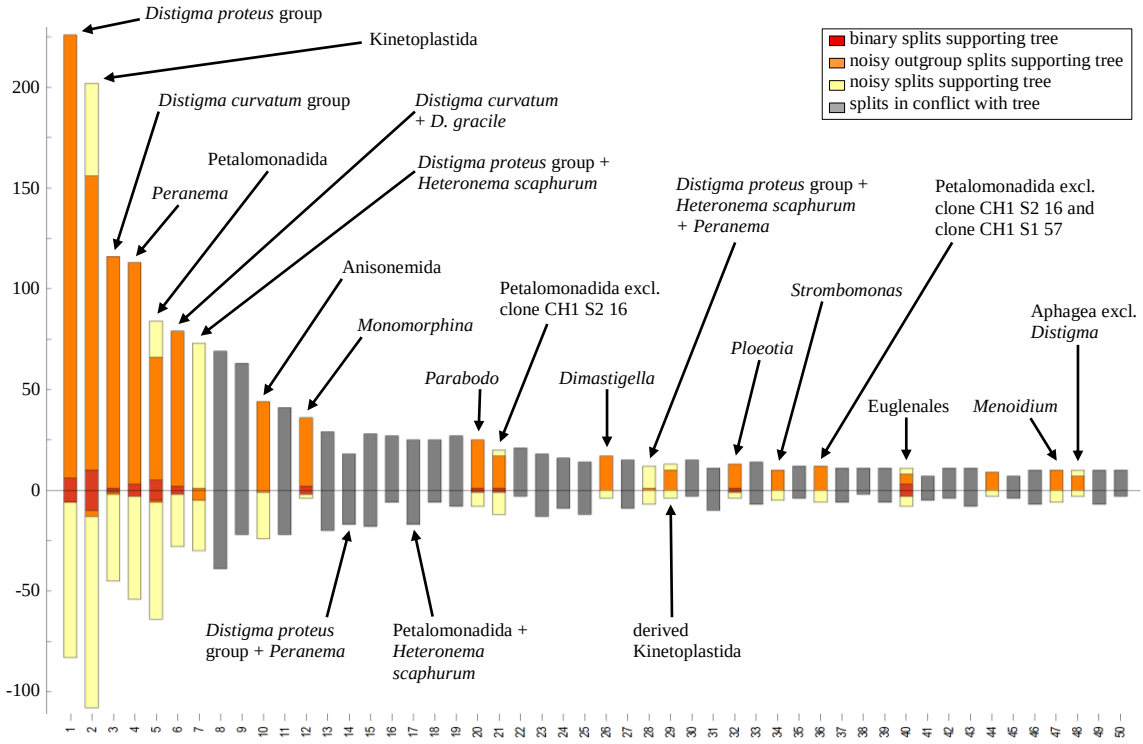


Fig. 5.9: Split support spectrum for modified dataset II accordant with Fig. 5.7 B. Splits of relevant freshwater euglenozoan taxa are marked by black arrows. Conflicting splits are related to supposedly nonsense combinations of taxa, most of which were represented by Kinetoplastida branching with single phagotrophic euglenid taxa.

Phylogenetic analyses in which marine and freshwater Euglenozoa have been examined individually were performed for the first time in the context of this work. The results confirmed significant findings regarding the SSU rDNA genealogy of euglenids, for instance (1) the monophyly of phagotrophic groups like Petalomonadida and Anisonemida, (2) the monophyly of marine and freshwater Helicales, (3) the sister group relationship of the mixotroph *Rapaza viridis* with Eutreptiales, (4) the lack of phylogenetic signal for a sister group relationship of Diplonemida with Kinetoplastida, and (5) the lack of phylogenetic signal supporting monophyletic Euglenida. Importantly, the latter finding corroborated results

from an earlier study which included a spectral analysis of euglenozoan SSU rDNA sequences (Busse & Preisfeld 2003a). Albeit these results, sister group relationships of major euglenozoan groups still remained unclear, because of weak approval for alleged sister clades or undetermined positions of phagotrophic euglenids within possibly paraphyletic Euglenida. However, a paraphyly of Euglenida would have phylogenetic implications which affected the Euglenozoa as a whole and an examination of these implications would be vital to understand euglenozoan diversity.

5.1.3 Combined dataset

To address these problems and further investigate persisting incongruences, marine and freshwater datasets were combined, recently published SSU rDNA sequences added (e.g. *Notosolenus urceolatus* from Lee & Simpson, 2014b), and merged into a third dataset comprising a more extensive taxon sampling of Euglenozoa (dataset III, see Tab. 3.5). The ML tree of dataset III recovered strong monophyletic Euglenozoa, with Diplonemida and Kinetoplastida as poorly supported sister groups, but with strong support in the BI tree (Fig. 5.10 A). Newly obtained *Ploeotia edaphica* gained robust support as sister taxon of *Keelungia pulex*, and *Peranema* sp. retrieved maximum support as sister to *Peranema trichophorum*. Major and minor euglenozoan groups formed very good supported monophyla, i.e. Diplonemida, Kinetoplastida, Petalomonadida, Symbiontida, Aphagea, Anisonemida and Euglenea. However, the ML and BI analyses including long-branching *Entosiphon* sequences resulted in different tree topologies. In the ML tree, *Entosiphon* was found as deepest-branching euglenozoan, being sister to a weakly supported Diplonemida/Kinetoplastida clade and all remaining euglenids including symbiontids with strong support, thereby rendering the Euglenida paraphyletic. In the BI tree, *Entosiphon* nested into the euglenid crown group and formed the sister group to the *Peranema* clade, though with low support. The Helicales appeared as strongly supported monophyletic clade, thus representing the euglenid crown group in the ML tree. *Dinema* sequences branched as sister group to Anisonemida with good support. Though *Rapaza viridis* appeared as strongly supported sister taxon of the Euglenea, positions of other phagotrophic euglenids within the Helicales were not resolved properly, for sister group relationships of *Heteronema scaphurum* with *Peranema* and *Neometanema* cf *exaratum* with the Anisonemida/*Dinema* clade were weakly supported. Petalomonadida and Symbiontida resolved as sister groups, but without statistical support. Although all *Ploeotia* sequences formed a clade including *Keelungia pulex*, the monophyly of this presumptive

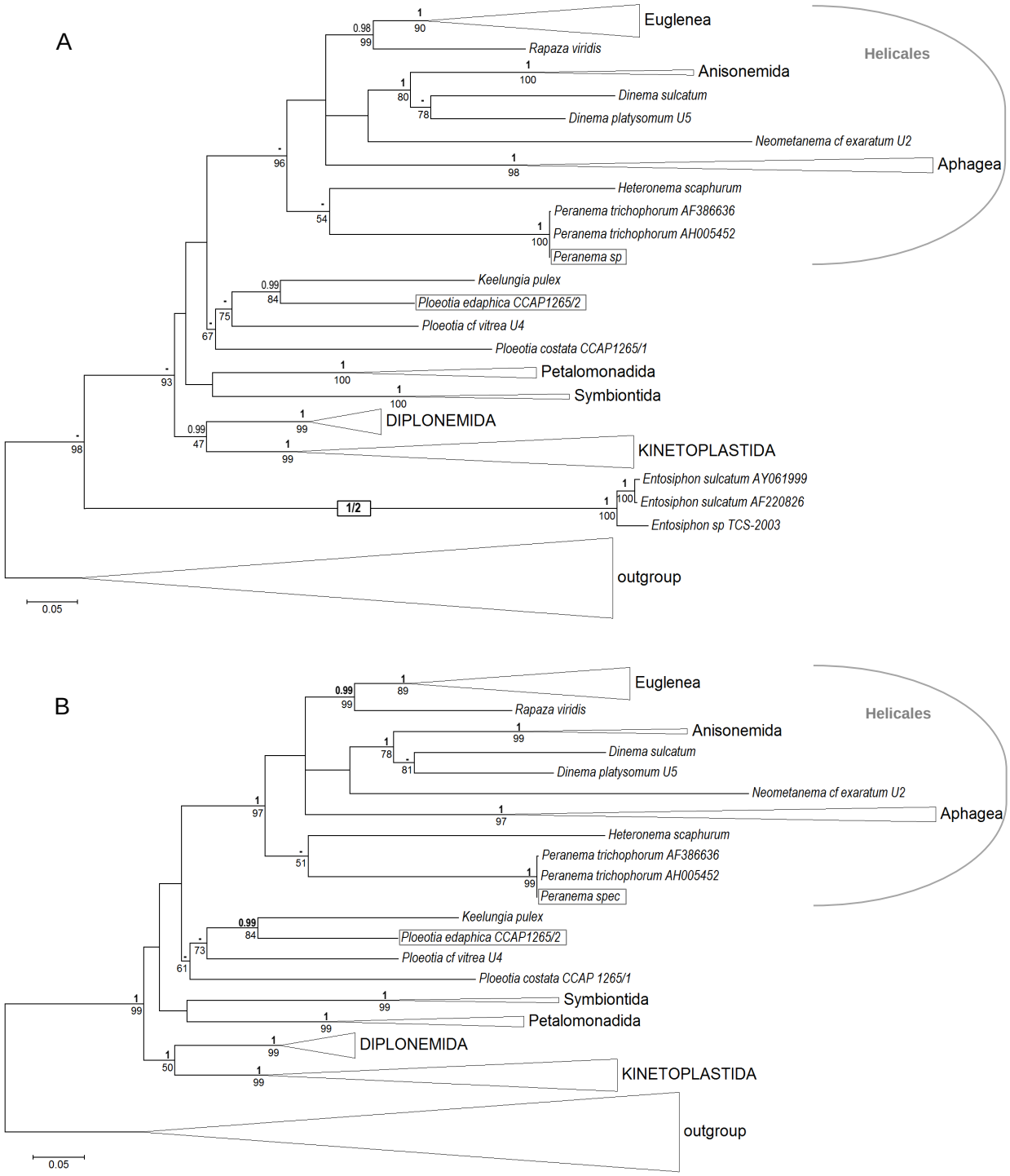


Fig. 5.10: Consensus trees obtained from ML and BI analyses of dataset III comprising 241 taxa with new sequences boxed. Congruent posterior probability values > 0.50 were mapped onto the ML trees and are depicted above, ML bootstrap support values > 47 below corresponding nodes; discrepancies are hyphenated. Scale bars represent 5% sequence divergence. **A**: ML tree containing 180 euglenozoan taxa including long-branching *Entosiphon* (half of the original branch length depicted; GTR+Γ+I, -lnL = 59395.03, gamma shape = 0.673, p-invar = 0.106). **B**: ML tree from reiterated analyses inclosing 177 euglenozoan taxa but excluding *Entosiphon* sequences (GTR+Γ+I model, -lnL = 58194.02, gamma shape = 0.663, p-invar = 0.107).

Ploeotiida clade gained only mediocre support in the ML tree, but intriguingly, *Keelungia pulex* appeared as sister taxon of newly obtained *Ploeotia edaphica* with good to very good support in ML and BI trees, respectively. In results from reiterated analyses excluding long-

branching *Entosiphon* sequences, some discrepancies between ML and BI tree topologies were resolved, and the monophyly of Euglenozoa gained very strong support (Fig. 5.10 B). Removal of *Entosiphon* also slightly improved the very weak bootstrap support for a Diplonemida/Kinetoplastida clade, which represented the sister to monophyletic Euglenida. Nevertheless, statistical support regarding the monophyly of Euglenida and corresponding 'backbone' branches remained poor, and although Petalomonadida and Symbiontida formed sister clades, this combination completely lacked statistical support as the deepest branch of putatively monophyletic Euglenida. Major euglenozoan groups appeared as monophyla, each with very good support. Monophyly of alleged Ploeotiida remained weakly supported, but support for a sister group relationship of *Ploeotia edaphica* and *Keelungia pulex* maintained to be robust. While the monophyly of Helicales gained very good support in both trees, positions of Aphagea and some phagotrophic taxa within the Helicales, i.e. *Heteronema scaphurum* and *Neometanema* cf *exaratum* lacked statistical support and thus remained unclear. The mixotroph *Rapaza viridis* obtained strong support as sister taxon of Euglenea.

The neighbor-net graph of dataset III exhibited a rather high diversity of monophyletic Euglenozoa, which considerably exceeded that of the discoban outgroup taxa (Fig. 5.11 A). *Entosiphon* sequences branched within the ingroup, but as sister taxon of the outgroup, and thereby weakened splits that supported the monophyly of Euglenozoa: splits supporting Euglenozoa excluding *Entosiphon* were longer than those that supported monophyletic Euglenozoa including *Entosiphon*. Similar to earlier results, no splits were found that supported the monophyly of Euglenida, and intriguingly, no splits supported a sister group relationship of Diplonemida and Kinetoplastida either. Both findings contradicted the results obtained from phylogenetic tree reconstruction of dataset III (Fig. 5.10). Network analysis of the combined dataset also recovered major euglenozoan groups as monophyla, i.e. Kinetoplastida, Diplonemida, Petalomonadida, Symbiontida, Aphagea and Anisonemida (Fig. 5.11 B). But no splits supported the monophyly of Euglenea, for two sets of splits embraced this group: the basal splits set embraced all Euglenea but excluded *Trachelomonas grandis*, the other splits set combined all Euglenales to the exclusion of deep-branching Eutreptiales (bluish-green and yellowish-green in Fig. 5.11 B). *Rapaza viridis* branched in the proximity of Euglenea, but not as sister taxon, probably due to a lack of splits supporting the monophyly of Euglenea. *Dinema* sequences formed a clade with Anisonemida which was supported by common splits, and *Neometanema* cf *exaratum* branched between this clade and Aphagea. *Peranema* and *Heteronema scaphurum* grouped together and shared common

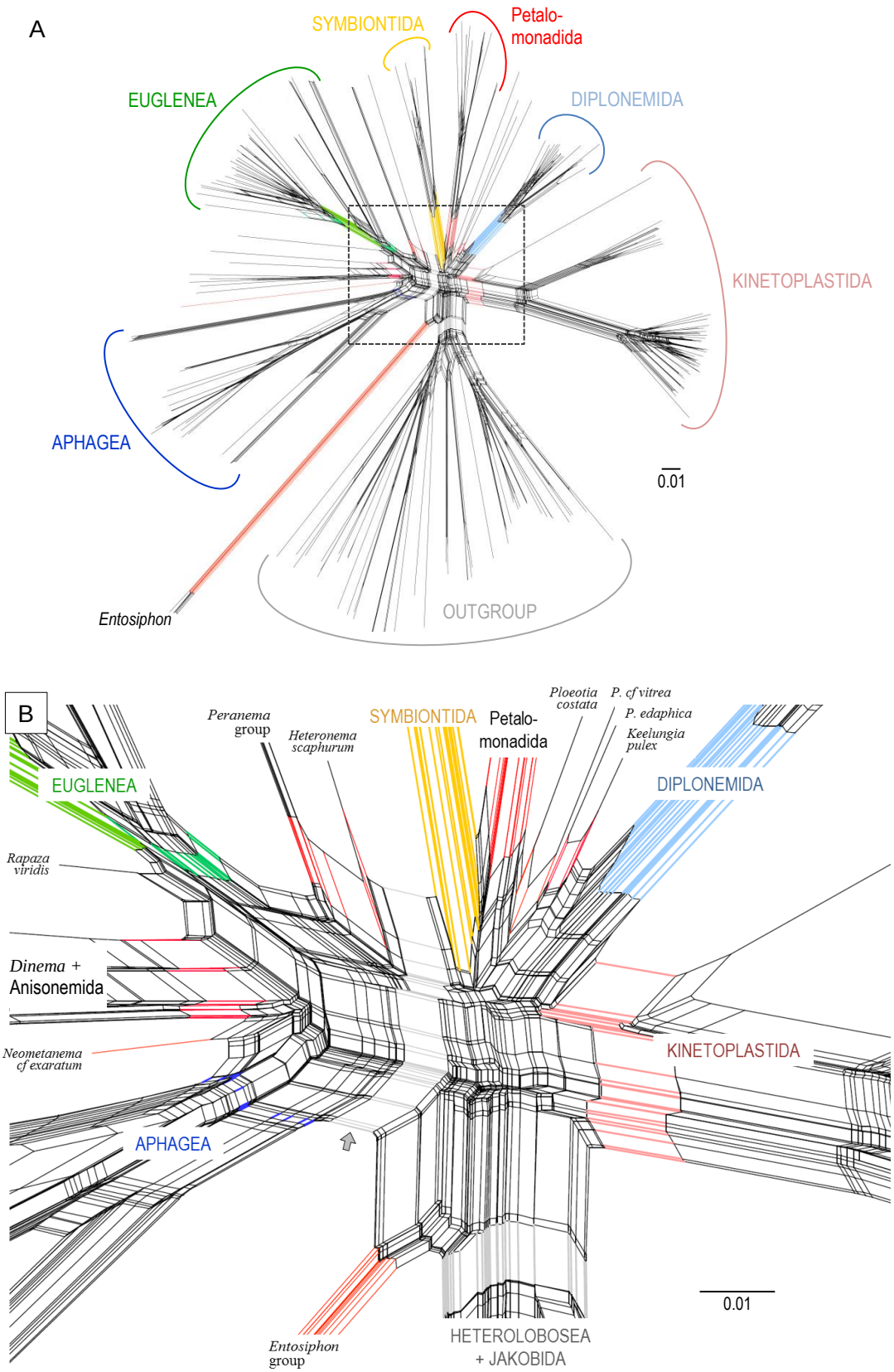


Fig. 5.11: Neighbor-net graph obtained from network analysis of dataset III comprising Euglenozoa and Heterolobosea plus Jakobida as outgroup (i.e. subset A1, see Tab. 5.3). Network splits of monophyletic clades are colored. Scale bars represent 1 % sequence divergence. **A**: Network overview containing terminal splits. A box with dashed lines depicts the scale of the network center shown below. **B**: Detailed center view of the same network. *Entosiphon* branches near the outgroup. A grey arrow highlights splits supporting the monophyly of Helicales.

splits. The monophyly of Helicales was also confirmed by supporting splits (grey arrow in Fig. 5.11 B). *Keelungia pulex* and *Ploeotia* sequences were situated between Petalomonadida and Diplonemida, but no splits supported monophyly for this assemblage. While *Ploeotia costata* shared common splits with *Ploeotia* cf *vitrea*, *Keelungia pulex* grouped together with *Ploeotia edaphica*, the latter corroborated results from ML and BI analyses.

To validate findings from network analysis of the combined dataset that could have been a result of the strong long-branch attraction effect of *Entosiphon* sequences, e.g. the distortion of splits that supported monophyletic Euglenea, the analysis had to be reiterated without them. However, additional network analyses of the combined dataset demanded a double strategy, as taxon sampling also revealed an extraordinary impact on network splits, which has been observed in earlier network results and corresponding spectral analyses; the influence of highly diverse heterolobosean SSU rDNA sequences decreasing split support for well-known major euglenozoan groups is a good example (Figs. 5.4 and 5.5). To follow this double strategy, the taxon sampling of dataset III was gradually reduced in three steps from outgroup to ingroup, thereby also decreasing phylogenetic complexity, and additionally, network analyses were reiterated without *Entosiphon* sequences for each of the four subsets, resulting in eight subsets overall (Tab. 5.3), thus enabled an examination of *Entosiphon's* long-branch attraction effect on euglenozoan SSU rDNA based phylogenetic networks.

Tab. 5.3: Subsets of dataset III derived from stepwise reduction of original taxon sampling. Number and name of subset, phylogenetic scope of each modified taxon sampling, number of taxa therein, and number of euglenozoan subgroups are given (S: symbiontids; K: kinetoplastids; D: diplonemids).

No.	Name	phylogenetic scope of modified taxon sampling	No. of taxa	phagotrophic euglenids	Euglenea	Aphagea	S	K	D
1	A1	Discoba incl. *Entosiphon*	241	33	41	25	9	51	23
2	A2	Discoba excl. *Entosiphon*	238	30	41	25	9	51	23
3	B1	Discicristata incl. *Entosiphon*	217	33	41	25	9	51	23
4	B2	Discicristata excl. *Entosiphon*	214	30	41	25	9	51	23
5	C1	Euglenozoa incl. *Entosiphon*	182	33	41	25	9	51	23
6	C2	Euglenozoa excl. *Entosiphon*	179	30	41	25	9	51	23
7	D1	Euglenida incl. *Entosiphon*	108	33	41	25	9	-	-
8	D2	Euglenida excl. *Entosiphon*	105	30	41	25	9	-	-

Neighbor-net analyses of the subsets derived from stepwise reduction of taxa from dataset III confirmed several results from subsequent ML and BI analyses, the network graphs of subsets A1, A2, B1, C1 and C2 are depicted in Figs. 5.11 to 5.14 (for graphs of other subsets see Figs. 8.11 and 8.12 in the Appendix). Monophyly of the Euglenozoa was confirmed in

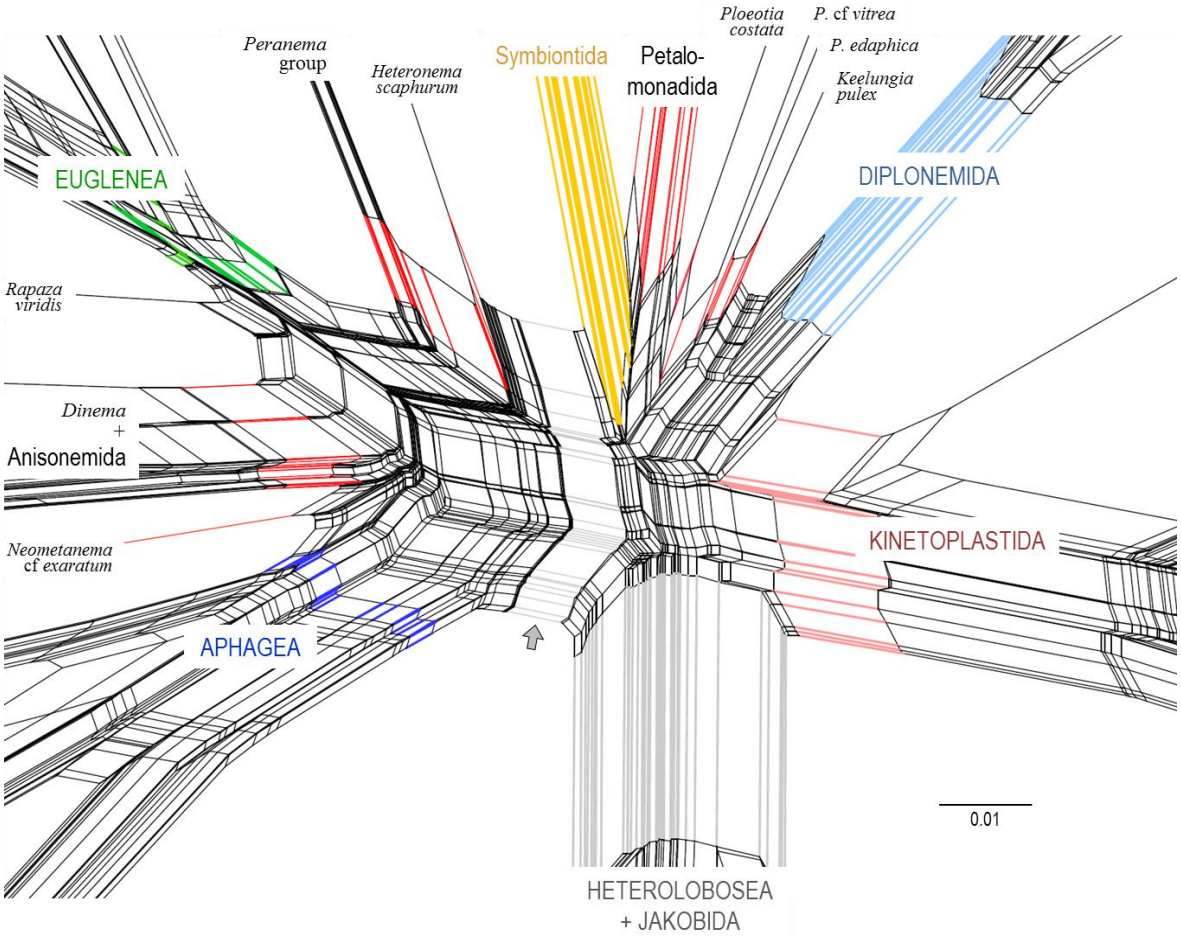


Fig. 5.12: Neighbor-net graph of dataset III comprising Euglenozoa and Heterolobosea plus Jakobida as outgroup, but excluding long-branching *Entosiphon* sequences (i.e. subset A2, see Tab. 5.3). Network splits supporting monophyletic clades are color coded. Splits supporting the monophyly of Helicales are marked by a grey arrow. The scale bar depicts 1 % sequence divergence.

relevant subsets A and B, for both subset types contained taxon samplings including outgroup taxa. While subsets C embraced the Euglenozoa, subsets D constituted artificially monophyletic Euglenida, and both subset types were merely used to examine internal branches of the ingroup and the effects of *Entosiphon* sequences on them. As a result, monophyly of major groups was verified by common splits in all subset analyses, i.e. of Petalomonadida, Symbiontida, Aphagea, Anisonemida and Euglenea. The Symbiontida were positioned next to Petalomonadida in all network graphs, but no splits were found to support a sister group relationship. Sequences of *Ploeotia* and *Keelungia pulex* clustered together in pairs in most networks and even appeared as a group united by common splits in three out of eight subsets analysed (A1, A2, and C2). When clustered in pairs, mostly *Ploeotia edaphica* and *Keelungia pulex* shared common splits, as did sequences of *Ploeotia* cf *vitrea* and *Ploeotia costata*. As an exception, *Ploeotia edaphica* branched with *Ploeotia costata* in the neighbor-net graph of subset B1 (Fig. 5.13). Strikingly, the sequences of *Ploeotia* and *Keelungia pulex* shared common splits with the Diplonemida in the graph of subset B1, and in

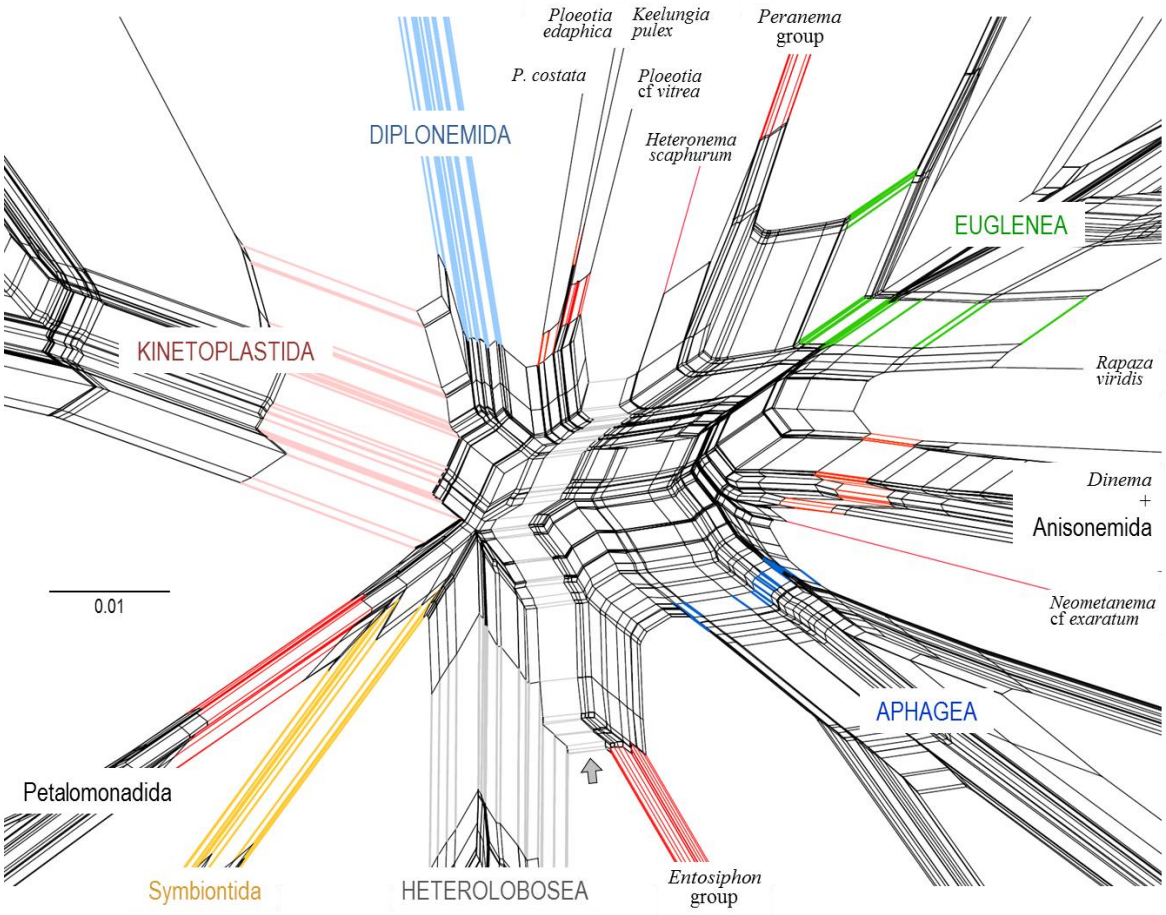


Fig. 5.13: Neighbor-net graph of subset B1 comprising Euglenozoa including long-branching *Entosiphon* sequences and Heterolobosea as outgroup. Network splits supporting monophyletic clades are colored. The scale bar represents 1 % sequence divergence. Splits supporting the monophyly of Helicales are marked by a grey arrow.

network subset types A and C to the exclusion of *Ploeotia costata*. Monophyly of Diplonemida and Kinetoplastida was confirmed in all relevant subset analyses, but only one subset revealed a set of splits which supported a sister group relationship of Kinetoplastida and Diplonemida (subset C1, black arrow in Fig. 5.14 A), and this set of splits represented the only exception, in which Euglenida were found to be a monophylum. But after removal of *Entosiphon* sequences from this subset, no splits were recovered which supported a sister group relationship of Diplonemida and Kinetoplastida, nor splits which supported monophyletic Euglenida (Fig. 5.14 B). Additionally, none of the other results revealed splits supporting the monophyly of Euglenida, even at higher levels of diversity, i.e. within the Discoba and Discicristata networks (subsets A and B), where taxon sampling of the outgroup exceeded that of Euglenida by factors 2.01 and 2.18. *Entosiphon* branched as sister to outgroup taxa in subsets A1 and B1. Exceptionally, a joint set of splits amalgamated *Entosiphon* with Aphagea in the analysis of subset C1, thus distorting the monophyly of Aphagea (splits highlighted by a purple arrow in Fig. 5.14 A). The long-branching *Entosiphon*

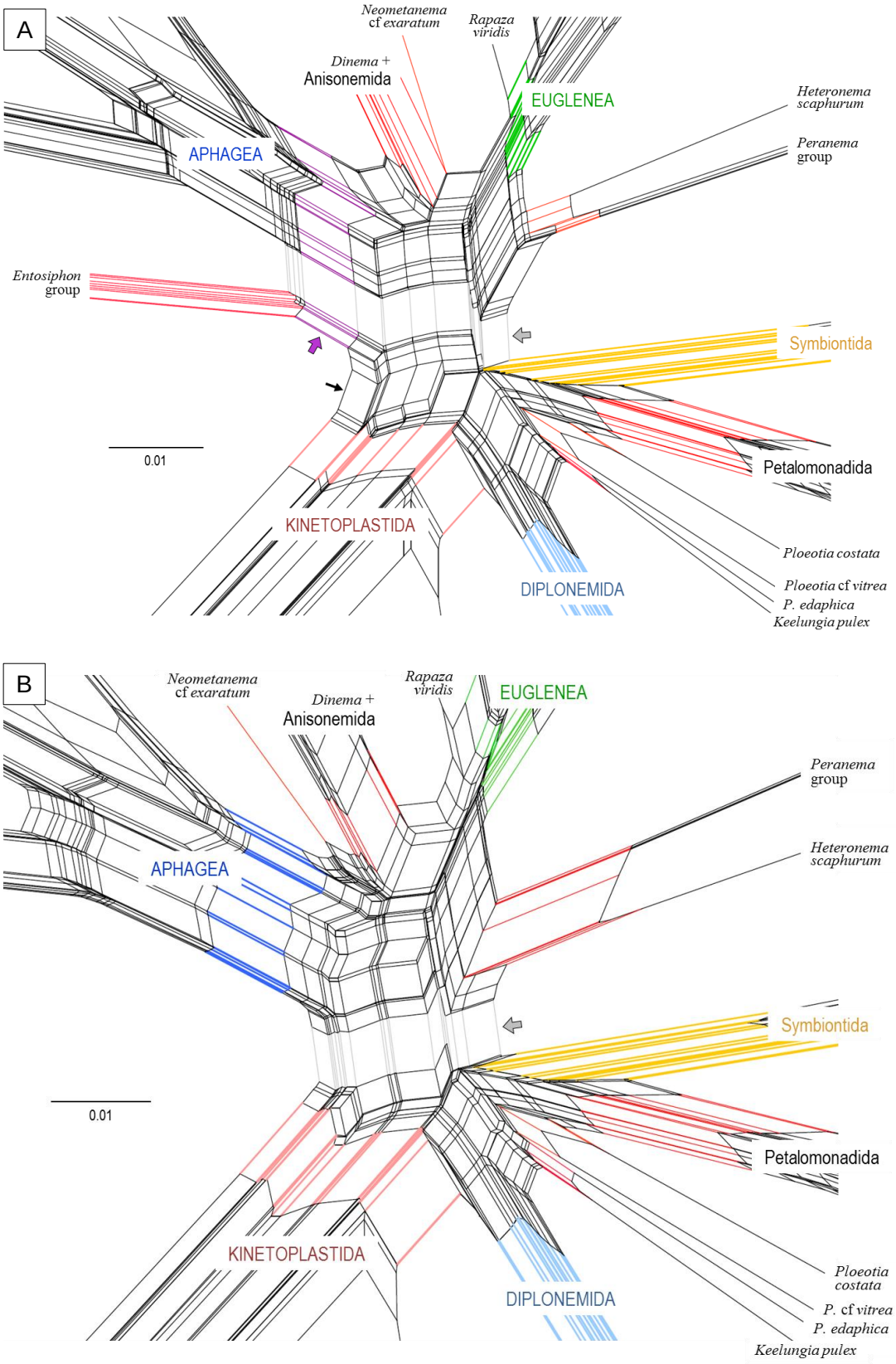


Fig. 5.14: Neighbor-net graphs of reduced dataset III representing type C subsets comprising the Euglenozoa. Network splits of monophyletic clades are colored and splits supporting the monophyly of Helicales are marked by a grey arrow. Scale bars depict 1 % sequence divergence. **A**: Detailed center view on subset C1 including *Entosiphon* sequences. A purple arrow highlights common splits of Aphagea and *Entosiphon*. Splits falsely supporting a sister group relationship of Diplonemida and Kinetoplastida are depicted by a small black arrow. **B**: Detailed center view on subset C2 excluding *Entosiphon* sequences.

sequences were also situated near monophyletic Aphagea in the neighbor-net graph of subset D1, but without a distortion effect or sharing common splits. The monophyly of the Euglenea was sequestered by two internal splits in some networks investigated (in subsets D2, C2, A1 and A2); however, *Rapaza viridis* branched independently from these splits, and always separated from Euglenea. Similar to subsequent ML and BI analyses, euglenids morphologically characterized by a helical pellicle were united by common splits in most network analyses. In the splits graph of subset B1, *Entosiphon* sequences integrated into Helicales combined by a joint split, but the result from subset B2 showed Helicales sharing a common split together with *Ploeotia* cf *vitrea*. If not for the effect of *Entosiphon*, this could have been caused by the diverse sequences of the Heterolobosea, for most network analyses corroborated common splits support for the monophyly of Helicales. Within the Helicales, monophyletic Anisonemida recovered mutual splits together with both *Dinema* sequences in all network graphs. The sequence of *Neometanema* cf *exaratum* branched between monophyletic Aphagea and the *Dinema*/Anisonemida clade in most results. Interestingly, *Heteronema scaphurum* and sequences of the *Peranema* group shared common splits in most subsets examined, the only exception was found in subset D1 (containing artificially monophyletic euglenids), wherein no uniting splits were present, but both grouped together nonetheless.

Results from phylogenetic tree reconstruction and modified network analyses of combined dataset III which comprised the most extensive taxon sampling of euglenozoan SSU rDNA sequences confirmed important findings: (1) the monophyly of Euglenida completely lacked statistical support as well as a phylogenetic signal, (2) statistical support for a sister group relationship of Diplonemida and Kinetoplastida was poor in tree reconstructions and inexistent in network analyses, (3) the euglenid crown group Helicales was monophyletic, (4) SSU rDNA sequences of *Entosiphon* caused a long-branch attraction artefact with a strong influence on in- and outgroup taxa, (5) the phagotrophic euglenid lineages Petalomonadida and Anisonemida were monophyletic. These results imply serious consequences for the phylogeny of Euglenozoa. Paraphyly of the Euglenida would involve that diplonemids and kinetoplastids derived from a common euglenid ancestor. Furthermore, the lack of support for a sister group relationship of Diplonemida and Kinetoplastida could implicate that one of these major groups or even both derived from a phagotrophic euglenid lineage. For additional argumentation and discussion on these findings see chapter 6.

5.2 SSU rDNA nucleotide sequence analyses

Unambiguously aligned, homologous nucleotide sequences were a prerequisite not only for dataset formation prior to phylogenetic analyses, but also for the study of more parameters involving SSU rDNA sequence data information content. Influence of nucleotide distribution and putative secondary structure of sequences on SSU rDNA phylogeny had been examined earlier (Moreira et al. 2001, Marin et al. 2003). In the scope of this work, further examinations of base composition, identity matrix and deduced secondary structures of nucleotide sequences were performed to substantiate or falsify findings from phylogenetic analyses.

5.2.1 Base composition

The composition of nucleotides adenine (A), guanine (G), cytosine (C) and thymine (T) in euglenozoan SSU rDNA sequences displayed only few differences between major groups when contemplated in tabular form. For instance, Euglenea and Aphagea exhibited the highest percentage values of guanine and cytosine, while the lowest guanine and cytosine values were found in outgroup taxa, but nucleotide percentage values of phagotrophic euglenids, diplonemids and kinetoplastid taxa were seemingly too similar to demonstrate differences properly (a table containing base composition data from taxon- rich combined dataset III is given in the Appendix, see Tab. 8.1). However, differences in base composition between phagotrophic euglenids and diplonemid taxa became more apparent when displayed in a base frequency graph, which simultaneously allowed a correlation of nucleotide percentage values (Fig. 5.15). Intriguingly, SSU rDNA sequences of phagotrophic euglenids *Notosolenus ostium*, *Ploeotia edaphica* and *Keelungia pulex* revealed a similar base frequency pattern as the diplonemids *Diplonema ambulator* and *Rhynchopus euleeides*, in which adenine and guanine displayed relatively high percentage values ($\leq$ 28 %), while cytosine was the least frequent nucleotide (< 22 %), resulting in the pattern C < T < G < A. Frequency values for guanine and adenine were almost identical in sequences of *Notosolenus ostium*, *Ploeotia edaphica* and *Diplonema ambulator*. Interestingly, the enigmatic phagotrophic euglenid *Entosiphon* displayed the same pattern of base frequency, i.e. C < T < G < A (see Tab. 8.1). Both most primordial kinetoplastid clones AT4-103 and LC103 5EP 19 as well as the deep-branching symbiontid *Bihospites bacati* exhibited a different base frequency pattern, in which guanine was the most frequent nucleotide (> 28 %) and thymine was the least frequent base

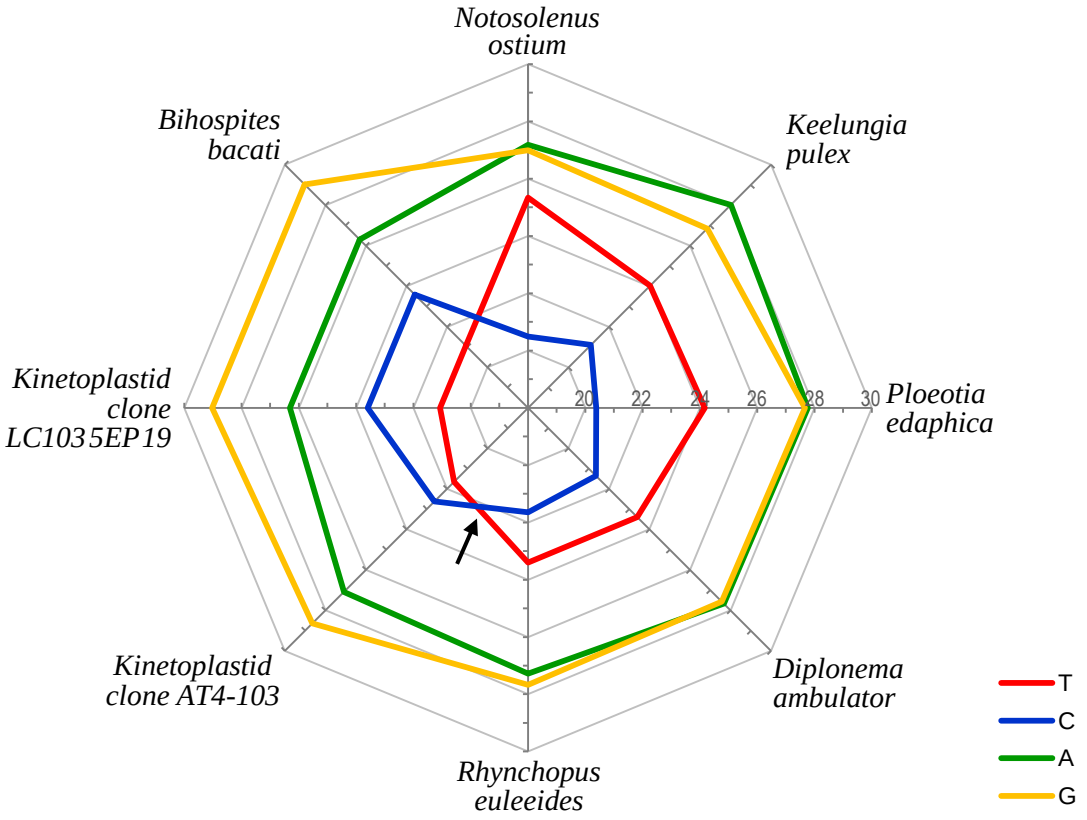


Fig. 5.15: Octagonal base frequency graph visualizing differences in SSU rDNA base composition of primordial euglenids, diplonemids and kinetoplastids. Each base is color coded as shown in the legend to the lower right, and base frequency percentages are depicted by connected lines in a spider web-like arrangement along radials which represent different SSU rDNA sequences each. The scale along the radius of *Ploeotia edaphica* indicates a base frequency percentage of 20 % to 30 % (from inner to outer octagon). The black arrow highlights a distinct shift in T-C frequencies between diplonemid *Rhynchopus euleeides* and the primordial kinetoplastid clone AT4-103.

(< 22 %), as demonstrated by a shift in the frequency of cytosine and thymine (black arrow in Fig. 5.15), resulting in the pattern T < C < A < G.

Likewise, the correlation of base frequencies was used to search for different patterns in SSU rDNA nucleotide composition of phagotrophic Helicales, primordial *Distigma proteus* as substitute of the Aphagea, and *Rapaza viridis*, the mixotrophic sister taxon of Euglenea (Fig. 5.16). The base frequencies of phagotrophic *Peranema trichophorum*, *Neometanema* cf *exaratum*, Anisonemida and *Distigma proteus* shared the same pattern, in which adenine and guanine exhibited the highest frequency values (≤ 28 %), while cytosine showed the lowest value (≤ 22 %). Despite this similarity, *Peranema trichophorum* and *Distigma proteus* displayed the pattern C < T < G < A, whereas *Neometanema* cf *exaratum* and Anisonemida showed the pattern C < T < A < G. However, a noticeable shift in T-C base frequency distinguished aforementioned sequences from the others. *Dinema sulcatum*, *Heteronema scaphurum* and the mixotroph *Rapaza viridis* differed clearly in base composition, guanine was the predominant nucleotide (> 28 %), the second frequent nucleotide was adenine which

exhibited more than two percent points difference to guanine, and finally, thymine showed the lowest frequency (≤ 22 %), resulting in the pattern G < A < C < T.

Although the pattern C < T < G < A was observed in SSU rDNA sequences of primordial, non-helical phagotrophic euglenids and diplonemids as well as phagotrophic Helicales and a primary osmotrophic representative of Aphagea, it does not unite major euglenozoan groups necessarily, for morphological characteristics would contradict such a coherence, e.g. possession of an ingestion apparatus would reject *Distigma*, or helical pellicle strips would exclude *Notosolenus* and *Ploeotia*. Nonetheless, if not for major groups, observed similarities and differences in base composition patterns at least constitute an argument for relatedness of taxa. Furthermore, in the case that these related taxa were primordial representatives of major groups, base composition pattern similarity would indeed be relevant for the phylogeny of major euglenozoan lineages.

