Large-scale, Multi-gene Phylogenetic Analyses of Previously Overlooked Microeukaryotes: Toward Better Understanding of the Evolution and Diversity of Eukaryotes

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Science
(Doctoral Program in Biological Sciences)

January 2017

Euki Yazaki

Table of Contents

Abstract

One of the most intriguing questions in biology is how eukaryote was emerged from one of archaeal lineages. To answer that question, we need to understand the phylogenetic relationship among major lineages in eukaryotes, as well as the root position in the tree of eukaryotes, in detail. To infer the tree of eukaryotes precisely, the diversity of eukaryotes needs to be understood. Therefore, both phylogeny and diversity of eukaryotes are indispensable for the precise phylogeny of eukaryotes. It is widely accepted that unicellular eukaryotes (protists) occupy the vast majority of the diversity of eukaryotes. Moreover, studies incorporating metagenome approaches, which examine the nucleotide samples extracted from diverse environmental samples, suggested that our current knowledge on protist diversity is highly limited. In other words, novel protists have been left unknown in natural environments, and some of those are likely key species/linages to depict the origin and evolution of eukaryotes. In recent years, it became easy to generate a large scale sequence data (i.e. genomic and/or transcriptomic data) from a small quantity of nucleotide samples. The advance in sequencing technology allows us to infer the phylogenetic relationship amongst organisms from a large alignment comprising multiple genes. Such “phylogenomic analyses” are anticipated to resolve both relationship among anciently separated species/lineages and that among closely related species, which are often unresolved in phylogenies based on small scale alignments of a single gene or small number of genes. Therefore, We are in a position to subject novel protists and “previously known but overlooked protists,” of which phylogenetic positions have yet to be clarified, to large-scale sequencing analyses followed by phylogenomic analyses.

In this study, I described the results of three distinct phylogenomic analyses. Firstly, for better understanding of the evolution of parasitism in Kinetoplastea, I

generated transcriptomic data of two parasitic kinetoplastids, *Azumiobodo hoyamushi* and *Trypanoplasma borreli*, and conducted a phylogenomic analysis of 43 genes. Secondly, I investigated the precise phylogenetic position of strain PAP020, a previously undescribed protist, by a phylogenomic analysis based on the transcriptomic data generated in this study. Finally, the same experimental procedure was repeated on strain SRT308, another previously undescribed protist. I successfully elucidated the precise positions of the two novel protists in the tree of eukaryotes. Each of them turned out to be the earliest branch in the known protist assemblage or represent a novel assemblage. Combined, the results presented here contribute to unveil a part of eukaryotic diversity, and provide solid bases for future studies challenging the origin and evolution of eukaryotes.

Chapter 1 General Introduction

1-1 The Divergence of Eukaryote Have Been Undescribed

All the living organisms on Earth are classified into three domains, namely Bacteria, Archaea, and Eukaryota (Woese, Kandlert, and Wheelis 1990; Spang et al. 2015). The first eukaryote likely emerged within the extant archaeal lineages, and recent studies suggested that Lokiarchaeota was designated as the closest group of eukaryotes amongst the archael phyla known so far (Guy and Ettema 2011; Spang et al. 2015). The first eukaryote was most likely heterotrophic with the ability to engulf and digest other organisms for nutrients (phagocytosis). The phagocytotic capacity is critical to establish the mitochondrion through the endosymbiosis of an alpha-proteobacterium, which took place in the common ancestor of the extant eukaryotes and allowed the descendants to produce ATP molecules effectively via oxidative phosphorylation (Gray 1999; Gray, Burger, and Lang 2001). The mitochondrial endosymbiosis can be regarded as a large evolutionary step toward the current diversity of eukaryotes. Importantly, the mitochondria in eukaryotes adopted to anaerobic/microaerophilic environments are known to discard subsets of the canonical functions, and, in an extreme case, the organelle was seemingly eliminated from the cell (Tovar et al. 2003; Embley and Martin 2006; Muller et al. 2012; Makiuchi and Nozaki 2014). Another large step in early eukaryotic evolution is ‘primary endosymbiosis’ of a cyanobacterium (so-called primary endosymbiosis) that gave rise to the first plastid, which introduced a photoautotrophic lifestyle into eukaryotes (Gould, Waller, and McFadden 2008; Archibald 2009; Keeling 2010). Modern glaucophytes, red algae, and green plants (green algae plus land plants), which are collectively called Archaeplastida, are believed as the descendants of the first photosynthetic eukaryote with ‘primary plastid’ (Palmer 2003; Reyes-Prieto, Weber, and