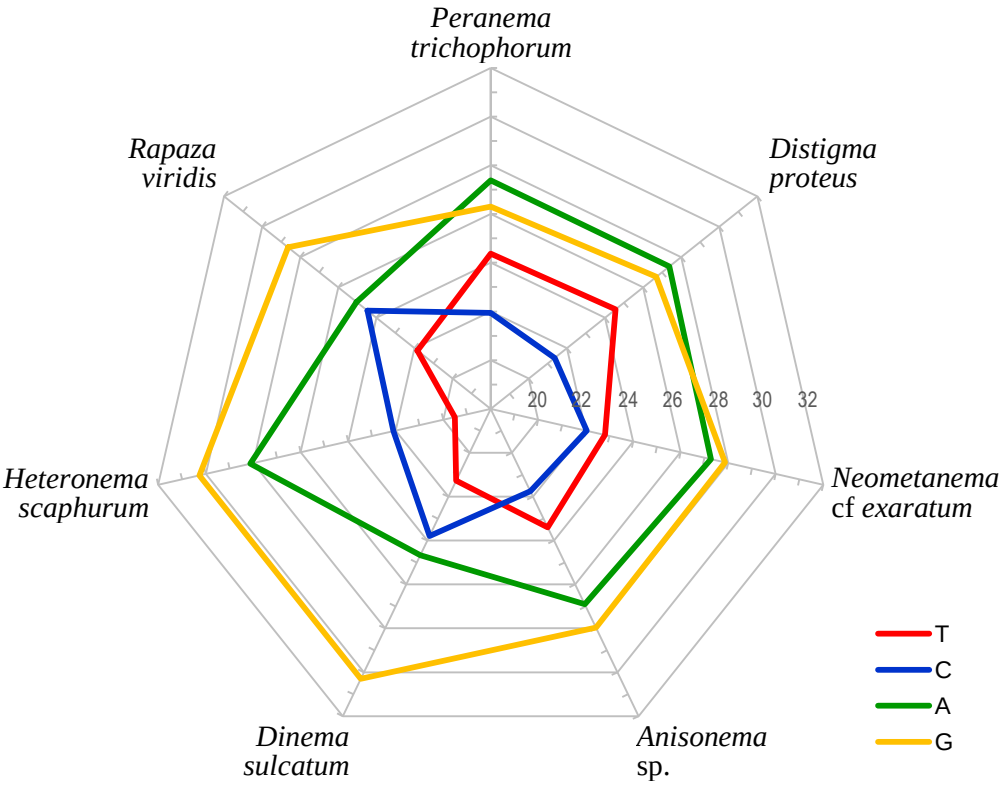


Fig. 5.16: Heptagonal base frequency graph illustrating differences in SSU rDNA base composition of phagotrophic Helicales, primordial *Distigma proteus* representing Aphagea and the mixotroph *Rapaza viridis*. Each base is color coded and base frequency percentages are depicted as described in Fig. 5.15. Scale numbers near the spoke of *Neometanema* indicate base frequencies from 20 % to 32 % (from inner to outer heptagon).

5.2.2 Identity matrix

Comparison of whole SSU rDNA sequences was performed using an identity matrix, which measured similarities of each pair of sequences within taxon-rich dataset III comprising 1,030 nucleotide positions in numbers and percentages of identical nucleotides; a table containing the identity matrix for all taxa of dataset III is given in the supplementary data (see xlsx-file 'IdSSU' in folder 'Supplement' on the CD).

The identity matrix was searched for the most similar SSU rDNA sequences which would connect primordial representatives of major euglenozoan lineages best, i.e. the sequence of primordial kinetoplastid clone AT4-103 was compared with each sequence in the dataset, then the sequence of primordial petalomonad *Notosolenus ostium* etc., and these sequences were combined in an identity matrix graph (Fig. 5.17). As expected, both diplonemid sequences displayed the highest value, i.e. 1,022 of 1,030 identical nucleotides which equated to 99.2 % similarity. Although the SSU rDNA of kinetoplastid clone AT4-103 shared 802 and 801

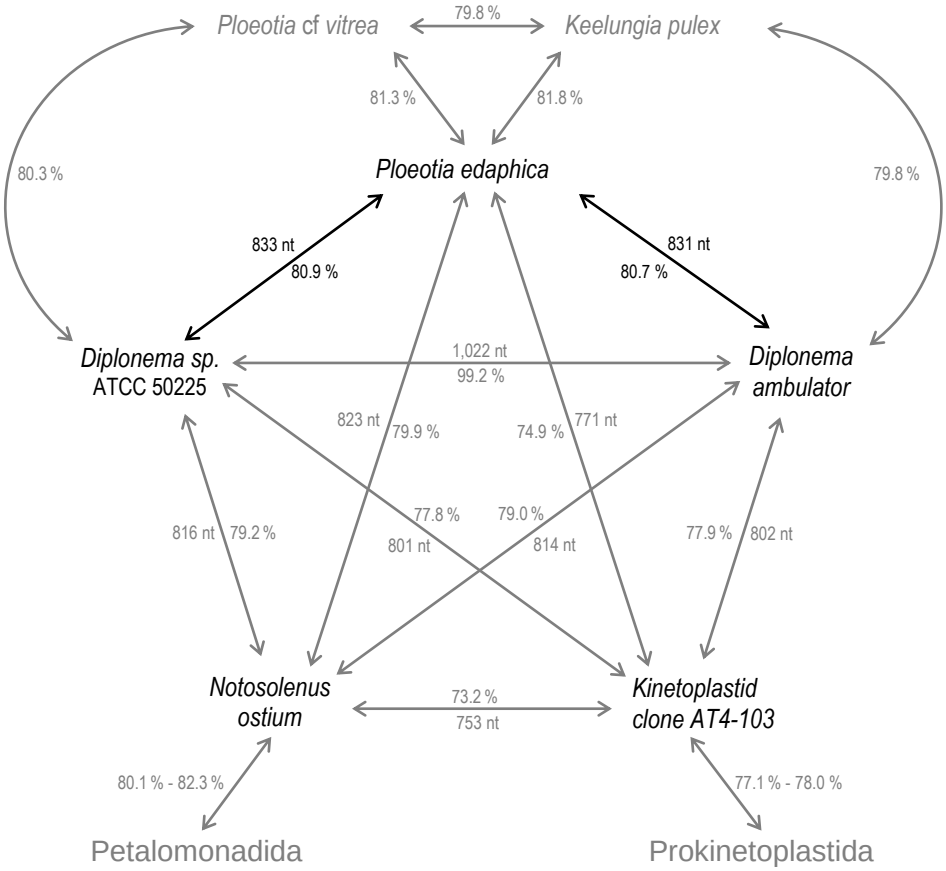


Fig. 5.17: Comparative identity matrix graph illustrating SSU rDNA sequence similarity of *Ploeotia edaphica*, *Ploeotia* cf *vitrea*, *Keelungia pulex*, deepest-branching diplonemids, the petalomonad *Notosolenus ostium*, and the primordial kinetoplastid clone AT4-103 based on dataset III. Similarity values are given in percentage of identical nucleotides, and in numbers within the central pentagon. Petalomonadida and Prokinetoplastida are depicted for reasons of comparability.

identical nucleotides with diplonemid sequences, which corresponded to 77.8 % and 77.9 %, similarity values of the latter compared to sequences of *Ploeotia* and *Keelungia pulex* were higher. Intriguingly, the sequence of *Ploeotia edaphica* revealed the highest similarity to diplonemids, for it shared 833 and 831 identical nucleotides with both diplonemid sequences, which equated to 80.9 % and 80.7 %. Even similarity values combining sequences of phagotrophic *Ploeotia* cf *vitrea*, *Keelungia pulex* and *Notosolenus ostium* with diplonemid sequences (80.3 %, 79.8 % and 79.2 %) surpassed the values of primordial kinetoplastid clone AT4-103 considerably. Interestingly, these findings clearly contradicted a sister group relationship of Diplonemida and Kinetoplastida which had been observed previously, though weakly supported, in maximum likelihood tree reconstructions. Moreover, this outcome corroborated results obtained from phylogenetic network analyses, in which monophyletic Diplonemida shared common splits with *Ploeotia edaphica, Ploeotia* cf *vitrea* and *Keelungia pulex*.

5.2.3 Secondary structure

For this work, euglenozoan and outgroup SSU rDNA sequences have been aligned according to helix numbering and secondary structure information of *Saccharomyces cerevisiae* which was provided by Petrov et al. (2014). Generally, ribosomal DNA is transcripted into complementary ribosomal RNA in which the nucleotide uracil (U) is the complementary base of adenine instead of thymine. After posttranscriptional modifications the remaining coding regions (*or more correctly*: structural regions) of this rRNA build a secondary structure, which then interacts with proteins to become a part of the nascent ribosome. Formation of the secondary structure partly depends on base interactions, therefore base changes in ribosomal DNA, i.e. transitional or transversional nucleotide substitutions can have profound effects on the secondary structure of ribosomal RNAs. Consequently, SSU rDNA sequence differences between taxa can be reflected substantially by unique nucleotide substitutions in their corresponding secondary structure. A thorough examination of deduced secondary structure properties from euglenozoan SSU rDNA sequences confirmed differences and similarities which have been observed in the identity matrix analysis (5.2.2).

A comparison of unique base changes in the putative secondary structure of helix 24 revealed that the primordial petalomonad *Notosolenus ostium* shared a unique substitution with Diplonemida, Symbiontida and *Ploeotia edaphica*, a transitional nucleotide change from

cytosine (C) to uracil, which resulted in an uracil/guanine (U/G) base pairing, that differed from the complementary cytosine/guanine (C/G) base pairing observed in the outgroup. Additional transitions were present in *Keelungia pulex* and *Ploeotia* cf *vitrea*, *Keelungia pulex* featured a base change from C to U in the apex of helix 24 and *Ploeotia* cf *vitrea* exhibits a base change from guanine (G) to adenine (A), which results in a complementary U/A base pairing. Intriguingly, representatives of the Kinetoplastida possess all three described transitions without exception. Kinetoplastida varied in two base changes from Diplonemida and in three transitional base changes from the outgroup, while *Notosolenus ostium*, *Ploeotia edaphica*, the Diplonemida and Symbiontida exclusively shared an identical secondary structure of helix 24 (Fig. 5.18; for more taxa see Fig. 8.5). Though occurrences of silent substitutions or back mutations are possible, other explanations of these findings would necessarily dispute the principle of parsimony.

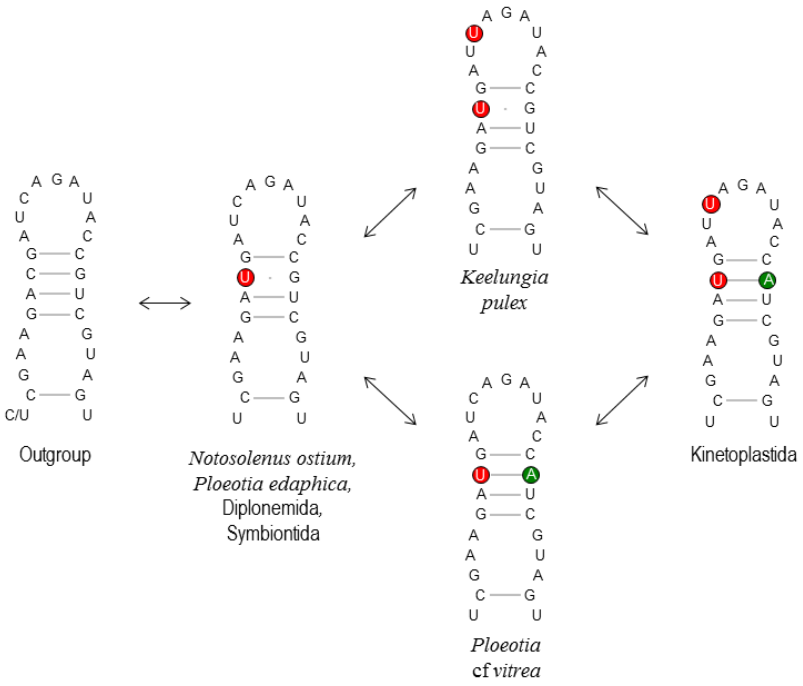


Fig. 5.18: Unique base substitutions specific to major euglenozoan groups and primordial phagotrophic euglenids in deduced SSU rRNA secondary structure of helix 24. Black arrows depict single base changes between corresponding taxa. *Malawimonas jakobiformis*, jakobid *Andalucia incarcerata* and deep-branching *Heterolobosea* sp. BB2 exemplify outgroup taxa.

The analysis of deduced SSU rRNA secondary structure of helix 27 revealed that *Notosolenus ostium* and *Petalomonas cantuscygni* differed from the outgroup in a transversional (G to U) and a transitional (G to A) base change. While the primordial kinetoplastid clone AT4-103 shared the same transversion, it exhibited a transitional base change in another position

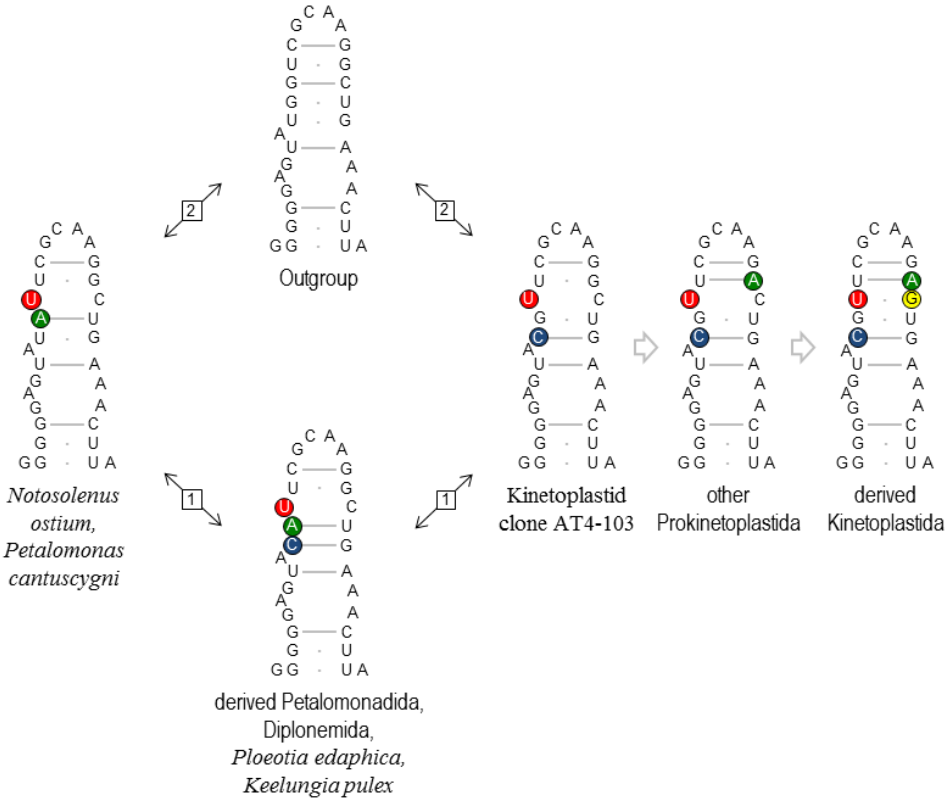


Fig. 5.19: Nucleotide substitution differences between primordial representatives of major euglenozoan groups and outgroup taxa in the putative SSU rRNA secondary structure of helix 27. Substitution changes are colored and boxed numbers on black arrows depict the number of base changes between taxa. The outgroup consists of deep-branching heteroloboseans and the Jakobida.

(U to C), but it also varied in two nucleotides from outgroup taxa. The transversion as well as both transitional base changes were present in the Diplonemida, *Ploeotia edaphica*, *Keelungia pulex* and derived Petalomonadida, which all shared an identical secondary structure of helix 27 that showed three nucleotide alterations compared to the outgroup. Other Prokinetoplastida and more derived kinetoplastids were distinguished from the primordial kinetoplastid clone AT4-103 by other transitional (G to A) and transversional (G to C) base changes in completely different positions on the 5'- strand (Fig. 5.19). Indeed, silent or back mutations could have possibly occurred that would contradict the observed equality of Diplonemida, *Ploeotia edaphica*, *Keelungia pulex* and the derived Petalomonadida, but nonetheless, identical helix 27 corroborated results obtained from phylogenetic network analyses (5.1.3) and found in comparison of the SSU rDNA identity matrix (5.2.2). Similar results were observed in other helices, e.g. helix 13 and helix 33 (see Figs. 8.2 to 8.7).

One analysis produced two results

Comparative secondary structure analysis not only enabled to identify homologous positions, but also to pinpoint boundaries of variable regions within SSU rDNA sequences. For instance,

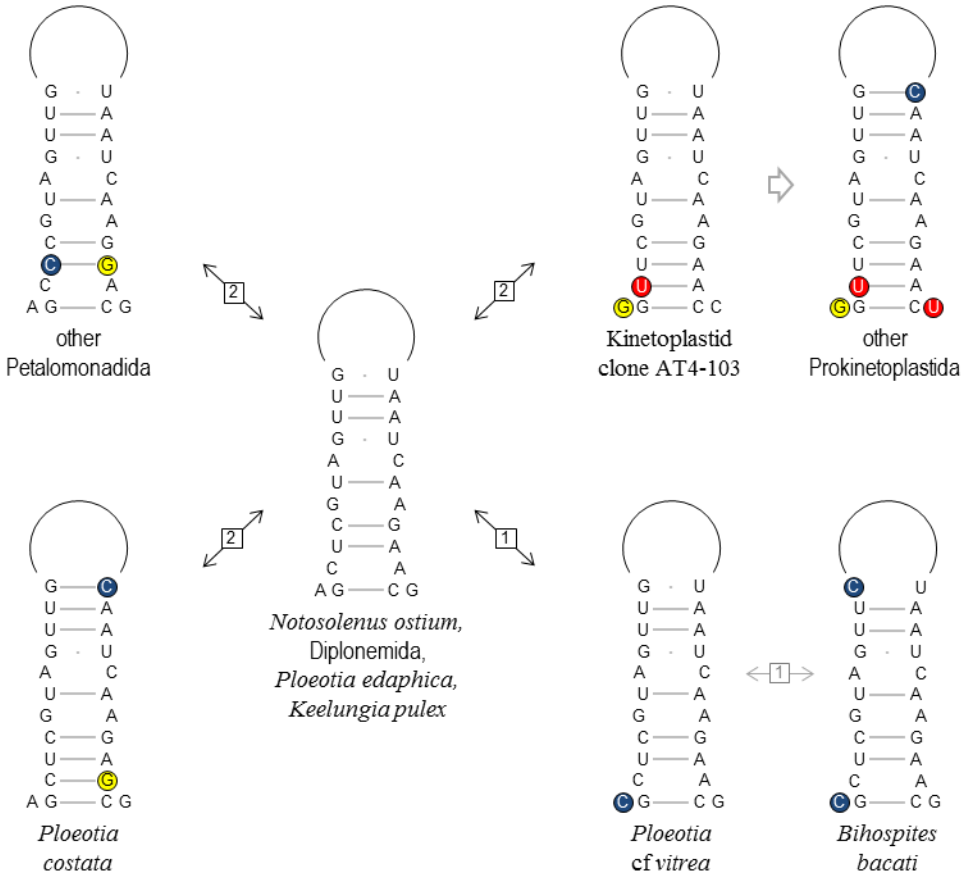


Fig. 5.20: Nucleotide substitution differences between primordial representatives of major euglenozoan groups in deduced SSU rRNA secondary structure of helix 20 that constitutes the basis of variable region 4, which is illustrated by a simplified circular apex. Nucleotide changes are colored and boxed numbers in arrows depict number of base changes between corresponding taxa.

helix 20 embraces the next three helices and also variable region 4, therefore an analysis allowed for comparative examination of nucleotide substitutions and simultaneously for a measurement of variable region length variation between taxa (see 5.2.4). Similar to earlier results, analysis of helix 20 revealed that *Notosolenus ostium*, Diplonemida, *Ploeotia edaphica* and *Keelungia pulex* share an identical secondary structure which diverged from outgroup taxa by one substitution (most jakobids and primordial heteroloboseans, not shown). More derived Petalomonadida, *Ploeotia costata* and primordial kinetoplastid clone AT4-103 each varied in two transitional base changes, all of which involved different nucleotide positions (Fig. 5.20). More derived prokinetoplastids deviated by two individual transitions from primordial kinetoplastid clone AT4-103. Interestingly, *Ploeotia* cf *vitrea* exhibited one transversion (A to C), which represented the only base change compared to *Ploeotia edaphica*, *Keelungia pulex*, the Diplonemida and *Notosolenus ostium*. The same transversion was also present in the assumed primordial symbiontid *Bihospites bacati*, which differed from *Ploeotia* cf *vitrea* by a secondary transversional base change (G to C). Similar to earlier

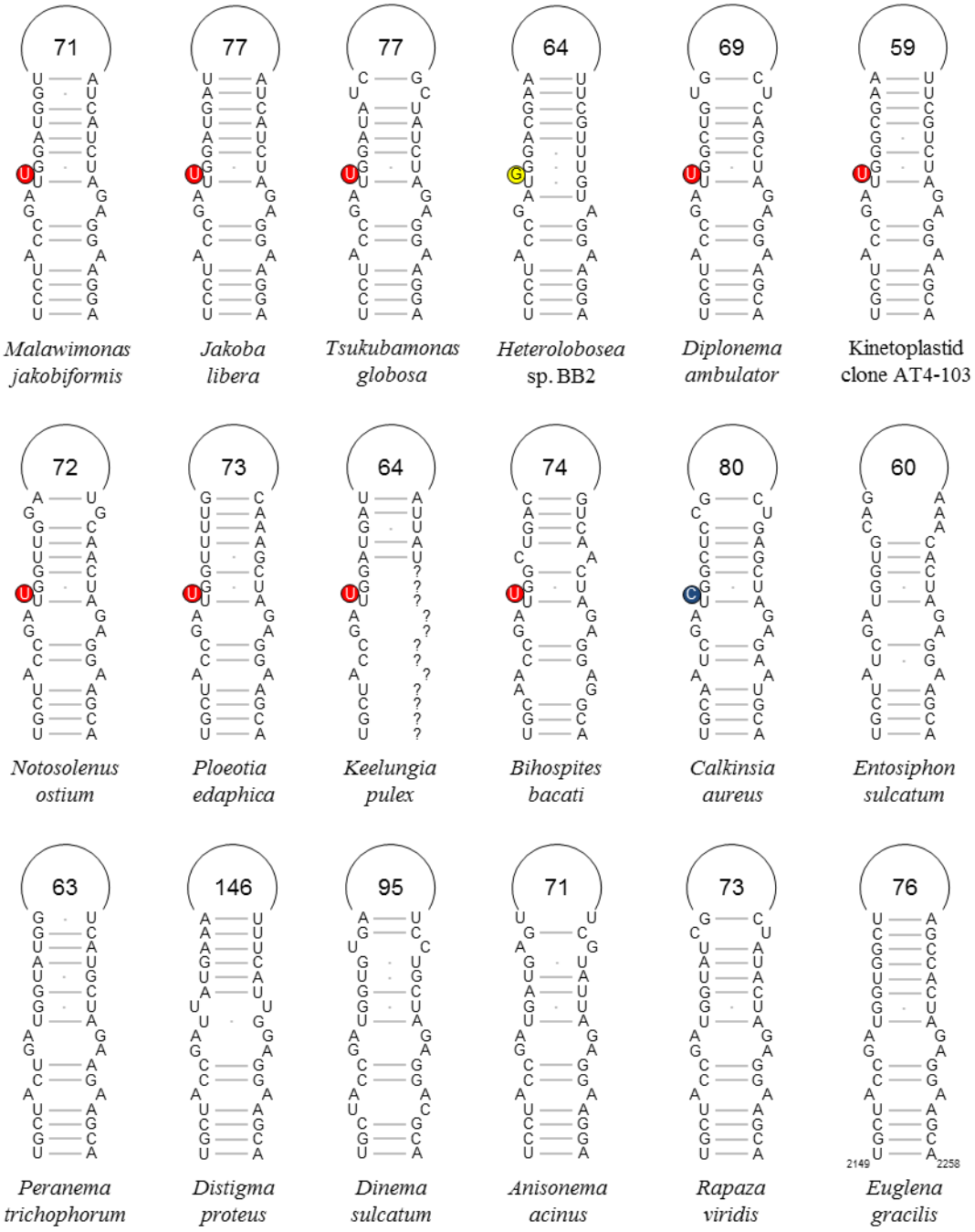


Fig. 5.21: Occurrence of individual nucleotides in outgroup taxa and primordial representatives of major euglenozoan groups in the putative secondary structure of helix 44, which embraces SSU rDNA variable region 9. Where present, individual nucleotides are colored, a number in the apex of each helix summarizes remaining nucleotides of helix 44 including variable region 9.

results, these findings contradicted a sister group relationship between Kinetoplastida and Diplonemida, but confirmed the affiliation of Diplonemida with *Ploeotia edaphica* and *Keelungia pulex* exemplified by another identical SSU rDNA secondary structure.

An astonishing feature was discovered in the secondary structure of helix 44, both strands of which encompassed the SSU rDNA variable region 9. Distant and close outgroup taxa as well as Diplonemida, Symbiontida, primordial kinetoplastid clone AT4-103, and most primordial

non-helical euglenids, i.e. deep-branching petalomonad *Notosolenus ostium*, representatives of genus *Ploeotia* and *Keelungia pulex*, possessed a unique nucleotide in the proximal region of the 3'- strand, which seemed to lack any structural relevance for helix 44 (Fig. 5.21). No corresponding nucleotide was present on the complementary 5'-strand of helix 44 that would have constituted an appropriate or reasonable binding partner. Within Euglenozoa, this individual nucleotide was neither present in the SSU rDNA sequence of *Entosiphon sulcatum* nor in that of Helicales, i.e. in sequences of *Peranema*, *Neometanema*, *Dinema*, Anisonemida, Aphagea and Euglena, respectively (for more taxa see Fig. 8.7 in the Appendix). It was absent in derived Petalomonadida and derived Kinetoplastida, but it was present in all Diplonemida and Symbiontida. It was found in SSU rDNA sequences of discoban outgroup lineages Jakobida and deep-branching Heterolobosea as well as in other excavates (e.g. *Malawimonas jakobiformis*) and in representatives of distant outgroup lineages, e.g. Amoebozoa (*Amoeba proteus, Chaos nobile*), Opisthokonta (e.g. *Monosiga brevicollis, Saccharomyces cerevisiae*), Alveolata (e.g. *Paramecium tetraurelia, Tetrahymena thermophila*), Rhizaria (e.g. *Chlorarachnion reptans*), Stramenopiles (e.g. *Heterosigma akashiwo*) and Cryptophyta (e.g. *Chilomonas paramecium*).

As a phylogenetic implication, the deep-branching position of *Entosiphon* in Euglenozoa, as found in most phylogenetic tree reconstructions, would consequently demand a primary absence (in *Entosiphon*) and an apomorphic evolvement of this individual nucleotide in Diplonemida, Symbiontida, primordial Petalomonadida and Kinetoplastida, plus in both the latter a secondary absence, and additionally an absence in Helicales. Such a scenario would not only be imparsimonious, it would neglect the fact that the individual nucleotide was also present in the outgroup. If the occurrence of this individual nucleotide constituted a plesiomorphic feature, multiple losses exclusively within derived Petalomonadida, derived Kinetoplastida and *Entosiphon sulcatum* plus Helicales would have followed. Moreover, Helicales were the only monophyletic group that primarily lacked this character, and the absence of this presumably non-structural nucleotide in *Entosiphon* would imply relatedness to Helicales, which would be the most parsimonious explanation. Therefore, this finding indicates affiliation of Diplonemida, Symbiontida, primordial kinetoplastid clone AT4-103 with non-helical phagotrophic euglenids to the exclusion of *Entosiphon* and all helical phagotrophic euglenids, Aphagea and Euglenea.

5.2.4 SSU rDNA variable regions

As mentioned before, secondary structure analysis enabled to identify exact boundaries of SSU rDNA variable regions, which was a prerequisite for the measurement of variable region lengths between major groups of the Euglenozoa and outgroup taxa (5.2.3). Interestingly, an examination of euglenozoan variable regions has been done only once before, in the context of a study which investigated the unusually expanded SSU rDNA sequences of primary osmotrophic euglenids (Busse & Preisfeld 2002b). Since that time, the number of available euglenozoan SSU rDNA sequences had increased considerably, therefore it was worth to reexamine variable regions of SSU rDNA (i.e. V1, V2, V3, V4, V5, V7, V8 and V9) with a bigger taxon sampling, especially in respect to phagotrophic euglenids and including symbiontids for the first time.

Initially, length values of taxa related to major euglenozoan groups overlapped in most variable regions, which made it problematic to clearly distinguish between these groups (see Tab. 8.2 in the Appendix). To circumvent this problem, the data was divided following a method of descriptive statistics. Measured variable region length values were sorted corresponding to group relatedness, then ranked according to their range and divided into quartiles and/or percentiles. For an even number of data points, the lower and upper quartiles were determined, thus separating the lowest and highest 25 % from a ‘centered’ 50 % of group related measured values. For small and uneven numbers of values, an approximate calculation was applied to determine the outliers from a narrowed majority of up to 67 % of data values. This approach was not based on an exact mathematical framework, for the number of taxa differed considerably between subgroups, nonetheless application of this quartile (percentile) method facilitated to at least roughly distinct between majority and outlier values of group related taxa for reasons of comparability.

Individual variable regions

The outgroup taxon Heterolobosea exhibited an extraordinary internal variance in the comparative analysis of SSU rDNA variable region 1 (V1) length values, which resembled the high diversity of heterolobosean sequences observed in results from spectral analyses of SSU rDNA sequences excluding variable regions (see 5.1.2). Most heterolobosean V1 length values were > 17 nucleotides (12 out of 16), which noticeably exceeded measured values of most Euglenozoa, but significantly those of Petalomonadida, Kinetoplastida, Diplonemida and Symbiontida (Fig. 5.22). Though *Entosiphon* sequences showed the largest V1 length

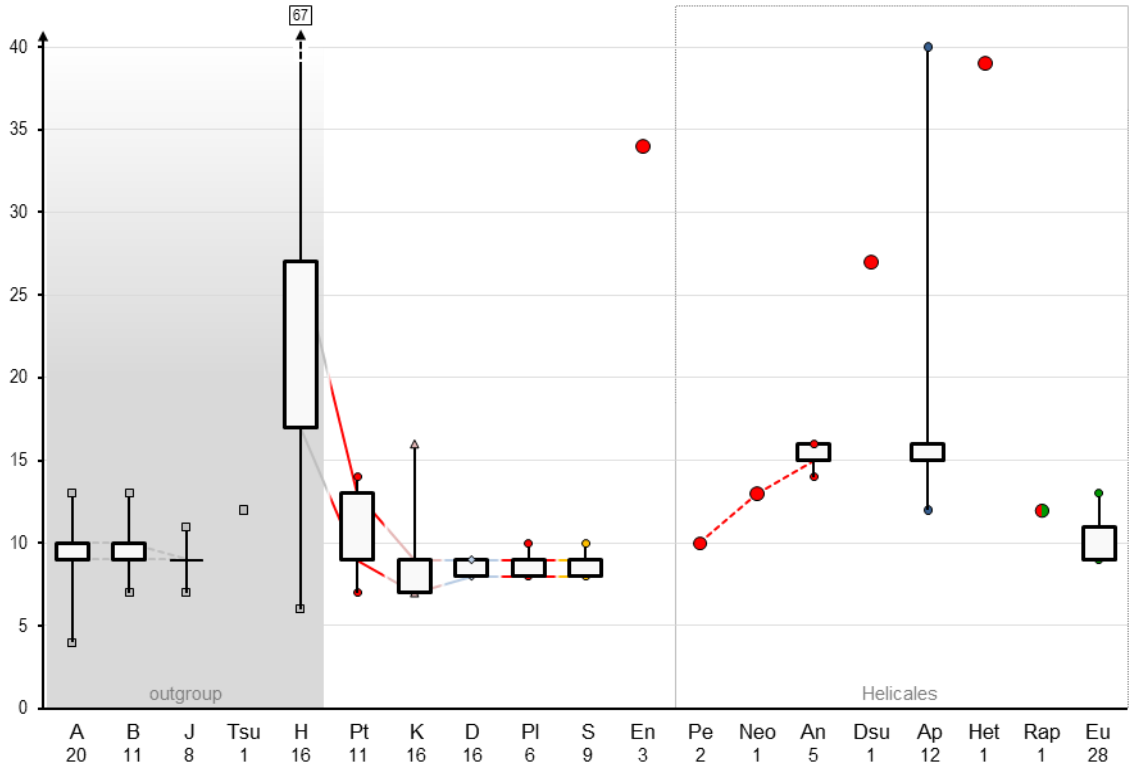


Fig. 5.22: Length variations of SSU rDNA variable region 1 among euglenozoan and outgroup taxa. Subgroups containing more than 3 sequences are represented by quartiles, i.e. rectangulars contain 50 % of group related sequences, and upper and lower quartiles are shown as lines above and below rectangulars. Single and subgroup taxa are marked by symbol and color: outgroup taxa ☐, kinetoplastids △, diplonemids ◇, euglenids ○; phagotrophic euglenids in red, primary osmotrophic euglenids in dark blue, phototrophic euglenids in green, symbiontids in yellow. Abbreviations for group names and numbers of representatives are given below the x-axis: **A**: outgroup, Eukaryota (20 sequences); **B**: outgroup, Excavata; **J**: outgroup, Jakobida; **Tsu**: outgroup, *Tsukubamonas globosa*; **H**: outgroup, Heterolobosea; **Pt**: Petalomonadida; **K**: Kinetoplastida; **D**: Diplonemida; **Pl**: Ploeotiida; **S**: Symbiontida; **En**: *Entosiphon*; **Pe**: *Peranema*; **Neo**: *Neometanema* cf *exaratum*; **An**: Anisonemida; **Dsu**: *Dinema sulcatum*; **Ap**: Aphagea; **Het**: *Heteronema scaphurum*; **Rap**: *Rapaza viridis*; **Eu**: Euglenea. Colored lines connecting rectangulars and dashed lines connecting single data points are used to better depict length differences between subgroups and taxa.

values among non-helical euglenids (34 nucleotides), other non-helical groups revealed length variances limited to the range of lower outliers of the Heterolobosea (6 to 17 nucleotides). Upper outliers of Kinetoplastida presented V1 sequence lengths > 10 nucleotides, while the majority of Petalomonadida exhibited length values between 9 and 14 nucleotides (8 out of 11 taxa). Interestingly, all Diplonemida as well as most representatives of Symbiontida and Ploeotiida displayed smaller V1 length values within a similarly limited range (8 and 9 nucleotides). Single taxa and major groups of the Helicales exhibited much greater V1 length variations than most primordial euglenozoan groups. Within the Helicales, Aphagea showed the greatest V1 length variance which ranged from 12 to 40 nucleotides, though half of all representatives (6 out of 12) displayed length values of 15 and 16 nucleotides, which matched the V1 length variance of most Anisonemida. *Peranema* sequences displayed the smallest V1 length value of phagotrophic Helicales, which matched the upper outliers of Ploeotiida and

Symbiontida. While length values of *Heteronema scaphurum* and *Dinema sulcatum* clearly surpassed that of Anisonemida, interjacent V1 sequence of *Neometanema* cf *exaratum* was situated between that of Anisonemida and *Peranema*. The V1 length value of *Rapaza viridis* resided within range of upper outliers of the Euglenea. Some taxa could not be included into all variable region length analyses due to 5'-truncated SSU rDNA sequences, e.g. *Dinema platysomum* (V2 to V9) and *Neometanema parovale* (V4 to V9).

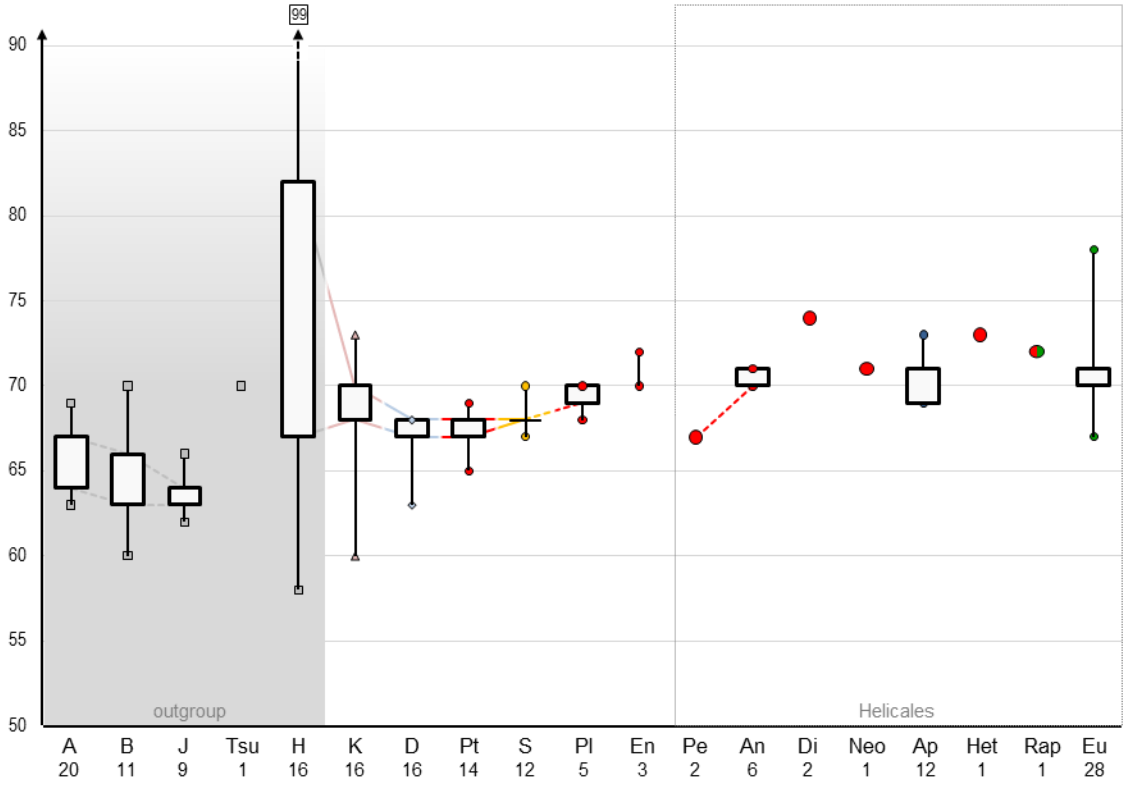


Fig. 5.23: Length variations of euglenozoan SSU rDNA variable region 3 in contrast to different outgroup clades. For detailed description of graph and symbols, see Fig. 5.22. Note that the ordinate begins with value 50. **A**: outgroup, Eukaryota; **B**: outgroup, Excavata; **J**: outgroup, Jakobida; **Tsu**: outgroup, *Tsukubamonas globosa*; **H**: outgroup, Heterolobosea; **K**: Kinetoplastida; **D**: Diplonemida; **Pt**: Petalomonadida; **S**: Symbiontida; **Pl**: Ploeotiida; **En**: *Entosiphon*; **Pe**: *Peranema*; **An**: Anisonemida; **Di**: *Dinema*; **Neo**: *Neometanema* cf *exaratum*; **Ap**: Aphagea; **Het**: *Heteronema scaphurum*; **Rap**: *Rapaza viridis*; **Eu**: Euglenea.

Examination of length variations in SSU rDNA variable region 3 sequences (V3) provided similar results as observed in V1 length analysis, for Heterolobosea displayed an extraordinary wide range of length values and by far the highest internal variability. The range of kinetoplastid length values exceeded that of all other euglenozoan groups, even that of Euglenea, which comprised more taxa than any other group (Fig. 5.23). The centered 50 % of Diplonemida and of Petalomonadida exhibited an identical distribution, and the majorities of measured length values of Diplonemida and of Petalomonadida matched those of lower

outliers of the Kinetoplastida. An overall increase in V3 length of the lower outliers was observed in Heterolobosea (58 nucleotides) and among non-helical euglenozoan taxa, i.e. from 60 nucleotides in Kinetoplastida, 63 nucleotides in Diplonemida, 65 nucleotides in Petalomonadida, 67 nucleotides in Symbiontida, 68 nucleotides in Ploeotiida to 70 nucleotides in *Entosiphon*. Only in variable region 3 the Helicales presented smaller variations of length values compared to other Euglenozoa. Among Helicales, *Peranema* displayed the shortest V3 length, and Anisonemida, *Dinema*, *Neometanema* cf *exaratum*, *Heteronema scaphurum* as well as *Rapaza viridis* showed length values within range of the upper outliers of Euglenea, the latter exhibited a higher V3 length variability than Aphagea.

In length variation comparison of SSU rDNA variable region 9 (V9), *Entosiphon* and the majority of Prokinetoplastida showed smaller values than other non-helical Euglenozoa and most Helicales (Fig. 5.24). Moreover, the majority of Petalomonadida and Diplonemida shared a rather narrow distribution of similar V9 length values, and although Symbiontida and Ploeotiida exhibited a higher variability, length values of most Petalomonadida, Diplonemida and Ploeotiida clearly surpassed even those of the upper outliers of Prokinetoplastida. Within

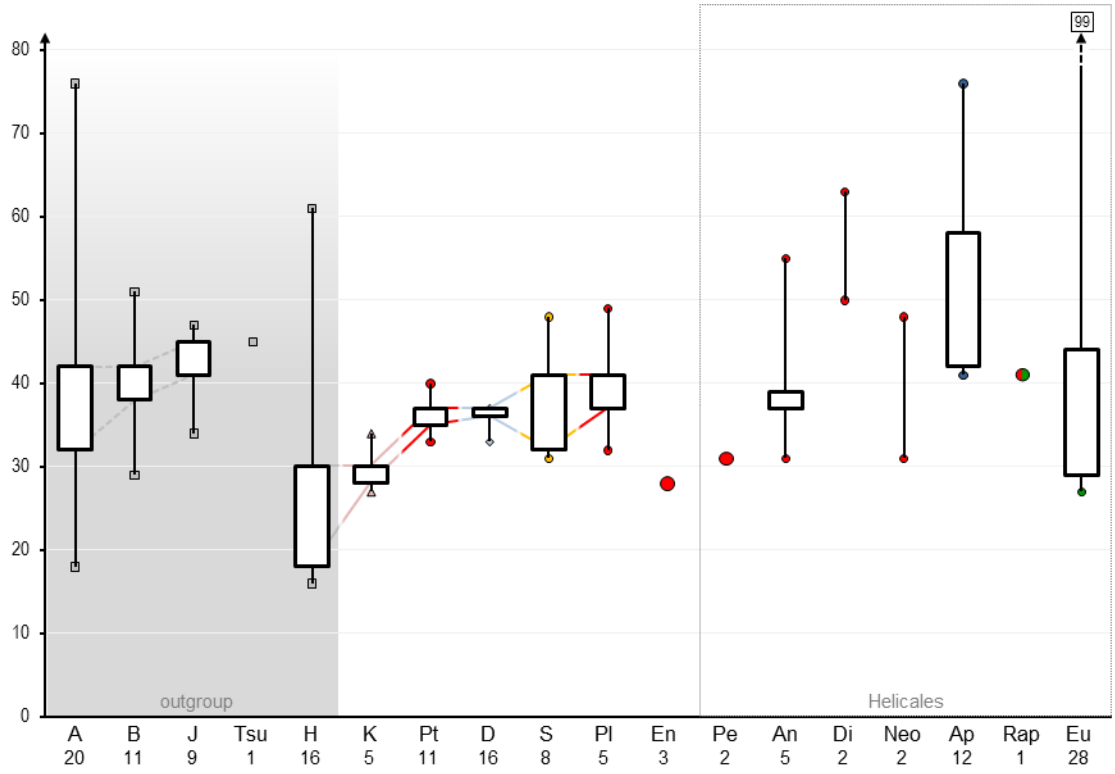


Fig. 5.24: SSU rDNA variable region 9 length variations of euglenozoan and outgroup taxa. For detailed description of graph and symbols, see Fig. 5.22. **A**: outgroup, Eukaryota; **B**: outgroup, Excavata; **J**: outgroup, Jakobida; **Tsu**: outgroup, *Tsukubamonas globosa*; **H**: outgroup, Heterolobosea; **K**: Prokinetoplastida; **Pt**: Petalomonadida; **D**: Diplonemida; **S**: Symbiontida; **Pl**: Ploeotiida; **En**: *Entosiphon*; **Pe**: *Peranema*; **An**: Anisonemida; **Di**: *Dinema*; **Neo**: *Neometanema*; **Ap**: Aphagea; **Rap**: *Rapaza viridis*; **Eu**: Euglenea.

Helicales, Euglenea showed the highest length variability, for corresponding lower and upper outliers displayed V9 length values of 27 and 99 nucleotides, respectively. *Peranema*, *Neometanema* cf *exaratum* and lower outliers of Anisonemida presented the smallest V9 length values, while *Dinema sulcatum* showed the longest V9 sequence of phagotrophic Helicales. Unfortunately, length values of *Heteronema scaphurum* for variable regions V7, V8 and V9 could not be analysed due to its 3'-truncated SSU rDNA sequence.

Comparative examination of length variations in individual SSU rDNA variable regions revealed profound differences not only between outgroup and euglenozoan taxa, but also between major euglenozoan groups, for instance, the high diversity of Heterolobosea which had been observed in earlier phylogenetic network and spectral analyses (5.1.2) was also found in SSU rDNA variable regions. As a result, in most variable regions examined, length values of Kinetoplastida considerably differed in range and variability from length values of the Diplonemida. Furthermore, diplonemid variable region length values usually resembled those of most Petalomonadida, i.e. in V2, V3, V4 and V9, Symbiontida in V7 and V8, as well as Ploeotiida in V1 (for graphs of other variable regions see Figs. 8.8 to 8.10). While most taxa of the Helicales generally exhibited longer variable region sequences than other Euglenozoa, length variability of the Aphagea exceeded those of other examined groups in most, but not all variable regions, i.e. in V2, V4, V5, V7 and V8. Mostly, *Heteronema scaphurum* displayed the longest variable region sequences among phagotrophic Helicales, i.e. in V1, V2, V4 and V5. While variable region sequences of *Peranema* were the shortest among phagotrophic Helicales in V1, V2 and V3, lower outliers of Anisonemida showed smallest length values in V5 and V7, whereas *Neometanema* and *Dinema* sequences exhibited smallest length values in V4 and V8, respectively. However, even the smallest length values related to representatives of the Helicales were situated in range of non-helical taxa in all cases.

Concatenated variable regions

Group specific differences of variable region length became even more apparent when surveyed in summary. In the context of this work, variable region sequences of individual taxa were concatenated to gain insight into absolute length values of euglenozoan SSU rDNA variable regions and appendant phylogenetic implications.

A comparison of absolute length values of Aphagea with those of other Helicales revealed that *Heteronema scaphurum* was the only phagotrophic taxon among Helicales which

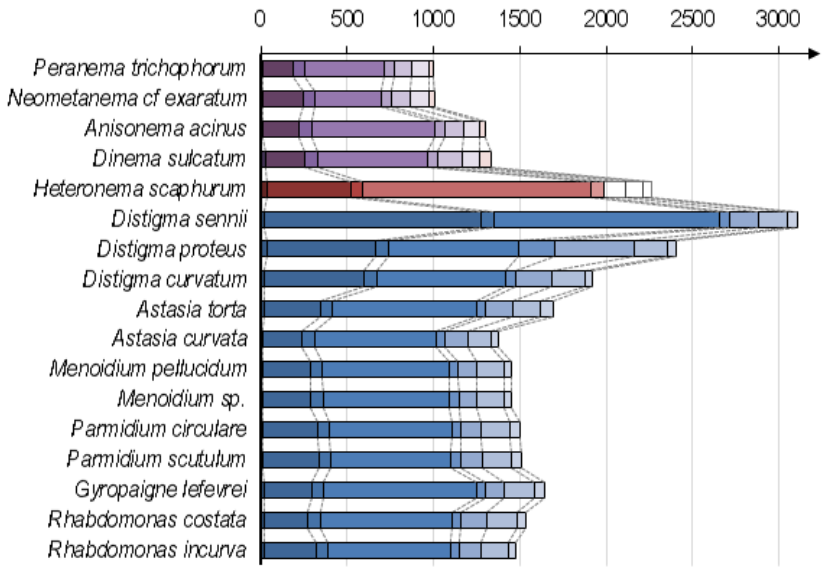


Fig. 5.25: Concatenated SSU rDNA variable region graph illustrating absolute variable region sequence length values of Aphagea in comparison to phagotrophic taxa of the Helicales. Variable regions are displayed by equivalent segments (from dark to light-colored), taxa of the Aphagea in blue, phagotrophic Helicales in purple and *Heteronema scaphurum* in red. The arithmetic means of V7, V8 and V9 length values from other phagotrophic Helicales were used to hypothetically display missing values of *Heteronema scaphurum* (uncolored segments).

possessed a variable region sequence of nearly similar length to those of primordial aphagean taxa *Distigma sennii* and *Distigma proteus*. Although the truncated SSU rDNA sequence of *Heteronema scaphurum* lacked nucleotides covering regions V7, V8 and V9, its sequences related to V1, V2, V3, V4 and V5 surpassed by far corresponding length values of any other phagotrophic taxon within the Helicales (Fig. 5.25). This finding was inconsistent with results obtained from identity matrix comparison, in which *Neometanema* cf *exaratum* exhibited a higher SSU rDNA sequence similarity to taxa from the *Distigma proteus* group (presumed primordial representatives of Aphagea) than *Heteronema scaphurum* (see Fig. 8.1). Though phylogenetic implications are not beyond question due to incompleteness of the *Heteronema scaphurum* SSU rDNA sequence, its extraordinary long variable regions put the aforementioned representative of phagotrophic Helicales closest to Aphagea.

The survey of concatenated SSU rDNA variable regions also revealed a striking similarity between sequence length values of Diplonemida and Petalomonadida, in fact concerning absolute as well as standardized length values, which both differed clearly from those of Kinetoplastida. While summarized diplonemid variable region sequences shared very similar length values in a narrow distribution, kinetoplastid taxa showed wider distributed length values which varied from shorter sequences of Prokinetoplastida, i.e. primordial kinetoplastid clone AT4-103, to much longer variable regions of more derived Kinetoplastida, i.e. *Crithidia*

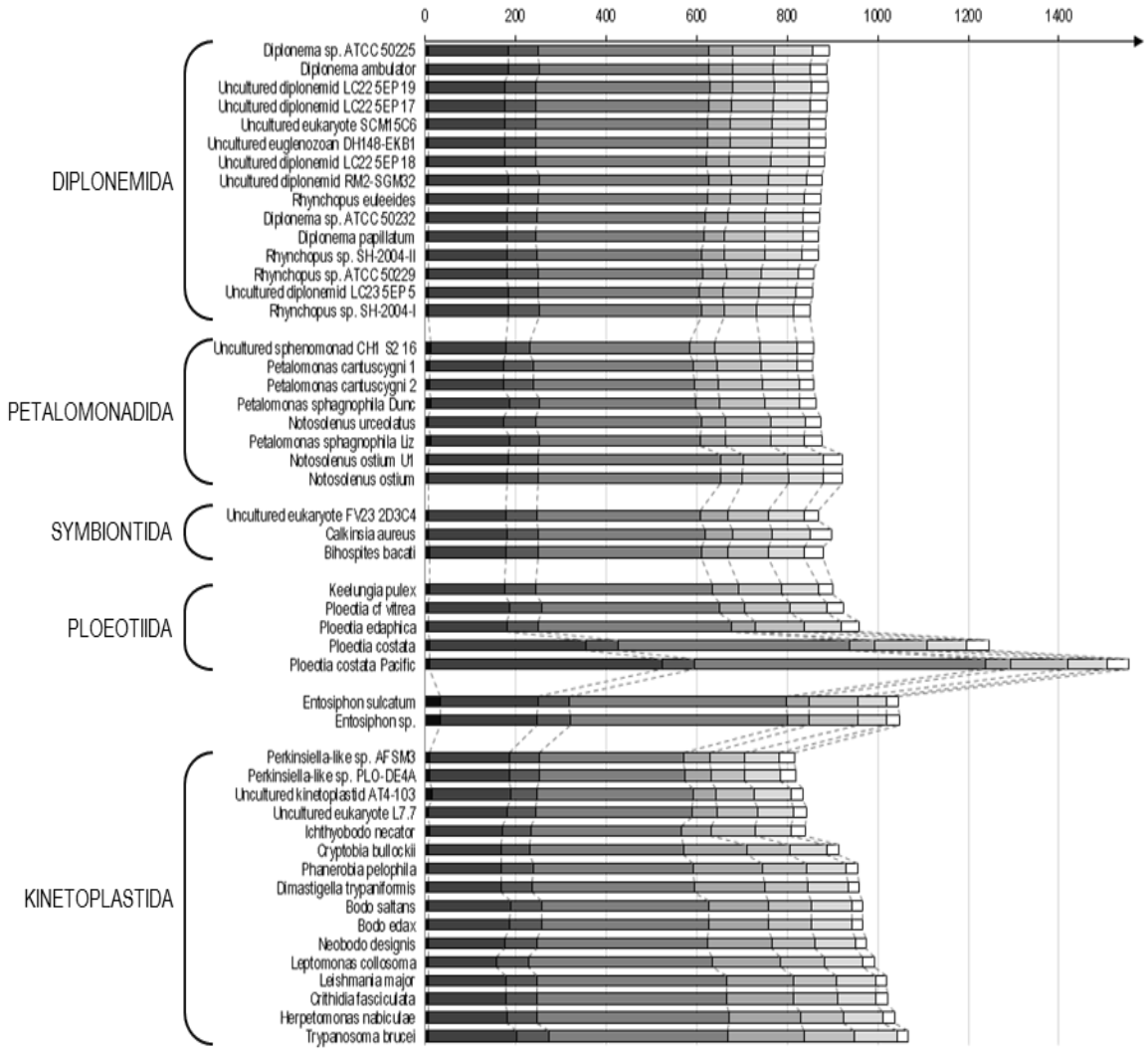


Fig. 5.26: SSU rDNA variable regions graph demonstrating absolute variable region lengths of non-helical phagotrophic euglenids compared to Diplonemida, Symbiontida and Kinetoplastida. Analogous variable regions V1 to V9 are displayed in greyish segments (from dark to light-colored).

fasciculata and *Trypanosoma brucei* (Fig. 5.26). As a result, the Kinetoplastida displayed an overall increase in variable region length considering widely accepted hypotheses on the phylogeny of Kinetoplastida, namely that Prokinetoplastida are primordial to free-living taxa and human-parasitic genera like *Trypanosoma* and *Leishmania* (Deschamps et al. 2011, Doležel et al. 2000, Moreira et al. 2004, Simpson et al. 2002). Ploeotiida exhibited the highest variability, for variable region length of *Ploeotia costata* widely surpassed those of *Trypanosoma brucei* and *Entosiphon* sequences. Additionally, concatenated variable region sequences were standardized to investigate relative proportions of variable region length values. Likewise to absolute variable region length values, standardized variable regions also showed nearly identical proportions of individual variable region length in Diplonemida and Petalomonadida (Fig. 8.13). These findings corroborated results from phylogenetic network, identity matrix and secondary structure analyses of SSU rDNA sequences which contradicted a sister group relationship of Kinetoplastida and Diplonemida.

5.3 Phylogenetic analyses of LSU rDNA sequences

5.3.1 Broad dataset IV

Results from Maximum likelihood and Bayesian inference analyses of LSU rDNA dataset IV showed strong discrepancies between ML and BI tree topologies. Most of the newly obtained sequences fitted adequately well into the trees, though some of which branched in positions that differed to those previously observed in results from SSU rDNA analyses (Fig. 5.27). Monophyly of Euglenozoa was firmly supported, *Notosolenus ostium* represented the deepest-branching taxon in both trees, but *Petalomonas cantuscygni* formed the sister taxon of Diplonemida, thus rendering Petalomonadida paraphyletic. Kinetoplastida as well as Diplonemida appeared as monophyla with maximum statistical support, the latter including new LSU rDNA sequences of *Rhynchopus euleeides* (ATCC 50226) and *Diplonema ambulator* (ATCC 50223). As observed in SSU rDNA based trees, the LSU rDNA sequence of *Entosiphon sulcatum* exhibited the longest branch of all taxa, but interestingly it branched as strongly supported sister taxon of *Ploeotia costata* in the ML tree (Fig. 5.27 A). A tetrafurcation including *Petalomonas cantuscygni*, *Entosiphon sulcatum*, Diplonemida and Kinetoplastida represented the predominant deviation in the BI trees. *Ploeotia costata* appeared as very good supported sister taxon of robustly supported monophyletic Helicales. Within the Helicales, Aphagea and Euglenea both formed monophyla with maximum support while *Peranema trichophorum* appeared as firmly supported sister taxon of Euglenea. Exclusion of long-branching *Entosiphon sulcatum* had no effect on the ML tree topology, but as a result, statistical support for the position of Kinetoplastida as well as the *Petalomonas*/Diplonemida clade was considerably weakened in the reiterated ML tree and *Ploeotia costata* was confirmed as sister taxon of Helicales in the reiterated BI tree with average support (Fig. 5.27 B). Curiously, phylogenetic network analysis of unmodified dataset IV confirmed paraphyletic Petalomonadida and the corresponding split support spectrum showed nonsense groupings of euglenozoan taxa supported by compatible splits, e.g. *Peranema trichophorum* with *Notosolenus ostium* or outgroup taxa with *Petalomonas cantuscygni* (Figs. 8.14 and 8.15). These findings revealed that the rather broadly sampled outgroup taxa had a remarkably strong impact on the relatively few euglenozoan LSU rDNA sequences. For instance *Naegleria gruberi*, representing a relatively derived member of Heterolobosea, exemplified the only heterolobosean sequence in the dataset and therefore an inept outgroup taxon. To complicate matters, any other outgroup taxon represented evolutionary more distant clades without group-specific substitution patterns (see ‘DItestLSU’

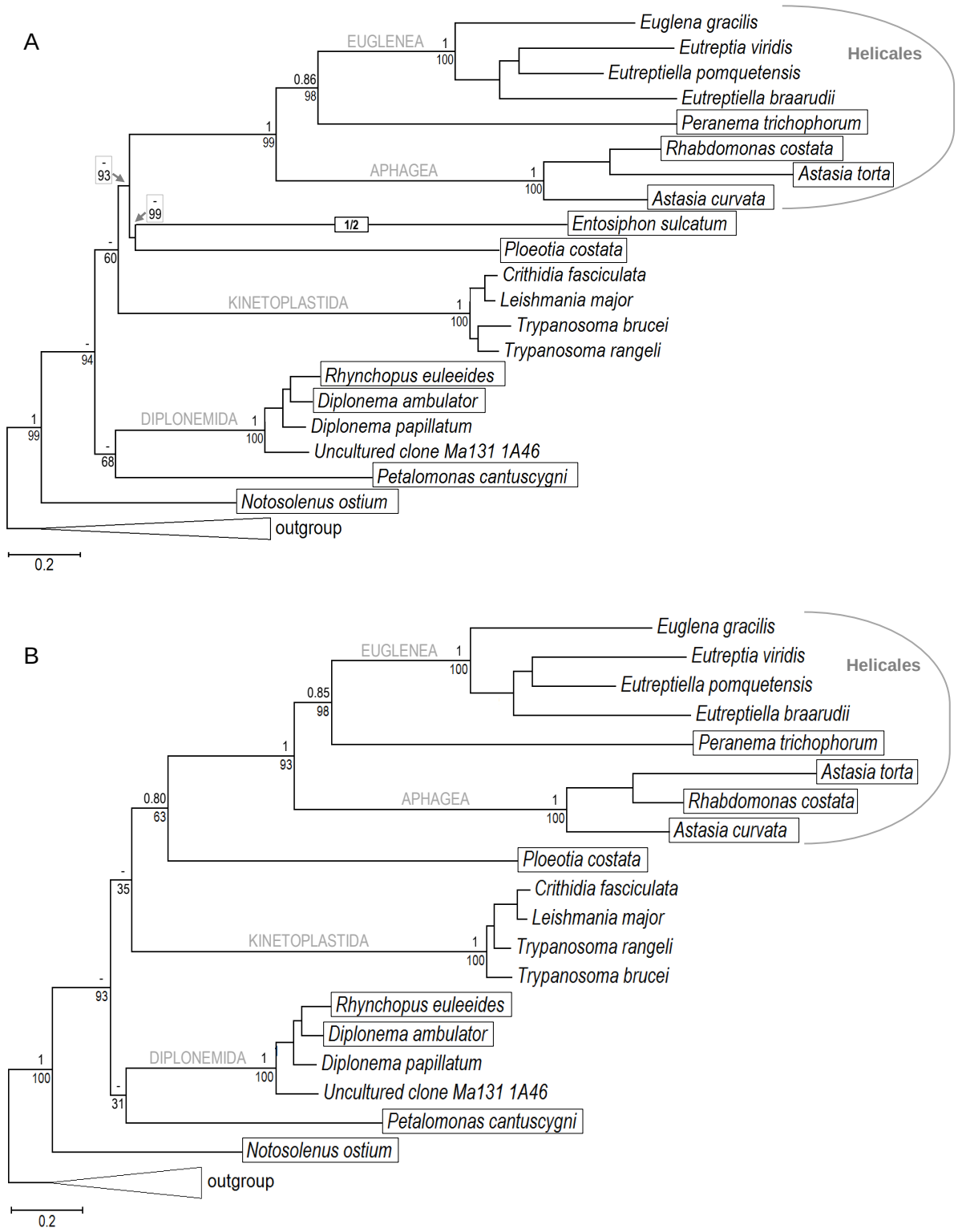


Fig. 5.27: Consensus trees obtained from analyses of LSU rDNA dataset IV comprising 2,406 nucleotides with new sequences boxed. For taxon sampling see Tab. 4.5. Congruent posterior probabilities are mapped onto the Maximum likelihood trees and displayed above, ML bootstrap support values below corresponding nodes; discrepancies to Bayesian inference are hyphenated. Scale bars represent 20 % sequence divergence. **A**: Results involving 44 taxa including *Entosiphon*, half of the original branch length depicted (GTR+Γ+I, -lnL = 53487.22, gamma shape = 0.411, p-invar = 0.101). **B**: Results from reiterated analyses excluding *Entosiphon* (GTR+Γ+I, -lnL = 48439.24, gamma shape = 0.435, p-invar = 0.101).

in folder ‘Supplement’ on the CD). To avoid undesirable effects on ingroup taxa which may have been caused by unsuitably distant, i.e. evolutionary broadly sampled outgroup taxa, phylogenetic network and spectral analyses were reiterated with the most stable outgroup, i.e. the Opisthokonta, and without outgroup taxa. As a result, the splits support spectrum demonstrated coherent compatible splits referring to euglenozoan taxa and incompatible splits which exclusively represented nonsense correlations and stood in conflict with a binary tree (Fig. 5.28). Intriguingly, a combination of Kinetoplastida and *Ploeotia costata* recovered compatible splits. Reiterated phylogenetic network analysis of modified dataset IV produced monophyletic Diplonemida, Kinetoplastida, Aphagea and Euglenea (Fig. 5.29 A). *Notosolenus ostium* and *Petalomonas cantuscygni* shared common splits, thus resulting in monophyletic Petalomonadida. Interestingly, Petalomonadida and Diplonemida shared common splits as well in the spectral analysis (column 26 in Fig. 5.28). *Ploeotia costata* and *Entosiphon sulcatum* grouped together supported by mutual splits (red arrow in Fig. 5.29 B). This network topology was replicated when using Opisthokonta as a stable outgroup (not shown), for this outgroup branched between Petalomonadida and Kinetoplastida without any

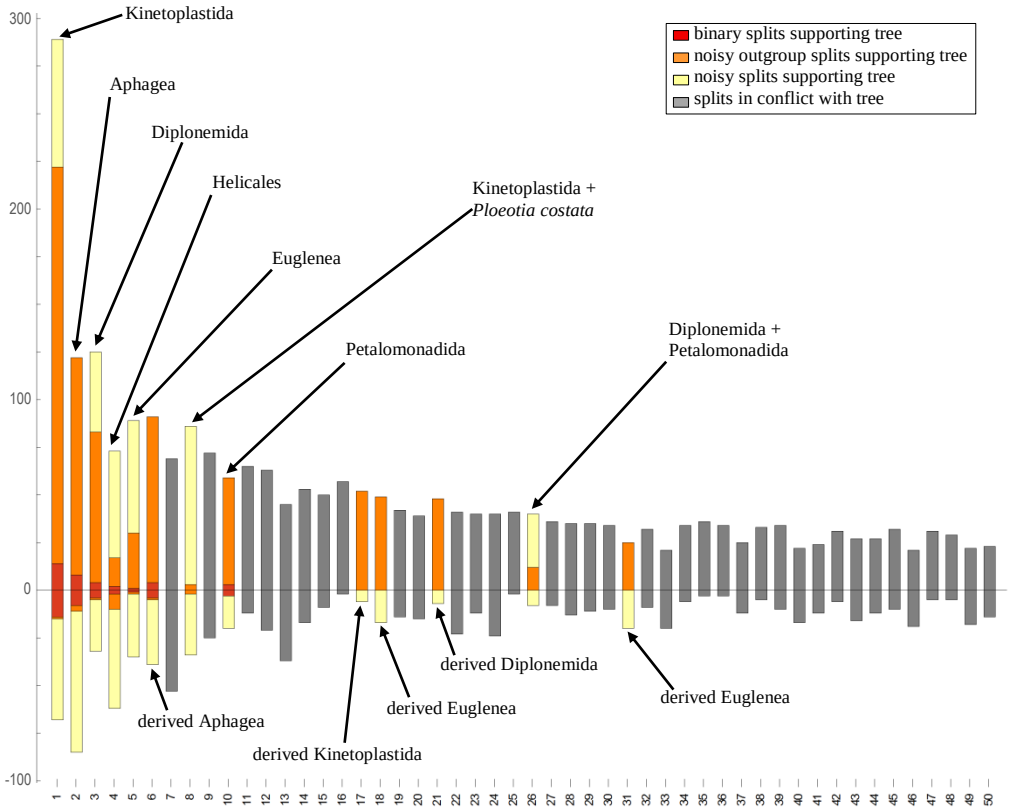


Fig. 5.28: Split support spectrum comprising the 50 best splits of modified dataset IV with Opisthokonta as outgroup. Compatible splits referring to euglenozoan groups and taxa are marked by black arrows above, those referring to derived euglenozoan groups below the graph. Conflicting splits represent nonsense correlations of outgroup or euglenozoan taxa, e.g. *Peranema trichophorum* and *Notosolenus ostium* (column 14).

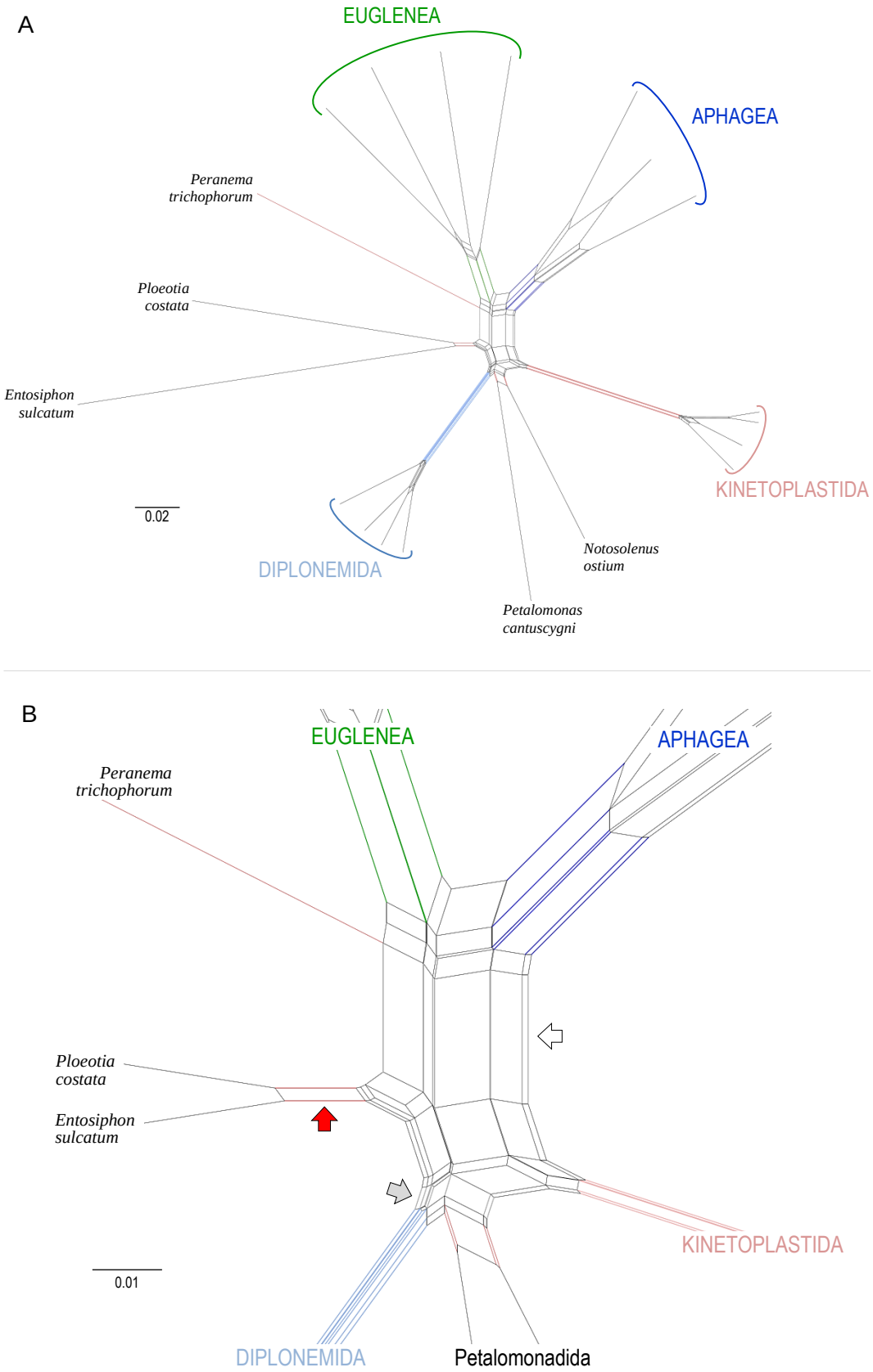


Fig. 5.29: Neighbor-net graph of modified LSU rDNA dataset IV after exclusion of outgroup taxa. Network splits supporting major groups are colored. **A**: Network overview displaying terminal splits. The scale bar represents 2 % sequence divergence. **B**: Detailed center view. The red arrow marks common splits of *Ploeotia costata* and *Entosiphon sulcatum*, a grey arrow accentuates splits which unite Diplonemida with Petalomonadida, the white arrow highlights splits supporting monophyletic Helicales. The scale bar depicts 1 % divergence.

changes to network topology or splits support for monophyletic ingroup taxa, as demonstrated in the corresponding spectral analysis (Fig. 5.28).

These findings revealed problems regarding choice of outgroup taxa, but of dissentient quality as observed in results from SSU rDNA analyses (see section 5.1.2). Similar to SSU rDNA results, analyses of LSU rDNA dataset IV confirmed the monophyly of major euglenozoan groups, i.e. Diplonemida, Kinetoplastida, Aphagea and Euglenea. Additionally, the monophyly of Helicales, representing a euglenid crown group, was confirmed in network and spectral analyses. More importantly, no splits were found which supported a sister group relationship of Diplonemida and Kinetoplastida, nor a monophyletic assemblage of Euglenida, for LSU rDNA sequences of Petalomonadida demonstrated a higher affinity to Diplonemida than to other Euglenida, which was affirmed by network analyses and splits support (Fig. 5.29 B).

5.3.2 Strict dataset V

To further investigate euglenozoan LSU rDNA genealogy, another possible taxon sampling was used which included more primordial representatives of Kinetoplastida than trypanosomes, i.e. *Bodo saltans*, *Neobodo saliens* and *Dimastigella mimosa*. Hence data was limited for these taxa, LSU rDNA regions had to be taken into consideration, which were not available for all euglenozoan subgroups. In particular, nucleotide sequences of Aphagea could not be included into the new dataset. Subsequent to formation of LSU rDNA dataset V, criteria regarding dataset construction which have been formulated by Castresana (2000) were modified and applied to aligned LSU rDNA sequences: firstly, alignment positions in which any sequence contains a gap, and secondly, alignment positions in which all sequences included identical nucleotides were strictly eliminated from the alignment. Though Castresana originally conceptualized much more restrictive criteria, the unmodified employment of those would have had too constrictive consequences, since dataset V covered merely a small region of LSU rDNA and included 862 nucleotides.

Maximum likelihood and Bayesian inference analyses of dataset V resulted in maximal supported monophyletic Euglenozoa, but the basal euglenozoan radiation formed a tetra-furcation which produced four separate branches of unassured position: (1) monophyletic Petalomonadida appeared as sister clade to monophyletic Diplonemida with average support,

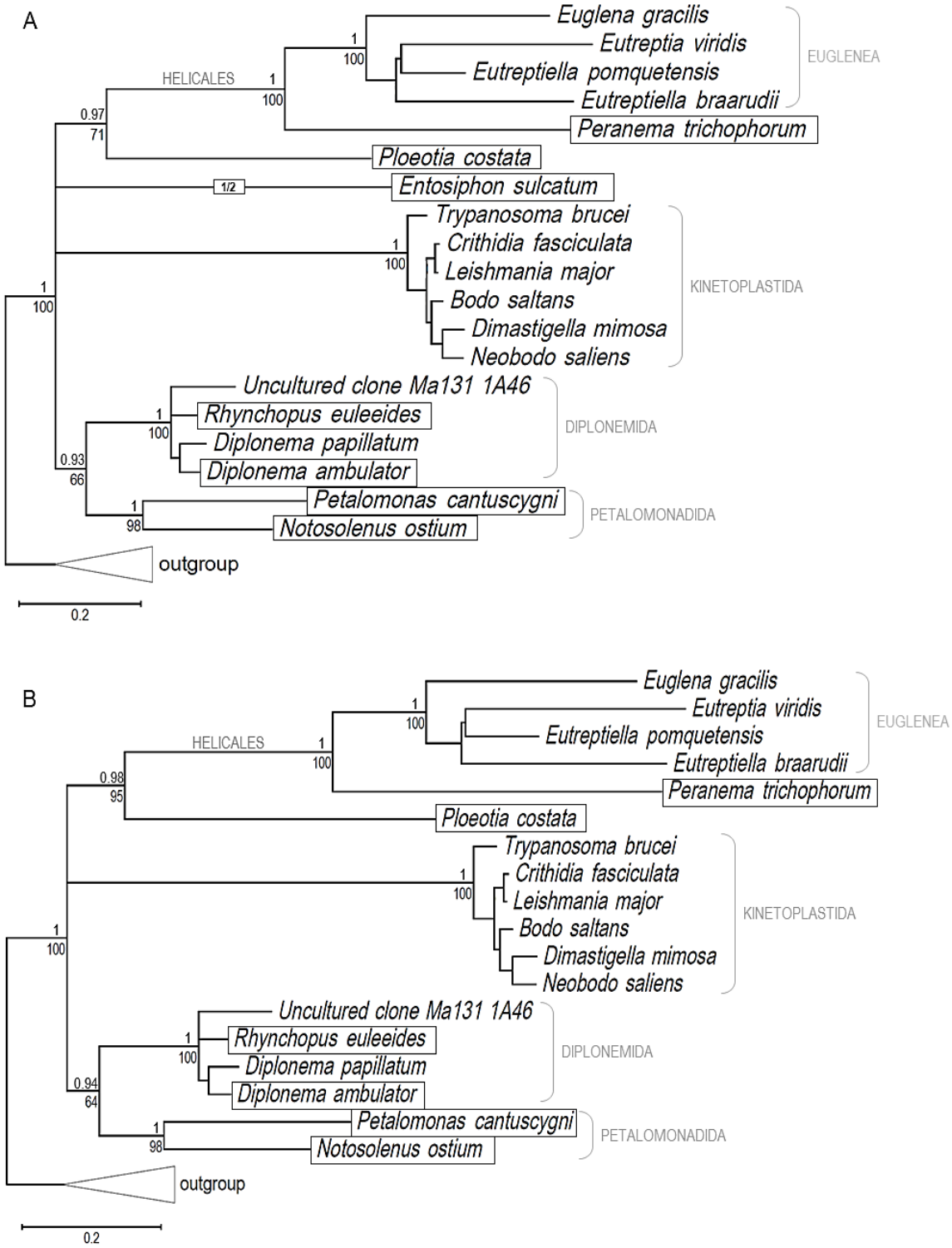


Fig. 5.30: Consensus trees obtained from ML and BI analyses of dataset V comprising 862 nucleotide positions with new sequences boxed. Congruent posterior probabilities are mapped onto ML trees and displayed above, ML bootstrap support values below corresponding nodes; discrepancies to Bayesian inference are hyphenated. Scale bars represent 20 % sequence divergence. **A**: Results including *Entosiphon sulcatum*, half of the original branch length depicted (GTR+Γ+I, -lnL = 11016.92, gamma shape = 0.586, p-invar = 0.113). **B**: Results from reiterated analyses excluding *Entosiphon sulcatum* LSU rDNA sequence (GTR+Γ+I, -lnL = 10250.45, gamma shape = 0.599, p-invar = 0.131).

(2) maximal supported monophyletic Kinetoplastida, (3) *Entosiphon sulcatum*, and (4) *Ploeotia costata* formed the sister taxon of a *Peranema*/Euglenea clade with good support (Fig. 5.30 A). As a result, initial tree topology from reiterated ML and BI analyses without

long-branching *Entosiphon sulcatum* was retained, while statistical support for the fourth clade increased substantially, i.e. *Ploeotia costata* was confirmed as sister taxon of the *Peranema*/Euglenea clade with a posterior probability (PP) of 0.98 and 95 bootstrap support (BS) instead of 0.97 PP and 71 BS (compare Fig. 5.30 A/B). Taxon sampling of dataset V included presumably primordial kinetoplastid taxa (*sensu* Deschamps et al. 2011 and Doležel et al. 2000), but those branched as derived kinetoplastids from paraphyletic trypanosomatids in ML and BI trees. Though the positions of major euglenozoan clades remained unclear due to a tetrafurcation in the original and a trifurcation in reiterated trees, the sister group relationship of Petalomonadida and Diplonemida received fair bootstrap support (66 and 64) and very good posterior probability values (0.93 and 0.94). Furthermore, evolutionary divergence estimates between groups revealed that Petalomonadida and Diplonemida are the nearest relatives among examined major groups of Euglenozoa.

Tab. 5.4: Estimates of average evolutionary divergence over sequence pairs between euglenozoan taxa and the outgroup in dataset V. Divergence estimates are shown below, corresponding standard deviations above the diagonal.

No.	Taxon	1	2	3	4	5	6	7	8	9
1	Euglenea	-	0,044	0,078	0,044	0,049	0,041	0,058	0,058	0,045
2	*Peranema trichophorum*	0,437	-	0,086	0,054	0,048	0,050	0,063	0,063	0,049
3	*Entosiphon sulcatum*	0,673	0,676	-	0,072	0,070	0,058	0,084	0,083	0,064
4	*Ploeotia costata*	0,430	0,492	0,574	-	0,045	0,040	0,058	0,057	0,044
5	Petalomonadida	0,480	0,464	0,602	0,412	-	0,029	0,048	0,048	0,034
6	Diplonemida	0,409	0,450	0,514	0,356	0,295	-	0,042	0,041	0,025
7	bodonids	0,543	0,529	0,666	0,501	0,454	0,393	-	0,008	0,041
8	trypanosomatids	0,535	0,529	0,659	0,498	0,460	0,389	0,050	-	0,040
9	outgroup	0,453	0,462	0,561	0,406	0,347	0,257	0,389	0,375	-

Neighbor-net analyses of strict LSU rDNA dataset V approved the monophyly of Euglenozoa and major euglenozoan clades, i.e. Diplonemida, Petalomonadida, Kinetoplastida and Euglenea (Fig. 5.31 A). Interestingly, *Ploeotia costata* and *Entosiphon sulcatum* shared common splits in the network graph, which emerged as the third best set of splits in the corresponding spectral analysis (split 3 in Fig. 5.32 A). *Peranema trichophorum* formed the sister taxon of Euglenea and this clade represents monophyletic Helicales in a light taxon sampling supported by common splits. Intriguingly, Diplonemida and Petalomonadida shared mutual splits (Fig. 5.31 B), which corroborated findings from ML as well as BI trees, phylogenetic network and spectral analyses of dataset IV. No splits were found that would have supported a sister group relationship of Kinetoplastida and Diplonemida. After exclusion of *Entosiphon's* LSU rDNA sequence, *Ploeotia costata* solely shared common splits with the *Peranema*/Euglenea clade, in the network graph and splits support spectrum (Fig. 5.32 B, split

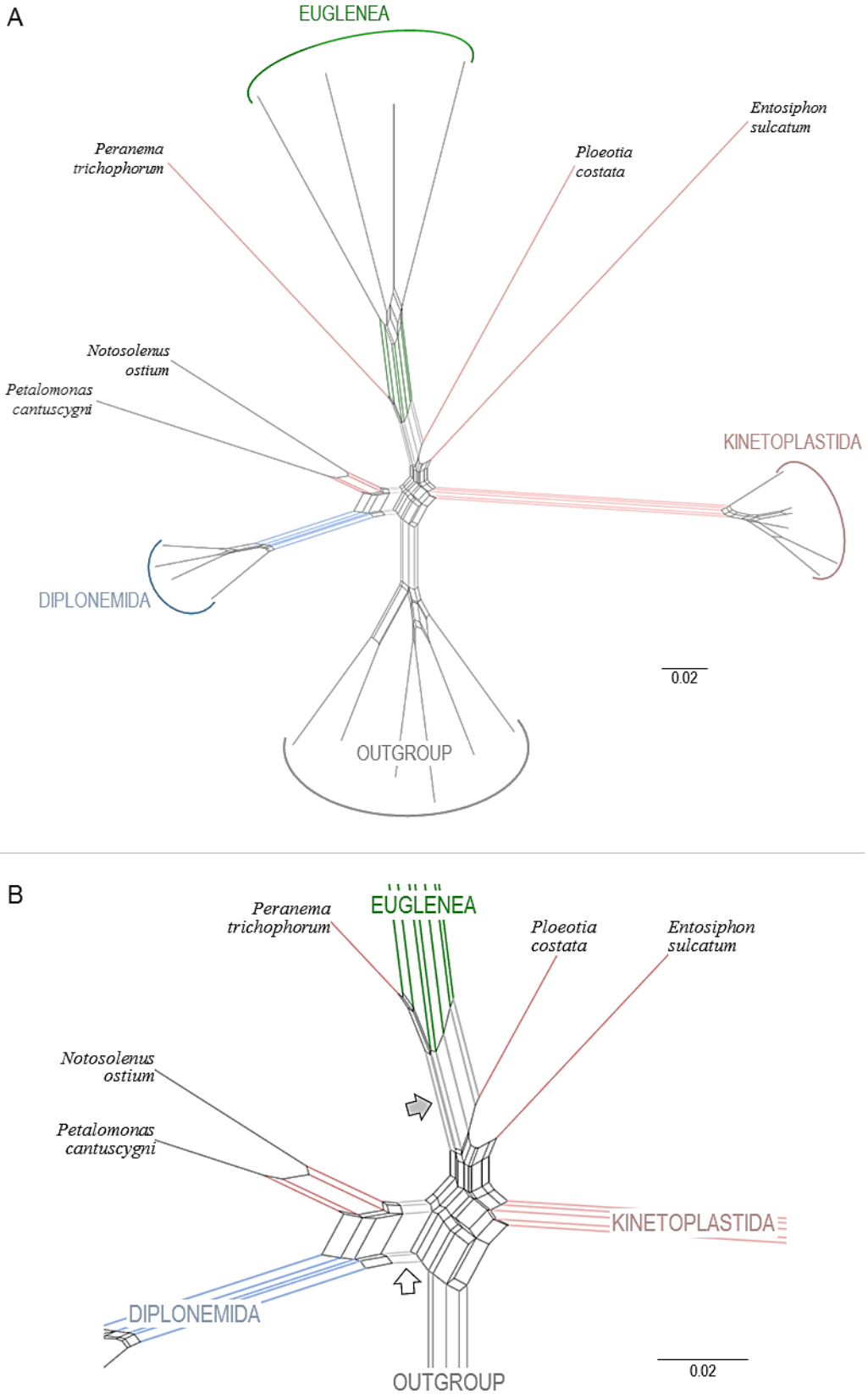


Fig. 5.31: Neighbor-net graph of dataset V comprising euglenozoan and outgroup taxa. Network splits supporting monophyletic clades are colored. The scale bar represents 2 % sequence divergence. **A**: Network overview displaying terminal splits. **B**: Detailed center view. Splits supporting monophyly of the *Peranema*/Euglenea clade (Helicales) are marked by a grey arrow, and a white arrow highlights common splits uniting Petalomonadida and Diplonemida.

21). Exclusion of the outgroup exhibited no changes to network topology, i.e. Diplonemida grouped together with Petalomonadida, supported by mutual splits (Fig. 8.16).

Finally, these findings revealed a deforming impact of distant outgroup taxa on LSU rDNA-based euglenozoan phylogeny, which was conditioned by an unsuitable outgroup taxon sampling, for only limited LSU rDNA sequences of Euglenozoa and inappropriately few, phylogenetically highly derived outgroup sequences were available at present time.

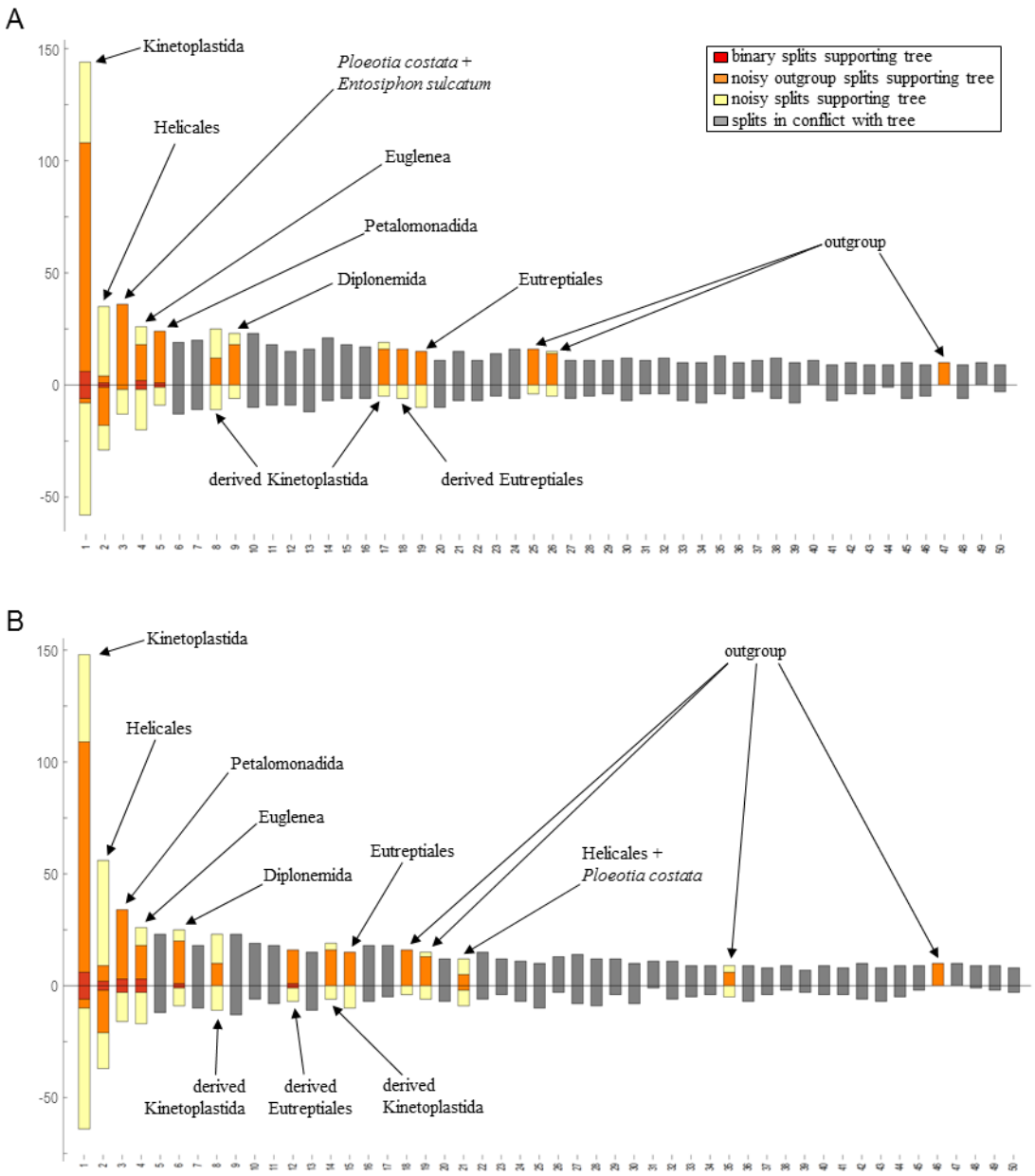


Fig. 5.32: Split support spectrum of LSU rDNA dataset V. Compatible splits supporting monophyletic euglenozoan clades and outgroup taxa are marked by black arrows above, those referring to internal branches below the graph. All other splits are related to nonsense groupings of euglenozoan taxa with outgroup or euglenozoan taxa only. Helicales comprises the *Peranema*/Euglenea clade from Fig. 5.31. **A**: Spectrum of unmodified dataset V. **B**: Splits support spectrum of dataset V excluding long-branching *Entosiphon sulcatum* sequence.

Nonetheless, results from LSU rDNA analyses confirmed significant findings concerning the understudied LSU rDNA genealogy of Euglenozoa: (1) the monophyly of major euglenozoan groups, i.e. Diplonemida, Kinetoplastida, Petalomonadida, Aphagea and Euglenea, confirmed findings from prior SSU rDNA analyses; (2) monophyletic Helicales represented the euglenid crown group in LSU rDNA and in SSU rDNA genealogy; (3) the absence of phylogenetic signal in euglenozoan LSU rDNA sequences, which would have provided support for a sister group relationship of Diplonemida and Kinetoplastida; (4) Diplonemida and Petalomonadida formed a clade in phylogenetic tree and network analyses, validated by statistical and splits support; (5) the absence of a phylogenetic signal which would have supported the monophyly of Euglenida corroborated results from preliminary SSU rDNA analyses and those found by Busse & Preisfeld (2003a). Modified criteria after Castresana (2000) were used as prerequisite for formation of a strict LSU rDNA-based dataset including presumably primordial bodonids, the analysis of which confirmed aforementioned findings, but also created second thoughts about the position of bodonids and trypanosomatids within Kinetoplastida. Phylogenetic implications of these important results combined with results from SSU rDNA analyses affect the Euglenozoa as a whole and are concluded in chapter 6.