Bhattacharya 2007) In latter eukaryotic evolution, multiple endosymbioses between green/red algae and diverse heterotrophic eukaryotes (so-called secondary endosymbioses) added new layers to the diversity of eukaryotes, yielding diverse photosynthetic lineages that are distantly related to Archaeplastida (Gould, Waller, and McFadden 2008; Archibald 2009; Keeling 2010). The resultant photosynthetic lineages bearing red/green alga-derived plastids (e.g., diatoms) occupy the current aquatic environment as the major primary producer. Similar to the evolution of mitochondria, secondary losses of photosynthetic capacity occurred frequently in eukaryotic evolution . Besides heterotrophic and photoautotrophic lifestyles, parasitic and symbiotic lifestyles have been emerged numerous times in the tree of eukaryotes, and increased the diversity of eukaryotes.

Combining both morphological and molecular data accumulated to date, eukaryotes described so far were proposed to coalesce into approximately twenty groups (Adl et al. 2012). Some of these groups show clear evolutionary affinity to each other, forming larger assemblages such as SAR composed of Stramenopiles, Alveolata, and Rhizaria (Burki et al. 2007), Archaeplastida composed of glaucophytes, red algae, and green plants (Adl et al. 2005, 2012), and Opisthokonta composed of Metazoa, Fungi (Cavalier-Smith T 1987; Adl et al. 2005), and other unicellular eukaryotes (protists). Some of the large eukaryotic assemblages are still controversial and need further examination. For instance, Excavata was proposed to comprise Metamonada, Discoba, Preaxostyla, and Malawimonadida, but it is difficult to recover their monophyly in molecular phylogenetic analyses (Hampl et al. 2009). Likewise, a putative assemblage including cryptophytes, haptophytes, and their close relatives (so-called Hacrobia) has yet to be fully confirmed (Burki et al. 2016). Most importantly, there are many groups of

which phylogenetic positions remain uncertain (Philippe and Laurent 1998; Roger 1999; Cavalier-Smith 2002; Yabuki et al. 2012; Yabuki, Ishida, and Cavalier-Smith 2013). There are two possibilities for these 'orphan' groups. Firstly, some orphan groups can be a part (probably a deep branch) of a large assemblage that has already defined (Yabuki, Inagaki, and Ishida 2010; Yabuki et al. 2014). Alternatively, some of them can belong to a large assemblage, which we have not overlooked to date (Yabuki et al. 2011; Kamikawa et al. 2014). Thus, depicting the precise phylogenetic positions of orphan groups is indispensable to understand the diversity and evolution of eukaryotes sufficiently.

1-2 Importance for phylogenetic analyses of large-scale multi-gene alignments

Morphological data including ultrastructural data have been used to examine the 'relatedness' between organisms. After nucleotide sequence data became available for diverse organisms, molecular phylogenetic techniques have been used to infer the evolutionary paths underlid the current organismal diversity. Phylogenetic analyses based on molecular data (i.e. nucleotide and amino acid sequences) is particularly useful for inferring the relationship amongst microbes that are poor in and/or difficult to extract morphological data. Genes for phylogeny need to be conserved among broad range of organism. For instance, small and large subunits ribosomal DNAs (SSU and LSU rDNAs), tubulins, translation elongation factors (EF-1α and EF-2), and heat shock proteins (HSP70 and HSP90) are considered as major phylogenetic markers for eukaryotes, and pioneering analyses based on these markers successfully provided insights into eukaryotic evolution (Baldauf et al. 2000; Philippe et al. 2000; Sakaguchi et al. 2005; Simpson, Stevens, and Lukeš 2006). However, phylogenetic studies based on single or small number of highly conserved genes failed to resolve the precise relationships among anciently separated

species, as phylogenetic information in the gene/genes analyzed were most likely limited (Lopez, Forterre, and Philippe 1999; Roger 1999; Stiller and Hall 1999; Cavalier-Smith 2002; Lartillot et al. 2007). The shortage of phylogenetic information can be also problematic in resolving the relationship among recently separated (i.e. closely related) species (Hasegawa, Kishino, and Yano 1989; Huynen and Bork 1998; Doolittle 1999; Forterre and Philippe 1999). In theory, the above issue can be overcome by analyzing a larger number of genes. Fortunately, it became realistic to analyze alignments, which are composed of a large number (typically more than 100) of genes, as genomic and/or transcriptomic data from diverse eukaryotes have been available in recent years (Brown et al. 2001; Zhao, Davis, and Lee 2005).