5.4 LSU rDNA nucleotide sequence analyses

5.4.1 Base composition

Analysis of euglenozoan LSU rDNA nucleotide composition revealed striking similarities between Diplonemida and Petalomonadida: both groups exhibited congeneric nucleotide percentages as well as identical base frequency patterns, which considerably differed from that of other euglenozoans, e.g. Kinetoplastida and Eutreptiales. Outgroup LSU rDNA sequences displayed the highest percentages of adenine (A) and thymine (T) and reciprocally the lowest values of cytosine (C) and guanine (G), while the sequences of genus *Eutreptiella* , *Euglena gracilis* and *Ploeotia costata* showed obverse base percentages (Tab. 8.3). Base frequencies of *Notosolenus ostium*, *Petalomonas cantuscygni*, *Diplonema ambulator*, *Diplonema papillatum* and *Rhynchopus euleeides* were composed of the nucleotide pattern T < C < A < G. Nucleotide frequencies in LSU rDNA sequences of the supposedly primordial kinetoplastids *Bodo saltans* and *Dimastigella mimosa* featured a G-A shift and followed the pattern T < C < G < A, with a comparatively narrow distribution of T/C and G/A percentages, whereas *Neobodo saliens* exhibited the pattern C < T < G < A (Fig. 5.33). This finding

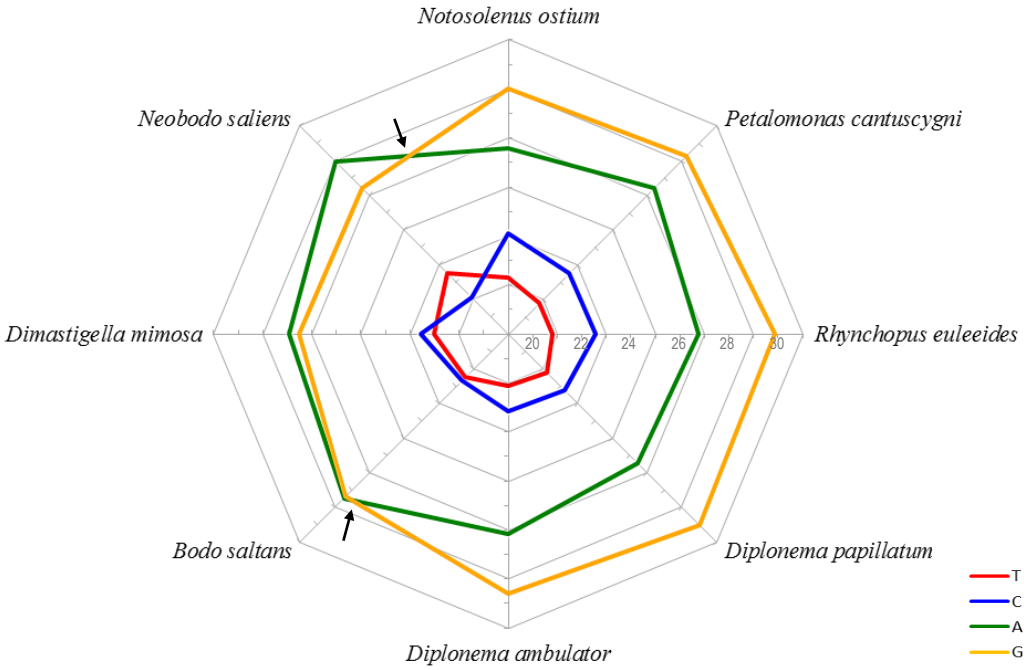


Fig. 5.33: Octagonal base frequency graph visualizing base composition differences in LSU rDNA of Petalomonadida, Diplonemida and Kinetoplastida. Each base is color coded as shown in the legend to the lower right, and base frequency percentages are depicted by connected lines in a spider web-like arrangement along radials which represent LSU rDNA sequences of different taxa. A scale along the radius of *Rhynchopus euleeides* indicates a base frequency percentage of 20 % to 30 % (from inner to outer octagon). Black arrows highlight the distinct shift in G-A frequencies observed in presumably primordial kinetoplastid sequences *sensu* Deschamps et al. (2011) and Doležel et al. (2000).

corroborated results obtained by phylogenetic analyses of LSU rDNA sequences, which proposed a close relationship of Petalomonadida and Diplonemida.

Multiple nucleotide shifts were observed in base frequencies of other euglenozoan taxa, i.e. *Ploeotia costata*, *Entosiphon sulcatum*, *Peranema trichophorum* and Eutreptiales compared to Petalomonadida. *Ploeotia costata* revealed the highest cytosine percentage and *Peranema trichophorum* the highest thymine percentage of all LSU rDNA sequences examined (Tab. 8.3). Intriguingly, *Peranema trichophorum* and *Eutreptia viridis* exhibited an identical base pattern, i.e. C < T < A < G, which differed completely from those observed in LSU rDNA sequences of other Eutreptiales, for *Eutreptiella braarudii* and *E. pomquetensis* both constituted pattern T < A < C < G (Fig. 5.34). However, if *Peranema trichophorum* was a relative of *Eutreptia viridis*, this would contradict previous results from phylogenetic analyses. These findings demonstrate a comparatively high variability of base frequencies between primordial petalomonad and presumably more derived LSU rDNA sequences of phagotrophic *Ploeotia costata*, *Entosiphon sulcatum*, *Peranema trichophorum* and phototrophic Eutreptiales.

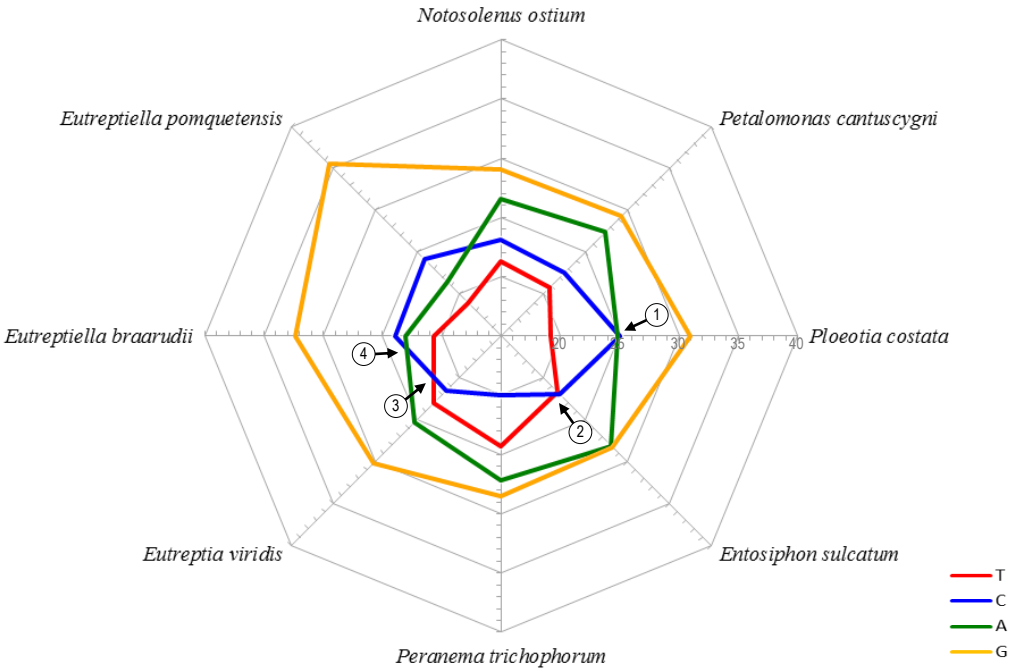


Fig. 5.34: Octagonal base frequency graph displaying multiple shifts in LSU rDNA base frequencies of *Ploeotia costata*, *Entosiphon sulcatum*, *Peranema trichophorum* and Eutreptiales compared to Petalomonadida. A scale along the radius of *Ploeotia costata* indicates a base frequency percentage of 20 % to 40 % (from inner to outer octagon). Black arrows highlight individual shifts: **1**: of A-C frequencies in *Ploeotia costata*; **2** and **3**: of T-C frequencies in *Peranema* and *Eutreptia* sequences; **4**: of A-C frequencies in the genus *Eutreptiella*.

5.4.2 LSU rDNA length comparison

Nucleotide sequence alignment according to secondary structure features was a prerequisite for identification of LSU rDNA domain boundaries and thereby allowed for exact length measurement of homologous sequences. New LSU rDNA sequences of phagotrophic euglenids and diplonemids which were obtained in the scope of this work largely increased the number of available euglenozoan taxa. A comparative survey of partial euglenozoan and outgroup LSU rDNA sequences revealed group-specific lengths of deduced LSU rRNA domains II to V (Fig. 5.35). Diplonemida exhibited rather uniform LSU rDNA sequence lengths of about 2,250 nucleotides (nt), which represented the smallest value of all Euglenozoa examined. Kinetoplastids showed more variable length values of about > 2,700 nt, i.e. *Crithidia fasciculata* presented the smallest and *Leishmania major* the largest value of 3,000 nt. Sequence lengths of phagotrophic euglenids lay in between, *Petalomonas cantuscygni* unveiled the most similar LSU rDNA sequence length to that of diplonemids. Phototrophic euglenids overall exhibited larger LSU rDNA sequences and *Eutreptiella braarudii* showed the longest LSU rDNA of all examined taxa (> 3,400 nt).

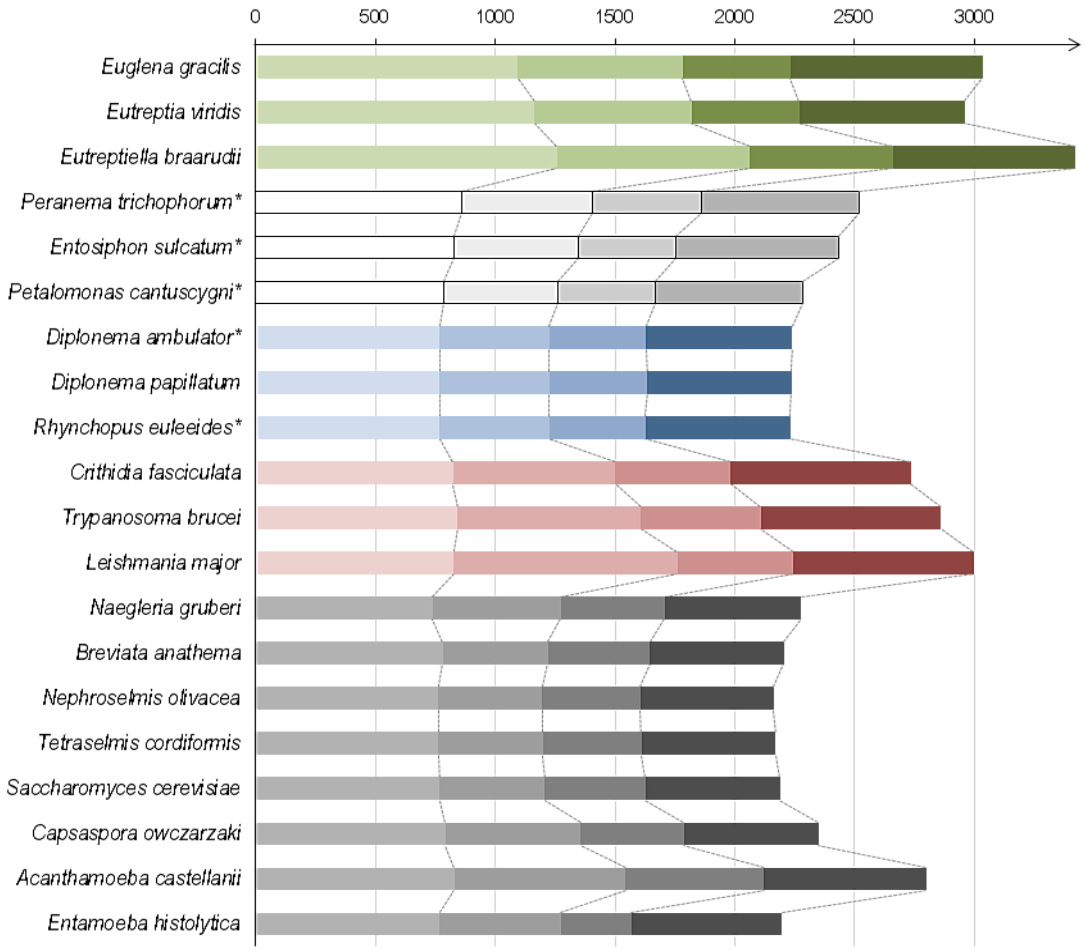


Fig. 5.35: Group-specific length polymorphism of euglenozoan LSU rDNA sequences compared to outgroup taxa. Taxa which LSU rDNA sequences were obtained in the scope of this work are marked by an asterisk. Sequences of deduced LSU rRNA domains II, III, IV and V were concatenated and are depicted in analogously colored bars, representatives of major groups are color-coded: phototrophic euglenids green, phagotrophic euglenids light-grey, Diplonemida blue, Kinetoplastida red and outgroup taxa dark-grey.

A comparison of LSU rDNA sequence proportions also disclosed group-specific length polymorphisms regarding the size of individual LSU rDNA domains, e.g. relative domain I and III length of kinetoplastids exceeded those of diplonemids and *Petalomonas cantuscygni*, while reciprocally, relative domain II and V lengths of *Petalomonas cantuscygni* and Diplonemida surpassed that of Kinetoplastida (Fig.8.17).

Although these findings clearly constituted a major group-specific length polymorphism of euglenozoan LSU rDNA sequences, this matter must be treated with caution, for present taxon sampling cannot be regarded satisfying: the Kinetoplastida are solely represented by (presumably derived) trypanosomatids and more taxa from other euglenozoan groups, e.g. Aphagea, Anisonemida and Ploeotiida, should be investigated to validate or falsify this hypothesis. Albeit findings from LSU rDNA phylogenetic analyses could challenge widely accepted SSU rDNA based hypotheses on kinetoplastid evolution, results from sequence

length comparison nonetheless affirmed a close relationship of Petalomonadida and Diplonemida as observed in previous phylogenetic and nucleotide sequence analyses.

5.4.3 Secondary structure

Euglenozoan and outgroup LSU rDNA sequences have been aligned according to secondary structure information of *Saccharomyces cerevisiae* provided by Petrov et al. (2013). Like in SSU rDNA (5.2.3), differences in LSU rDNA sequences are substantially reflected by unique nucleotide substitutions in their corresponding secondary structure. Therefore, secondary structure properties of deduced euglenozoan LSU rDNA sequences were investigated in the search for unique features. But LSU rDNA sequences include more components than SSU rDNA, because two internal transcribed spacers separate 5.8S and 28S structural regions. Consequently, LSU rDNA sequences contain more measurable information than SSU rDNA sequences. An incisive anomaly was discovered in examination of deduced LSU rRNA helix 10, which embodies the downstream coalescence of 5.8S and 28S ribosomal DNAs. The 5'-strand of helix 10 represent the 3'-end of 5.8S rDNA, whereas the 3'-strand of helix 10 is formed by the 5'-end of 28S rDNA and both rDNA strands encompass internal transcribed spacer 2. Discrepancies regarding the boundaries of this transition zone were found in annotations of some outgroup LSU rDNA sequences as well as in those of *Euglena gracilis* and *Diplonema papillatum*, which were available for download on NCBI's nucleotide database homepage. Since deduced LSU rRNA secondary structure of helix 10 includes complementary base pairs, nucleotides of ITS2 region usually do not form matching base pairs. For instance, following given annotation of *Besnoitia besnoiti*'s LSU rDNA, the 3'-strand of its assigned ITS2 would prevent the existence of helix 10, although deduced secondary structure of that sequence revealed a stable pairing of at least 15 base pairs with the complementary 3'-end of its 5.8S region (Fig. 5.36). Such discrepancies between annotated and deduced boundaries have been found in other taxa as well, and could arise when comparison of aligned sequences is used for annotation without regarding secondary structure information. But since the ITS2 sequence is not present in the nascent ribosomal RNA, this transition zone is posttranscriptionally processed in many steps. Therefore, when sequenced from RNA, such discrepancies could possibly represent different stages of ribosomal RNA maturation, which have been isolated in dissentient moments of posttranscriptional processing of the RNA molecule. This unsettling finding needs further examination to assure reliability of sequence data for molecular research.

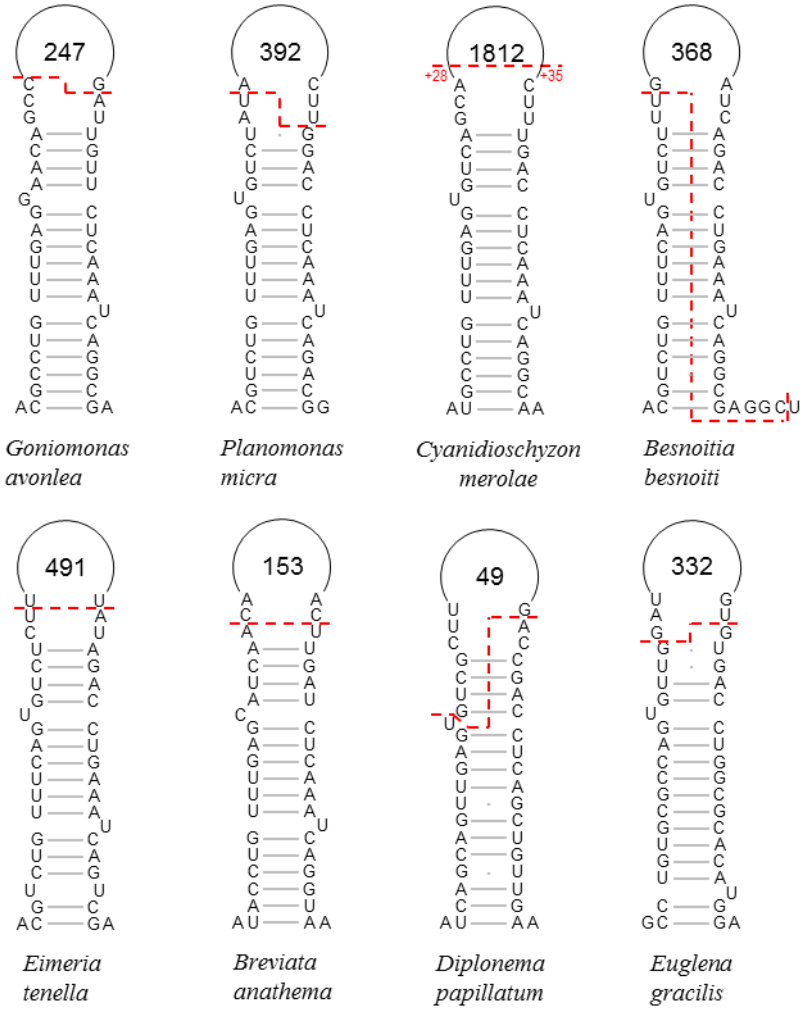


Fig. 5.36: Discrepancies between concluded secondary structure of LSU rRNA helix 10 (after Petrov et al. [2013]) and given annotations of the ribosomal 5.8S – ITS2 – 28S transition zone as observed in outgroup and euglenozoan taxa. Dashed red lines depict annotated boundaries of the transition zone. Numbers in apices represent individual length of ITS2 in nucleotides.

While helix 10 encompassed ITS2, inferred LSU rRNA helix 2 represented the ribosomal transition zone involving ITS1 which separates 18S from 5.8S rDNA. Limitations of both transition zones allowed for length comparison of ITS sequences. *Euglena gracilis* displayed by far the longest ITS1 sequence of all taxa investigated, more than two times longer than those of *Ploeotia costata* or *Peranema trichophorum*, and the phagotrophic euglenid *Notosolenus ostium* possessed the shortest ITS1 of all examined taxa (Fig. 5.37 A). The ITS1 length of *Bodo saltans* lay within length variance of diplonemids, while that of *Notosolenus ostium* was shorter and that of *Petalomonas cantuscygni* significantly longer. Like phagotrophic euglenids, kinetoplastid taxa exhibited a high variability of ITS lengths, e.g. trypanosomatids showed the longest ITS2 sequences of all taxa examined, while that of *Bodo saltans* was much shorter (Fig. 5.37 B). The shortest ITS2 sequence was found in *Diplonema*

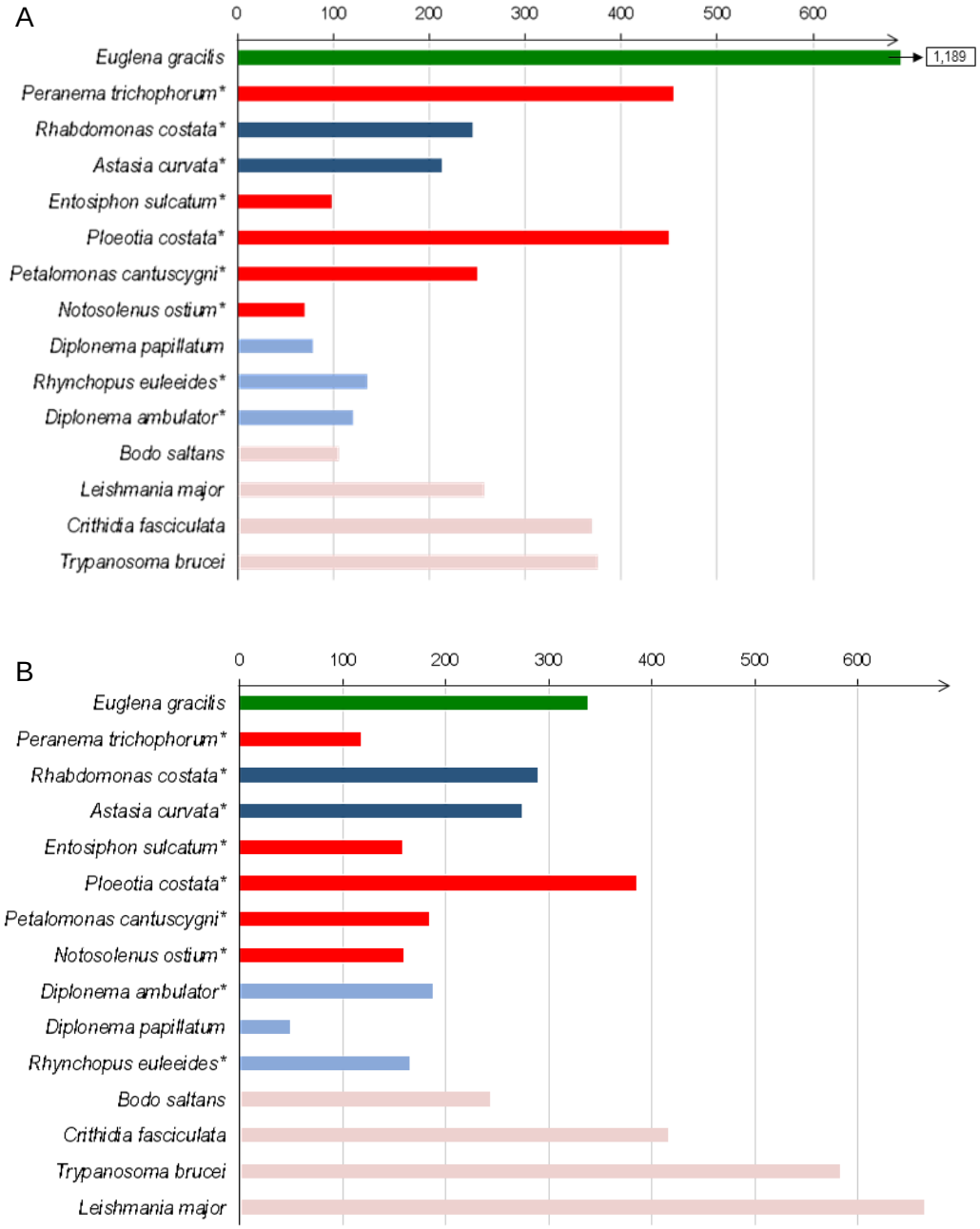


Fig. 5.37: Internal transcribed spacer sequences graph illustrating different absolute ITS sequence length of euglenozoan taxa. Nucleotide sequences obtained in the scope of this work are marked by an asterisk. Representatives of major groups are color-coded: *Euglena gracilis* green, phagotrophic euglenids red, Aphagea dark-blue, Diplonemida light-blue and Kinetoplastida light-red. **A**: Length values of ITS1 sequences, note that the ITS1 sequence of *Euglena gracilis* was shortened. **B**: Absolute length values of ITS2 sequences.

papillatum, but those of other diplonemids taxa were significantly larger. This conspicuous deviation from uniformity, which usually characterized Diplonemida, could be a result of aforementioned discrepancy conditioned by diverse stages of ribosomal RNA maturation and further sequencing of more taxa could rectify this ambiguity. Disregarding *Diplonema papillatum*, ITS2 lengths of Petalomonadida are more similar to that of residual Diplonemida than those of kinetoplastids. Concatenation of ITS sequences revealed, that *Notosolenus*

ostium and *Entosiphon sulcatum* exhibited ITS length values which lay within length variance of Diplonemida (Fig. 8.18 A). While ITS lengths of trypanosomatids by far exceeded that of *Bodo saltans*, the latter was merely longer than that of diplonemids. Although proportions of ITS length values showed an unexpectedly high variability, no group-specificity could be found in proportion comparison (Fig. 8.18 B).

Unique substitutions

Conversion of euglenozoan LSU rDNA sequences into their putative secondary structures led to the discovery of group-specific unique nucleotide substitutions. For instance, a single thymine (uracil in RNA) was found in the LSU rDNA of *Bodo saltans* which had no complementary binding partner nucleotide in the deduced secondary structure of helix 10. Moreover, *Crithidia fasciculata* and *Leishmania major* both possessed each another and *Trypanosoma brucei* even two additional thymines which apparently had no binding partner

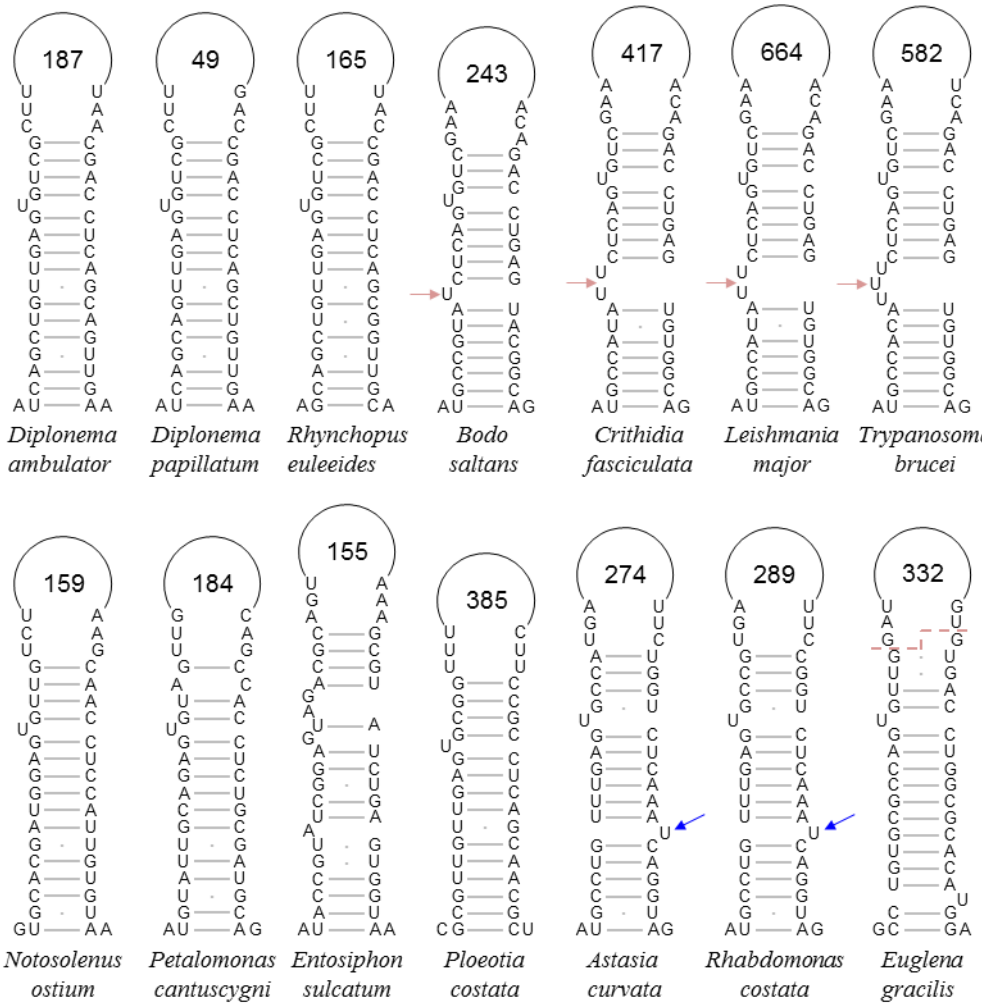


Fig. 5.38: Deduced LSU rRNA secondary structure of helix 10. Group-specific unique substitutions of Kinetoplastida are marked by light-red arrows, those of Aphagea by blue arrows. Numbers in apices represent individual ITS2 sequence lengths.

nucleotides on the complementary 3’-strand of helix 10 (Fig. 5.38). These unique nucleotide additions present in the 5’-strand of helix 10 represented group-specific base substitutions within Kinetoplastida, which were inexistent at that homologous position in any other taxon examined. One more potentially unique substitution for Kinetoplastida was found in putative LSU rDNA helix 61 (Fig. 8.19). Another nucleotide addition was detected in *Astasia curvata* and *Rhabdomonas costata* both exhibited an additional thymine on the 3’-strand of helix 10 without a binding partner nucleotide on the complementary 5’-strand (Fig. 5.38). As a result, this substitution can be regarded as potentially group-specific nucleotide addition for the Aphagea, but further sequencing of more taxa will be needed to validate this finding.

LSU rDNA sequences of Helicales contained exclusive nucleotide substitutions in their inferred secondary structure as well. A thymine (uracil in RNA) was detected on the 5’-strand of helix 91 of *Peranema trichophorum*, and in the very same position (homologous) cytosines were found in the LSU rDNA of Aphagea and Euglenea, all of which had no feasible binding

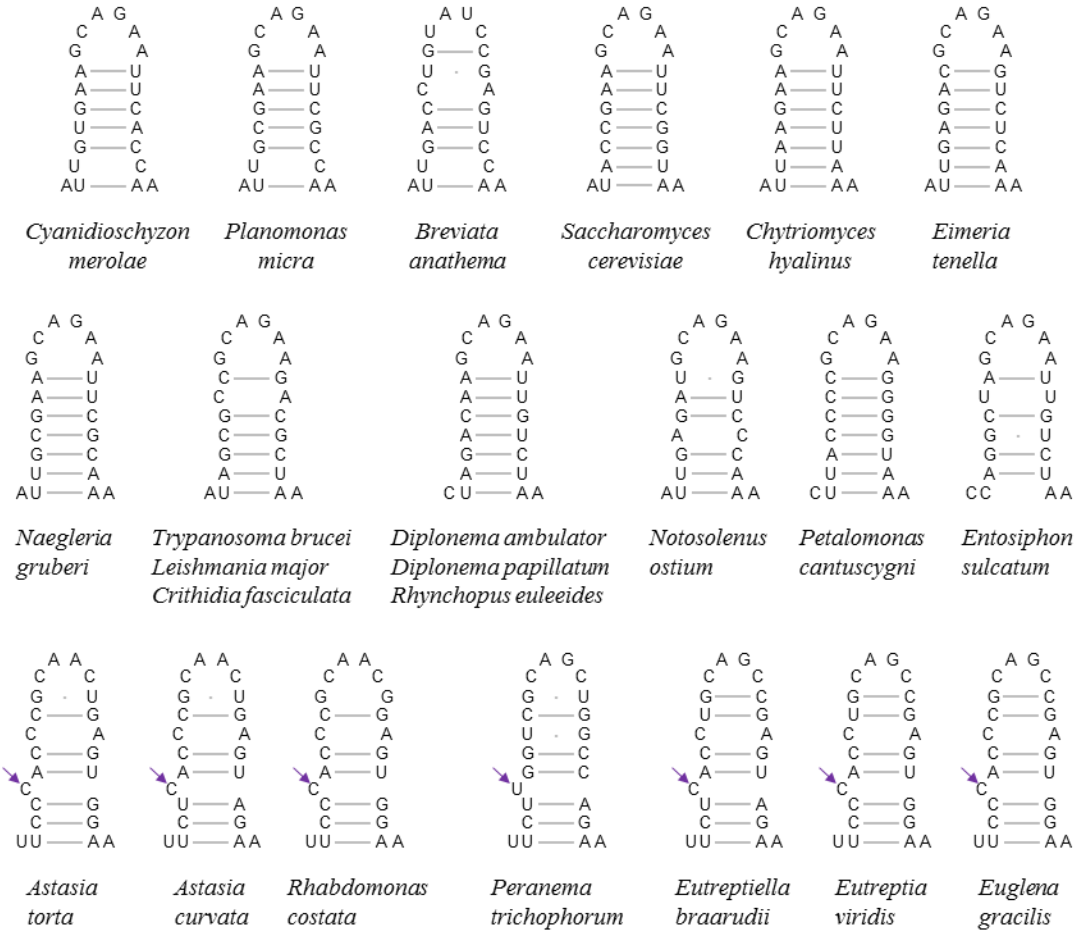


Fig. 5.39: Group-specific nucleotide substitutions for the Helicales as discovered in concluded secondary structure of LSU rRNA helix 91. Unique substitutions are highlighted by purple arrows and were found in *Peranema trichophorum*, Aphagea, Eutreptiales and *Euglena gracilis*, which all together represent the Helicales.

partner nucleotides on the complementary 3'-strand of helix 91 (Fig. 5.39). Although these exclusive nucleotide substitutions in helix 91 were group-specific for Helicales, more LSU rDNA sequences of Helicales, especially those of phagotrophic Helicales like Anisonemida or genera *Dinema*, *Heteronema* and *Neometanema*, need to be sequenced to validate this finding. Nonetheless, results obtained from secondary structure analysis of helix 91 revealed homologous additional nucleotide additions in taxa which all belong to diverse subgroups of the Helicales, i.e. Aphagea, Eutreptiales, Euglenales and *Peranema trichophorum* as individual phagotrophic representative, which corroborated previous findings from phylogenetic as well as spectral analyses and validated the identity of Helicales.

Conserved positions

Finally, secondary structure analysis was used to search for homologous positions in the bigger part of euglenozoan LSU rDNA. Therefore, highly conserved nucleotides in euglenozoan LSU rDNA were identified during the alignment process and then mapped onto a chart of the LSU rRNA secondary structure model of *Saccharomyces cerevisiae*, which was provided by Petrov et al. (2013). Thereby, ITS insertion points of kinetoplastid taxa were compared with those of *Euglena gracilis* in the bigger part of LSU rDNA, i.e. downstream sequence regions involving ITSs 13 and 14 of *Euglena gracilis* were not considered due to alignment ambiguities which were caused by an insufficient number of taxa. Kinetoplastid sequences included LSU rDNA sequences of *Crithidia fasciculata*, *Trypanosoma brucei* and *Leishmania major*, examined diplonemid LSU rDNA sequences were those of *Diplonema ambulator*, *Diplonema papillatum* and *Rhynchopus euleeides*. As a result, besides ITS1 and ITS2 regions, (1) no other ITSs were found in examined LSU rDNA sequences of diplonemids, (2) two exclusive ITS insertion points were existent in kinetoplastid sequences, i.e. ITSs 3 and 4 located in LSU rDNA domain III, (3) seven ITS insertion points were exclusively present in LSU rDNA of *Euglena gracilis*, i.e. ITS3, ITS4 and ITS5 in domain I, ITS6 in domain II, ITS7 and ITS8 in domain III and ITS9 in domain V, (4) three insertion points of *Euglena gracilis*' ITSs matched those of kinetoplastid LSU rDNA sequences, i.e. Euglena gracilis' ITS10, ITS11 and ITS12 inserted at the same points as kinetoplastid ITS5, ITS6 and ITS7, respectively. These findings showed that the dissimilar fragmentation present in LSU rDNA of kinetoplastids and *Euglena gracilis* might be the result of a continuous evolutionary process, for five out of seven kinetoplastid ITS regions share identical LSU rDNA insertion points with *Euglena gracilis*. Since *Euglena gracilis* represents a highly derived euglenid, LSU rDNA sequences of a lot more taxa settled between Euglenales and

Kinetoplastida should be investigated to validate or falsify this perception. Furthermore, the identification of conserved LSU rDNA regions can be used as a blueprint for primer design to start future investigations.

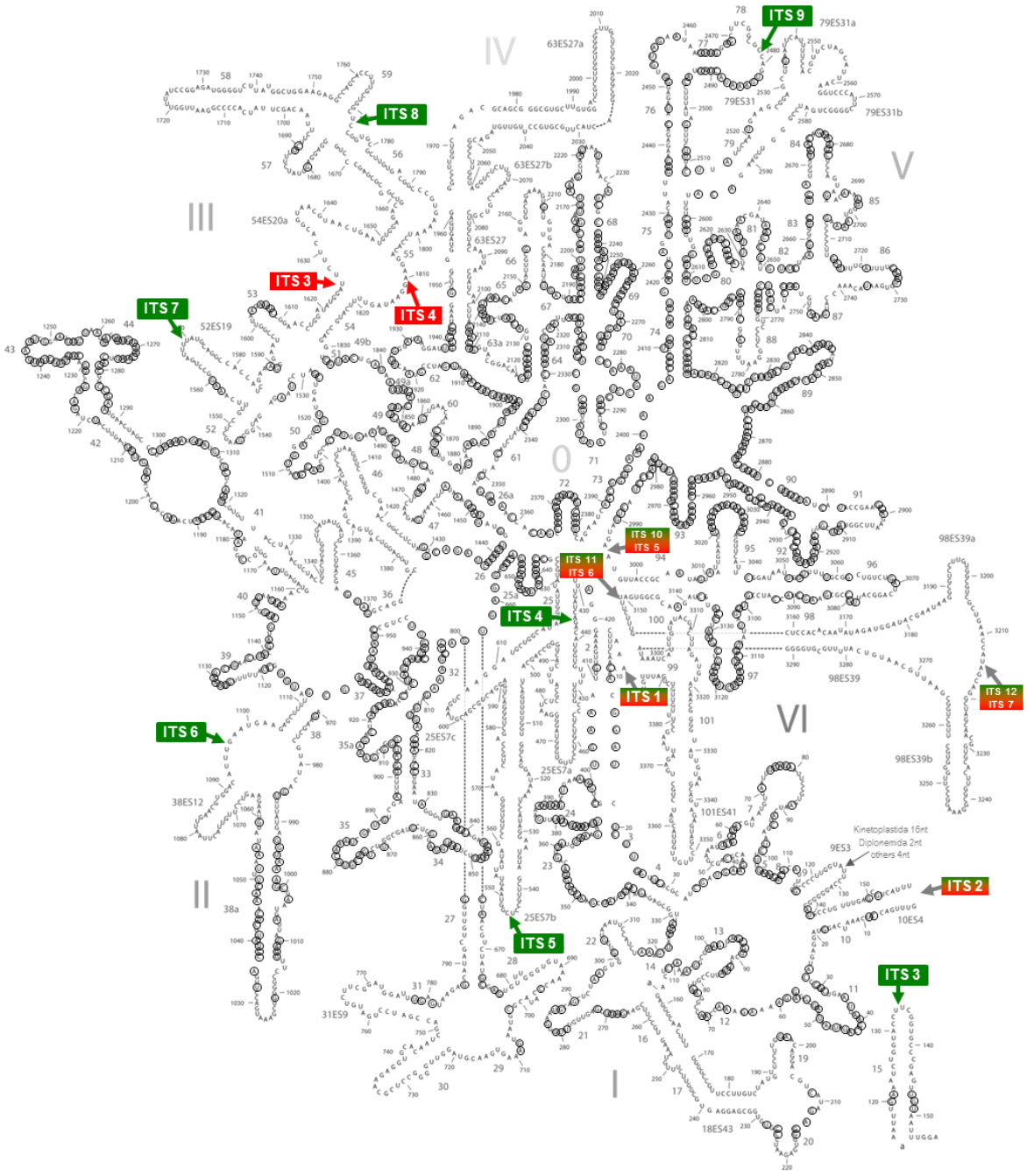


Fig. 5.40: Modified LSU rRNA secondary structure model of *Saccharomyces cerevisiae* including domain and helix numbering after Petrov et al. (2013) illustrating conserved nucleotide positions of euglenozoan LSU rDNA sequences which were mapped onto the graph and are represented by black circles. Individual and common insertion points of internal transcribed spacer (ITS) sequences of *Euglena gracilis* and kinetoplastid sequences are highlighted by colored boxes and arrows: consecutively numbered ITSs of *Euglena gracilis* are green, those of derived Kinetoplastida are red and congruent ITSs variegated (ITS13 and ITS14 of *Euglena gracilis* were not mapped).

5.5 Phylogenetic analyses of ribosomal operon sequences

Results from ML and BI analyses of dataset VI which contained concatenated SSU and LSU rDNA sequences confirmed monophyly of Euglenozoa and of major euglenozoan groups, i.e. Kinetoplastida, Petalomonadida, Diplonemida, Aphagea, Euglenea (including monophyletic subgroups Eutreptiales and Euglenales) and Helicales with maximum or very good support (Fig. 5.41 A). As observed in results from SSU rDNA-based analyses of preliminary dataset 0 as well as datasets II and III, *Entosiphon sulcatum* was the deepest-branching euglenozoan taxon and also produced the longest branch in the trees. This was expectedly due to its long-branch attraction effect, therefore analyses had to be reiterated without it. In the ML tree, position of Petalomonadida recovered moderate support, they branched between Kinetoplastida and Diplonemida, the positions of which were weakly supported. *Ploeotia costata* appeared as sister taxon of the Helicales with very good support. Aphagea formed the very good supported deepest branch within Helicales and Peranema trichophorum appeared as good supported sister taxon of Euglenea. These findings affirmed results from previous phylogenetic analyses of separate ribosomal genes, but some considerable topological discrepancies were found in the BI tree (see colored arrows in Fig. 5.41): *Peranema trichophorum* formed the deepest-branching taxon of Helicales with maximum support and thus changed branching position with Aphagea (orange double arrow), Kinetoplastida appeared as very good supported sister taxon of Diplonemida, albeit no bootstrap support (red arrow), and lastly Petalomonadida moved towards the root of the tree to a position near *Entosiphon sulcatum* (green arrow). As a result from reiterated analyses, topologies of ML and BI trees as well as discrepancies between both trees remained unchanged (colored arrows in Fig. 5.41 B). Kinetoplastida formed the deepest-branching euglenozoan clade in the ML tree, while Petalomonadida took that position in the BI tree and each alternative received maximum support, respectively. Slightly increased statistical support confirmed *Ploeotia costata* as sister taxon of monophyletic Helicales.

Additional analyses were performed to test whether ML or BI tree topology would be reflected in phylogenetic networks, but network graphs of unmodified dataset VI identified no support for Kinetoplastida neither for Petalomonadida as deepest-branching euglenozoan clade (Fig. 8.20). This resembled results from network analyses of separate SSU and LSU rDNA-based datasets, but as appertaining spectral analyses had shown, an unsuitable taxon sampling of outgroup representatives could have a distorting effect on euglenozoan sequences (see sections 5.1.2 and 5.3.1 for relevant split spectra). Since the comparatively small taxon

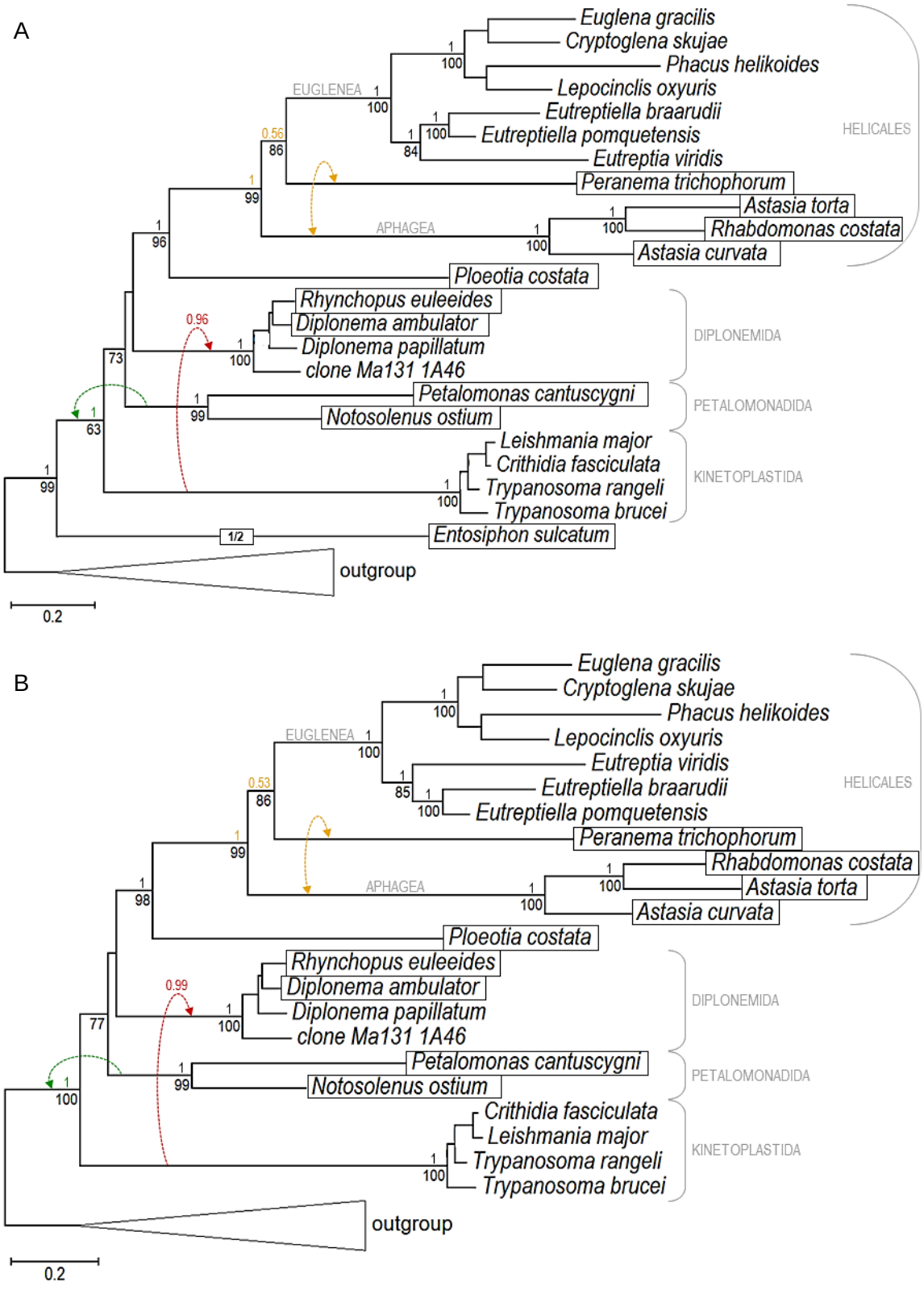


Fig. 5.41: Consensus trees obtained from analyses of operon dataset VI comprising 3,741 nucleotides with new sequences boxed. For taxon sampling see Tab. 4.5. Posterior probabilities are mapped onto ML trees and displayed above, bootstrap support values below corresponding nodes; topological differences to the Bayesian tree are highlighted by colored arrows, for explanation see text. Scale bars represent 20 % sequence divergence. **A**: Results involving 56 taxa including *Entosiphon sulcatum*, half of the original branch length depicted (GTR+Γ+I, -lnL = 91181.88, gamma shape = 0.465, p-invar = 0.119). **B**: Results from reiterated analyses excluding the sequence of *Entosiphon sulcatum* (GTR+Γ+I, -lnL = 87907.41, gamma shape = 0.467, p-invar = 0.126).

sampling of LSU rDNA sequences was the major limiting factor for composition of operon dataset VI, it seemed very likely that an inept sampling of outgroup taxa was the reason for this observation regarding the origin of Euglenozoa. However, results from a finally reiterated network analysis without outgroup taxa confirmed the monophyly of Euglenozoa and major subgroups, i.e. Petalomonadida, Diplonemida, Kinetoplastida, Aphagea, Euglenea and Helicales (Fig. 5.42). Interestingly, common splits united Petalomonadida and Diplonemida with *Ploeotia costata*, which reflected an amalgamation of results from SSU rDNA analyses (Diplonemida shared mutual splits with Ploeotiida) and LSU rDNA analyses (Diplonemida shared splits with Petalomonadida). These findings corroborated previous results from phylogenetic, spectral as well as statistical surveys of separate ribosomal genes, and were additionally confirmed by molecular apomorphies of major euglenozoan groups, which are finally discussed in the following chapter.