In the last decade, phylogenetic analyses of large-scale alignments including more than 100 genes offered the opportunities of assessing previously unresolved issues in eukaryotic phylogeny and novel insights into eukaryotic evolution. For instance, Hampl et al. (2009) assembled an alignment of 143 genes to examine the monophyly of Excavata, which were proposed mainly from a series of shared morphological characteristics. The 143-gene analyses failed to verify or reject the monophyly of Excavata as a whole, but provided a solid basis for latter studies on the origin and evolution of excavate protists. A 250-gene alignment prepared and analyzed by Burki et al. (2016) proposed a clade comprising both photosynthetic groups (i.e. cryptophytes and haptophytes) and heterotrophic groups (e.g., centrohelids and *Telonema*), of which positions in the tree of eukaryotes were uncertain prior to the particular study. If both cryptophytes and haptophytes truly share the most recent common ancestry with heterotrophic protists in the tree of eukaryotes, this large assemblage (so-called Hacrobia) is significant to retrace the evolution of photosynthesis in eukaryotes. Besides the two

studies described above, many phylogenetic studies based on large-scale alignments are available in the literature, and currently regarded as an indispensable approach to assess the relationship among anciently separated species (Burki et al. 2007; Brown et al. 2012; Pánek et al. 2016; Janouškovec et al. 2017).

1-3 Importance of protists previously overlooked in natural environments

Phylogenetic analyses of large-scale multi-gene alignments are seemingly the most reliable way to reconstruct the tree of eukaryotes among the methods currently available, and indeed resolved ancient splits with confidence (see the examples described in the previous section). Nevertheless, there is another major issue to be solved to understand the true diversity and evolution of eukaryotes. It is too naïve to consider that the previously described species sufficiently represent the organismal diversity on Earth. Previously undescribed organisms in environments have been (and still are) identified and isolated under the microscope to establish laboratory cultures, and some of these strains appeared to represent phylogenetically novel lineages (Takishita et al. 2007; Marande, López-García, and Moreira 2009; Bork et al. 2015). There is an alternative approach to evaluate the organismal diversity in the environment of interest. This approach does not need to establish laboratory cultures, and amplifies major genes for phylogenetic analyses (e.g., SSU rDNA) from the nucleotide sample extracted from natural environments. Such 'environmental PCR' approach often detects the sequences bearing no clear evolutionary affinity to any sequence of previously described species, albeit the cellular identifies related to the amplified sequences remain unclear (Takishita et al. 2007; Marande, López-García, and Moreira 2009; Bork et al. 2015). The results from both culture-dependent and culture-independent approaches consistently suggest

that a substantial proportion of the organismal diversity has been overlooked in environments, and some of the species overlooked in natural environments may provide important clues to depict unresolved issues in eukaryotic evolution. Nevertheless, the precise phylogenetic positions of enigmatic protists, which may contribute to understand the diversity and evolution of eukaryotes, tend to be unresolved in single-gene and small-scale multigene phylogenetic analyses, and need to be subjected to large-scale multigene phylogenetic analyses (see the next paragraph for the examples).

Tsukubamonas globosa is a heterotrophic excavate protist branching at the base of the clade of heterolobosean and euglenozoans (Kamikawa et al. 2014). The deep branching nature of *Tsukubamonas* in the Discoba clade was inferred by analyzing a 157-gene alignment, and is recognized as a key species to understand mitochondrial genome evolution (Kamikawa et al. 2014). Significantly, the precise position of *Tsukubamonas* has not been pinpointed by analyzing single-gene (SSU rDNA) or 5-gene phylogeny(Yabuki et al. 2011). *Palpitomonas bilix* is another heterotrophic protist that was overlooked until recent in environments and appeared to occupy an interesting position in the tree of eukaryotes (Yabuki et al. 2010; Yabuki et al. 2014). The ultrastructural characteristics of *Palpitomonas* were unique and showed no clear phylogenetic affinity to any previously described eukaryotes(Yabuki, Inagaki, and Ishida 2010). The position of *Palpitomonas* remained unclear in SSU rDNA and six-gene phylogenetic analyses (Yabuki, Inagaki, and Ishida 2010), but settled by analyzing a 157-gene alignment (Yabuki et al. 2014). In the 157-gene phylogeny, *Palpitomonas* branched at the base of the Cryptista clade comprising photosynthetic species (cryptophytes) and heterotrophic species (gonimonads and kathablepharids) (Yabuki et al. 2014) suggesting that *Palpitomonas* holds clues to retrace the origin of the plastids in cryptophytes.

As described above, phylogenetic analyses of large-scale multigene alignments is a powerful method to elucidate the precise positions of the organisms of interest (see chapter 1-2), and previously overlooked organisms are most likely indispensable for better understanding of the diversity and evolution of organism on Earth (see chapter 1-3). In this thesis, I reported the results of large-scale multi-gene phylogenetic analyses on three of previously overlooked protists, and discuss the implications deduced from their phylogenetic positions. In chapter 2, the global phylogeny of Kinetoplastea was revisited by incorporating new sequence data from three parasitic species (*Azumiobodo hoyamushi*, *Trypanoplasma borreli*, and, *Perkinsella* sp.), and discuss the evolution of parasitic lifestyles in this particular protist group. In chapters 3 and 4, the phylogenetic analyses of previously undescribed protists, strains PAP020 and SRT308, which were recently isolated from two distinct locations in Republic of Palau, are described. My analyses revealed strain PAP020 as a basal branch of Fornicata (Metamonada, Excavata), and strain SRT308 represent a lineage that is closely related to but clearly distict from Euglenozoa.

Chapter 2 Global Kinetoplastea phylogeny inferred from a large-scale, multi-gene alignment including parasitic species and transitions from a free-living to a parasitic lifestyle

2-1 Abstract

All members of Trypanosomatida known to date are parasites that are most likely descendants of a free-living ancestor. Trypanosomatids are one of the models to assess the transition from a free-living to a parasitic lifestyle, because a large amount of experimental data has been accumulated for well-studied members that are harmful to humans and livestock (*Trypanosoma* spp. and *Leishmania* spp.). However, recent advances in my understanding of the diversity of trypanosomatids and their close relatives (i.e. members of the class Kinetoplastea) suggested that the change in lifestyle took place multiple times independently from that gave rise to the extant trypanosomatid parasites. In the current study, transcriptomic data of two parasitic kinetoplastids belonging to orders other than Trypanosomatida, namely *Azumiobodo hoyamushi* (Neobodonida) and *Trypanoplasma borreli* (Parabodonida), were generated. I here re-examined the transition from a free-living to a parasitic lifestyle in the evolution of kinetoplastids by combining (i) the relationship among the five orders in Kinetoplastea and (ii) that among free-living and parasitic species within the individual orders. The former relationship was inferred from a large-scale multi-gene alignment including the newly generated data from *Azumiobodo* and *Trypanoplasma*, as well as the data from another parasitic kinetoplastid *Perkinsela* sp. deposited in GenBank database, and the latter was inferred from a taxon-rich small subunit ribosomal DNA alignment. Finally, I discuss the potential value of parasitic kinetoplastids identified in Parabodonida and Neobodonida for studying the evolutionary process that turned a free-living species into a parasite.

2-2 Introduction

Trypanosomatid flagellates have been studied extensively, as some of them are causative agents of human African trypanosomasis (sleeping sickness), Chagas disease, and leishmaniasis. Besides their clinical importance, these flagellates possess intriguing properties that are shared by only a few or none of other eukaryotes. Trypanosomatids are known to possess a unique peroxisome-derived organelle (glycosomes) that encloses glycolytic enzymes (Opperdoes and Borst, 1977; Gualdrón-López et al., 2012). Mitochondria of trypanosomatids contain a complex network of two types of circular DNA molecules, maxicircles and minicircles, and their mitochondrial mRNAs undergo complex and distinctive editing prior to translation (Lukeš et al., 2002, 2005). The 5′ termini of mRNAs from trypanosomatid nuclear genomes also undergo post-transcriptional modification (Campbell et al., 2003; Michaeli, 2011). Although the properties described above were observed in phylogenetic relatives of trypanosomatids (i.e. members of other orders in the class Kinetoplastea; see below), trypanosomatids, for which various experimental techniques in molecular and cell biology are available, have been the center of the research on Kinetoplastea.