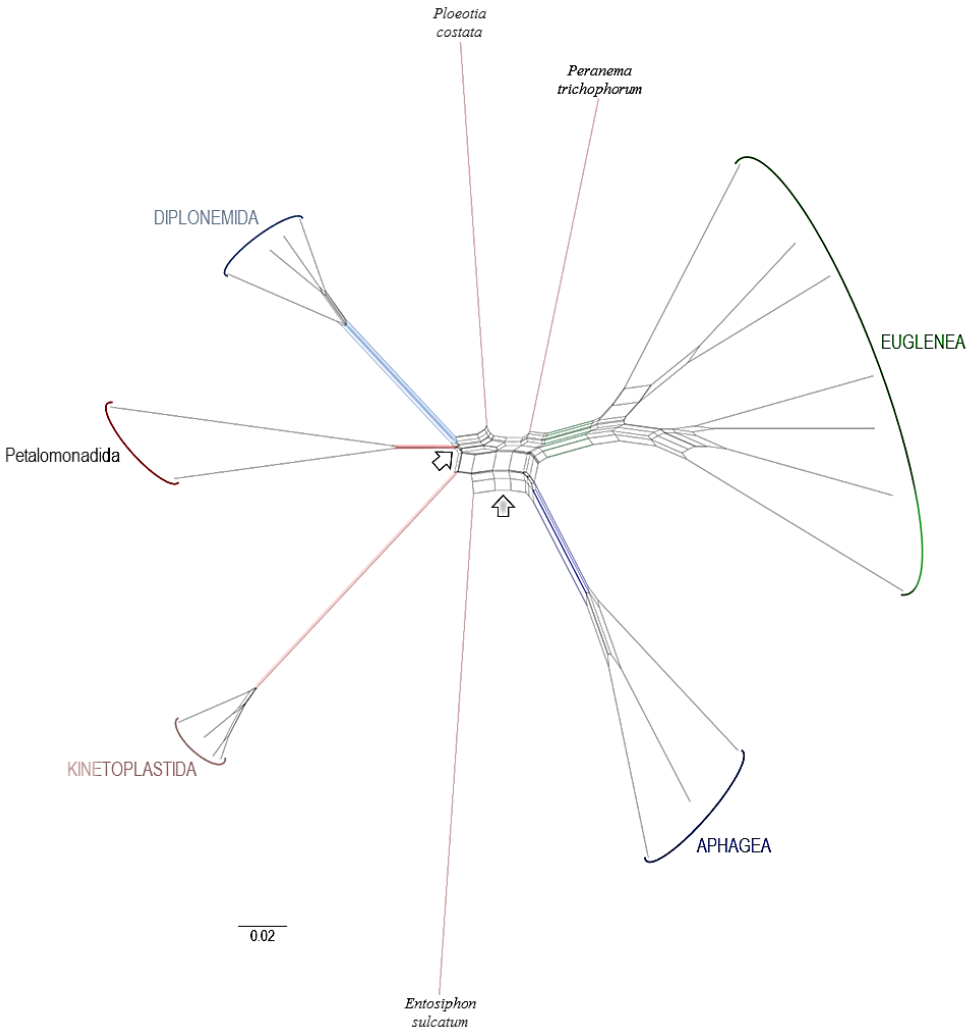


Fig. 5.42: Neighbor-net graph of modified operon dataset VI after exclusion of presumably unsuitable outgroup taxa. Network splits supporting monophyletic clades are colored. The scale bar represents 2 % sequence divergence. Network splits supporting monophyly of Helicales are marked by a grey arrow, common splits uniting Petalomonadida, Diplonemida and Ploeotia costata are highlighted by a white arrow.

6 Conclusion

In this chapter, the effects of obtained results on the phylogeny of euglenozoan taxa, groups and Euglenozoa as a whole, as well as their taxonomic implications are finally discussed and future prospects given.

6.1 Methodological approach

While most studies concerning the molecular phylogeny of Euglenida used SSU rDNA sequence data, only early published studies utilized secondary structure information as a requirement for alignment of homologous positions and dataset formation (Busse & Preisfeld 2002b and 2003b, Linton et al. 1999, Montegut-Felkner & Triemer 1997, Müllner et al. 2001, Preisfeld et al. 2000 and 2001). Without such a blueprint, alignments and datasets become highly individual and recognition of homologous positions more and more subjective, e.g. a manually masked alignment with "well-aligned sites suitable for phylogenetic analysis" or "reasonably well-aligned positions" is rather idiosyncratic than objective in regard to homology of nucleotide sequence positions (Lax & Simpson 2013, Lee & Simpson 2014b). Of course, taxon sampling may be a limiting factor for dataset width (i.e. number of positions therein), but variation would still be restricted by sequence length. For instance, two recently published phylogenetic studies of the Euglenida have been conducted with datasets including 39 and 49 taxa with similar euglenozoan taxon samplings, which consisted of 636 and 1,950 aligned nucleotide positions of SSU rDNA sequences. The latter contained three times the information content from the same gene (Breglia et al. 2013 and Chan et al. 2013, see also Tab. 1.2). Such dissimilarities could have resulted from differing methodologies concerning the treatment of gaps or from individually arranged alignments and would produce conflicting tree topologies in the end (Lake 1991, Morrison & Ellis 1997). Wong et al. (2008) elegantly demonstrated that even computed alignments are prone to ambiguities and produce diverse datasets, which consequently lead to different tree topologies in the worst case. To minimize ambiguities that could arise already during alignment procedure, and to maximize the recognition of correct homologous positions, alignment of ribosomal DNA sequences was consequently performed according to secondary structure information and helix numbering of *Saccharomyces cerevisiae* provided by Petrov et al. (2013 and 2014) as a precondition for alignment, dataset formation and basis for phylogenetic analyses. Certainly, only molecular

datasets which contain veritable homologous nucleotides can produce reliable results from phylogenetic analyses.

Of course, differing tree topologies that derived from the same gene must not be a result of different gap treatment or alignment size alone, for choice of ingroup and outgroup taxa reflect a methodological approach which is most crucial for phylogenetic analyses. In some cases regarding deeper phylogenies, certain groups could not have been recognized accurately, since a subordinate taxon was represented by solely one sequence or was not included into the alignment at all, which likely resulted in differing sister group relationships and consequently lead to incongruous tree topologies (see Parfrey et al. 2006 for a summary on Excavata). Whether intended or not, some recent SSU rDNA-based phylogenetic studies regarding euglenids utilized ingroup taxa as outgroup (e.g. Diplonemida or Kinetoplastida) and therefore contained no reliable information about the phylogeny of Euglenozoa (Tab. 1.2). To complicate matters, studies which investigated protein-based phylogenies as well as so-called multigene studies utilized poor taxon samplings regarding Euglenozoa, in most cases represented the very derived phototroph *Euglena gracilis* the only euglenid taxon (e.g. Simpson & Roger 2004); of course, such a phylogeny is of limited value, for it reflects only a distorted view on euglenozoan protein evolution, which might be another reason for the tessellated character of actual euglenozoan molecular phylogeny. Therefore, a taxon sampling as extensive as possible and as equilibrated as necessary was used for SSU rDNA-based contrasting juxtaposition analyses in this thesis.

Another methodological problem that produces incongruent tree topologies is the treatment of acquired molecular data, for about half of recent studies concerning gene evolution of Euglenozoa have been conducted without model testing of datasets prior to phylogenetic analyses and therefore should be regarded as unreliable (Tab. 1.2). Since phylogenetic reconstructions can result in incorrect trees when performed under the wrong model (Johnson & Omland 2004, Posada & Crandall 1998), the best-fit model of evolution was calculated for all datasets prior to phylogenetic analyses in the present work.

Most phylogenetic studies concerning the Euglenida have been conducted inferring gene evolution from tree-like reconstructions with different models of sequence evolution applied, though the reliability of reconstruction methods, e.g. Maximum likelihood and Bayesian inference, has not been undisputed for some time (Douady et al. 2003, Pol & Siddall 2001, Simmons et al. 2004, Steel & Penny 2000). Therefore, to circumvent biases that could possibly occur from a unidirectional methodological approach, both Maximum likelihood and

Bayesian inference were utilized for tree reconstructions in this work. In addition, phylogenetic network calculations were successfully combined with analyses of corresponding splits support spectra to test obtained tree topologies and to examine the phylogenetic signals within datasets which could not be pictured by a tree alone.

6.2 Phylogenetic inferences

6.2.1 Long-branching *Entosiphon*

Ribosomal DNA sequences of *Entosiphon* exhibited in almost all results an extraordinary long-branch and emerged as deepest-branching euglenozoan or within the outgroup, except for the LSU rDNA sequence of *Entosiphon sulcatum*, which branched as sister taxon of *Ploeotia costata* in trees and phylogenetic networks, confirmed by results from spectral analysis (Figs. 5.27, 5.29B, 5.31 and 5.32A). Position of the *Entosiphon* clade as deepest-branching euglenozoan was highly questionable, because of its long-branch effect which had been found in other SSU rDNA-based studies as well (Busse et al. 2003, Chan et al. 2013, Lax & Simpson 2013, von der Heyden et al. 2004, Yamaguchi et al. 2012), and due to its remarkable evolutionary divergence estimates (Tabs. 5.1, 5.2 and 5.4). Unfortunately, the class or type of *Entosiphon*'s long-branch effect could not be specified any further, probably due to its extraordinary individual rDNA sequences which for the bigger part are not synapomorphic to any other euglenozoan representative examined in the present work. This divergence might reflect a relatively high degree of signal erosion which would represent a class II effect according to the classification of long-branch effects proposed by Wägele & Mayer (2007). Nonetheless, there are compelling morphological and physiological aspects confirming the notion that a deep-branching of *Entosiphon* within Euglenozoa is unacceptable.

Euglenozoan feeding apparatus

The first reason is revealed by *Entosiphon*'s elaborate feeding apparatus, which consists of four vanes and two or three rods present in *Entosiphon applanatum* and *E. sulcatum* (Triemer & Farmer 1991): the assumption that the most complex ingestion device among euglenids could be primordial to a comparably simple cytopharynx like the MTR/pocket present in early-branching Petalomonadida (and Bodonids), or to a feeding apparatus structured by vanes and supportive rods which is present in most other phagotrophic euglenids, would

command imparsimonious multiple losses and reoccurrences of complex substructures. Results obtained by reiterated analyses without *Entosiphon* pointed to another possible evolutionary trend within Euglenozoa regarding their ingestion device: from a non-rod-bearing MTR/pocket type present in Petalomonadida (Triemer & Farmer 1991, therein 'Type I') to a rod-bearing type owned by *Ploeotia costata* (Leander et al. 2001), *Keelungia pulex* (Chan et al. 2013), Diplonemida (Roy et al. 2007, Triemer & Ott 1990), symbiontid *Bihospites bacati* (Breglia et al. 2010), *Entosiphon applanatum* (Triemer & Farmer 1991), *Entosiphon sulcatum* (Triemer & Fritz 1987), and phagotrophic Helicales like *Peranema trichophorum* (Nisbet 1974), *Dinema sulcatum* (Triemer & Farmer 1991), *Heteronema scaphurum* (Breglia et al. 2013), and probably *Rapaza viridis* (Yamaguchi et al. 2012, see 6.2.4). This most parsimonious scenario would assume the unique occurrence of a MTR/pocket-like feeding apparatus and an evolutionary trend in increasing complexity of substructures, combined with a loss and a reduction: lost in the last common ancestor of Aphagea, i.e. primary osmotrophic euglenids, and reduced in the last common ancestor of Euglenea (i.e. phototrophic euglenids including secondary osmotrophic forms, not *Rapaza viridis*) which own ingestion apparatus that are reduced to the MTR/pocket type (Shin et al. 2002, Surek & Melkonian 1986).

Paramylon

A rarely considered character is the possession of paramylon, a beta-1,3-glucan carbohydrate (Bäumer et al. 2001, Kiss et al. 1987), which is commonly shared by representatives of the euglenid crown group possessing helical pellicle striations (Leander et al. 2001, therein clade G). Although *Entosiphon* bears a longitudinal pellicle strip organization, which represents a plesiomorphic character state in euglenids according to Leander & Farmer (2001), it is known to possess paramylon (Vollmer and Preisfeld, unpublished data). Thus, an early branching of *Entosiphon* would demand convergent evolvement of paramylon in a presumably primordial *Entosiphon* lineage and additionally within derived Helicales, which would be most imparsimonious, for a physiological accouterment involves not only presence of a storage carbohydrate, but many more associated, specialized enzymes (Bäumer et al. 2001). These considerations, together with findings from SSU and LSU rDNA analyses, render a primordial position of *Entosiphon* within the Euglenozoa invalid. The possession of paramylon, a rod-bearing ingestion device and a longitudinal pellicle strip organization indicate that *Entosiphon* is more realistically situated between Ploeotiida, which also own longitudinal pellicle strips, a rod-bearing feeding apparatus, but do not contain paramylon,

and the Helicales, which phagotrophic representatives also possess a rod-bearing feeding apparatus and contain paramylon. This derivation of compatible morphological characters is strongly confirmed by results obtained from phylogenetic analyses of LSU rDNA and the nucleotide reduction in deduced secondary structure of SSU rRNA helix 44, which *Entosiphon* exclusively shares with Helicales, as well as the presence of paramylon.

6.2.2 Most recent common ancestor of Euglenozoa

Recent phylogenetic studies considering representatives of all euglenozoan lineages mainly aimed to clarify the identity of newly described species, rather than investigating euglenozoan sister group relationships, they employed imbalanced taxon samplings, utilized subordinate euglenozoan taxa as improper outgroup or even included no outgroup at all (e.g. Lee & Simpson 2014a and see Tab. 1.2). In this study, monophyly of Euglenozoa was found strongly supported in results from SSU and LSU rDNA analyses, concurrent with findings from other studies based on SSU rDNA (e.g. Cavalier-Smith 2004 and all in Tab. 1.2).

After exclusion of *Entosiphon* from datasets, the Petalomonadida emerged as recent representatives of a putative euglenozoan common ancestor with maximized support in contrasted juxtaposition of SSU rDNA and in results from LSU rDNA-based phylogenetic analyses. Only few studies based on SSU rDNA data recognized *Ploeotia costata* as deepest branching euglenozoan together with *Petalomonas cantuscygni* (Busse & Preisfeld 2003c, Fig. 2 in von der Heyden et al. 2004). But these studies employed imbalanced taxon samplings regarding Euglenozoa, where subsidiary lineages superimposed each other two- to threefold. In fact, far more studies based on SSU rDNA data already have identified a representative of the Petalomonadida, i.e. *Notosolenus ostium*, *Petalomonas cantuscygni* or both, respectively, as deepest-branching euglenozoan (Breglia et al. 2010, Cavalier-Smith and Nikolaev 2008, Lara et al. 2006, Moreira et al. 2004, Yubuki et al. 2009). Interestingly, this important finding was regarded with caution and suspected to be caused by imbalanced taxon sampling or interpreted as an effect of high nucleotide sequence divergence among euglenids or was sometimes not mentioned at all. Many other studies found a representative of Petalomonadida as deepest-branching taxon of monophyletic Euglenida, though this monophyly received weak to no statistical support, whether containing reliable (Busse & Preisfeld 2002a, 2002b and 2003b, Moreira et al. 2004) or rather unreliable results regarding euglenozoan SSU rDNA genealogy constituted by choice of ingroup taxa and applied

methodology (Breglia et al. 2013, Lax & Simpson 2013, Preisfeld et al. 2001, Fig. 1 in von der Heyden et al. 2004).

Furthermore, thorough phylogenetic network and spectral analyses of SSU and LSU rDNA in this work revealed that the monophyly of Euglenida received neither splits support nor any phylogenetic signal, which firmly corroborates earlier findings (Busse & Preisfeld 2003a) and has dire consequences for ribosomal gene evolution of Euglenozoa (see 6.2.3). Another study which based on hsp90 gene sequences also identified *Petalomonas cantuscygni* as earliest branching euglenozoan (Breglia et al. 2007). In accordance to these findings, Euglenida appear to be polyphyletic, i.e. the allegedly monophyletic taxon Euglenida represents an assemblage of primordial and derived euglenozoan groups which have no common ancestor, and Petalomonadida are primordial to all other Euglenozoa in most SSU and LSU rDNA genealogies. As a consequence Diplonemida, Kinetoplastida as well as Symbiontida (presumably also Ploeotiida) and Helicales are monophyletic euglenozoan lineages that altogether derived from putatively phagotrophic euglenid ancestors, which are nowadays represented by the Petalomonadida. Similarities between Petalomonadida and Diplonemida found in deduced secondary structure elements (5.2.3) and variable regions of SSU rDNA (5.2.4) as well as in length comparison of ITS2 (5.4.3) and LSU rDNA sequences (5.4.2) corroborate this conclusion. Moreover, compelling results from other studies also support these findings, i.e. similarities in morphological characters like pellicle organization and structural composition of the feeding apparatus (Leander et al. 2001 and 2007) as well as occurrence of the elongation factor-like protein instead of EF-1α in *Petalomonas* and Diplonemida (Gile et al. 2009).

6.2.3 Major group relationships of Euglenozoa

The complex inter-relationships of euglenozoan lineages were disputed long before erection of the Symbiontida and all possible sister group variants have been produced in preceding studies by applying diverse methodologies and choosing various genetic markers and taxon samplings (Triemer and Farmer 2007). According to the classification *sensu* Adl et al. (2012), the Euglenozoa actually embrace four monophyletic major groups of equal rank: Euglenida, Kinetoplastida, Diplonemida and Symbiontida, but this grouping of Euglenozoa became outdated, since symbiontids have been classified as Euglenida in a study based on morphological data (Yubuki et al. 2013).

Diplonemida is not the sister taxon of Kinetoplastida

Some authors found Diplonemida and Kinetoplastida to be sister groups to the exclusion of Euglenida, but this opinion was not undisputed (Marande et al. 2005). For instance, Simpson & Roger (2004) hypothesized a sister group relationship of diplonemids and kinetoplastids found in phylogenies of heat shock proteins, but the derived phototroph *Euglena gracilis* represented the only euglenid taxon in these analyses, which is an unacceptably poor taxon sampling for an inference of euglenozoan major group relationships. In the same study SSU rDNA-based phylogenies included two phototrophic and one derived phagotrophic euglenid, which involved the very same problem. A zoological example to elucidate this matter: when investigating phylogenetic relationships of tortoises and crocodiles, the addition of a bird to the dataset will be of no use (unless regarding dinosaurs and their fossil record). In another case Makiuchi et al. (2011) hypothesized the compartmentalization of a glycolytic enzyme as synapomorphic feature which united kinetoplastids and diplonemids, but again the derived phototrophic *Euglena gracilis* was the only representative of Euglenida, i.e. more primordial euglenid representatives need to be tested before conclusions concerning sister group relationships of major euglenozoan groups become reliable. Strikingly, a sister group relationship of Diplonemida and Kinetoplastida was clearly rejected in the summary of all results in the present work: it was at most weakly supported in SSU rDNA-based tree reconstructions, inexistent in all phylogenetic networks as well as corresponding split support spectra, and moreover it was neglected in results from analyses of base composition (5.2.1 and 5.4.1), identity matrix (5.2.2), SSU rDNA variable region length (5.2.4), LSU rDNA sequence length (5.4.2), ITS2 sequence length (5.4.3) and deduced secondary structure analyses of ribosomal DNA sequences (5.2.3 and 5.4.3).

Six major groups of Euglenozoa

The existence of a sister group relationship of Diplonemida and Kinetoplastida could have made paraphyly of Euglenida imaginable, but since Petalomonadida (and probably also Ploeotiida) are close relatives of Diplonemida rather than Kinetoplastida, six monophyletic major groups constitute the Euglenozoa: Petalomonadida, Diplonemida, Kinetoplastida, Symbiontida, Ploeotiida and Helicales. Note that Ploeotiida presumably include *Keelungia pulex* and *Entosiphon sulcatum* here, which would be concordant with Adl et al. (2012). Cognition of the polyphyletic nature of phagotrophic euglenids plays a key role in understanding the complex phylogeny of Euglenozoa.

6.2.4 Phagotrophic euglenids are polyphyletic

The assemblage ‘phagotrophic euglenids’ comprises, as indicated by Adl et al. (2012), monophyletic lineages, i.e. Petalomonadida and Ploeotiida, but “many traditional genera are probably polyphyletic”. Anisonemida were also found to be monophyletic in SSU rDNA-based tree reconstructions, phylogenetic networks and split support spectra (5.1.2). Nonetheless, aforementioned monophyla do not share a common ancestor, since Petalomonadida and Ploeotiida diverge separately prior to Helicales. In addition, Aphagea and Euglenea as well as representatives of several other phagotrophic genera branch within Helicales, e.g. *Dinema*, *Heteronema*, *Neometanema* and *Peranema*, not to mention the presumed mixotroph *Rapaza viridis*. The latter shares a common ancestor with Euglenea, while Aphagea branch amidst several phagotrophic Helicales, the exact positions of which are not secured yet.

Notwithstanding, tree reconstructions and phylogenetic network analyses confirmed that *Peranema* is affiliated with *Heteronema* and that *Dinema* and Anisonemida share a common ancestor (5.1.1 and 5.1.3). Variable region length of *Heteronema scaphurum* strongly implies a relation to Aphagea (5.2.4), though results from base composition analyses favor *Peranema* to share a common ancestor with Aphagea (5.2.1). However, the common ancestor of the *Rapaza viridis*/Euglenea clade is not identical to the common ancestor of Aphagea/*Heteronema* (or even Aphagea/*Peranema*). Furthermore, as close relatives of Diplonemida, the Ploeotiida (including *Keelungia* and *Entosiphon*) are primordial to Helicales. Finally, Petalomonadida are primordial to all Euglenozoa, diverging in a stem line separate from that of Ploeotiida or Helicales (compare Figs. 5.2 and 5.42). Consequently, these findings constitute the polyphyly of the ‘phagotrophic euglenids’ and implicate inferences for euglenozoan taxonomy.

6.3 Taxonomic implications

Leedale (1967) classified phagotrophic euglenids into two groups, the Heteronematina, which own a specialized ingestion apparatus, and the Sphenomonadina, with a simple ingestion apparatus. Heteronematina *sensu* Adl et al. (2012) included all phagotrophic genera, but it was noted that “there is no phylogenetic taxonomy for phagotrophic euglenids as a whole”. Since phagotrophic euglenids emerge prior to and in between non-euglenid monophyla like

Diplonemida and Kinetoplastida as well as non-phagotrophic euglenid monophyla diverging in a monophyletic crown group, i.e. Aphagea and Euglenea, the term ‘phagotrophic euglenids’ describes a polyphylum. Consequently from a taxonomic point of view, the term ‘phagotrophic euglenids’ is invalid as are the terms ‘Heteronematina’ and ‘Sphenomonadina’ (including related terms with other endings, e.g. -ida, -idea, -ales), for the latter would include *Anisonema*, which already has been shown to belong to the euglenid crown group (Busse et al. 2003, Leander et al. 2001).

Ambiregnal Helicales

This euglenid crown clade, which unites euglenid taxa that share the autapomorphy of a helical pellicle, appeared earlier in many molecular studies (Busse & Preisfeld 2002a and 2002b, Müllner et al. 2001, Preisfeld et al. 2000 and 2001), whether without being named or with different names. It was termed ‘clade G’ in a study based on morphological characters (Leander et al. 2001), in a molecular study based on SSU rDNA data it was named ‘clade H’ (Busse et al. 2003) and recently recurred as ’HP grouping’ or ‘HP clan’ (Lee & Simpson 2014a and 2014b). In the present work, denomination of this clade as ‘Helicales’ promotes a better understanding of euglenid phylogeny, for it provides a unique name for a naturally evolved clade comprising derived phagotrophic euglenids, primary osmotrophic euglenids and phototrophic euglenids (including secondary osmotrophic taxa) which all share the helical pellicle as a distinct morphological character, i.e. an autapomorphy, that is reflected by the name chosen. The rather inappropriate denominations ‘clade G’, ‘clade H’, ‘HP grouping’ and ‘HP clan’ are therefore converted into the clade name ‘Helicales’ at least within this work.

Phylogenetic (apomorphy-based) diagnosis

Euglenozoa CAVALIER-SMITH 1981

Helicales *taxon nov.* PAERSCHKE & PREISFELD 2015

Natural clade comprising derived phagotrophic euglenids, primary osmotrophic Aphagea, mixotrophic *Rapaza viridis*, and primary phototrophic Euglenea including secondary osmotrophic euglenids, characterized by a euglenid-specific pellicle with primary helical organization (which can be secondarily altered, e.g. in phototrophic euglenids with lorica) as inferred from analyses of nuclear ribosomal gene sequences.

Apomorphy

Helical pellicle.

The Euglenozoa represent an ambiregnal group, partly present in the International Code of Zoological Nomenclature (ICZN) as well as in the International Code of Nomenclature for algae, fungi, and plants (ICN), and this ambiregnal status has been discussed critically for a long time (Lahr et al. 2012, Patterson & Larsen 1992). Helicales would also receive an ambiregnal status, for they include phagotrophic, osmotrophic and phototrophic euglenids, therefore it remains to be seen whether (or when) the Helicales will find acceptance. Alternatively the Helicales could constitute a ribogroup according to Adl et al. (2012). An interesting approach was presented with the PhyloCode (International Code of Phylogenetic Nomenclature), which current version is specifically designed to regulate the naming of clades rather than species. Since the term 'Helicales' describes a natural clade representing the crown group of Euglenozoa as inferred from ribosomal gene evolution, it could be coherently nominated as *nomen cladi conversum* in terms of the PhyloCode (Article 9, Version 4c, 2010).

Symbiontida

The euglenid monophylum Symbiontida has been described as novel euglenozoan subclade consisting of uncharacterized cells living in low-oxygen environments (Yubuki et al. 2009). Hence rod-shaped epibiotic bacteria represent the corresponding apomorphy for this taxon, the name 'Symbiontida' could provoke a connotation in the sense of symbiosis, which could be intended, but would be somewhat misleading, for phagotrophic and phototrophic euglenids are known to be also associated with bacteria, e.g. endobiotic bacteria in *Petalomonas sphagnophila* (Kim et al. 2010, Schnepf et al. 2002), rod-shaped bacteria on *Euglena helicoideus* (Leander & Farmer 2000) and ecto- and endobiotic bacteria on *Eutreptiella* sp. (Kuo & Lin 2013). Since those euglenids, which represent descendants from completely different clades than Symbiontida, exhibited also a not yet fully understood, but nonetheless close relationship with bacteria, the name 'Symbiontida' would in a strict sense not describe a valid apomorphy for this clade, for it would also include the aforementioned phagotrophic and phototrophic euglenids. Originally the name 'Anox clade' was used for this group in an early molecular study (Zuendorf et al. 2006) and as those organisms exclusively (for euglenids at least) live in anoxic or suboxic habitats, perhaps the denomination 'Anoxida' would better reflect the more fitting apomorphy 'anoxic'.

6.4 Future prospects

The present work sought to combine established approaches with new ideas to shed new light on the phylogeny of Euglenozoa, but a lot of work still needs to be done. For instance, analysis of euglenozoan LSU rDNA is still in its infancy because the present taxon sampling marks only the beginning of LSU rDNA genealogy of Euglenozoa. Novel LSU rDNA sequences especially of key taxa such as *Keelungia pulex* or *Rapaza viridis* and of genera *Dinema*, *Entosiphon*, *Ploeotia*, *Heteronema* and *Neometanema* would greatly amend the molecular phylogeny of Euglenozoa. Furthermore, representatives of some genera still await sequencing, for they have been examined only morphologically by now, e.g. *Anehmia*, *Atraktomonas*, *Bordnamonas*, *Calycimonas*, *Dolium*, *Dylakosoma*, *Jenningsia*, *Peranemopsis*, *Scytomonas*, *Sphenomonas*, *Tropidoscyphus* , *Urceolopsis* and *Urceolus*. Though secondary structure analysis seemed to be out of fashion, a recent study found that evolutionary rates vary among ribosomal RNA structural elements (Smit et al. 2007), which could be a stepping stone to liven up this field of research also beyond the Euglenozoa.

7 References

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S *et al.* (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. Journal of Eukaryotic Microbiology 52: 399–451.

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V *et al.* (2012) The revised classification of eukaryotes. Journal of Eukaryotic Microbiology 59: 429–514.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J, Guhl F, Arriaza B, Renier C, Wittmers LE, Fornaciari G, Allison M (2004) A 9,000-year record of Chagas' disease. Proceedings of the National Academy of Science USA 101(7): 2034–2039.

Bäumer D, Preisfeld A, Ruppel HG (2001) Isolation and characterization of paramylon synthase from *Euglena gracilis* (Euglenophyceae). Journal of Phycology 37: 38–46

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290: 972–977.

Behnke A, Bunge J, Barger K, Breiner HW, Alla V, Stoeck T (2006) Microeukaryote community patterns along an O2/H2S gradient in a supersulfidic anoxic Fjord (Framvaren, Norway). Applied and Environmental Microbiology 72(5):3626–3636.

Belhadri A, Bayle D, Brugerolle G (1992) Biochemical and immunological characterization of intermicrotubular cement in the feeding apparatus of phagotrophic euglenoids: *Entosiphon*, *Peranema*, and *Ploeotia*. Protoplasma 168: 113–124.

Boenigk J & Arndt H (2002) Bacterivory by heterotrophic flagellates: community structure and feeding strategies. Antonie van Leeuwenhoek 81: 465–480.

Breglia SA, Slamovits CH, Leander BS (2007) Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from hsp90 gene sequences. Journal of Eukaryotic Microbiology 54: 86–92.

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

Breglia SA, Yubuki N, Leander BS (2013) Ultrastructure and molecular phylogenetic position of *Heteronema scaphurum*: a eukaryovorous euglenid with a cytoproct. Journal of Eukaryotic Microbiology 60: 107–120.

Brosnan S, Shin W, Kjer KM & Triemer RE (2003) Phylogeny of the photosynthetic euglenophytes inferred from the nuclear SSU and partial LSU rDNA. International Journal of Systematic and Evolutionary Microbiology 53: 1175–1186.

Brown MW, Silberman JD, Spiegel FW (2012) A contemporary evaluation of the acrasids (Acrasidae, Heterolobosea, Excavata). European Journal of Protistology 48: 103–123.

Bryant D & Moulton V (2004) NeighborNet: an agglomerative algorithm for the construction of planar phylogenetic networks. Molecular Biology and Evolution 21: 255–265.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. PloSone 2: e790

Bütschli O (1884) Protozoa. In: Bronn's Klassen und Ordnungen des Thier-Reichs 2, Abt. II Mastigophora. Bronn HG (Hrsg.), C.F. Winter'sche Verlagsbuchhandlung, Leipzig, S. 617-1097

Busse I (2003) Die SSU rDNA und rRNA der Euglenida: Phylogenetische Analyse und molekulare Evolution. (Dissertation) Universität Bielefeld, Germany.

Busse I, Patterson DJ, 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 (2002a) Phylogenetic position of *Rhynchopus sp*. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. Gene 284: 83–91.

Busse I & Preisfeld A (2002b) Unusually expanded SSU ribosomal DNA of primary osmotrophic euglenids: molecular evolution and phylogenetic inference. Journal of Molecular Evolution 55: 757–767.

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

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

Busse I & Preisfeld A (2003c) Discovery of a group I intron in the SSU rDNA of *Ploeotia costata* (Euglenozoa). Protist 154: 57–69.

Cachon J, Cachon M, Cosson M-P, Cosson J (1988) The paraflagellar rod: a structure in search of a function. Biology of the Cell 63: 169–181.

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 (1981) Eukaryote kingdoms: seven or nine? BioSystems 14: 461–481.

Cavalier-Smith T (1993) Kingdom Protozoa and its 18 phyla. Microbiological Reviews 57: 953–994.

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

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

Cavalier-Smith T (2003) The excavate protozoan phyla Metamonada GRASSE emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa. International Journal of Systematic and Evolutionary Microbiology 53: 1741–1758.

Cavalier-Smith T (2004) Only six kingdoms of life. Proceedings of the Royal Society Biological Sciences 271: 1251–1262.

Cavalier-Smith T & Nikolaev S (2008) The zooflagellates *Stephanopogon* and *Percolomonas* are a clade (class Percolatea: phylum Percolozoa). Journal of Eukaryotic Microbiology 55: 501–509.

Chan Y, Moestrup O, 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: 15–31.

Chen M, Chen F, Yu Y, Ji J, Kong F (2008) Genetic diversity of eukaryotic microorganisms in Lake Taihu, a large shallow subtropical lake in China. Microbial Ecology 56(3): 572–583.

Ciugulea I, Nudelman MA, Brosnan S, Triemer RE (2008) Phylogeny of the euglenoid loricate genera *Trachelomonas* and *Strombomonas* (Euglenophyta) inferred from nuclear SSU and LSU rDNA. Journal of Phycology 44: 406–418.

Clark CG & Cross GAM (1987) rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. Molecular and Cellular Biology 7(9): 3027–3031.

Countway PD, Gast RJ, Dennett MR, Savai P, Rose JM, Caron DA (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). Environmental Microbiology 9(5): 1219–1232.

DeGraaf RM, Duarte I, van Alen TA, Kuiper JWP, Schotanus K, Rosenberg J, Huynen MA, Hackstein JHP (2009) The hydrogenosomes of *Psalteriomonas lanterna*. BMC Evolutionary Biology 9: 287.

Derelle R & Lang BF (2012) Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Molecular Biology and Evolution 29: 1277–1289.

Deschamps P, Lara E, Marande W, López-García P, Ekelund F, Moreira D (2011) Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids. Molecular Biology and Evolution 28(1): 53–58.

Doležel D, Jirků M, Maslov DA, Lukeš J (2000) Phylogeny of the bodonid flagellates (Kinetoplastida) based on small-subunit rRNA gene sequences. International Journal of Systematic and Evolutionary Microbiology 50(5): 1943–1951.

Dooijes D, Chaves I, Kieft R, Dirks-Mulder A, Martin W, Borst P (2000) Base J originally found in Kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. Nucleic Acids Research 28(16): 3017–3021.

Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP (2003) Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Molecular Biology and Evolution 20: 248–254.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C *et al.* (2005) Comparative genomics of trypanosomatid parasitic Protozoa. Science 309: 404–409.

Fenchel T (1993) There are more small than large species? Oikos 68(2): 375–378.

Fenchel T, Bernard C, Esteban G, Finlay BJ, Hansen PJ, Iversen N (1995) Microbial diversity and activity in a Danish Fjord with anoxic deep water. Ophelia 43(1): 45–100.

Flavin M & Nerad TA (1993) *Reclinomonas americana* n.g., n.sp., a new freshwater heterotrophic flagellate. Journal of Eukaryotic Microbiology 40: 172–179.

Frantz C, Ebel C, Paulus F, Imbault P (2000) Characterization of *trans*-splicing in euglenoids. Current Genetics 37: 349–355.

Garrison DL & Buck KR (1989) The biota of Antarctic pack ice in the Weddell Sea and Antarctic peninsula regions. Polar Biology 10: 211–219.

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Canadian Journal of Botany 56(22): 2883–2889.

Gile GH, Faktorova D, Castlejohn CA, Burger G, Lang BF, Farmer MA, Lukeš J, Keeling PJ (2009) Distribution and phylogeny of EFL and EF-1alpha in Euglenozoa suggest ancestral co-occurrence followed by differential loss. PLoS one 4: e5162

Gommers-Ampt JH, Van Leeuwen F, de Beer ALJ, Vliegenthart JFG, Dizdaroglu M, Kowalak JA, Crain PF, Borst P (1993) B-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. Cell 75: 1129–1136.

Greenwood SJ, Schnare MN, Cook JR, Gray MW (2001) Analysis of intergenic spacer transcripts suggests ‘read-around’ transcription of the extrachromosomal circular rDNA in *Euglena gracilis*. Nucleic Acids Research 29(10): 2191–2198.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proceedings of the National Academy of Science USA 106: 3859–3864.

Harding T, Brown MW, Plotnikov A, Selivanova E, Park JS, Gunderson JH, Baumgartner M, Silberman JD, Roger AJ, Simpson AGB (2013) Amoeba stages in the deepest branching heteroloboseans, including *Pharyngomonas*: evolutionary and systematic implications. Protist 164: 272–286.

He D, Fiz-Palacios O, Fu C-J, Fehling J, Tsai C-C, Baldauf SL (2014) An alternative root for the eukaryote tree of life. Current Biology 24: 465–470.

Honigberg BM (1963) Evolutionary and systematic relationships in the flagellate order Trichomonadida KIRBY. Journal of Protozoology 10: 20–63.

Huelsenbeck JP & Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.

Huson DH (1998) SplitsTree: a program for analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.

Huson DH & Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254–267.

Jebaraj CS, Raghukumar C, Behnke A, Stoeck T (2010) Fungal diversity in oxygen-depleted regions of the Arabian Sea revealed by targeted environmental sequencing combined with cultivation. FEMS Microbiology Ecology 71(3): 399–412.

Johnson JB & Omland KS (2004) Model selection in ecology and evolution. Trends in Ecology and Evolution 19(2): 101–108.

Kadereit JW, Körner C, Kost B, Sonnewald U (2014) Strasburger – Lehrbuch der Pflanzenwissenschaften. Springer Verlag, Berlin, Heidelberg.

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: 1647–1649.

Kim E, Park JS, Simpson AGB, Matsunaga S, Watanabe M, Murakami A, Sommerfeld K, Onodera NT, Archibald JM (2010) Complex array of endobionts in *Petalomonas sphagnophila*, a large heterotrophic euglenid protest from *Sphagnum*-dominated peatlands. ISME Journal 4(9): 1108–1120.

Kim JI, Shin W, Triemer RE (2013) Cryptic speciation in the genus *Cryptoglena* (Euglenaceae) revealed by nuclear and plastid SSU and LSU rRNA gene. Journal of Phycology 49: 92–102.

Kiss JZ, Vasconcelos AC, Triemer RE (1987) Structure of the euglenoid storage carbohydrate, paramylon. American Journal of Botany 74(6): 877–882.

Kivic PA & Walne PL (1984) An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. Origins of Life 13: 269–288.

Kumar S & Gadagkar SR (2001) Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158: 1321–1327.

Kuo RC & Lin S (2013) Ectobiotic and endobiotic bacteria associated with Eutreptiella sp. Isolated from Long Island sound. Protist 164(1): 60–74.

Lahr DJG, Lara E, Mitchell EAD (2012) Time to regulate microbial eukaryote nomenclature. Biological Journal of the Linnean Society 107: 469–476.

Lake JA (1991) The order of sequence alignment can bias the selection of tree topology. Molecular Biology and Evolution 8: 378–385.

Lara E, Chatzinotas A, Simpson AGB (2006) *Andalucia* (n. gen.)–the deepest branch within jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil. Journal of Eukaryotic Microbiology 53: 112–120.

Lara E, Mitchell EA, Moreira D, López-García P (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. Protist 162: 14–32.

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

Lax G & Simpson AGB (2013) Combining molecular data with classical morphology for uncultured phagotrophic euglenids (Excavata): a single-cell approach. Journal of Eukaryotic Microbiology 60: 615–625.

Leander BS & Farmer MA (2000) Epibiotic bacteria and a novel pattern of strip reduction on the pellicle of *Euglena helicoideus* (BERNARD) LEMMERMANN. European Journal of Protistology 36(4): 405–413.

Leander BS & Farmer MA (2001) Comparative morphology of the euglenid pellicle. II. Diversity of strip substructure. Journal of Eukaryotic Microbiology 48: 202–217.

Leander BS, Triemer RE, Farmer MA (2001) Character evolution in heterotrophic euglenids. European Journal of Protistology 37: 337–356.

Leander BS, Esson HJ, Breglia SA (2007) Macroevolution of complex cytoskeletal systems in euglenids. BioAssays 29: 987–1000.

Lee WJ & Simpson AGB (2014a) Ultrastructure and molecular phylogenetic position of *Neometanema parovale* sp. nov. (*Neometanema* gen. nov.), a marine phagotrophic euglenid with skidding motility. Protist 165: 452–472.

Lee WJ & Simpson AGB (2014b) Morphological and molecular characterization of *Notosolenus urceolatus* LARSEN AND PATTERSON 1990, a member of an understudied deep-branching euglenid group (petalomonads). Journal of Eukaryotic Microbiology 61: 463–479.

Leedale GF (1967) Euglenoid flagellates. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Linton EW, Hittner D, Lewandowski C, Auld T, Triemer RE (1999) A molecular study of euglenoid phylogeny using small subunit rDNA. Journal of Eukaryotic Microbiology 46: 217–223.

Linton EW, Nudelman MA, Conforti V, Triemer RE (2000) A molecular analysis of the euglenophytes using SSU rDNA. Journal of Phycology 36: 740–746.

López-García P, Philippe H, Gail F, Moreira D (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. Proceedings of the National Academy of Science USA 100(2): 697–702.

López-García P, Rodríguez-Valera F, Pedrós-Alió, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409: 603–607.

López-García P, Vereshchaka A, Moreira D (2006) Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. Environmental Microbiology 9(2): 546–554.

Makiuchi T, Annoura T, Hashimoto M, Hashimoto T, Aoki T, Nara T (2011) Compartmentalization of a glycolytic enzyme in *Diplonema*, a non-kinetoplastid Euglenozoan. Protist 162: 482–489.

Marande W, Lukeš J, Burger G (2005) Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. Eukaryot Cell 4: 1137–1146.

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: 99–145.

Maruyama S & Nozaki H (2007) Sequence and intranuclear location of the extrachromosomal rDNA plasmidof the amoeba-flagellate *Naegleria gruberi*. Journal of Eukaryotic Microbiology 54(4): 333–337.

Maslov DA, Yasuhira S, Simpson L (1999) Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. Protist 150: 33–42.

May RM (1988) How many species are there on Earth? Science 241: 1441–1449.

Meyer H (1968) The fine structure of the flagellum and the kinetoplast chondriome of *Trpanosoma* (*Schizotrypanum*) *cruzi* in tissue culture. Journal of Protozoology 15(3): 614–621.

Montegut-Felkner AE & Triemer RE (1994) Phylogeny of *Diplonema ambulator* (Larsen and Patterson): 1. Homologies of the flagellar apparatus. European Journal of Protistology 30(2): 227–237.

Montegut-Felkner AE & Triemer RE (1996) Phylogeny of *Diplonema ambulator* (Larsen and Patterson): 2. Homologies of the feeding apparatus. European Journal of Protistology 32: 64–76.

Montegut-Felkner AE & Triemer RE (1997) Phylogenetic relationships of selected euglenoid genera based on morphological and molecular data. Journal of Phycology 33: 512–519.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biol 9(8): e1001127.

Moreira D & López-García P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. Trends in Microbiology 10(1): 31–38.

Moreira D, López-García P, Rodríguez-Valera F (2001) New insights into the phylogenetic position of diplonemids: G+C content bias, differences of evolutionary rate and a new environmental sequence. International Journal of Systematic and Evolutionary Microbiology 51: 2211–2219.