Trypanosomatida, together with Eubodonida, Parabodoida, Neobodonida, and Prokinetoplastida, comprise the class Kinetoplastea (Moreira et al., 2004; Simpson et al., 2006). All the known members of Trypanosomatida and those of Prokinetoplastida are parasites (Simpson et al., 2006; Lukeš et al., 2014). However, the remaining three orders are dominated by free-living members, and only a few or none of the members are known to be parasitic. Previously published phylogenies of small subunit ribosomal DNA (SSU rDNA) sequences constantly and robustly united Neobodonida, Parabodonida, Eubodonida, and Trypanosomatida, excluding Prokinetoplastida (Simpson et al., 2002;

Heyden et al., 2004; Moreira et al., 2004). This tree topology suggested that Trypanosomatida and Prokinetoplastida acquired parasitic lifestyles separately (Moreira et al., 2004; Simpson et al., 2006; Lukeš et al., 2014). Owing to their apparent importance in public health, the origin of parasitism in the extant trypanosomatids was one of the major questions in the evolution of Kinetoplastea. To address this question, the precise relationship among Neobodonida, Parabodonida, Eubodonida, and Trypanosomatida has been explored mainly by analyzing SSU rRNA genes or genes encoding highly conserved proteins, but has not been resolved with high statistical support (Dolezel et al. 2000; Simpson et al., 2002; Heyden et al., 2004; Moreira et al., 2004; Simpson et al., 2004; Deschamps et al., 2011). A phylogenetic analysis of 64 genes encoding highly conserved proteins successfully designated Eubodonida as the closest relative of Trypanosomatida (Deschamps et al., 2011).

Compared to pathogenic trypanosomatids, other parasitic members in Kinetoplastea have received less research attention. There are two types of parasites belonging to Prokinetoplastida: fish parasites (e.g., causative agent of ichthyobodosis, *Ichthyobodo necator*: Callahan et al., 2002) and intracellular parasites in an amoebozoan *Paramoeba pemaquidensis* (i.e. *Perkinsela* sp. or *Ichthyobodo*-related organism: Dyková et al., 2003; Caraguel et al., 2007; Dyková et al., 2008; Feehan et al. 2013; Lukeš et al., 2014). Parabodonida includes fish parasites that cause cryptobiosis in salmonid and cyprinid fishes (e.g., *Cryptobia salmositica* and *Trypanoplasma borreli*: Woo and Poynton, 1995), as well as a snail parasite *Cryptobia helicis* (Leidy, 1846). Among the known neobodonids, there is a single parasitic member, *Azumiobodo hoyamushi*, which infects ascidians and causes soft tunic syndrome (Hirose et al., 2012). As the transition from a free-living to a parasitic lifestyle occurred after the divergence of the extant

parabodonids/neobodonids, the parasites in Parabodonida and Neobodonida are potentially useful to retrace the evolutionary path from a free-living to a parasitic lifestyle.

For a deeper understanding of the evolution of parasitism in Kinetoplastea, a well-resolved, taxon-rich phylogeny is indispensable. Deschamps et al. (2011) analyzed an alignment of 64 proteins and elucidated the relationship among Trypanosomatida, Eubodonida, Neobodonida, and Parabodonida. However, their analyses contained two potential limitations. Firstly, the alignment analyzed in Deschamps et al. (2011) contained no prokinetoplastid species. Secondly, each of Parabodonida and Neobodonida was represented by only a single free-living species but no parasitic member. In this study, I overcame these limitations by analyzing a new alignment of 43 proteins (43-gene alignment), which covered all the five orders in Kinetoplastea, and Parabodonida and Neobodonida were represented by both free-living and parasitic members. Combined the global Kinetoplastea phylogeny updated by analyzing 43-gene alignment with a taxon-rich SSU rDNA phylogeny, I discuss the transition from a free-living to a parasitic lifestyle in the evolution of Kinetoplastea.