Moreira D, López-García P, Vickerman K (2004) An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. International Journal of Systematic and Evolutionary Microbiology 54: 1861–1875.

Morrison DA & Ellis JT (1997) Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. Molecular Biology and Evolution 14: 428–441.

Müllner AN, Angeler DG, Samuel R, Linton EW, Triemer RE (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.

Nei M & Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford, New York.

Nicholas KB, Nicholas HB, JR., Deerfield DW (1997) GeneDoc. Analysis and visualization of genetic variation. http://www.nrbsc.org/gfx/genedoc/

Nisbet B (1974) An ultrastructural study of the feeding apparatus of *Peranema trichophorum*. Journal of Protozoology 21: 39–48.

O'Kelly CJ & Nerad TA (1999) *Malawimonas jakobiformis* n. gen., n. sp. (Malawimonadidae n. fam.): A *Jakoba*-like heterotrophic nanoflagellate with discoidal mitochondrial cristae. Journal of Eukaryotic Microbiology 46: 522–531.

Orsi W, Song YC, Hallam S, Edgcomb V (2012) Effect of oxygen minimum zone formation on communities of marine protists. ISME Journal 6: 1586–1601.

Page FC & Blanton RL (1985) The Heterolobosea (Sarcodina: Rhizopoda), a new class uniting the Schizopyrenida and the Acrasidae (Acrasida). Protistologica 21: 121–132.

Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson DJ, Katz LA (2006) Evaluating support for the current classification of eukaryotic diversity. PloS Genetics 2: e220.

Park JS, de Jonckheere JF, Simpson AGB (2012) Characterization of *Selenaion koniopes* n. gen., n. sp., an amoeba that represents a new major lineage within Heterolobosea, isolated from the Wieliczka salt mine. Journal of Eukaryotic Microbiology 59: 601–613.

Patterson DJ (1999) The diversity of eukaryotes. American Naturalist 154: S96-S124.

Patterson DJ & Larsen J (1992) A perspective on protistan nomenclature. The Journal of Protozoology 39(1): 125-131.

Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, Stepanov VG, Gaucher EA, Grover MA, Harvey SC *et al.* (2013) Secondary structure and domain architecture of the 23S and 5S rRNAs. Nucleic Acids Research 41:7522–7535.

Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, Williams LD (2014) Secondary structures of rRNAs from all three domains of life. PlOS One 9:e88222.

Pol D & Siddall ME (2001) Biases in Maximum likelihood and parsimony: a simulation approach to a 10-taxon case. Cladistics 17:266–281.

Posada D & Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Posada D (2008) jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256.

Preisfeld A, Berger S, Busse I, Liller S, Ruppel HG (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, Talke S, Ruppel HG (2001) Phylogenetic position and inter-relationships 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.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: a comprehensive online resource for quality-checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188–7196.

Rat'kova TN, Sazhin AF, Kosobokova KN (2004) Unicellular inhabitants of the White Sea underice pelagic zone during the early spring period. Oceanology 44(2): 240–246.

Ravel-Chapuis P (1988) Nuclear rDNA in *Euglena gracilis*: paucity of chromosomal units and replication of extrachromosomal units. Nucleic Acids Research 16(11): 4801–4810.

Roy J, Faktorová D, Benada O, Lukeš J, Burger G (2007) Description of *Rhynchopus euleeides* n. sp. (Diplonemea), a free-living marine euglenozoan. Journal of Eukaryotic Microbiology 54: 137–145.

Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, Cavalier-Smith T, Guiry MD, Kirk PM (2015) A higher level classification of all living organisms. PLoS one 10(4): e0119248.

Saburova M, Al-Yamani F, Polikarpov I (2009) Biodiversity of free-living flagellates in Kuwait's intertidal sediments. Biodiversity and Ecosystem Risk Assessment 3: 97–110.

Salani FS, Arndt H, Hausmann K, Nitsche F, Scheckenbach F (2012) Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. ISME Journal 6: 713–723.

Sauvadet AL, Gobet A, Guillou L (2010) Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific hydrothermal habitats. Environmental Microbiology 12: 2946–2964.

Scheckenbach F, Hausmann K, Wylezich C, Weitere M, Arndt H (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. Proceedings of the National Academy of Sciences of the United States of America 107(1): 115–120.

Scheckenbach F, Wylezich C, Mylnikov AP, Weitere M, Arndt H (2006) Molecular comparisons of freshwater and marine isolates of the same morphospecies of heterotrophic flagellates. Applied and Environmental Microbiology 72: 6638–6643.

Schnare MN, Cook JR, Gray MW (1990) Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis*. Journal of Molecular Biology 215: 85–91.

Schnare MN & Gray MW (1990) Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. Journal of Molecular Biology 215: 73–83.

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

Schönborn W, Dörfelt H, Foissner W, Krienitz L, Schäfer U (1999) A fossilized microcenosis in Triassic amber. Journal of Eukaryotic Microbiology 46(6): 571–584.

Shin W, Brosnan S, Triemer RE (2002) Are cytoplasmic pockets (MTR/pocket) present in all photosynthetic euglenoid genera? Journal of Phycology 38: 790–799.

Simmons MP, Pickett KM, Miya M (2004) How meaningful are Bayesian support values? Molecular Biology and Evolution 21: 188–199.

Simpson AGB (1997) The identity and composition of the Euglenozoa. Archiv der Protistenkunde 148: 318–328.

Simpson AGB & Patterson DJ (2001) On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*. Journal of Eukaryotic Microbiology 48: 480–492.

Simpson AGB & Roger AJ (2004) Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. Molecular Phylogenetics and Evolution 30: 201–212.

Simpson AGB, Gill EE, Callahan HA, Litaker RW, Roger AJ (2004) Early evolution within kinetoplastids (Euglenozoa), and the late emergence of trypanosomatids. Protist 155: 407–422.

Simpson AGB, Inagaki Y, Roger AJ (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Molecular Biology and Evolution 23: 615–625.

Simpson AGB, Lukeš J, Roger AJ (2002) The evolutionary history of kinetoplastids and their kinetoplasts. Molecular Biology and Evolution 12: 2071–2083.

Simpson AGB, Perley TA, Lara E (2008) Lateral transfer of the gene for a widely used marker, α-tubulin, indicated by a multi-protein study of the phylogenetic position of *Andalucia* (Excavata). Molecular Phylogenetics and Evolution 47: 366–377.

Smit S, Widmann J, Knight R (2007) Evolutionary rates vary among rRNA secondary structural elements. Nucleic Acids Research 35(10): 3339–3354.

Sommer JR (1965) The ultrastructure of the pellicle complex of *Euglena gracilis*. Journal of Cell Biology 24(2): 253–257.

Spencer DF, Collings JC, Schnare MN, Gray MW (1987) Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*. EMBO Journal 6(4): 1063–1071.

Steel M & Penny D (2000) Parsimony, likelihood, and the role of models in molecular phylogenetics. Molecular Biology and Evolution 17(6): 839–850.

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proceedings of The Royal Society Biological Sciences 272: 2073-2081.

Stoeck T & Epstein SS (2003) Novel eukaryotic lineages inferred from SSU rRNA analyses of oxygen-depleted marine environments. Applied and Environmental Microbiology 69:2657–2663.

Stoeck T, Kasper J, Bunge J, Leslin C, Ilyin V, Epstein SS (2007) Protistan diversity in the Arctic: a case of paleoclimate shaping modern biodiversity? PLOSOne 2: e728.

Stoeck T, Taylor GT, Epstein SS (2003) Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). Applied and Environmental Microbiology 69:5656–5663.

Stonik IV & Selina MS (2001) Species composition and seasonal dynamics of density and biomass of euglenoids in Peter the Great Bay, Sea of Japan. Russian Journal of Marine Biology 27: 174–176.

Sturm NR, Maslov DA, Grisard EC, Campbell DA (2001) *Diplonema* spp. possess spliced leader RNA genes similar to the Kinetoplastida. Journal of Eukaryotic Microbiology 48: 325–331.

Surek B & Melkonian M (1986) A cryptic cytostome is present in *Euglena*. Protoplasma 133: 39–49.

Suutari M, Majaneva M, Fewer DP, Voirin B, Aiello A, Friedl T, Chiarello AG, Blomster J (2010) Molecular evidence for a diverse green algal community growing in the hair of sloths and a specific association with *Trichophilus welckeri* (Chlorophyta, Ulvophyceae). BMC Evolutionary Biology 10: 86.

Takishita K, Kakizoe N, Yoshida T, Maruyama T (2010) Molecular evidence that phylogenetically diverged ciliates are active in microbial mats of deep-sea cold-seep sediment. Journal of Eukaryotic Microbiology 57(1): 76–86.

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

Tamura K & Kumar S (2002) Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Molecular Biology and Evolution 19: 1727–1736.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731–2739.

Tessier L-H, Keller M, Chan RL, Fournier R, Weil J-H, Imbault P (1991) Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena*. EMBO Journal 10: 2621–2625.

Thomas MC, Selinger LB, Inglis GD (2010) Seasonal diversity of planktonic protists in southwestern Alberta rivers over a 1-year period as revealed by terminal restriction fragment length polymorphism and 18S rRNA gene library analyses. Applied and Environmental Microbiology 78(16): 5653–5660.

Tikhonenkov DV, Mazei YA, Mylnikov AP (2006) Species diversity of heterotrophic flagellates in White Sea littoral sites. European Journal of Protistology 42: 191–200.

Torres-Machorro AL, Hernández R, Cevallos AM, López-Villaseñor I (2010) Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? FEMS Microbiological Reviews 34: 59–86.

Triemer RE & Farmer MA (1991) The ultrastructural organization of the heterotrophic euglenids and its evolutionary implications. In: Patterson DJ, Larsen J (eds) The biology of free-living heterotrophic flagellates. Oxford University Press, Oxford, pp 185–204.

Triemer RE & Farmer MA (2007) A decade of euglenoid molecular phylogenetics. In: Brodie J, Lewis J (eds) Unravelling the algae: The Past, Present and Future of Algal Systematics. Systematics Association Series. CRC Press, Boca Raton, p. 315–330.

Triemer RE & Fritz L (1987) Structure and operation of the feeding apparatus in a colorless euglenoid, *Entosiphon sulcatum.* The Journal of Protozoology 34(1): 39–47.

Triemer RE & Ott DW (1990) Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema.* European Journal of Protistology 25: 316–320.

Triemer RE, Linton E, Shin W, Nudelman A, Monfils A, Bennett M, Brosnan S (2006) Phylogeny of the Euglenales based upon combined SSU and LSU rDNA sequence comparisons and description of *Discoplastis* gen. nov. (Euglenophyta). Journal of Phycology 42: 731–740.

Vaerewijck MJM, Sabbe K, Baré J, Houf K (2008) Microscopic and molecular studies of the diversity of free-living protozoa in meat-cutting plants. Applied and Environmental Microbiology 74(18): 5741–5749.

Vaerewijck MJM, Sabbe K, Baré J, Houf K (2011) Occurrence and diversity of free-living protozoa on butterhead lettuce. International Journal of Food Microbiology 147(2): 105–111.

Von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T (2004) Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of Euglenozoa. Journal of Eukaryotic Microbiology 51(4): 402–416.

Wägele JW & Mayer C (2007) Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evolutionary Biology 7: 147.

Walker G, Dorrell RG, Schlacht A, Dacks JB (2011) Eukaryotic systematics: a user's guide for cell biologists and parasitologists. Parasitology 138: 1638–1663.

Walne PL & Dawson NS (1993) A comparison of paraxial rods in the flagella of euglenoids and kinetoplastids. Archiv der Protistenkunde 143: 177–194.

Wang Y, Zhang WP, Cao HL, Shek CS, Tian RM, Wong YH, Batang Z, Al-Suwailem A, Qian P-Y (2014) Diversity and distribution of eukaryotic microbes in and around a brine pool adjacent to the Thuwal cold seeps in the Red Sea. Frontiers in Microbiology 5: 37.

Wehner R & Gehring W (2013) Zoologie. Georg Thieme Verlag, Stuttgart, New York.

Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. Science 319: 473–476.

Wuyts J, Van de Peer Y, De Wachter R (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Research 30: 183–185.

Wylezich C & Jürgens K (2011) Protist diversity in suboxic and sulfidic waters of the Black Sea. Environmental Microbiology 13: 2939–2956.

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

Yamaguchi A, Yubuki N, Leander BS (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: 29.

Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, Cole JC, Logsdon JM, JR, Patterson DJ, Bhattacharya D, Katz LA (2008) Broadly sampled multigene trees of eukaryotes. BMC Evolutionary Biology 8: 14.

Yubuki N, Edgcomb VP, Bernhard JM, Leander BS (2009) Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria. BMC Microbiology 9: 16.

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

Zuendorf A, Bunge J, Behnke A, Barger K, Stoeck T (2006) Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. FEMS Microbiological Ecology 58: 476-491.

8 Appendix

This appendix section contains figures and tables which bear reference to chapters 5 and 6.

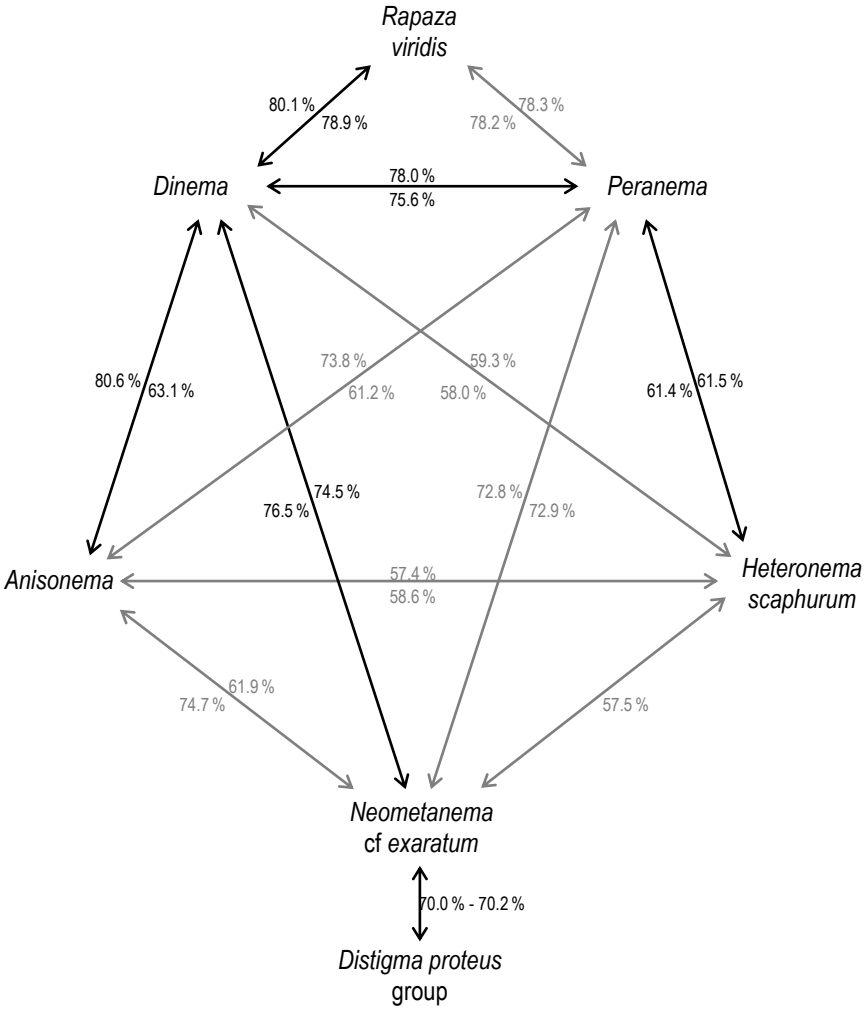


Fig. 8.1: Identity matrix graph based on SSU rDNA sequences from dataset III depicting highest and lowest similarity percentages between the Helicales, i.e. phagotrophic euglenid taxa *Dinema*, *Peranema*, *Heteronema scaphurum*, *Neometanema* cf *exaratum* and Anisonemida, as well as the *Distigma proteus* group (primordial Aphagea) and *Rapaza viridis*. For other taxa see Fig. 5.17.

Tab. 8.1: Nucleotide composition of SSU rDNA sequences in dataset III. Heated percentages for nucleotides and GC-summaries are given; lowest percentages of each nucleotide are colored green, highest values are red.

Taxon	T	C	A	G	GC
EUGLENEA					
Euglena gracilis	22,14	23,59	25,34	28,93	52,52
Euglena sp	22,06	23,62	25,36	28,96	52,58
Euglena gracilis var bacillaris	22,14	23,59	25,34	28,93	52,52
Astasia longa	22,04	24,08	25,34	28,54	52,62
Khawkinea quartana	21,30	24,12	25,29	29,28	53,40
Euglena cf mutabilis	20,87	23,50	25,63	30,00	53,50
Cyclidiopsis acus	21,55	24,85	23,59	30,00	54,85
Discoplastis spathirhyncha	22,23	24,08	25,15	28,54	52,62
Monomorphina megalopsis	22,14	23,69	25,05	29,13	52,82
Monomorphina rudicula	21,56	23,84	26,33	28,28	52,11
Strombomonas verrucosa	21,75	23,88	25,53	28,83	52,72
Trachelomonas grandis	22,91	23,11	25,92	28,06	51,17
Euglena stellata	21,36	24,66	24,66	29,32	53,98
Euglena tripteris	22,91	23,69	24,47	28,93	52,62
Colacium vesiculosum	21,55	23,69	26,12	28,64	52,33
Colacium mucronatum	21,94	23,59	26,21	28,25	51,84
Colacium sp	21,65	23,50	26,31	28,54	52,04
Cryptoglena pigra	22,33	24,17	25,83	27,67	51,84
Cryptoglena skujae	22,33	24,08	25,83	27,77	51,84
Cryptoglena sp	22,14	24,17	25,63	28,06	52,23
Discoplastis sp	21,55	24,85	24,08	29,51	54,37
Hyalophacus ocellata	24,17	23,39	24,37	28,07	51,46
Lepocinclis oxyuris	21,65	24,85	23,50	30,00	54,85
Lepocinclis spirogyroides	22,82	23,69	24,95	28,54	52,23
Monomorphina aenigmatica	21,46	25,15	25,24	28,16	53,30
Monomorphina pyrum	22,04	23,69	26,12	28,16	51,84
Phacus oscillans	23,01	23,40	25,34	28,25	51,65
Phacus pusillus	23,20	22,52	25,73	28,54	51,07
Strombomonas acuminata	21,75	23,88	25,63	28,74	52,62
Trachelomonas hispida	22,62	23,01	26,70	27,67	50,68
Trachelomonas volvocinopsis	22,43	23,79	25,05	28,74	52,52
Eutreptia pertyi	23,59	22,72	25,24	28,45	51,17
Eutreptia viridis AF157312	23,79	22,82	25,63	27,77	50,58
Eutreptia viridis AJ532395	23,69	22,82	25,53	27,96	50,78
Eutreptia sp AJ532396	22,72	23,98	24,85	28,45	52,43
Eutreptiella gymnastica	22,72	22,82	26,12	28,35	51,17
Eutreptiella eupharyngea	22,62	23,50	25,53	28,35	51,84
Eutreptiella sp AF112875	22,56	24,02	25,10	28,32	52,34
Eutreptiella sp JQ337867	22,62	23,98	24,85	28,54	52,52
Eutreptiella pomquetensis	22,91	23,50	25,34	28,25	51,75
Eutreptiella braarudii	22,84	23,52	25,46	28,18	51,70
Rapaza viridis	21,84	24,47	25,05	28,64	53,11
APHAGEA					
Gyropaigne lefevrei	22,82	23,30	24,37	29,51	52,82
Parmidium circulare	21,75	23,01	25,24	30,00	53,01
Parmidium scutulum	22,40	22,98	25,22	29,41	52,39
Menoidium pellucidum	21,84	23,59	24,47	30,10	53,69
Rhabdomonas costata	22,14	23,88	24,66	29,32	53,20
Rhabdomonas incurva	22,08	23,74	24,51	29,67	53,40
Rhabdomonas spiralis	21,36	23,59	24,08	30,97	54,56
Rhabdomonas intermedia	22,55	22,74	25,75	28,96	51,70
Menoidium bibacillatum	21,60	23,74	24,81	29,86	53,60
Menoidium cultellus	21,84	23,59	24,56	30,00	53,59
Menoidium gibbum	21,36	23,59	24,08	30,97	54,56

Tab. 8.1: continued.

Taxon	T	C	A	G	GC
Menoidium intermedium	21,84	23,30	24,56	30,29	53,59
Menoidium obtusum	21,94	23,50	24,27	30,29	53,79
Menoidium sp	21,94	23,40	24,37	30,29	53,69
Astasia curvata AY004245	22,72	22,14	25,92	29,22	51,36
Astasia curvata SAG1204-5b	22,82	22,23	25,92	29,03	51,26
Astasia sp	23,11	22,72	25,53	28,64	51,36
Astasia curvata AF403153	23,11	22,43	25,05	29,42	51,84
Astasia torta	22,72	22,62	25,34	29,32	51,94
Distigma curvatum	23,79	20,97	28,25	26,99	47,96
Distigma gracile	23,88	20,87	28,25	26,99	47,86
Distigma sennii	22,23	23,11	26,60	28,06	51,17
Distigma pringsheimii	24,56	21,36	27,28	26,80	48,16
Distigma gracilis	24,37	21,36	27,38	26,89	48,25
Distigma proteus	24,56	21,36	27,38	26,70	48,06
Peranema trichophorum AH005452	24,37	21,94	27,48	26,21	48,16
Peranema trichophorum AF386636	24,37	21,94	27,38	26,31	48,25
Peranema sp	24,42	21,89	27,53	26,17	48,05
ANISONEMIDA					
Anisonema sp syn Peranema AY048919	23,40	21,75	26,89	27,96	49,71
Anisonema acinus AF403160	22,43	22,33	25,83	29,42	51,75
Anisonema acinus isolateB1	22,34	22,21	26,46	28,99	51,20
Anisonema sp isolateW1	23,40	22,43	25,92	28,25	50,68
Anisonema/Dinema sp isolateU3	22,52	23,01	25,34	29,13	52,14
Dinema sulcatum	21,26	23,79	24,66	30,29	54,08
Dinema platysomum	21,44	23,10	26,32	29,14	52,24
Heteronema scaphurum	19,51	22,09	28,12	30,27	52,37
Neometanema cf exaratum	22,82	22,04	27,28	27,86	49,90
ANOXIDA/SYMBIONTIDA					
Uncultured marine eukaryote clone NA1 4H11	23,35	20,60	28,43	27,61	48,21
Uncultured marine eukaryote clone NA1 4B5	23,35	20,60	28,57	27,47	48,08
Uncultured marine eukaryote clone BLACKSEA 50	23,50	20,62	28,21	27,67	48,29
Uncultured marine eukaryote clone BLACKSEA 52	23,69	20,69	28,72	26,90	47,59
Uncultured marine eukaryote clone FV23 2D3C4	23,81	21,09	27,79	27,31	48,40
Uncultured marine eukaryote clone NA1 1G12	24,17	20,78	27,67	27,38	48,16
Calkinsia aureus	23,50	22,04	27,38	27,09	49,13
Bihospites bacati isolate1	21,07	23,30	26,50	29,13	52,43
Bihospites bacati isolate2	21,07	23,59	26,31	29,03	52,62
PLOEOTIIDA					
Ploeotia cf vitrea	24,85	20,97	28,74	25,44	46,41
Ploeotia costata	22,33	23,69	26,31	27,67	51,36
Keelungia pulex	24,03	21,11	28,02	26,85	47,96
Ploeotia edaphica	24,17	20,39	27,77	27,67	48,06
Entosiphon sulcatum AF220826	23,01	22,23	27,18	27,57	49,81
Entosiphon sulcatum AY061999	22,91	22,33	27,09	27,67	50,00
Entosiphon sp	22,72	22,14	27,48	27,67	49,81
PETALOMONADIDA					
Petalomonas cantuscygni U84731	24,20	21,28	26,72	27,79	49,08
Petalomonas cantuscygni AF386635	23,98	21,36	26,70	27,96	49,32
Petalomonas sphagnophila Liz	23,69	22,04	27,38	26,89	48,93
Petalomonas sphagnophila Dunc	23,79	22,04	27,18	26,99	49,03
Petalomonas sphagnophila HF	23,98	21,84	27,28	26,89	48,74
Uncultured Sphenomonadales clone PR3 3E 63	23,48	21,88	28,29	26,35	48,22
Uncultured marine eukaryote clone BLACKSEA 51	23,23	20,99	27,09	28,69	49,68
Uncultured sphenomonad euglenozoan clone CH1 S2 16	23,59	20,97	28,16	27,28	48,25
Uncultured sphenomonad euglenozoan clone CH1 S2 19	24,57	21,22	27,92	26,29	47,51
Uncultured sphenomonad euglenozoan clone CH1 S1 57	24,63	20,47	28,68	26,21	46,69
Uncultured eukaryote clone D3P06F06	22,39	20,49	28,63	28,49	48,98

Tab. 8.1: continued.

Taxon	T	C	A	G	GC
Notosolenus ostium isolateU1	25,34	20,49	27,38	26,80	47,28
Notosolenus ostium AF403159	25,34	20,49	27,18	26,99	47,48
DIPLONEMIDA					
Diplonema ambulator ATCC50223	23,40	21,36	27,67	27,57	48,93
Diplonema papillatum	22,52	21,84	27,57	28,06	49,90
Diplonema sp ATCC50232	22,72	22,04	27,18	28,06	50,10
Diplonema sp1 ATCC50224	23,40	21,75	27,57	27,28	49,03
Diplonema sp2 ATCC50224	23,40	21,75	27,57	27,28	49,03
Diplonema ambulator	23,40	21,36	27,67	27,57	48,93
Diplonema sp ATCC50225	23,50	21,46	27,38	27,67	49,13
Uncultured eukaryote clone RM2-SGM31	24,56	20,97	26,80	27,67	48,64
Uncultured diplonemid clone LC22 5EP 32	23,82	20,78	28,24	27,16	47,94
Uncultured marine diplonemid clone Ma131 1A46	23,55	21,59	27,67	27,18	48,77
Uncultured euglenid clone CCW85	22,55	21,30	28,13	28,02	49,32
Uncultured marine euglenozoan DH148-EKB1	24,44	21,23	28,24	26,10	47,32
Uncultured diplonemid clone LC22 5EP 17	23,59	20,87	27,96	27,57	48,45
Uncultured diplonemid clone LC22 5EP 18	23,88	20,97	28,16	26,99	47,96
Uncultured diplonemid clone LC22 5EP 19	23,59	21,17	28,06	27,18	48,35
Uncultured diplonemid clone LC23 5EP 5	23,98	20,97	27,28	27,77	48,74
Uncultured eukaryote clone SCM15C6	23,69	20,97	28,35	26,99	47,96
Uncultured eukaryote clone RM2-SGM32	23,79	21,55	27,28	27,38	48,93
Rhynchopus sp SH-2004-IV	23,20	21,46	27,57	27,77	49,22
Rhynchopus sp SH-2004-I	23,20	21,46	27,57	27,77	49,22
Rhynchopus euleeides ATCC50226	23,40	21,65	27,28	27,67	49,32
Rhynchopus sp SH-2004-II	23,88	21,36	26,70	28,06	49,42
Rhynchopus sp ATCC50229	23,50	21,17	27,96	27,38	48,54
KINETOPLASTIDA					
Uncultured kinetoplastid clone AT4-103	21,65	22,62	27,09	28,64	51,26
Uncultured kinetoplastid clone LC103 5EP 19	23,79	21,07	27,67	27,48	48,54
Perkinsiella-like sp AFSM3	23,40	21,94	27,48	27,18	49,13
Perkinsiella-like sp PLO-DE4A	23,40	22,14	27,48	26,99	49,13
Perkinsela-like organism AK-2011	24,25	21,52	27,75	26,48	48,00
Ichthyobodo necator AY028448	23,50	20,39	28,35	27,77	48,16
Ichthyobodo necator DK	23,01	20,68	28,64	27,67	48,35
Uncultured eukaryote clone L7.7	23,59	21,26	27,57	27,57	48,83
Angomonas deanei	23,98	20,58	28,64	26,80	47,38
Azumiobodo hoyamushi	23,11	21,46	28,06	27,38	48,83
Bodo edax	22,82	21,26	28,64	27,28	48,54
Bodo rostratus	23,01	21,94	28,16	26,89	48,83
Crithidia dedva	23,50	20,78	28,64	27,09	47,86
Cruzella marina	22,91	22,04	28,06	26,99	49,03
Cryptobia helicis	22,91	21,65	27,67	27,77	49,42
Cryptobia salmositica	23,11	21,17	28,64	27,09	48,25
Dimastigella mimosa	22,33	21,55	29,03	27,09	48,64
Dimastigella trypaniformis	22,33	21,55	29,03	27,09	48,64
Endotrypanum sp 889	23,59	20,87	28,54	26,99	47,86
Herpetomonas sp TCC247	23,69	20,78	28,54	26,99	47,77
Leishmania major	23,40	20,97	28,74	26,89	47,86
Leptomonas mirabilis	23,88	20,78	28,35	26,99	47,77
Neobodo designis	22,62	21,65	28,16	27,57	49,22
Neobodo saliens	23,01	21,94	27,67	27,38	49,32
Parabodo caudatus	22,91	20,87	28,64	27,57	48,45
Parabodo nitrophilus	23,20	20,68	28,64	27,48	48,16
Phanerobia pelophila	22,43	21,55	29,03	26,99	48,54
Bodo sorokini	22,72	21,75	28,06	27,48	49,22
Rhynchomonas nasuta	23,01	21,55	28,25	27,18	48,74
Sergeia podlipaevi	23,91	20,51	28,38	27,21	47,72

Tab. 8.1: continued.

Taxon	T	C	A	G	GC
Strigomonas culicis	22,62	21,55	28,06	27,77	49,32
Strigomonas galati	22,72	21,46	28,06	27,77	49,22
Trypanosoma cruzi	23,50	20,97	28,54	26,99	47,96
Soil flagellate AND31	23,13	21,77	27,79	27,31	49,08
Uncultured bodonid clone AT1-3	22,62	21,65	27,67	28,06	49,71
Uncultured kinetoplastid clone AT4-56	23,69	21,17	27,38	27,77	48,93
Uncultured bodonid clone AT5-9	23,11	21,65	27,77	27,48	49,13
Uncultured bodonid clone AT5-25	23,01	22,04	27,28	27,67	49,71
Uncultured bodonid clone AT5-48	23,11	21,46	27,96	27,48	48,93
Uncultured kinetoplastid clone Discovery IF R B	22,52	21,65	28,25	27,57	49,22
Kinetoplastida sp FV18-8TS	22,91	22,04	27,38	27,67	49,71
Uncultured kinetoplastid clone Kryos IF A3	22,91	21,26	28,25	27,57	48,83
Uncultured kinetoplastid clone Urania B B5	22,72	21,46	28,25	27,57	49,03
Uncultured eukaryote clone ZJ2007	23,50	21,07	28,54	26,89	47,96
Wallaceina sp	24,08	20,49	28,74	26,70	47,18
Bodo saltans	22,72	21,46	28,45	27,38	48,83
Cryptobia bullockii	23,01	21,26	28,54	27,18	48,45
Herpetomonas nabiculae	23,88	20,68	28,45	26,99	47,67
Leptomonas collosoma	24,17	20,68	28,45	26,70	47,38
Crithidia fasciculata	23,50	20,78	28,74	26,99	47,77
Trypanosoma brucei	23,01	21,84	28,06	27,09	48,93
OUTGROUP, HETEROLOBOSEA					
Pharyngomonas kirbyi BB2 JX509941	21,94	22,82	26,02	29,22	52,04
Heterolobosea sp strain SD1A	24,56	20,19	28,45	26,80	46,99
Heterolobosea sp strain AS12B	24,56	20,19	28,45	26,80	46,99
Macropharyngomonas halophila	24,68	20,21	28,38	26,72	46,94
Acrasis rosea	25,53	19,42	30,58	24,47	43,88
Allovahlkampfia spelaea	25,73	19,13	30,29	24,85	43,98
Euplaesiobystra hypersalinica	22,62	22,43	30,10	24,85	47,28
Harpagon descissus	28,99	15,86	34,05	21,11	36,96
Harpagon schusteri	28,64	16,50	33,30	21,55	38,06
Heteramoeba clara	22,23	22,62	29,71	25,44	48,06
Heterolobosea sp BB2	21,84	22,23	26,80	29,13	51,36
Heterolobosea sp HGG1	23,98	19,32	31,94	24,76	44,08
Heterolobosea sp LO	24,85	20,00	28,25	26,89	46,89
Pharyngomonas sp RL	25,73	20,00	28,35	25,92	45,92
Heterolobosea sp SAN2	23,50	19,51	30,58	26,41	45,92
Paravahlkampfia sp	24,56	20,00	30,29	25,15	45,15
Pleurostomum flabellatum	25,44	18,35	30,87	25,34	43,69
Psalteriomonas lanterna	29,51	15,92	35,73	18,83	34,76
Psalteriomonas magna	30,00	15,92	35,73	18,35	34,27
Pseudoharpagon pertyi	21,36	23,79	28,54	26,31	50,10
Sawyeria marylandensis	29,32	16,41	35,63	18,64	35,05
Stygamoeba regulata	26,72	18,46	28,86	25,95	44,41
Tetramitus thermacidophilus	25,15	18,64	31,07	25,15	43,79
Tetramitus thorntoni	25,15	18,45	31,17	25,24	43,69
Uncultured heterolobosean clone WIM43	23,59	19,03	32,23	25,15	44,17
Uncultured eukaryote clone CN207St155 8Be04F	24,95	20,10	28,25	26,70	46,80
Uncultured eukaryote clone CN207St70 8BBe08M	24,95	20,10	27,96	26,99	47,09
Uncultured marine eukaryote clone BLACKSEA 54	23,49	18,53	31,57	26,40	44,94
Selenaion koniopes	20,87	23,20	28,35	27,57	50,78
Plaesiobystra hypersalinica	22,45	21,87	30,52	25,17	47,04
Naegleria clarki	23,50	21,84	28,54	26,12	47,96
Naegleria gruberi	23,59	21,84	28,45	26,12	47,96
Stachyamoeba sp	23,59	19,32	31,07	26,02	45,34
Paravahlkampfia ustiana	24,56	19,90	30,29	25,24	45,15
Tsukubamonas globosa	22,62	22,14	27,57	27,67	49,81

Tab. 8.1: continued.

Taxon	T	C	A	G	GC
OUTGROUP, JAKOBIDA					
Uncultured marine eukaryote clone BLACKSEA 55	25,03	18,82	28,88	27,27	46,10
Uncultured eukaryote clone EN351CTD039 30mN9	26,31	18,64	29,42	25,63	44,27
Reclinomonas americana AF053089	22,82	23,20	27,57	26,41	49,61
Andalucia incarcerata MB1	24,33	20,08	27,89	27,70	47,77
Andalucia incarcerata AY117419	24,51	19,94	27,92	27,63	47,57
Andalucia godoyi AND19	24,56	20,00	27,96	27,48	47,48
Uncultured marine eukaryote clone cLA12C05	24,57	19,44	28,31	27,67	47,12
Uncultured marine eukaryote clone FV23 CilE10	24,47	19,61	28,93	26,99	46,60
Uncultured marine eukaryote clone M2 18G04	24,74	19,30	28,85	27,10	46,41
Uncultured marine eukaryote clone SA1 1D05	24,76	19,61	28,16	27,48	47,09
Uncultured Jakobida clone NKS105	23,79	20,78	27,48	27,96	48,74
Uncultured Jakobida clone NKS177	24,08	20,19	27,48	28,25	48,45
Jakoba libera AF411288	22,52	22,23	28,16	27,09	49,32
Jakoba libera AY117418	22,52	22,33	27,57	27,57	49,90
Reclinomonas americana AY117417	23,17	22,20	26,78	27,85	50,05
Seculamonas ecuadoriensis	20,39	23,59	27,18	28,83	52,43
OUTGROUP, DISTANT					
Malawimonas jakobiformis AY117420	25,56	18,95	28,77	26,72	45,68
Retortamonas sp	22,72	21,36	25,92	30,00	51,36
Carpediemonas membranifera	22,49	20,93	28,24	28,33	49,27
Ergobibamus cyprinoides	22,62	22,43	27,48	27,48	49,90
Dysnectes brevis	23,69	21,65	26,31	28,35	50,00
Hicanonectes teleskopos	24,66	20,29	27,57	27,48	47,77
Dinenympha exilis	24,66	19,51	28,54	27,28	46,80
Pyrsonympha grandis	24,37	19,61	28,64	27,38	46,99
Oxymonas sp	25,24	18,45	29,22	27,09	45,53
Trimastix pyriformis	24,27	19,22	29,32	27,18	46,41
Proteromonas lacertae	26,92	16,03	33,82	23,23	39,26
Paramecium tetraurelia	25,05	18,64	29,42	26,89	45,53
Gymnodinium sanguineum	26,41	18,93	28,35	26,31	45,24
Chilomonas paramecium	25,34	19,51	29,03	26,12	45,63
Dictyostelium discoideum	24,90	18,97	31,32	24,81	43,77
Palmaria palmata	22,62	21,55	26,99	28,83	50,39
Acanthamoeba castellanii	24,88	20,02	27,70	27,41	47,42
Heterosigma akashiwo	25,73	19,03	29,13	26,12	45,15
Chrysowaernella hieroglyphica	24,98	19,92	28,67	26,43	46,36
Sarcocystis alceslatrans	25,63	19,22	29,13	26,02	45,24
Symbiotic dinoflagellate BBSR 323	26,02	19,22	28,74	26,02	45,24
Prorocentrum rhathymum	26,31	18,93	28,54	26,21	45,15
Theileria annulata	26,50	18,54	28,35	26,60	45,15
Cyrtohymena shii	26,02	18,45	28,83	26,70	45,15
Salpingoeca infusionum	24,27	18,83	27,96	28,93	47,77
Monosiga brevicollis	25,73	18,45	29,03	26,80	45,24
Byssochlamys spectabilis	23,98	20,58	27,67	27,77	48,35
Paramicrosporidium vannellae	25,34	17,96	28,54	28,16	46,12
Blastocystis sp	26,99	17,09	29,81	26,12	43,20
Tetrahymena thermophila	26,02	18,45	30,29	25,24	43,69
Chaos nobile	24,17	19,51	28,83	27,48	46,99
Amoeba leningradensis	23,11	20,10	28,54	28,25	48,35
Amoeba proteus	22,43	21,65	27,48	28,45	50,10
Plasmodium ovale	27,48	17,38	32,14	23,01	40,39
Chlorarachnion reptans	23,03	21,87	26,53	28,57	50,44
Pessonella sp	26,85	17,12	31,13	24,90	42,02
Saccharomyces cerevisiae	25,73	18,54	28,64	27,09	45,63
Average	23,53	21,40	27,65	27,42	48,82

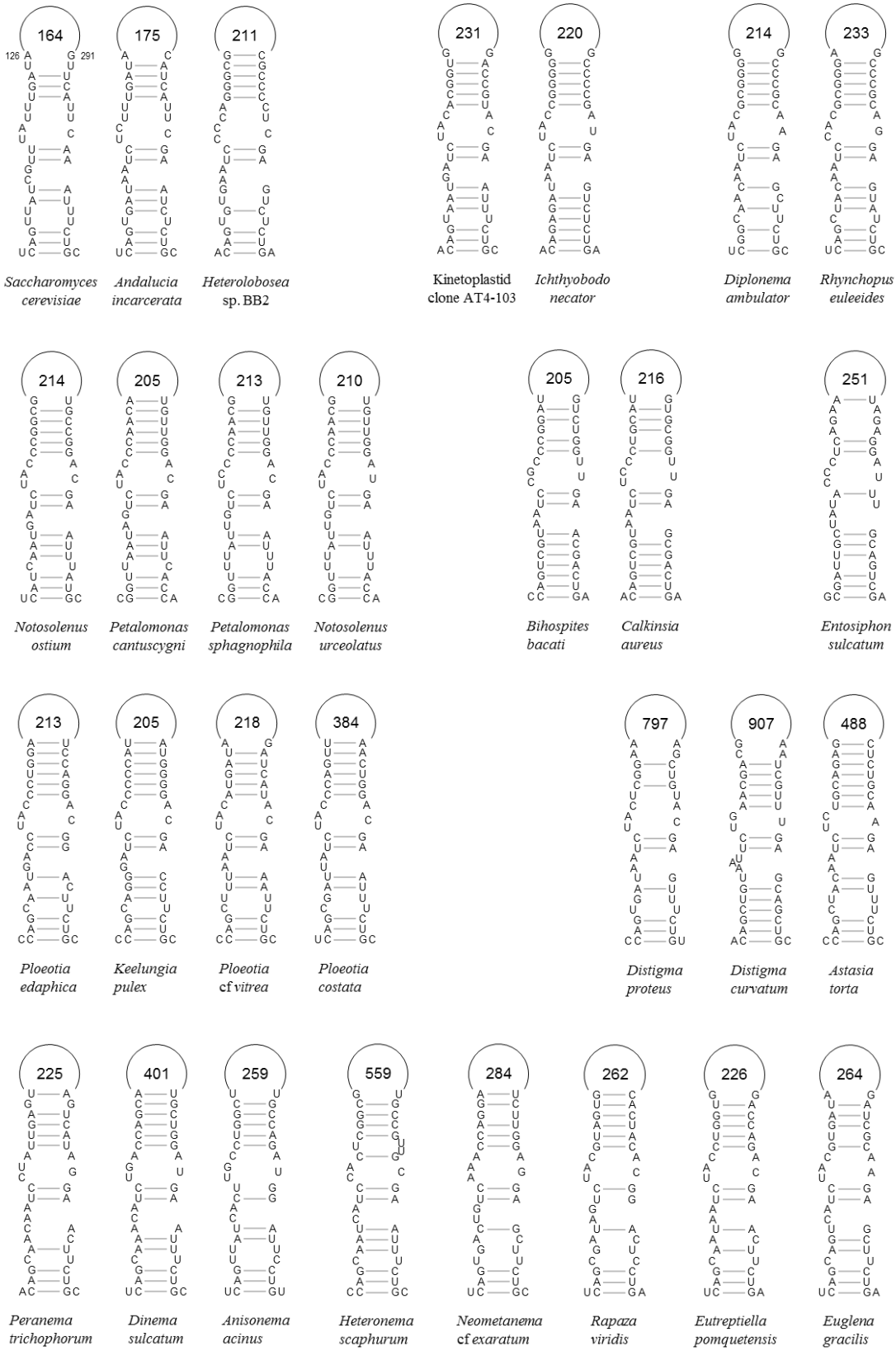


Fig. 8.2: Deduced SSU rRNA secondary structures of helix 7 illustrating nucleotide substitution differences between selected euglenozoan and outgroup taxa. Numbers centered in the apex of each helix represent taxon specific nucleotides associated with helices 8, 9, 10 and variable region 2, all of which are encompassed by helix 7. Small numbers indicate positions in the SSU rDNA nucleotide sequence of *Saccharomyces cerevisiae*.

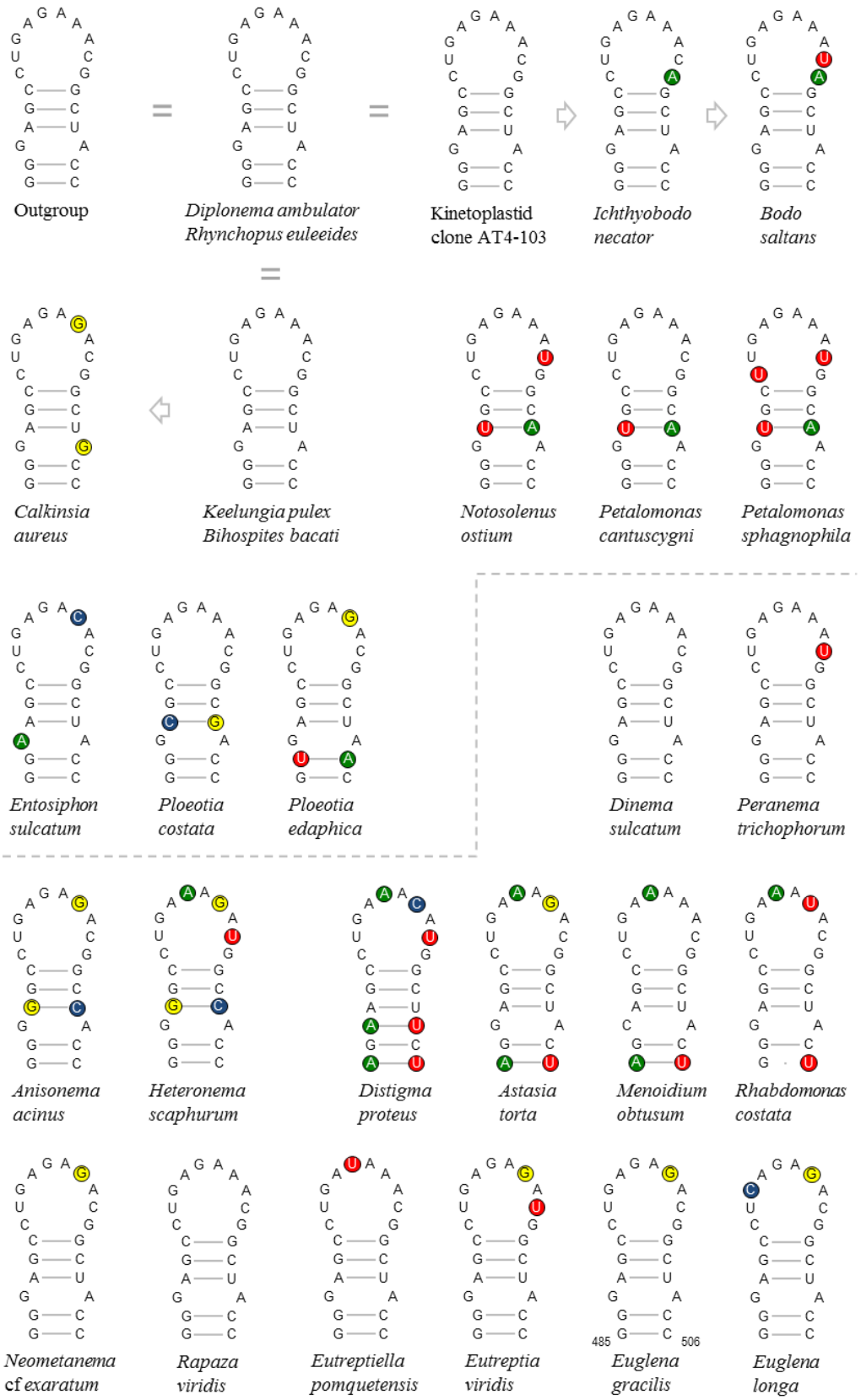


Fig. 8.3: Inferred SSU rRNA secondary structures of helix 13 exemplifying nucleotide substitution differences between selected euglenozoan and outgroup taxa. Unique nucleotide changes compared to the outgroup are colored, a dashed line separates outgroup and non-helical euglenozoan taxa from the Helicales. Small numbers indicate nucleotide coordinates in the SSU rDNA of *Euglena gracilis* (M12677). Helix 13 secondary structures of outgroup taxa, Diplonemida, primordial kinetoplastid clone AT4-103, *Keelungia pulex*, primordial symbiontid *Bihospites bacati*, *Dinema sulcatum* and *Rapaza viridis* are identical.