2-3 Materials and Methods

2-3-1 Cultures, RNA extraction, and Sequencing

The laboratory culture of *Azumiobodo hoyamushi* established by Hirose et al. (2012) was grown and maintained in sea water containing 2% heat-inactivated fetal bovine serum (Gibco) at 17 °C. *Trypanoplasma borreli* ATCC50836 was purchased from American Type Culture Collection (ATCC), and grown in live-infusion-tryptose medium (Fernandes and Castellani, 1966) at 17 °C. Total RNA was extracted from the harvested cells using Trizol (Thermo Fisher Scientific, Waltham, Massachusetts, USA) by following the manufacture's protocol. Construction of cDNA library and subsequent

sequencing by the Illumina HiSeq2500 system were performed at Hokkaido System Science (Hokkaido, Japan). I generated 401,725,240 and 433,374,224 paired-end, 100-base reads from the *Azumiobodo* and *Trypanoplasma* libraries, respectively (deposited in NCBI Sequence Read Archive under accession numbers SRX2210809 and SRX22109115, respectively). The two sets of the initial reads were separately assembled into 28,134 contigs (*Azumiobodo*) and 18,591 contigs (*Trypanoplasma*) by TRINITY (Grabherr et al., 2011; Haas et al., 2013). The transcriptomic data generated from *Perkinsela* by the Illumina HiSeq2000 system was retrieved from NCBI Sequence Read Archive (accession number ARX255943), and was assembled into 18,600 contigs by TRINITY. The contig data of *Azumiobodo* and *Trypanoplasma* are freely available at https://sites.google.com/site/eukiyazaki/home/data-archive/kinetoplastida.

2-3-2 Phylogenetic alignments

I prepared a multi-gene alignment including sequences from *Azumiobodo*, *Trypanoplasma*, and *Perkinsela* by following Deschamps et al. (2011), which assessed the phylogenetic relationship among seven kinetoplastids based on 64 protein-coding genes. Firstly, I prepared four sets of the 64 genes—those of *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, and *Leishmania infantum*—by surveying in NCBI (http://www.ncbi.nlm.nih.gov/), TriTrypDB (http://tritrypdb.org/tritrypdb), and Sanger Institute databases (http://www.sanger.ac.uk/resources/databases/). The putative *Trypanosome*/*Leishmania* proteins identified in the first survey were then used as queries for TBLASTX against the contig data of *Azumiobodo*, *Trypanoplasma*, and *Perkinsela*. I recovered the *Azumiobodo*, *Trypanoplasma*, and *Perkinsela* transcripts that matched the queries with *E*-values smaller than 10^{-10} as the candidates encoding the proteins of interest.

The candidate transcripts were subjected to BLASTX against NCBI nr protein database to confirm that these transcripts encode proteins that were homologous to the queries in the initial BLAST analysis (TBLASTX). Finally, the selected transcripts of *Azumiobodo*, *Trypanoplasma*, and *Perkinsela* were conceptually translated into amino acid sequences by EMBOSS TRANSEQ (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). I repeated the same procedure above for three kinetoplastids (*Rhynchomonas nasuta*, *Procryptobia sorokini*, and *Bodo saltans*: GenBank accession numbers, HO651677–HO651928), an amoebozoan *Dictyostelium discoideum* (NCBI BioProject: PRJNA201), a heterolobosean *Naegleria gruberi* (NCBI BioProject: PRJNA43691), a diplonemid *Diplonema papillatum* (TBestDB organism ID: DP), and a euglenid *Euglena gracilis* (NCBI SRA: ERX324117). *Dictyostelium* and *Naegleria* were included in the alignments to identify the sequences originated from an amoebozoan *Paramoeba pemaquidensis* (i.e. the host organism of *Perkinsela*) that potentially contaminate the *Perkinsela* transcriptomic data. If the *Paramoeba* sequences were misassigned as *Perkinsela* sequences, I anticipated that such sequences were most likely grouped with the *Naegleria* and *Dictyostelium* sequences with the specific affinity to the latter (i.e. amoebozoan) sequences in the phylogenetic analyses described below.