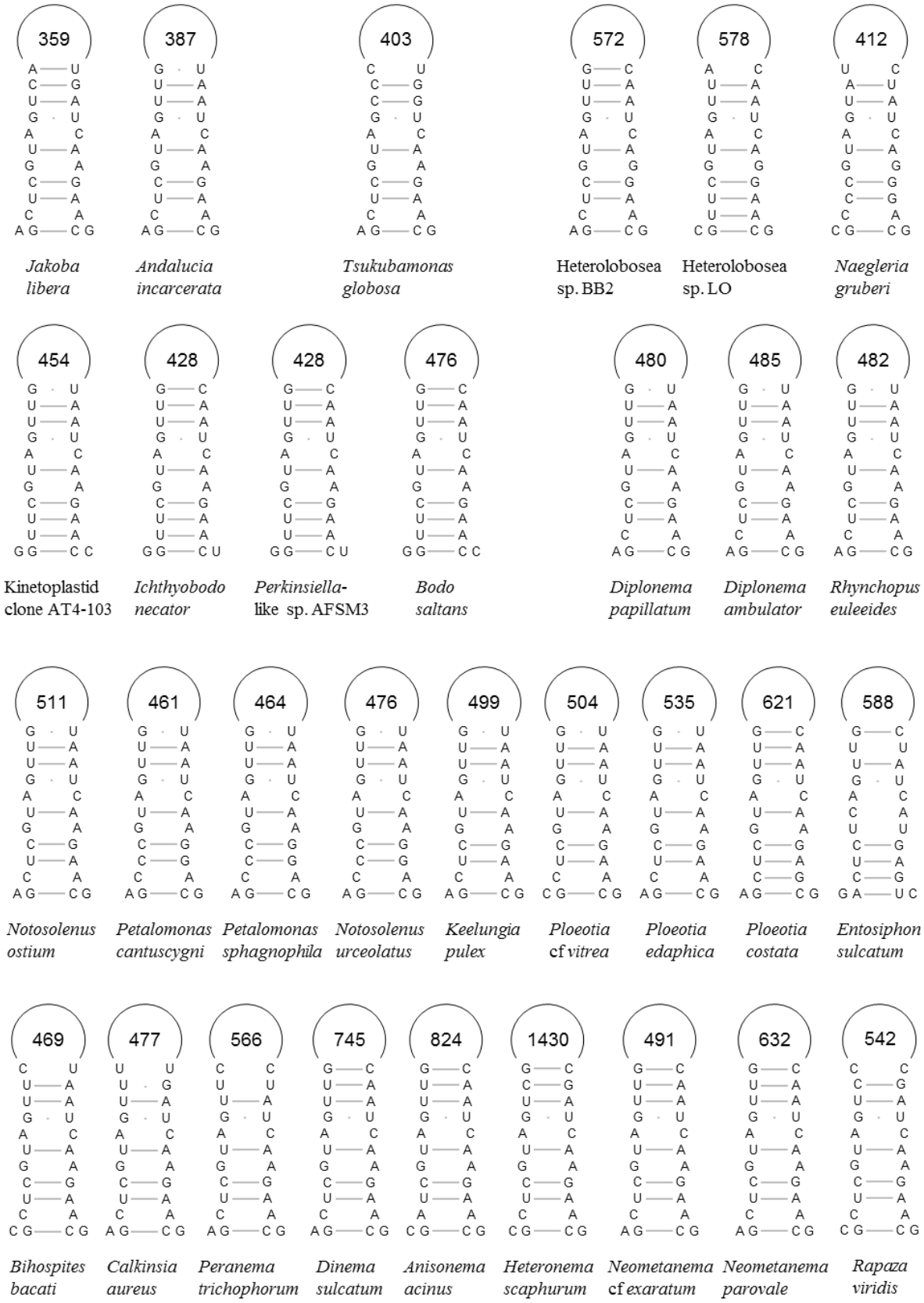


Fig. 8.4: Concluded secondary structures of SSU rRNA helix 20 illustrating differences in nucleotide substitution and variable region length between euglenozoan and outgroup taxa. A number in the apex of each helix represents residual nucleotides of helices 21, 22 and 23 including extended regions as well as variable region 4, for these are encompassed by helix 20. *This graph continues on the next page*. →

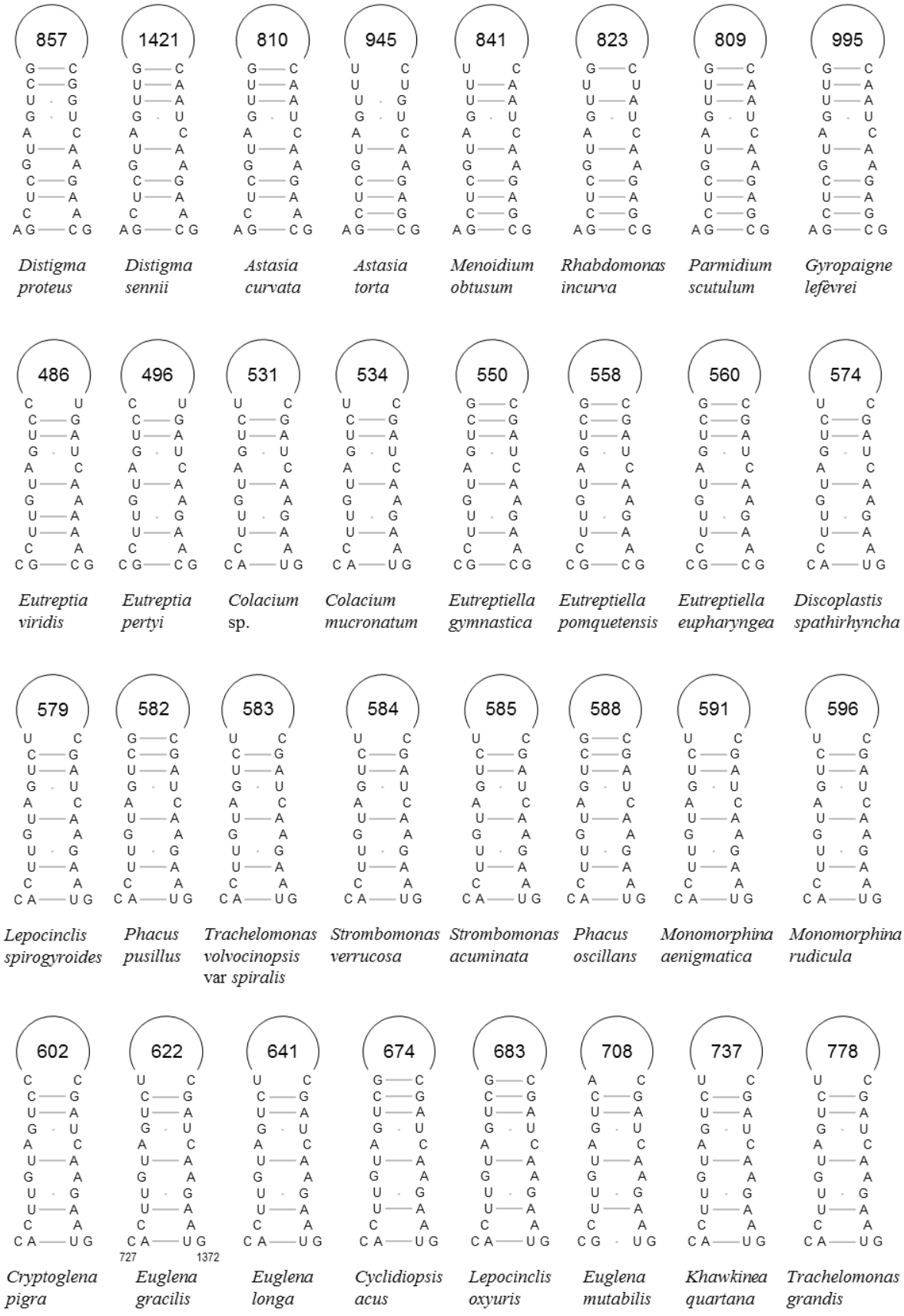


Fig. 8.4: Continued. Small numbers represent nucleotide coordinates in the SSU rDNA sequence of *Euglena gracilis* (to the lower left; referring to accession M12677). For a summary of primordial Euglenozoa see Fig. 5.20.

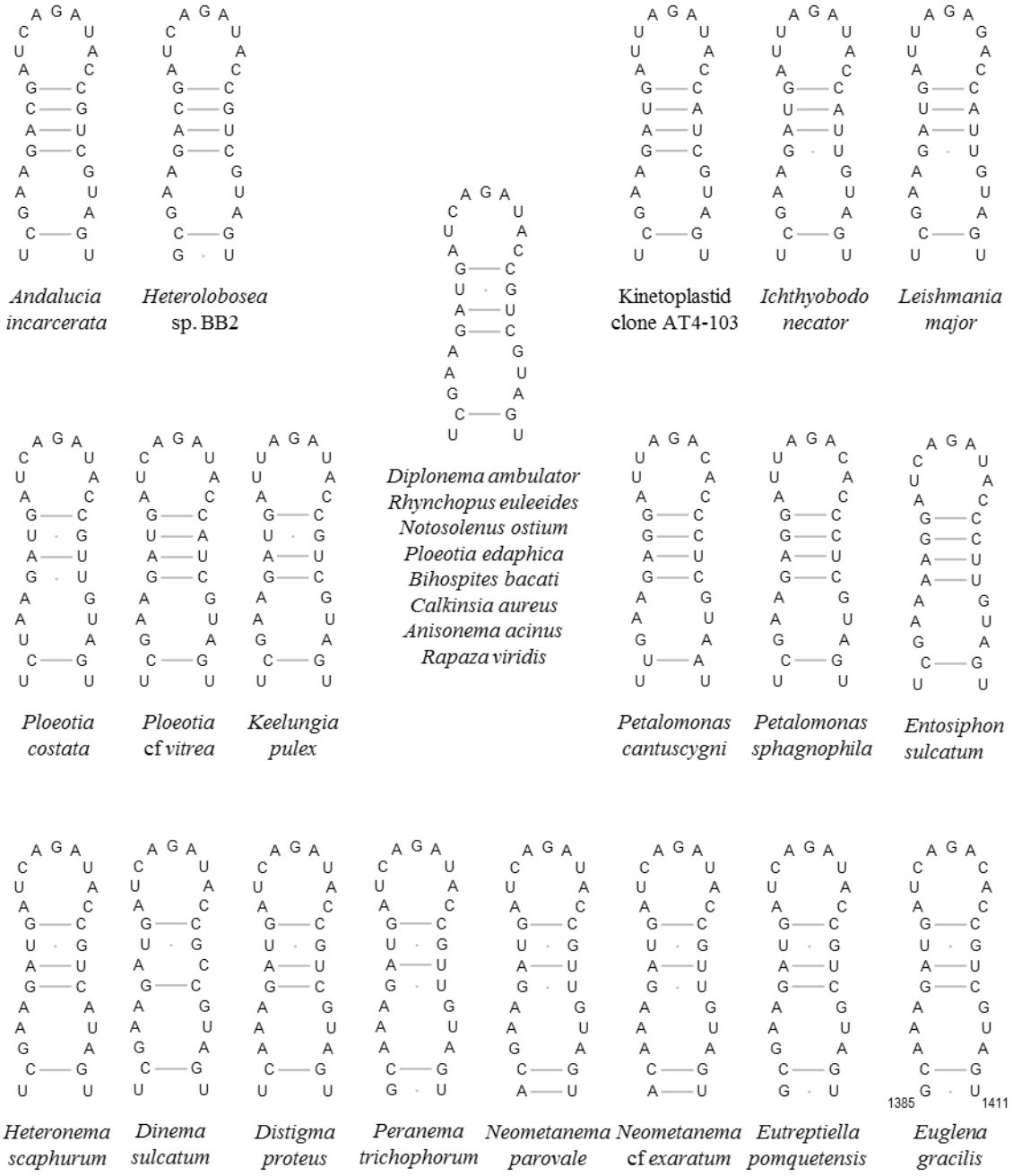


Fig. 8.5: Deduced SSU rRNA secondary structures of helix 24 showing nucleotide substitution differences between outgroup taxa and Euglenozoa. Small numbers represent nucleotide coordinates in the sequence of *Euglena gracilis* (lower right; accession M12677). For a summary see Fig. 5.18.

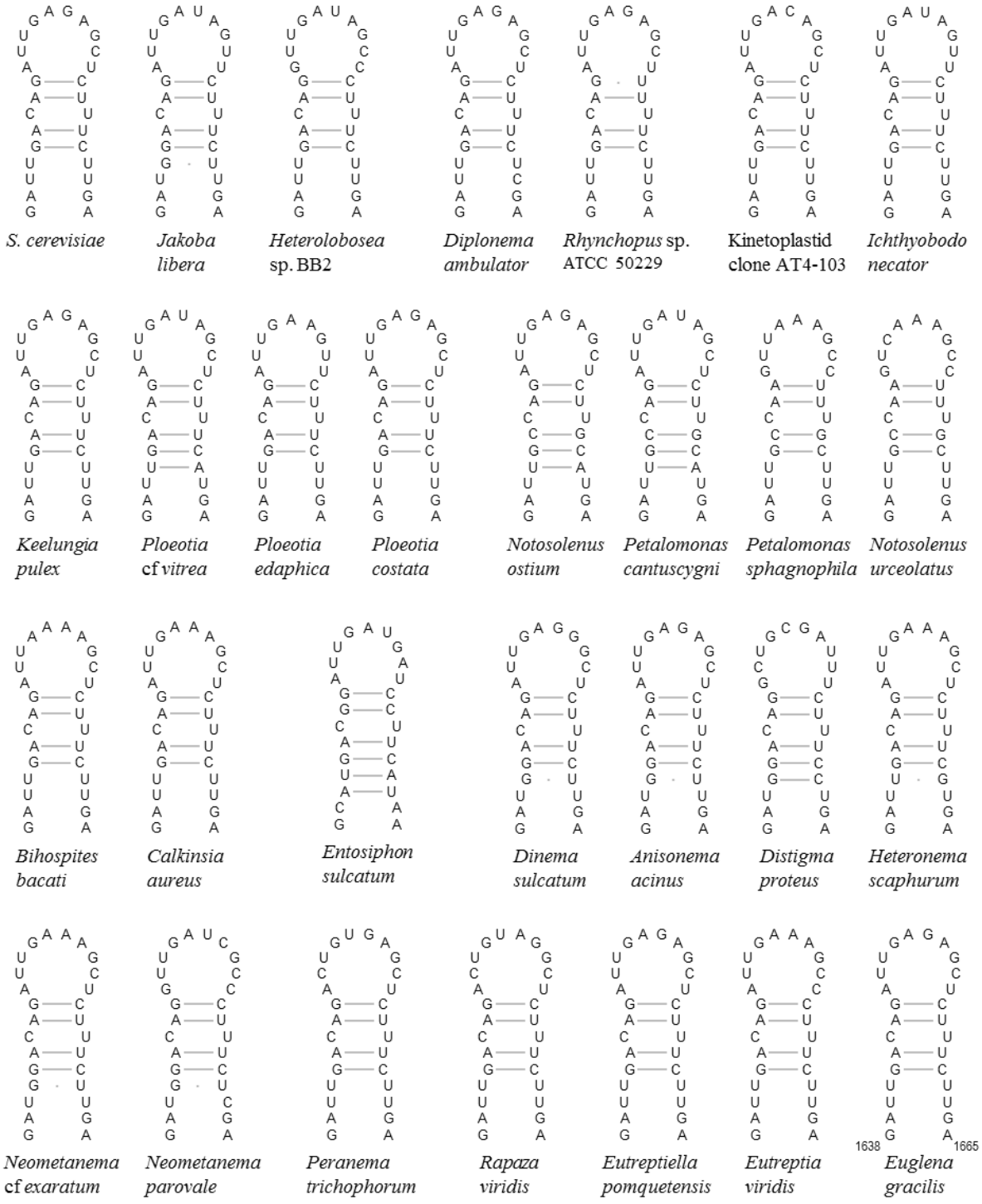


Fig. 8.6: Inferred SSU rRNA secondary structures of helix 33 exemplifying nucleotide substitution differences between outgroup taxa and Euglenozoa. Small numbers represent nucleotide coordinates in the sequence of *Euglena gracilis* (lower right; accession M12677).

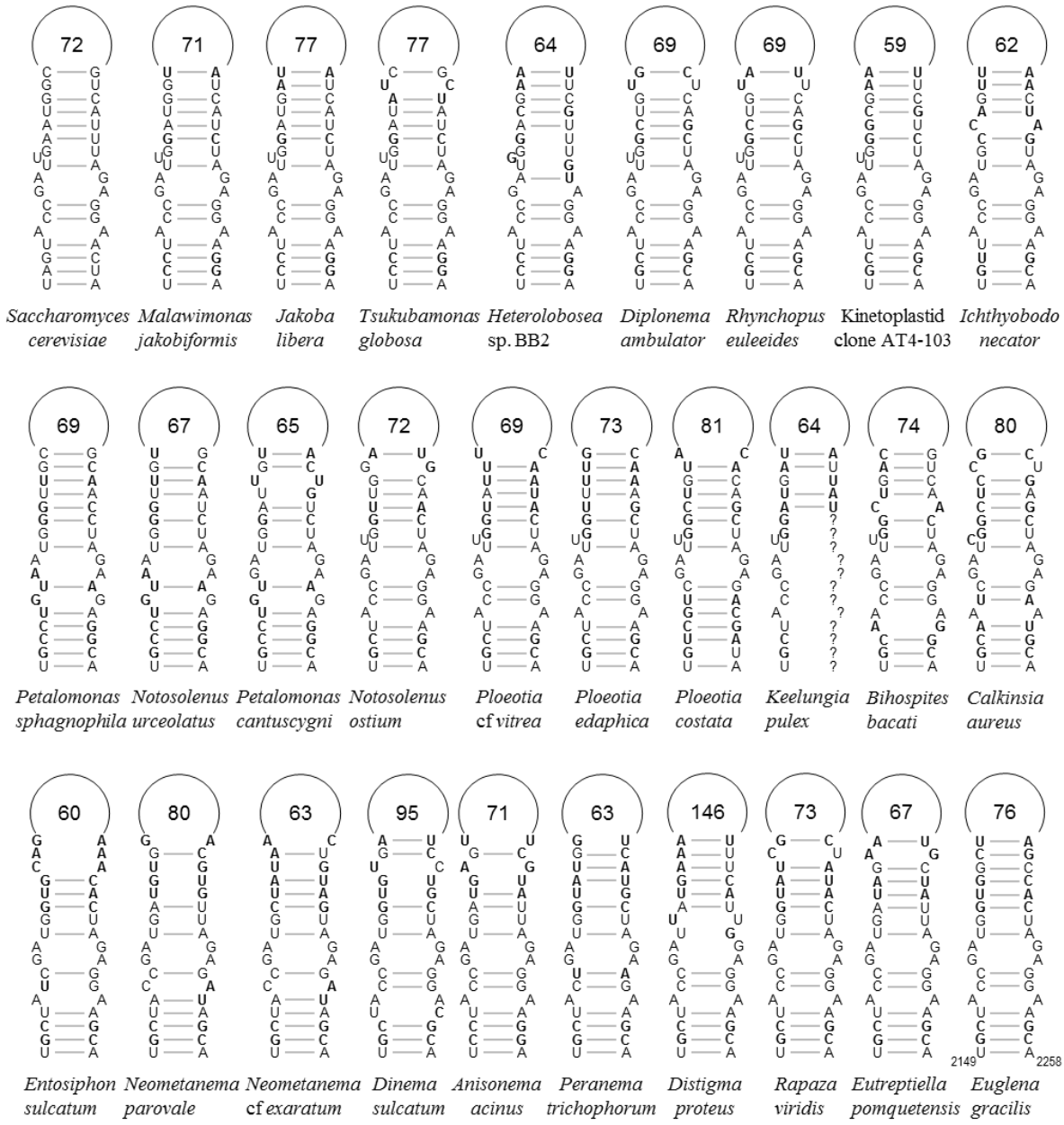


Fig. 8.7: Concluded secondary structures of SSU rRNA helix 44 illustrating nucleotide substitution differences between euglenozoan and outgroup taxa. A number in the apex of each helix represents remaining nucleotides of helix 44 and variable region 9. Small numbers represent nucleotide coordinates in the sequence of *Euglena gracilis* (lower right; referring to accession M12677). For a colored excerpt see Fig. 5.21.

Tab. 8.2: Table of SSU rDNA variable region length values of euglenozoan and outgroup taxa. * = V2 and V5 include additional helices, V2: h7-10 and V5: h26; see section 5.2.4.

No.	Taxon	V1	V2*	V3	V4	V5*	V7	V8	V9
	EUGLENEA								
1	*Khawkinea quartana*	11	288	71	626	60	167	95	48
2	*Euglena longa*	9	259	70	532	55	146	87	48
3	*Euglena gracilis*	9	216	70	513	54	121	86	44
4	*Euglena stellata*	9	202	70	503	53	131	87	34
5	*Euglena* cf *mutabilis*	12	305	71	594	55	147	95	27
6	*Colacium mucronatum*	9	195	70	425	53	107	84	29
7	*Colacium vesiculosum*	9	195	70	422	53	106	84	29
8	*Cryptoglena pigra*	10	199	71	493	52	145	83	28
9	*Cryptoglena skujae*	10	197	71	499	52	124	83	28
10	*Cyclidiopsis acus*	11	219	71	565	53	146	94	34
11	*Discoplastis spathirhyncha*	11	197	71	465	52	116	82	27
12	*Discoplastis* sp.	12	206	70	503	53	123	83	37
13	*Lepocinclis oxyuris*	13	213	70	574	53	143	89	85
14	*Lepocinclis spirogyroides*	10	200	70	470	54	115	88	31
15	*Monomorphina aenigmatica*	9	212	70	482	54	121	83	31
16	*Monomorphina pyrum*	9	224	78	494	61	164	91	33
17	*Phacus oscillans*	9	202	70	478	54	116	86	31
18	*Phacus pusillus*	9	202	71	473	54	116	83	31
19	*Strombomonas acuminata*	9	193	69	476	53	121	81	28
20	*Strombomonas verrucosa*	9	193	69	475	53	121	81	28
21	*Trachelomonas grandis*	12	330	71	669	59	447	97	87
22	*Trachelomonas hispida*	9	190	67	424	53	98	82	27
23	*Eutreptiella braarudii*	9	186	72	451	57	132	82	34
24	*Eutreptiella eupharyngea*	9	186	72	451	57	132	82	34
25	*Eutreptiella gymnastica*	9	185	72	441	57	130	81	81
26	*Eutreptiella pomquetensis*	9	186	72	449	57	132	82	35
27	*Eutreptia viridis*	10	267	71	387	66	100	142	78
28	*Eutreptia pertyi*	10	268	71	387	66	100	142	99
	APHAGEA								
29	*Rhabdomonas costata*	16	257	70	764	54	144	176	56
30	*Rhabdomonas incurva*	16	301	69	713	54	124	160	41
31	*Parmidium circulare*	14	312	69	713	54	114	169	60
32	*Parmidium scutulum*	14	320	70	699	55	126	167	58
33	*Menoidium pellucidum*	15	270	71	735	54	104	159	41
34	*Menoidium* sp.	15	275	71	731	54	103	160	42
35	*Gyropaigne lefevrei*	17	279	69	885	54	102	177	57
36	*Astasia curvata*	12	228	69	703	56	135	129	47
37	*Astasia torta*	16	329	70	835	54	153	163	76
38	*Distigma curvatum*	16	582	71	746	60	210	196	42
39	*Distigma sennii*	19	1258	70	1312	60	163	165	62
40	*Distigma proteus*	40	628	73	748	211	459	195	50
41	*Rapaza viridis*	12	191	72	432	74	108	89	41
	ANISONEMIDA								
42	*Anisonema* sp. *"Peranema"*	16	212	71	384	57	105	92	36
43	*Anisonema* sp. W1	15	212	71	478	52	130	93	55
44	*Anisonema* sp. U3	16	217	70	383	53	95	96	31
45	*Anisonema* sp. G1	-	-	71	384	58	105	92	37
46	*Anisonema acinus* B1	15	212	70	407	57	-	-	-
47	*Anisonema acinus*	14	208	70	712	58	112	91	39
48	*Dinema platysomum*	-	349	74	554	57	128	90	50
49	*Dinema sulcatum*	27	228	74	634	63	141	100	63
50	*Neometanema parovale*	-	-	-	523	60	112	123	48
51	*Neometanema* cf *exaratum*	13	232	71	382	57	107	113	31
52	*Heteronema scaphurum*	39	479	73	1321	77	-	-	-
53	*Peranema* sp. SP	10	180	67	457	55	101	98	31
54	*Peranema trichophorum*	10	180	67	457	57	101	98	31
55	*Entosiphon sulcatum*	34	216	70	479	49	107	64	28
56	*Entosiphon sulcatum*	34	216	70	479	49	107	64	28
57	*Entosiphon sp.*	34	215	72	479	49	107	64	28
	PLOEOTIIDA								
58	*Keelungia pulex*	10	167	68	390	59	95	80	32
59	*Ploeotia edaphica*	9	173	68	426	54	108	81	41
60	*Ploeotia* cf *vitrea*	8	180	69	395	54	100	81	37
61	*Ploeotia costata*	10	346	70	512	55	116	87	49
62	*Ploeotia costata* Pacific	10	515	70	644	55	125	87	49
63	*Ploeotia costata* Brackish	10	515	70	644	55	125	87	49

Tab. 8.2: Continued.

No.	Taxon	V1	V2*	V3	V4	V5*	V7	V8	V9
	PETALOMONADIDA								
76	*Notosolenus urceolatus*	7	168	69	367	52	102	75	35
77	*Petalomonas sphagnophila* Liz	13	173	68	355	54	100	76	37
78	*Petalomonas sphagnophila* HF	13	173	68	355	54	100	76	37
79	*Petalomonas sphagnophila* Dunc	13	173	68	344	54	100	76	37
80	Uncultured BLACKSEA cl 51	8	178	65	434	59	175	88	-
81	*Petalomonas cantuscygni* 2	11	164	66	354	53	97	81	33
82	*Petalomonas cantuscygni* 1	11	163	66	352	53	97	81	33
83	Uncultured sphenomonad CH1 S2 16	13	165	55	352	54	101	81	37
84	Uncultured sphenomonad CH1 S2 19	-	173	69	355	54	100	76	-
85	Uncultured sphenomonad CH1 S1 57	-	169	69	371	54	99	76	35
86	Uncultured eukaryote D2P04B10	-	-	68	358	54	96	82	33
87	Uncultured eukaryote D3P06F06	14	179	66	416	57	-	-	-
88	Uncultured sphenomonad PR3 3E 63	-	173	68	355	54	100	76	-
89	*Notosolenus ostium* U1	9	175	67	402	50	99	79	40
90	*Notosolenus ostium*	9	174	67	403	49	101	78	40
	SYMBIONTIDA								
91	Uncultured eukaryote SA2 3B11	-	-	68	360	61	88	80	32
92	Uncultured eukaryote M4 18H08	-	-	68	361	61	86	80	31
93	Uncultured eukaryote NA1 3E11	-	-	68	361	61	88	80	31
94	Uncultured eukaryote NA1 4B5	9	-	68	361	61	-	-	-
95	Uncultured eukaryote NA1 1G12	9	171	68	-	61	88	80	32
96	Uncultured eukaryote NA1 4H11	9	-	68	361	61	-	-	-
97	Uncultured eukaryote FV23 2D3C4	9	171	68	361	61	88	80	32
98	Uncultured BLACKSEA cl 50	8	171	68	360	61	87	80	-
99	Uncultured BLACKSEA cl 52	8	171	68	367	62	88	80	-
100	*Calkinsia aureus*	9	174	67	368	63	85	84	48
101	*Bihospites bacati* 2	8	170	70	360	59	89	80	41
102	*Bihospites bacati* 1	10	170	70	360	59	89	80	41
	DIPLONEMIDA								
103	*Diplonema ambulator*	9	176	67	376	52	90	82	37
104	*Diplonema papillatum*	9	174	63	371	45	90	82	36
105	*Diplonema sp.* ATCC 50225	9	175	67	377	52	93	82	37
106	*Diplonema sp.* ATCC 50232	9	174	66	369	51	83	82	37
107	Uncultured euglenozoan DH148-EKB1	8	169	68	380	49	92	82	36
108	Uncultured diplonemid LC22 5EP 17	8	169	68	382	50	93	82	36
109	Uncultured diplonemid LC22 5EP 18	8	169	68	378	50	92	82	36
110	Uncultured diplonemid LC22 5EP 19	8	169	68	385	50	92	82	37
111	Uncultured diplonemid LC23 5EP 5	8	176	66	357	52	79	82	36
112	Uncultured diplonemid PRTBE7438	8	169	68	-	50	92	82	36
113	Uncultured diplonemid RM2-SGM32	9	176	68	373	50	84	82	36
114	Uncultured eukaryote SCM15C6	8	169	68	380	50	91	82	36
115	*Rhynchopus* sp. SH-2004-I	9	175	68	359	51	71	82	37
116	*Rhynchopus* sp. SH-2004-II	8	174	67	363	50	88	82	37
117	*Rhynchopus* sp. ATCC 50229	9	174	67	363	53	77	82	33
118	*Rhynchopus euleeides*	9	176	66	373	50	82	82	37
	KINETOPLASTIDA								
119	Uncultured kinetoplastid AT4-103	16	173	60	345	48	85	81	27
120	*Perkinsiella*-like sp. AFSM3	11	175	68	319	57	76	77	34
121	*Perkinsiella*-like sp. PLO-DE4A	11	175	68	320	57	76	77	34
122	*Ichthyobodo necator*	10	162	63	332	66	98	79	30
123	Uncultured eukaryote L7.7	9	172	63	346	55	91	78	28
124	*Bodo edax*	7	179	71	370	131	95	89	25
125	*Bodo saltans*	7	182	70	367	134	94	89	25
126	*Crithidia fasciculata*	8	172	68	419	148	96	86	25
127	*Cryptobia bullockii*	7	161	65	340	139	93	84	26
128	*Dimastigella trypaniformis*	7	161	70	357	155	96	88	26
129	*Herpetomonas nabiculae*	7	174	68	422	158	96	87	26
130	*Neobodo designis*	7	169	73	375	143	95	89	25
131	*Leishmania major*	8	172	68	419	147	96	86	24
132	*Leptomonas collosoma*	9	149	71	405	152	96	85	25
133	*Phanerobia pelophila*	7	162	70	355	152	96	88	27
134	*Trypanosoma brucei*	7	195	71	396	170	109	95	25
	HETEROLOBOSEA								
135	*Heterolobosea* sp. BB2	64	145	65	463	47	230	144	31
136	*Heterolobosea* sp. RL	21	120	58	466	44	103	152	18
137	*Heterolobosea* sp. LO	57	121	58	469	45	104	151	18
138	*Pharyngomonas kirbyi* SD1A	31	123	58	458	45	103	149	21
139	*Acrasis helenhemmesae*	27	206	76	291	43	108	92	19

Tab. 8.2: Continued.

No.	Taxon	V1	V2*	V3	V4	V5*	V7	V8	V9
140	*Euplaesiobystra hypersalinica*	9	144	84	257	41	75	124	46
141	*Harpagon descissus*	21	90	83	314	43	76	84	23
142	*Heteramoeba clara*	6	107	74	194	42	69	102	21
143	*Naegleria gruberi*	25	196	82	300	45	104	82	20
144	*Pleurostomum flabellatum*	23	189	99	397	52	100	107	61
145	*Psalteriomonas lanterna*	8	115	88	327	44	84	92	16
146	*Stephanopogon minuta*	17	252	82	320	44	199	107	30
147	*Stygamoeba regulata*	11	139	67	357	45	123	70	45
148	*Tetramitus thermacidophilus*	39	217	79	338	43	117	94	28
149	*Paravahlkampfia ustiana*	20	114	81	270	45	94	96	16
150	*Vrihiamoeba italica*	20	88	78	238	49	83	76	16
151	*Tsukubamonas globosa*	12	152	70	294	43	112	75	45
	JAKOBIDA								
152	*Andalucia godoyi* 19	9	120	62	259	36	48	82	40
153	*Andalucia incarcerata*	9	148	63	282	42	122	75	44
154	*Jakoba libera*	10	128	63	254	45	59	69	46
155	Uncultured eukaryote EN351CTD039	9	135	66	225	43	52	70	41
156	Uncultured eukaryote FV23 CilE10	9	129	64	248	42	50	70	45
157	Uncultured eukaryote MA1 2H5L	9	129	64	-	42	50	70	45
158	Uncultured eukaryote SA1 1D05	9	126	64	233	43	57	67	44
159	*Seculamonas ecuadoriensis*	11	141	63	238	50	62	70	34
160	*Reclinomonas americana*	9	129	63	261	48	65	69	47
	OUTGROUP (Excavata)								
161	*Malawimonas jakobiformis*	9	147	65	227	42	109	69	39
162	*Pyrsonympha grandis*	10	153	63	478	46	63	68	48
163	*Dinenympha exilis*	10	149	63	452	46	63	68	50
164	*Oxymonas sp.*	11	158	66	518	42	92	69	42
165	*Trimastix pyriformis*	7	144	68	234	42	59	67	42
166	*Retortamonas sp.*	9	157	70	366	53	94	71	51
167	*Dysnectes brevis*	10	119	67	197	43	45	66	35
168	*Kipferlia bialata*	10	112	64	210	37	50	67	40
169	*Ergobibamus cyprinoides*	13	131	64	229	45	49	67	38
170	*Carpediemonas membranifera*	9	123	63	222	41	49	67	29
171	*Hicanonectes teleskopos*	9	122	60	201	36	46	64	31
	OUTGROUP (distant)								
172	*Proteromonas lacertae*	12	110	66	189	50	48	75	29
173	*Paramecium tetraurelia*	8	112	66	209	43	47	72	29
174	*Gymnodinium sanguineum*	9	136	65	223	44	51	69	37
175	*Chilomonas paramecium*	7	121	63	208	43	49	69	36
176	*Dictyostelium discoideum*	9	130	64	223	43	132	66	34
177	*Physarum polycephalum*	13	173	63	290	40	76	78	44
178	*Palmaria palmata*	10	110	67	221	42	53	64	39
179	*Acanthamoeba castellani*	11	196	69	361	128	154	133	60
180	*Cyrtohymena shii*	8	136	63	216	40	48	68	26
181	*Salpingoeca infusionum*	10	127	64	268	46	111	144	46
182	*Monosiga brevicollis*	10	122	66	220	43	49	67	37
183	*Byssoclamys spectabilis*	10	133	67	222	44	48	70	40
184	*Paramicrosporidium vannellae*	9	128	66	205	42	47	68	36
185	*Blastocystis* sp.	9	132	64	230	42	58	65	42
186	*Tetrahymena thermophila*	7	130	68	205	38	47	69	22
187	*Chaos nobile*	10	240	67	347	44	152	73	18
188	*Amoeba proteus*	10	221	68	143	44	38	74	62
189	*Plasmodium ovale*	12	144	67	287	76	187	120	36
190	*Chlorarachnion reptans*	4	201	67	230	45	64	67	32
191	*Pessonella* sp.	11	186	68	279	96	160	141	76

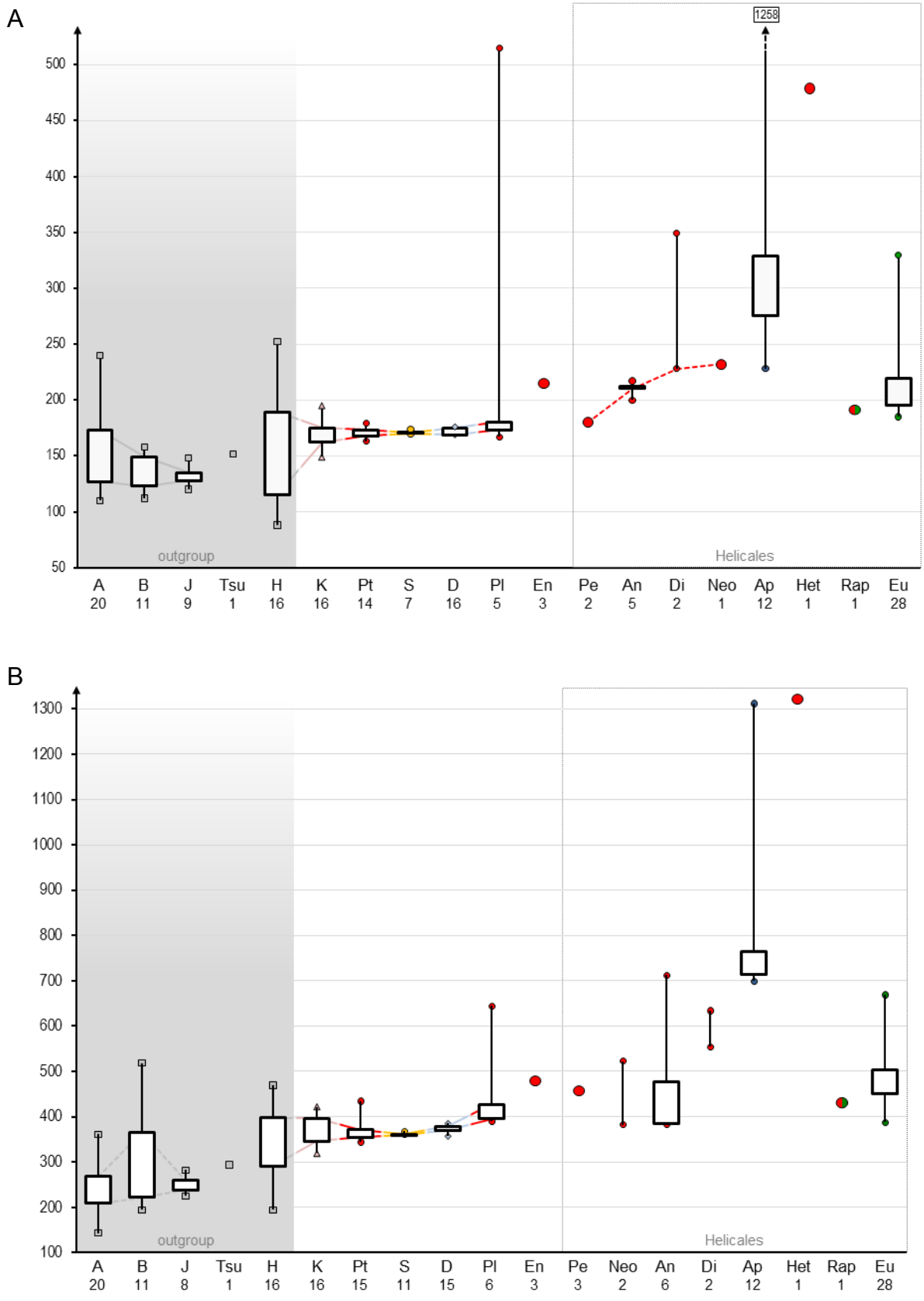


Fig. 8.8: SSU rDNA variable region graphs displaying length variations specific for major groups of the Euglenozoa and outgroups. For explanation of abbreviations for groups and taxa see Fig. 5.22 and text in 5.2.4. **A**: Length variation graph of variable region 2; note that the x-axis begins with value 50. **B**: Graph of variable region 4; note that the ordinate begins with value 100.

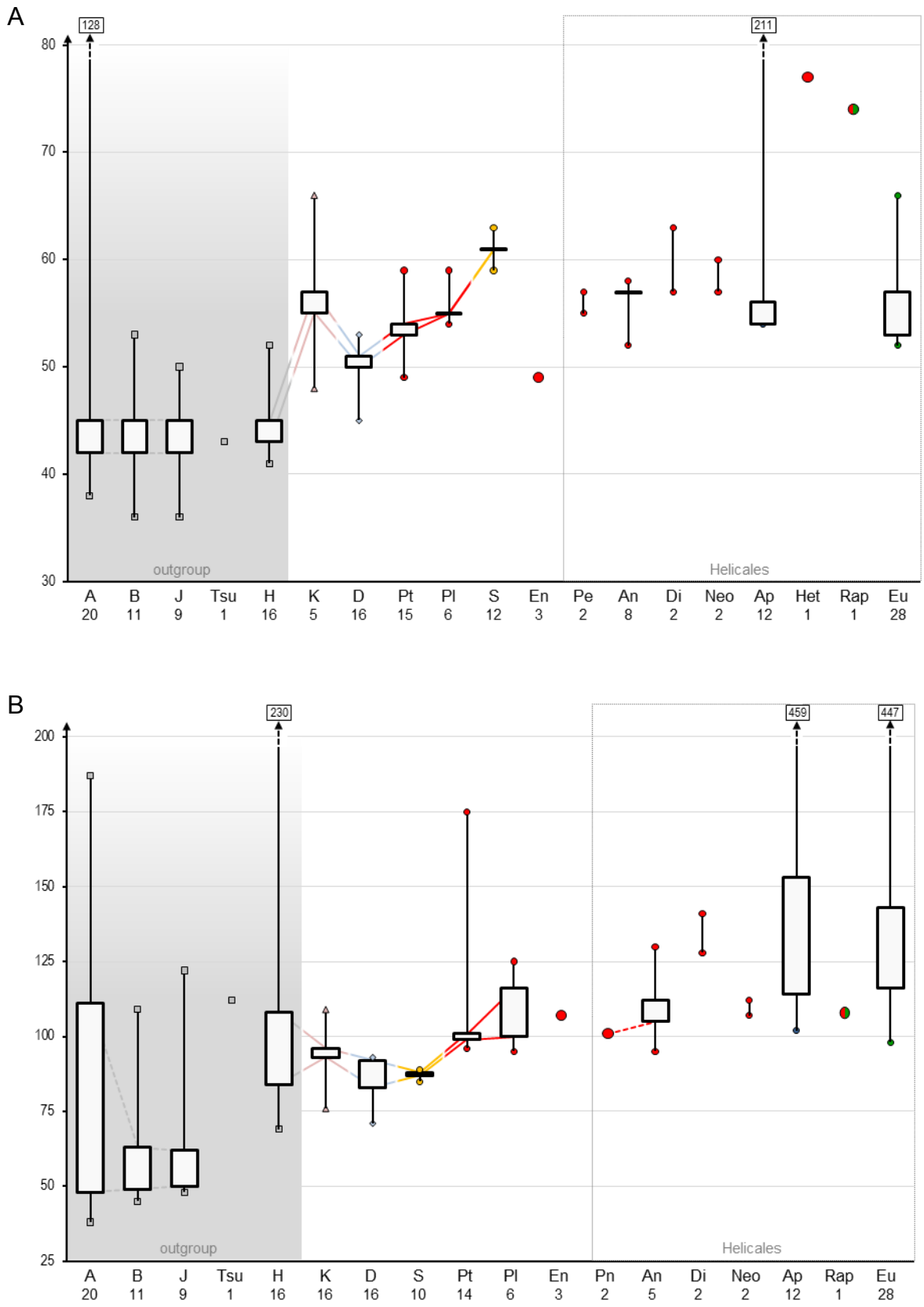


Fig. 8.9: SSU rDNA variable region graphs illustrating group specific length variations between euglenozoan and outgroup taxa. For explanation of group related abbreviations see Fig. 5.22 and text in 5.2.4. **A**: Length variation graph of variable region 5; note that the ordinate begins with value 30. **B**: Graph of variable region 7; note that the x-axis begins with value 25.