Sixty-four single-gene alignments (each containing the 14 species described above) were automatically aligned by MAFFT 7.205 (Katoh et al., 2002; Katoh and Standley, 2013). Ambiguously aligned positions were manually excluded prior to the phylogenetic analyses described below. The alignments were subjected separately to the maximum-likelihood (ML) phylogenetic analyses with the LG model (Le and Gascuel, 2008) incorporating empirical amino acid frequencies and among-site rate variation approximated by a discrete gamma distribution with four categories (LG + Γ + F model).

Tree search was started from ten maximum-parsimony (MP) trees, each of which was generated by random stepwise addition of sequences by RAxML 8.0.3 (Stamatakis, 2014). Although not shown here, the ML trees inferred from 21 out of the 64 single-gene alignments failed to reconstruct the monophyly of kinetoplastids, and were omitted from the multi-gene alignment described below. The 43 single-gene alignments listed in Supplementary Table 1 were then concatenated into a single alignment containing 6,842 amino acid positions (43-gene alignment). Prior to phylogenetic analyses, I excluded both *Dictyostelium* and *Naegleria* sequences, which were used as "probes" to detect the *Paramoeba* sequences potentially contaminating the *Perkinsela* transcriptomic data.

I sampled and aligned SSU rDNA sequences from 20 trypanosomatids, ten parabodonids, ten eubodonids, 15 neobodonids, five prokinetoplastids, three diplomenids, and a single euglenid (64 taxa in total). The 64 SSU rDNA sequences were aligned by MAFFT, and ambiguously aligned positions were manually excluded. The final SSU rDNA alignment contained 1,179 nucleotide positions.

Both 43-gene and SSU rDNA alignments are available at https://sites.google.com/site/eukiyazaki/home/data-archive/kinetoplastida.

2-3-3 Phylogenetic analyses

43-gene and SSU rDNA alignments were subjected to the ML phylogenetic analyses by RAxML 8.0.3. The detail settings of tree search were the same as the single-gene analyses described above, but I assigned the LG + Γ + F model for the ML analyses of 43-gene alignment, and the GTR + Γ model (Rodríguez et al., 1990) for those of SSU rDNA alignment. One hundred bootstrap replicates were generated from each alignment, and subjected to the tree search as described above. The resultant bootstrap trees were

used to calculate ML bootstrap support values (MLBPs).

The two alignments were also analyzed with Bayesian method. 43-gene alignment was subjected to PHYLOBAYSE 3.3 (Lartillot and Philippe, 2004) with the CAT + Poisson model. Two Markov chain Monte Carlo (MCMC) runs were conducted for 10,000 cycles with "burn-in" of 2,500 ("maxdiff" value was as low as 0.00013). Subsequently, the consensus tree with branch lengths and Bayesian posterior probabilities (BPPs) were calculated from the remaining trees. SSU rDNA alignment was subjected to MRBAYES 3.2.3 (Huelsenbeck and Ronquist, 2001). I assigned the same substitution model as in the ML method described above. The MCMC run was performed with one cold and three heated chains with default chain temperatures. I ran 1,000,000 generations, and sampled log-likelihood scores and trees with branch lengths every 1,000 generations. The first 25% generations were discarded as "burn-in." The consensus tree with branch lengths and BPPs were calculated from the remaining trees.

2-4 Results and Discussion

2-4-1 Global Kinetoplastea phylogeny revisited

I updated the large-scale multi-gene (phylogenomic) alignment used in Deschamps et al. (2011) by adding *Trypanoplasma*, *Azumiobodo*, and *Perkinsela*, and re-examined the global phylogeny of Kinetoplastea. The ML tree and MLBPs inferred from 43-gene alignment are presented in Fig. 1. As the topology inferred from Bayesian method was identical to that from the ML method, I only mapped BPPs on the ML tree shown in Fig. 1. Although the 43-gene phylogeny includes the three parasitic kinetoplastids that were absent in the previous phylogenomic analysis, the overall tree topology (Fig. 1) was not largely different from that presented in Deschamps et al. (2011). Trypanosomatida was found to be tied with other four taxon/clades in the following order;