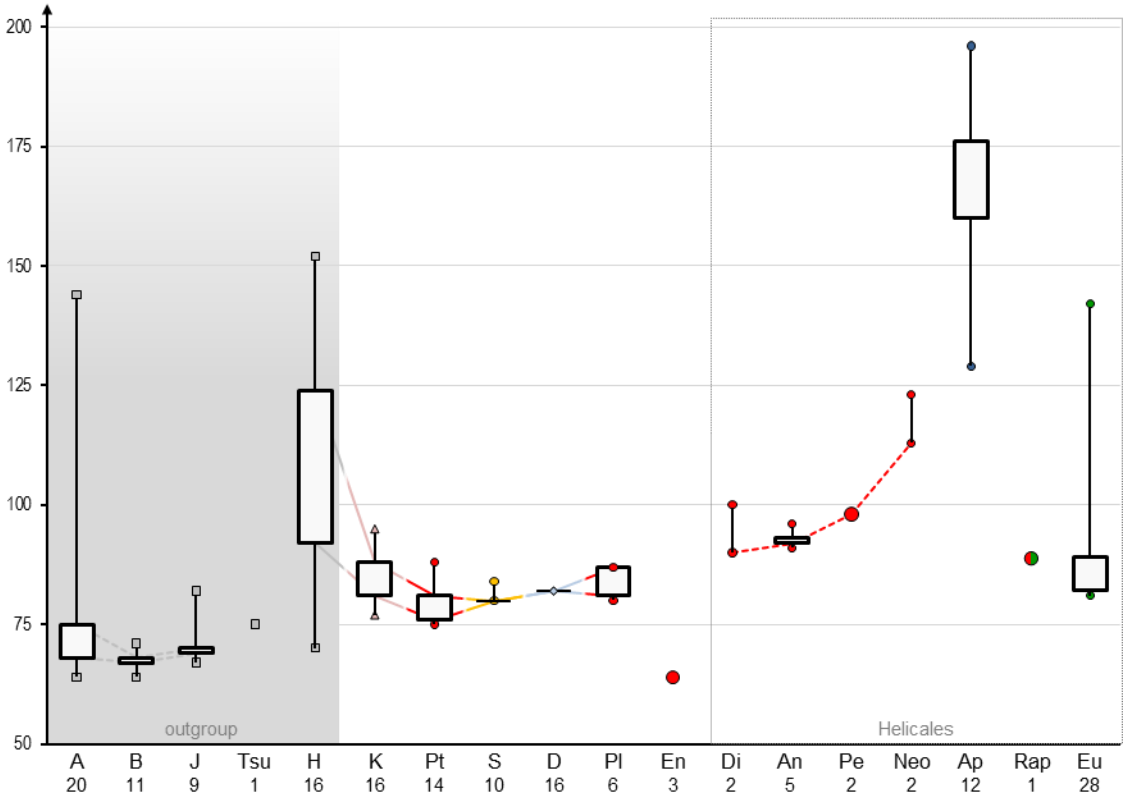


Fig. 8.10: SSU rDNA variable region graphs depicting length variations specific for major groups of the Euglenozoa and outgroup tax in variable region 8. For explanation of abbreviations for groups and taxa see Fig. 5.22 and text in 5.2.4.

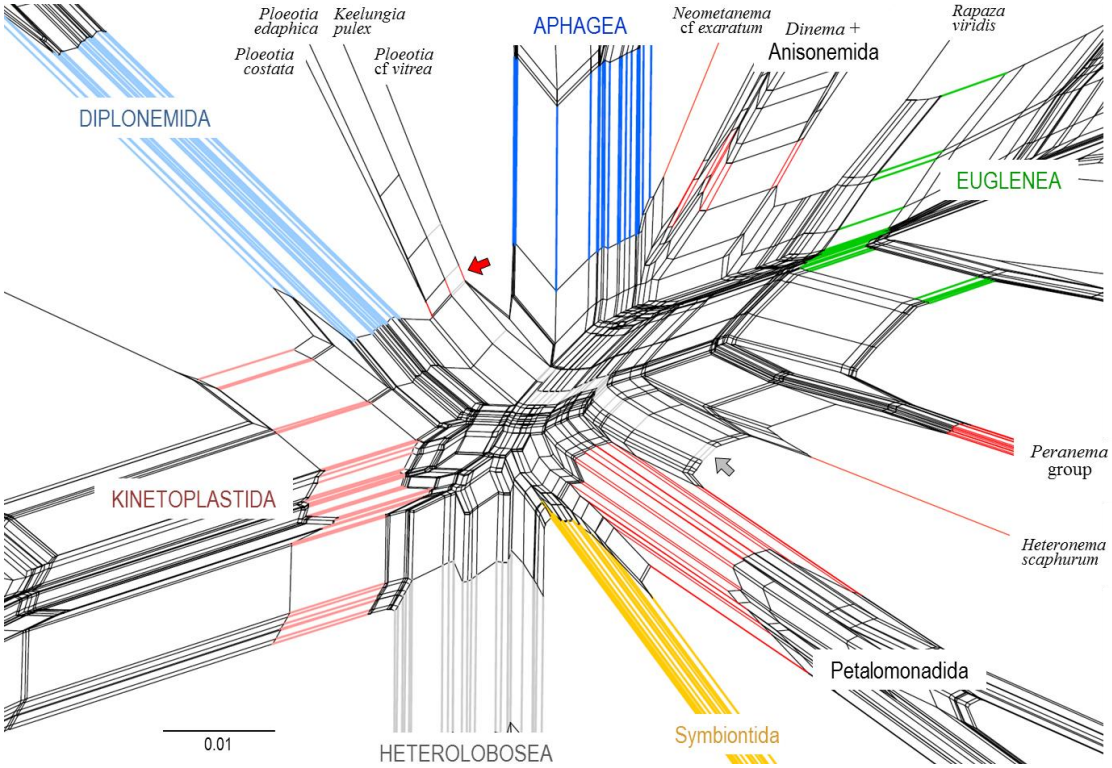


Fig. 8.11: Detailed center view on the neighbor-net graph obtained from analysis of modified dataset III comprising Euglenozoa and Heterolobosea as outgroup, but excluding *Entosiphon* sequences (i.e. subset B2, see Tab. 5.3). Network splits of monophyletic clades are colored, the scale bar represents 1 % sequence divergence. A red arrow marks common splits of the Ploeotiida; a grey arrow highlights splits which falsely support monophyletic Helicales including *Ploeotia* cf *vitrea*.

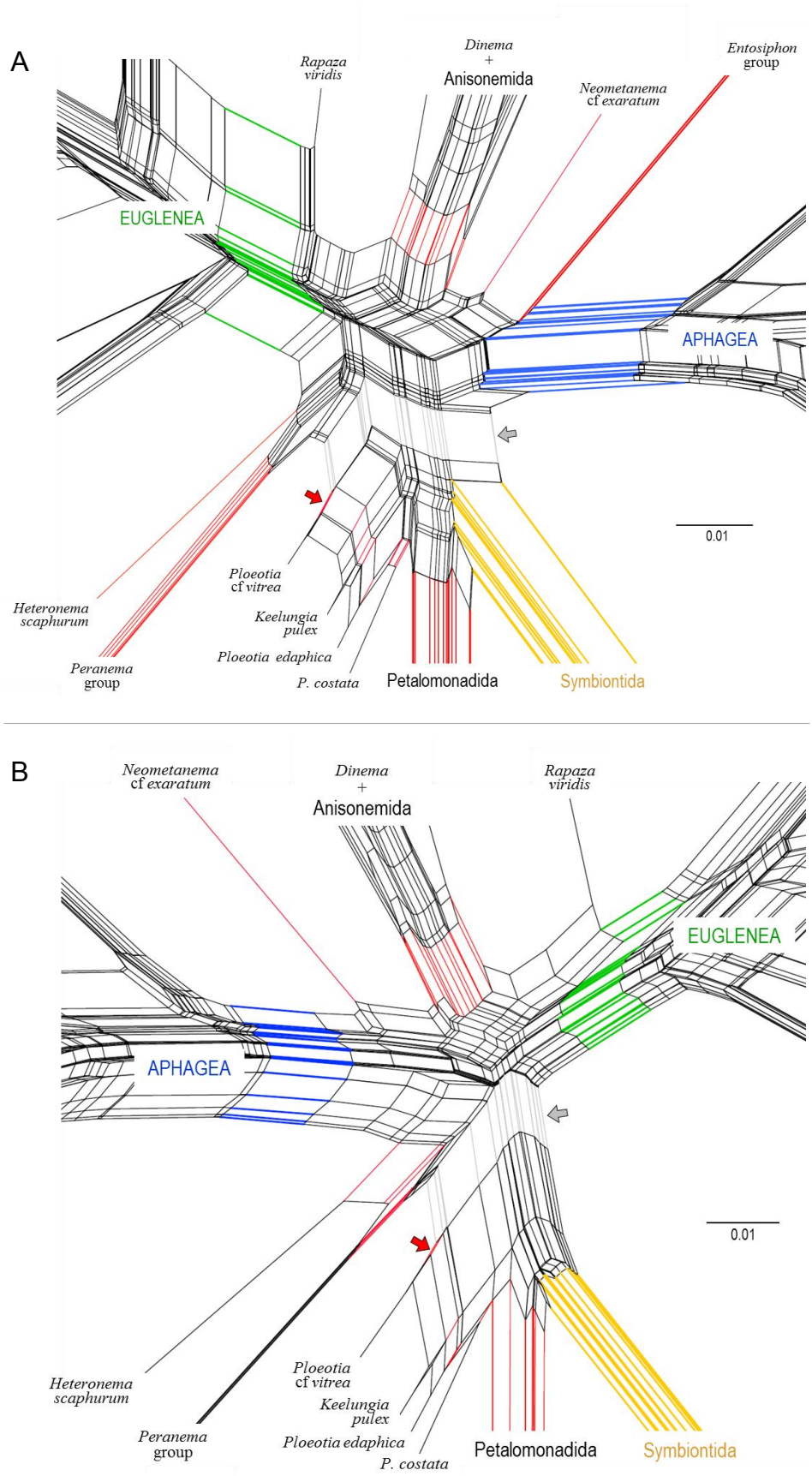


Fig. 8.12: Neighbor-net graphs obtained from network analysis of modified dataset III comprising Euglenida. Scale bars represent 1 % sequence divergence and network splits of monophyletic clades are colored. Red arrows mark common splits of monophyletic Ploeotiida and grey arrows highlight splits which support monophyletic Helicales. **A**: Subset D1 including *Entosiphon*. **B**: Subset D2 without *Entosiphon* (see Tab. 5.3).

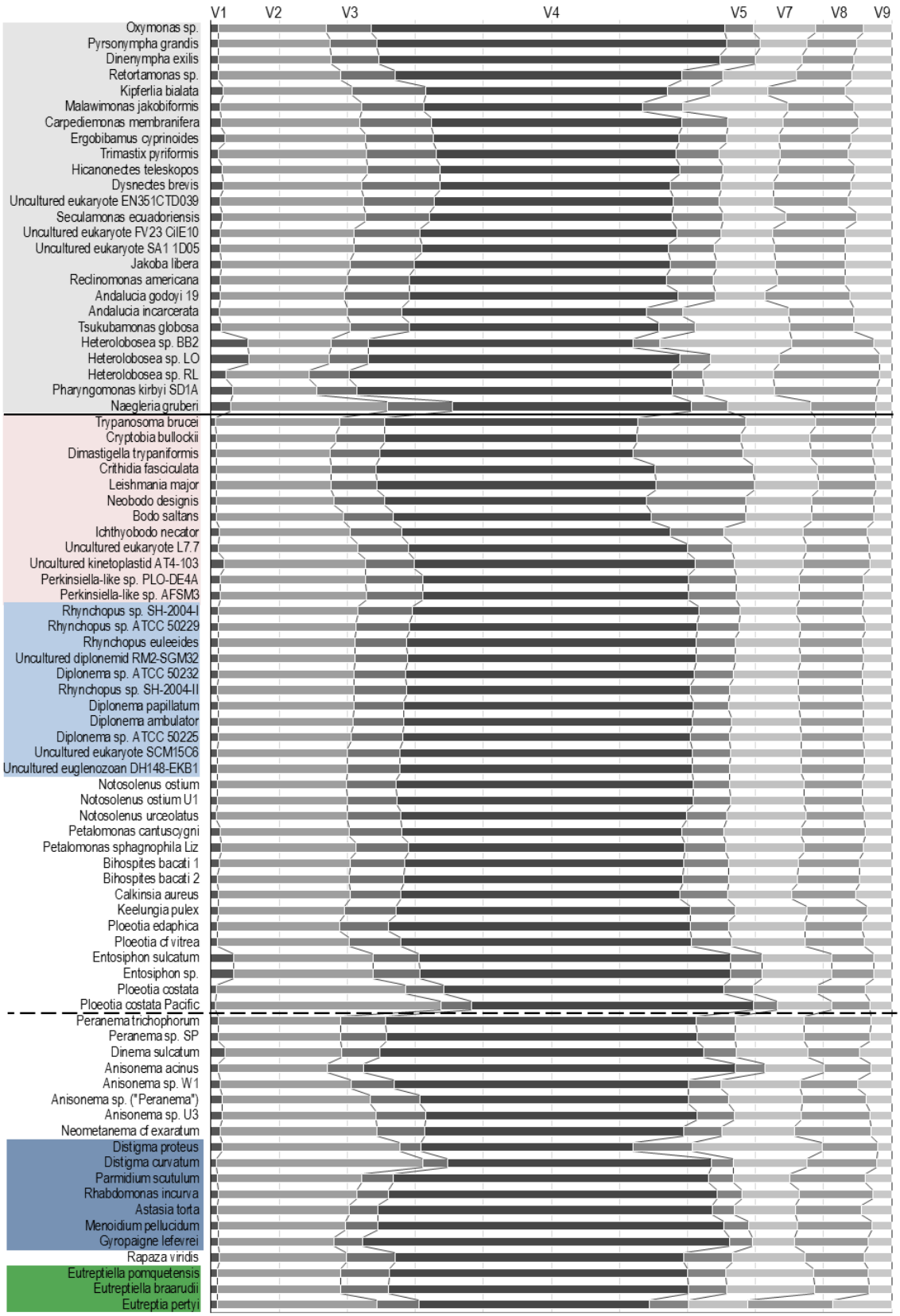


Fig. 8.13: Standardized variable regions graph illustrating relative proportions of SSU rDNA variable regions among major euglenozoan groups and excavate outgroup taxa. Nucleotide sequences of V1, V2, V3, V4, V5, V7, V8 and V9 were combined and normed to 100 % (each colored in analogous, varying grey tones from left to right). Taxa belonging to major groups differ in background colors: names of outgroup taxa are colored in light grey, Kinetoplastida in light red, Diplonemida in light blue, Aphagea in dark blue, Eutreptiales (i.e. primordial Euglenea) in green and phagotrophic euglenids in white. A black line divides outgroup from euglenozoan taxa, a dashed line separates Helicales from other Euglenozoa. For absolute variable region length values of primordial Euglenozoa see Fig. 5.26.

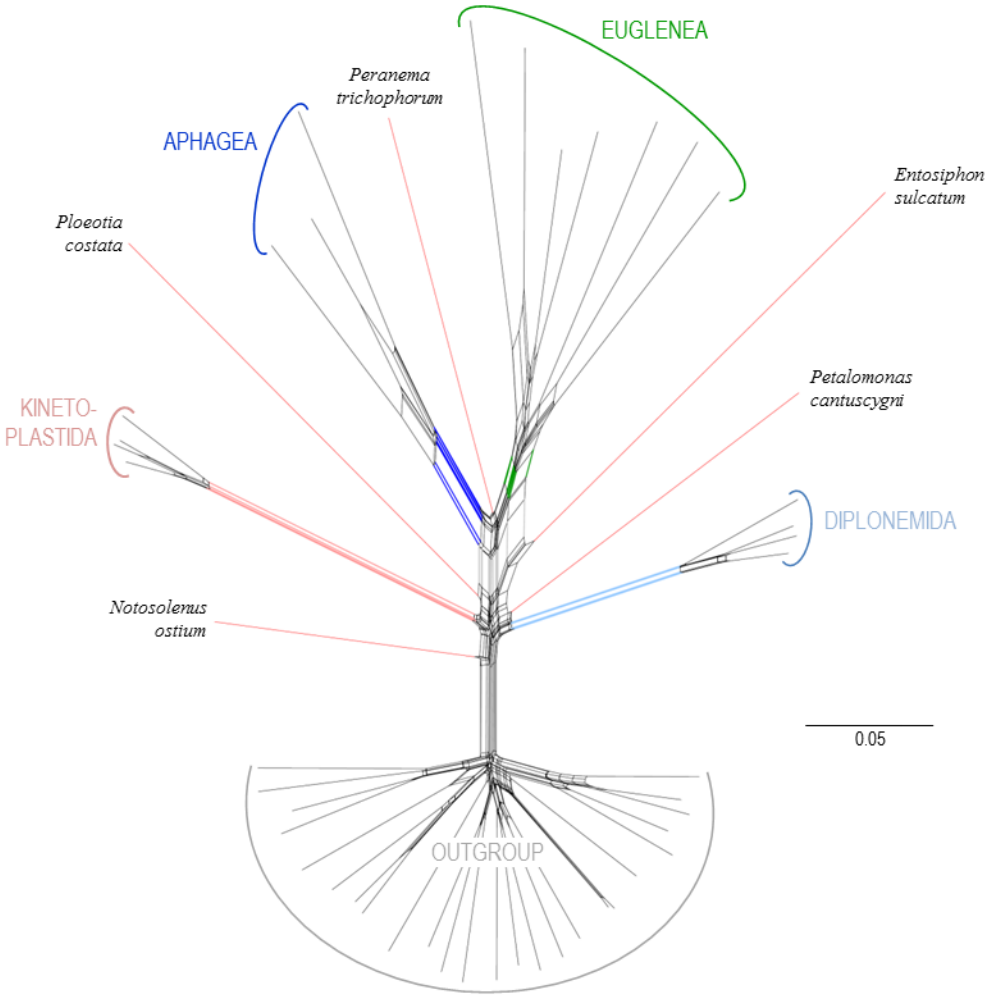


Fig. 8.14: Neighbor-net graph of LSU rDNA dataset IV comprising euglenozoan and outgroup taxa. Network splits supporting monophyletic euglenozoan clades are colored and the scale bar represents 5 % sequence divergence. Note paraphyly of Petalomonadida.

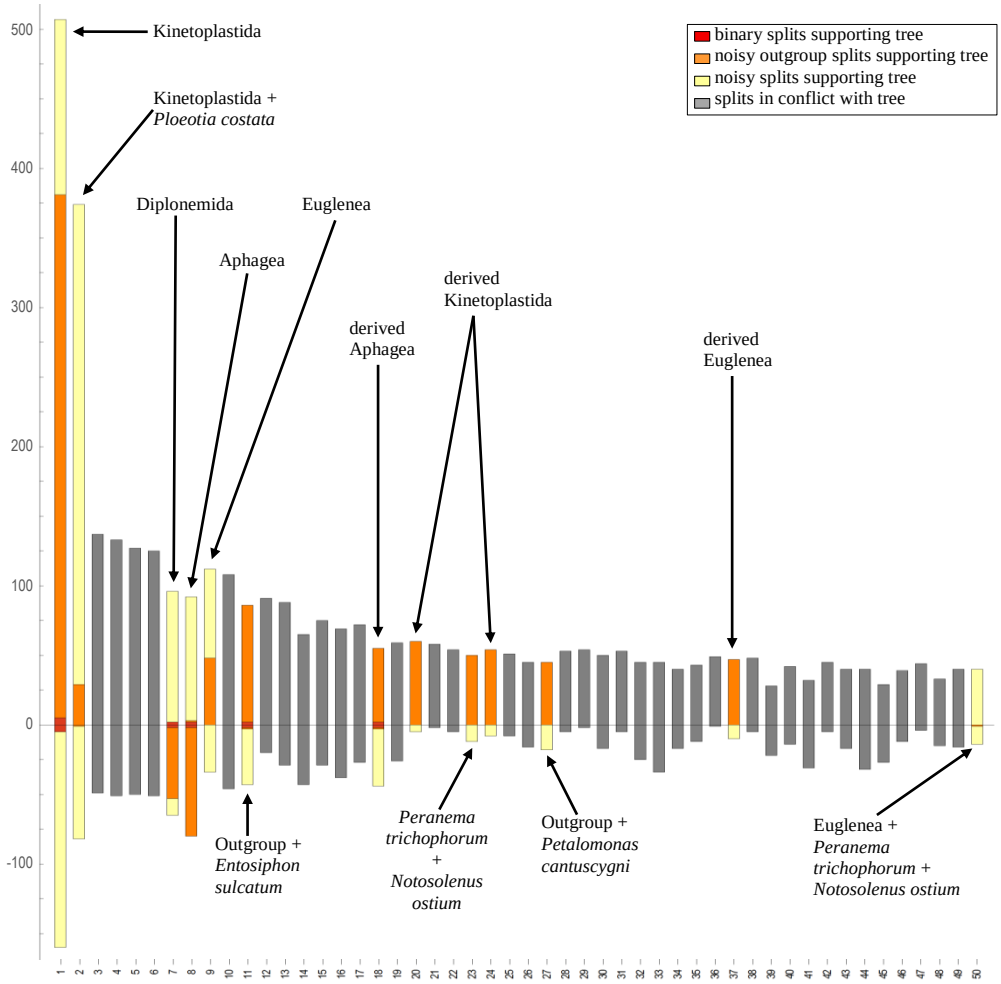


Fig. 8.15: Split support spectrum for LSU rDNA dataset IV displaying the 50 best splits accordant with neighbor-net graph in Fig. 8.14. Compatible splits referring to euglenozoan clades are marked by black arrows: well-known groups above, nonsense groupings below the splits graph. All conflicting splits show other nonsense correlations, e.g. outgroup with euglenozoan taxa.

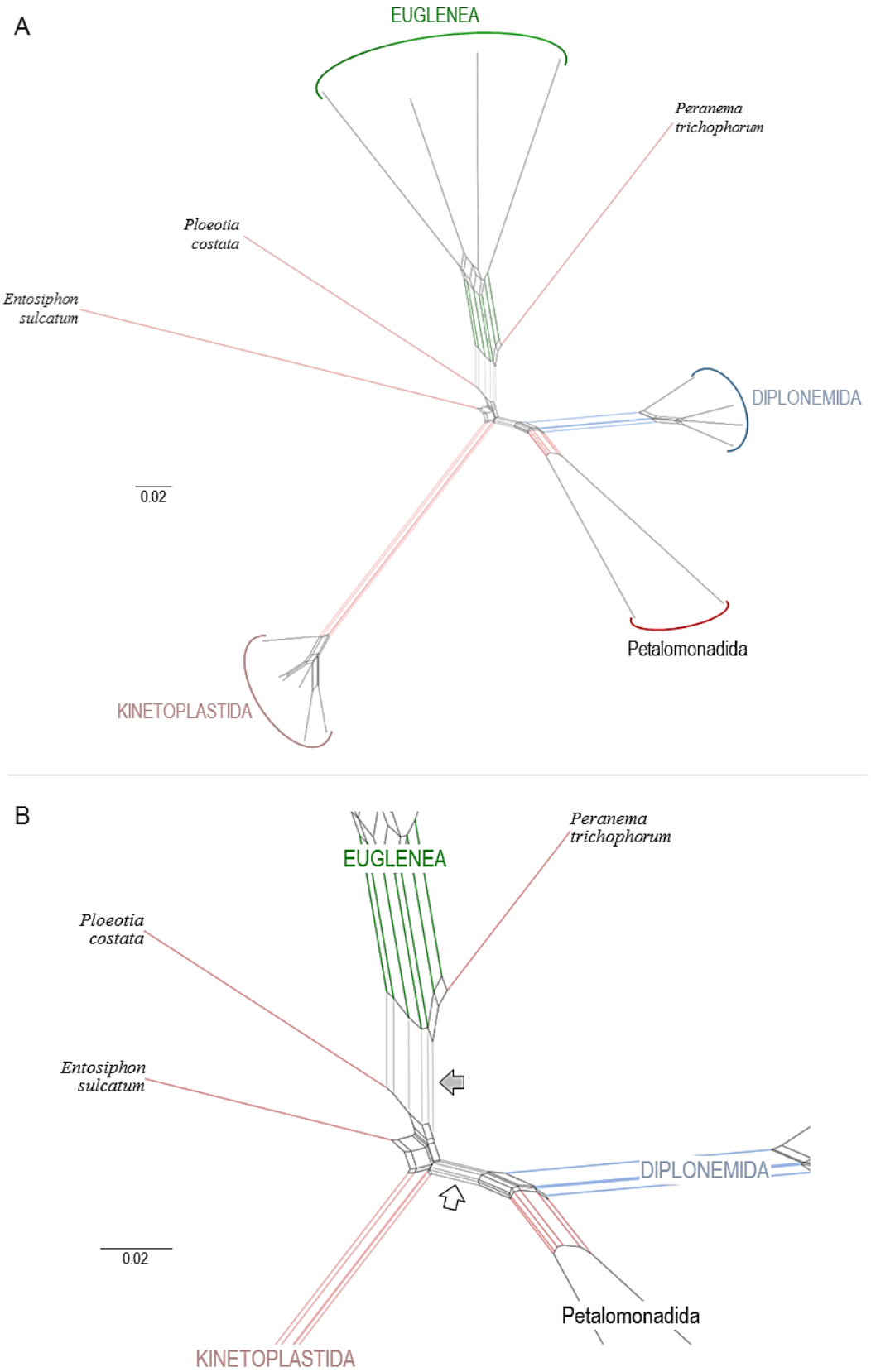


Fig. 8.16: Neighbor-net graph of modified LSU rDNA dataset V comprising Euglenozoa after exclusion of outgroup taxa. Network splits supporting monophyletic clades are colored. Scale bars represent 2 % sequence divergence. **A**: Splits graph overview displaying terminal splits. **B**: Detailed center view. Common splits of Diplonemida and Petalomonadida are marked by a white arrow. Splits supporting monophyletic Helicales, which comprise a *Peranema*/Euglenea clade, are highlighted by a grey arrow.

Tab. 8.3: Nucleotide composition of LSU rDNA sequences in dataset V. Percentages for nucleotides and GC-summaries are heated: highest percentages of each nucleotide are colored red, lowest values are green.

Taxon	T	C	A	G
Euglena gracilis	20,40	23,66	24,36	31,59
Eutreptia viridis	23,00	21,49	25,32	30,20
Eutreptiella pomquetensis	18,91	24,13	21,46	35,50
Eutreptiella braarudii	20,65	23,90	23,09	32,37
*Peranema trichophorum**	24,30	20,00	27,21	28,49
*Entosiphon sulcatum**	21,69	21,92	28,14	28,25
*Ploeotia costata**	19,14	25,06	24,83	30,97
*Petalomonas cantuscygni**	20,79	22,53	27,41	29,27
*Notosolenus ostium**	21,29	23,10	26,58	29,03
Uncultured clone Ma131 1A46	21,26	22,70	27,30	28,74
*Diplonema ambulator**	21,11	22,16	27,15	29,58
*Rhynchopus euleeides**	20,81	22,56	26,74	29,88
Diplonema papillatum	21,23	22,27	26,45	30,05
Bodo saltans	21,47	21,70	28,47	28,35
Dimastigella mimosa	22,02	22,54	27,92	27,52
Neobodo saliens	22,52	21,12	28,95	27,41
Trypanosoma brucei	21,58	22,62	26,68	29,12
Crithidia fasciculata	21,58	21,58	28,31	28,54
Leishmania major	21,46	21,81	28,19	28,54
Chytriomyces hyalinus	22,61	19,46	29,60	28,32
Saccharomyces cerevisiae	23,89	18,76	28,32	29,02
Codosiga gracilis	21,93	18,79	30,97	28,31
Capsaspora owczarzaki	21,35	20,77	29,93	27,96
Nephroselmis olivacea	21,81	21,58	27,73	28,89
Tetraselmis cordiformis	22,97	19,84	28,42	28,77
Average	21,58	21,84	27,17	29,41

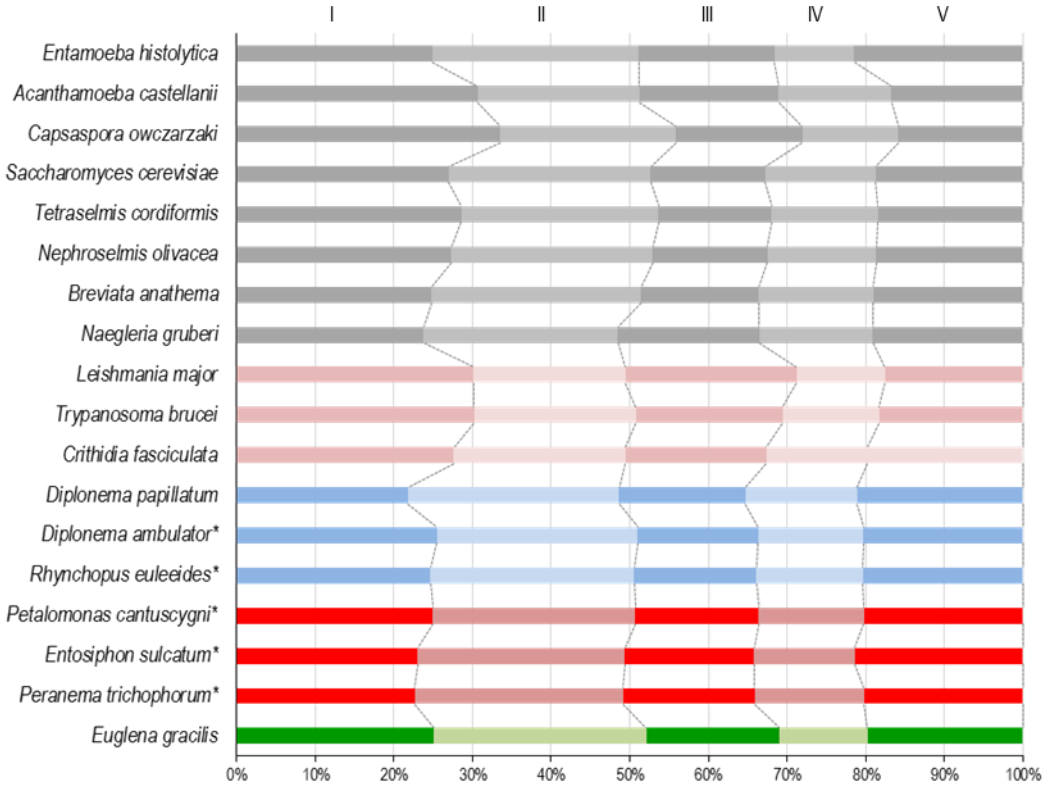


Fig. 8.17: Standardized LSU rDNA sequence length graph illustrating group specificity of LSU rDNA domain lengths among Euglenozoa compared to outgroup taxa. LSU rDNA domains I, II, III, IV and V are depicted as variegatingly colored bars, major groups are color-coded according to Fig. 5.37. Taxa which LSU rDNA sequence was obtained in the scope of this work are marked by an asterisk.

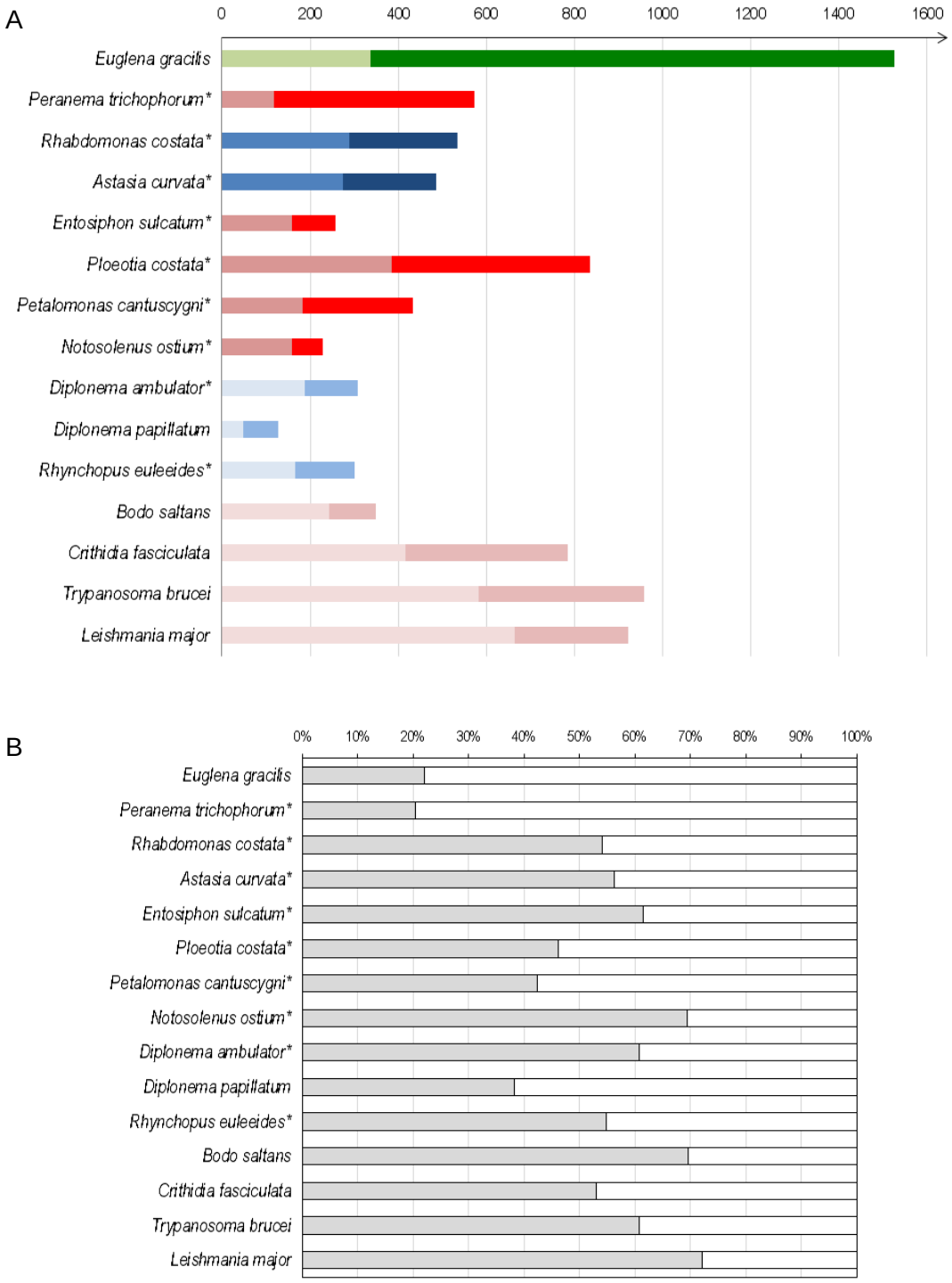


Fig. 8.18: **A**: Combined ITS sequences of euglenozoan taxa. Taxa which ITS sequences were obtained in this work are marked by an asterisk. ITS sequences were concatenated in taxon specific bars, ITS1 sequences are shown to the right in darker colors, corresponding ITS2 sequences to the left in lighter colors. Representatives of major groups are colored as in Fig. 5.37. **B**: ITS sequence proportions of selected Euglenozoa. ITS1 sequences depicted as white and ITS2 sequences as grey bars.

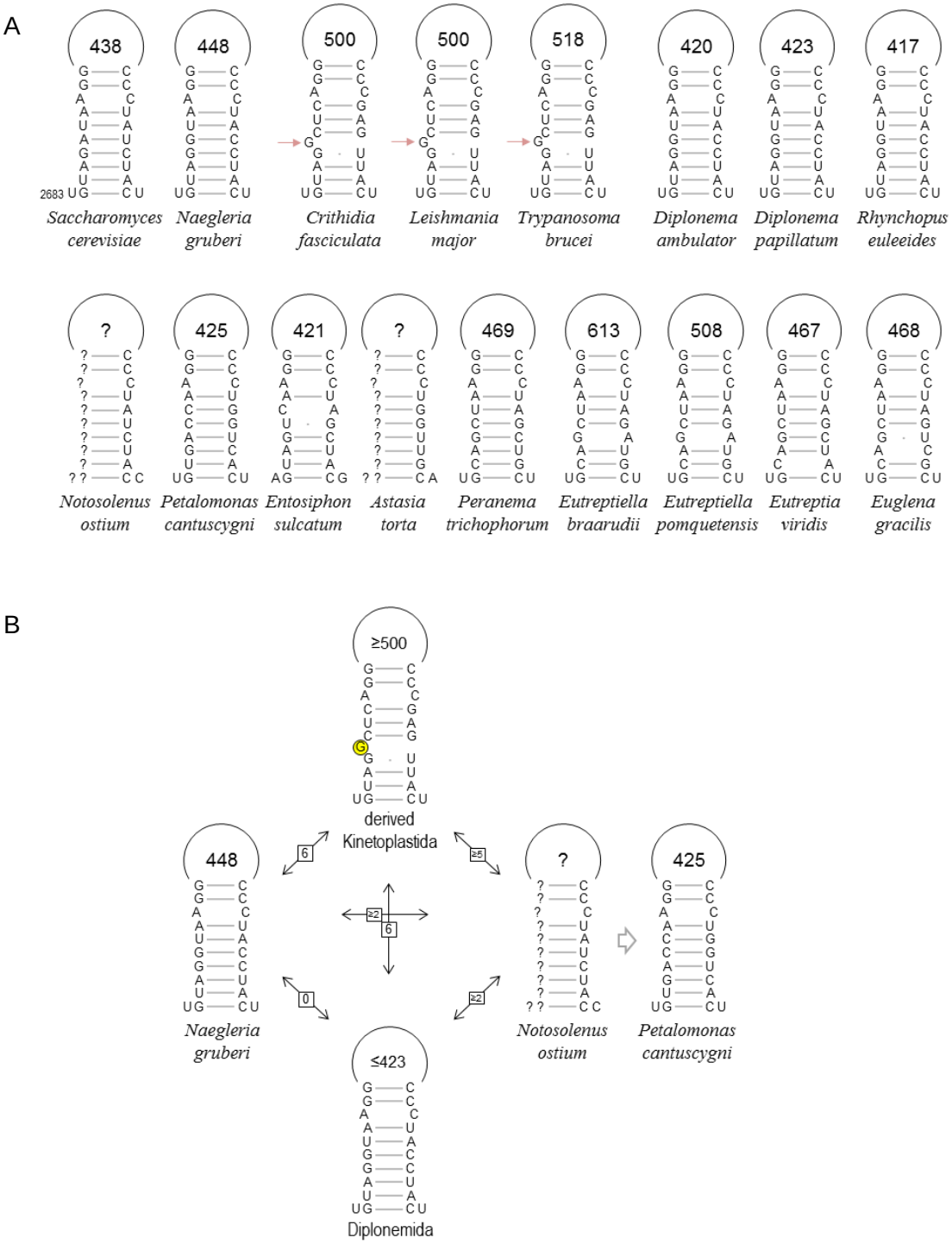


Fig. 8.19: Different nucleotide substitutions of euglenozoan and outgroup taxa in the inferred LSU rRNA secondary structure of helix 61. Unknown bases are depicted by question marks. **A**: Overview comprising all taxa examined. A group-specific unique substitution for Kinetoplastida, i.e. a single guanine, is marked by light-red arrows. Small numbers represent nucleotide coordinates in the LSU rDNA of *Saccharomyces cerevisiae*. **B**: Diagram illustrating nucleotide changes between derived Kinetoplastida, Diplonemida, Petalomonadida and *Naegleria gruberi*. Base changes between taxa are shown as numbers in boxed arrows. Group-specific unique substitution in derived kinetoplastids is highlighted by a yellow circle.

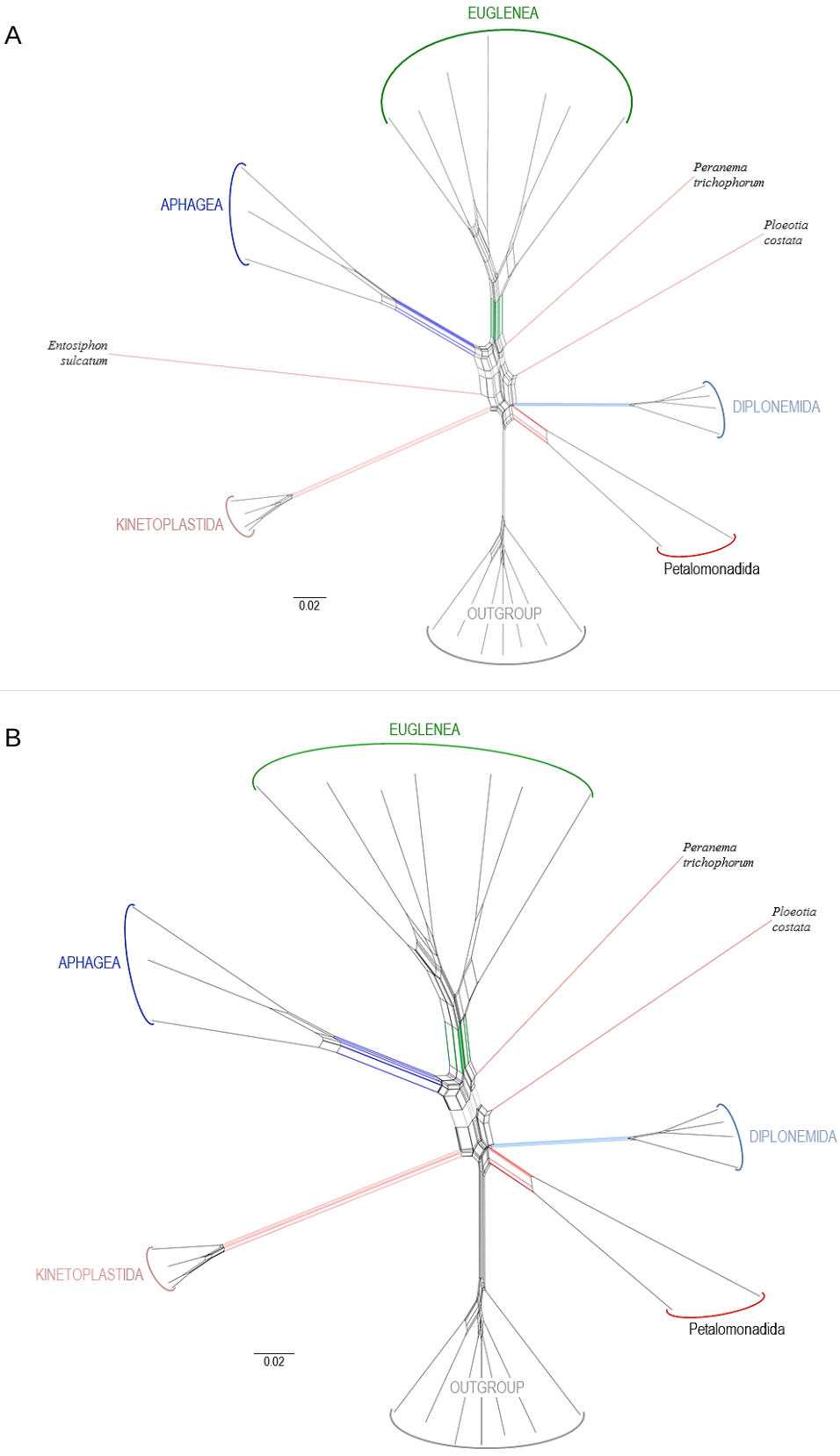


Fig. 8.20: Neighbor-net graphs of operon dataset VI comprising concatenated SSU and LSU rDNA sequences and displaying terminal splits. Network splits supporting monophyletic clades are colored. Scale bars represent 2 % sequence divergence. **A**: Network graph including the sequence of *Entosiphon sulcatum*. **B**: Network graph from reiterated analysis excluding *Entosiphon sulcatum*.

List of Abbreviations

A	Adenine
Amp	Ampicillin
BI	Bayesian inference
bp	Base pair
C	Cytosine
°C	Degree Celsius
cDNA	Complementary DNA
diH_2O	Purified (deionized) water
dsH_2O	Highly purified (deionized and sterilized) water
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine-tetraaceticacid
emend.	emended
Fig.	Figure
g	Gram
G	Guanine
GTR	General time reversible model
h	Hour
IGS	Intergenic spacer
ITS	Internal transcribed spacer
l	Litre
LB	Lysogeny broth
LSU	Large subunit
min	Minute
ML	Maximum likelihood
µ	Micro
m	Milli
n	nano
nt	Nucleotide
PCR	Polymerase chain reaction
pH	negative decadic logarithm of H^+ concentration
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rounds per minute
rDNA	Ribosomal DNA
rRNA	Ribosomal RNA
s	Second
S	Svedberg unit (sedimentation coefficient)
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SSU	Small subunit
T	Thymine

Tab.	Table
TAE	Tris-acetate-EDTA
Taq	DNA polymerase
Tris	Tris-(hydroxymethyl)-aminomethane
U	Unit
UV	Ultraviolet light
V	Volt
v/v	Volume per volume
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Acknowledgements

First of all I want to thank Prof. Dr. Gela Preisfeld for giving me the opportunity to work on this fascinating project with such independency. But never could I have finished this work without her reliance, generous advice and for her being in such good spirits (almost always).

I am very grateful to Prof. Dr. Heike Wägele for her kind willingness to review my thesis.

I want to thank ‘my lab students’, Olga Woltschanskaja and Martin Putsch, for their motivation and readiness. You vividly reminded me of my time as a fresh student – suddenly, I found myself ‘on the other side of the table’.

Also I want to express my deep gratitude to all colleagues (former and present) at the Preisfeld AG, your geniality and cooperativeness were responsible for creating a most friendly and motivating working experience for me.

I want to thank Tanja Eres wholeheartedly, for you raised this special feeling within me of having arrived safely.

Finally, and most importantly, I would like to thank my dear mother Bärbel Pärschke, without you none of this would have happened. Thank you for all your support, love, confidence and for everything.