(i) *Bodo saltans* representing Eubodonida, (ii) the Parabodonida clade comprising *Procryptobia sorokini* and *Trypanoplasma borreli*, (iii) the Neobodonida clade comprising *Rhynchomonas nasuta* and *Azumiobodo hoyamushi*, and then (iv) *Perkinsela* sp. representing Prokinetoplastida. All the internal nodes in the ingroup were supported by MLBPs of 100% and BPPs of 1.00, except the clade of *Rhynchomonas* and *Azumiobodo*, which received an MLBP of 97% and a BPP of 0.99 (Fig. 1). Prior to this study, the phylogenetic position of Prokinetoplastida has been assessed only by single-gene alignments including kinetoplastid species sampled from all the five orders (Callahan et al., 2002; Moreira et al., 2004; Simpson et al., 2004; Breglia et al., 2007; Hirose et al., 2012) or four-gene and 11-gene alignments with restricted taxon samplings (Tanifuji et al., 2011; Cenci et al 2016). Thus, the current study provides the first "phylogenomic" support for the earliest branching status of Prokinetoplastida in Kinetoplastea (Fig. 1).

2-4-2 Evolution of parasitism in Kinetoplastidea

I analyzed SSU rDNA alignment, of which taxon sampling was much richer than that of 43-gene alignment, to illustrate the sporadic distribution of parasitic and free-living species in Kinetoplastea (Fig. 2). Prokinetoplastida was excluded from the clade of Trypanosomatida, Eubodonida, Parabodonida, and Neobodonida with an MPBP of 100% and a BPP of 1.0. The monophylies of Eubodonida, Parabodonida, and Neobodonida were recovered with MLBPs of 63-91% and BPPs of 0.71-1.0, while that of Trypanosomatida was not positively supported. Although *Paratrypanosoma confusum* was excluded from the clade of other trypanosomatids in the SSU rDNA phylogeny (Fig. 2), I regard *Paratrypanosoma* as an ancestral branch in Trypanosomatida based on a series of multi-

gene phylogenetic analyses presented in Flegontov et al. (2013). I assume that the SSU rDNA phylogeny presented here failed to place *Paratrypanosoma* in the genuine position due to lack of phylogenetic signal. Thus, the combination of the 43-gene phylogeny, which resolved the backbone of the tree of Kinetoplastea (Fig. 1), and the taxon-rich SSU rDNA phylogeny (Fig. 2) enables us to depict how and when parasitic species emerged during the evolution of Kinetoplastea. I schematically illustrate the evolution of lifestyles in Kinetoplastea in Fig. 3 (see below for details). As life cycles of most of "free-living" kinetoplastids are not well understood, I cannot exclude the possibility that some of them have parasitic stages.

The 43-gene phylogeny united an obligatory parasitic order Trypanosomatida with Eubodonida, which comprises free-living members (Fig. 1). Thus, as discussed in Simpson et al. (2006), Deschamps et al. (2011), and Flegontov et al. (2013), a parasitic lifestyle was most likely established after the separation of Trypanosomatida and Eubodonida, but before the divergence of the extant trypanosomatids including *Paratrypanosoma* (Fig. 3). All trypanosomatids known so far are extracellular parasites, but intracellular stage has been also reported for *Leishmania* spp. and *Trypanosoma cruzi* (Tyler and Engman 2001; Handman and Bullen 2002). As *Leishmania* spp. and *T. cruzi* are distantly related in the SSU rDNA phylogeny (Fig. 2), the ability to invade host cytoplasm was likely acquired by the two separate lineages after the divergence of trypanosmatids (Fig. 3).

Parabodonida contains parasitic members, namely *Trypanoplasma borreli* and *Cryptobia* spp., as well as free-living members (Fig. 2). In the SSU rDNA phylogeny (Fig. 2), *T. borreli*, *C. catostomi*, *C. bullocki*, and *C. salmositica* formed a robust clade (MLBP of 100% and BPP of 1.0). The four parabodonids are commonly found in the blood stream,

albeit ectoparasitic forms have also been reported for *C. bullocki* and *C. salmositica* (Bower and Margolis, 1983; Woo and Wehnert, 1983). I can conclude that an extracellular parasitic lifestyle was established on the branch leading to *Trypanoplasma-Cryptobia* clade, as proposed in Simpson et al. (2006). However, I currently have no evidence to determine whether the ectoparasitic form is the ancestral trait of this clade. *C. helicis* (Leidy, 1846), which was found in the seminal receptacle of snails, most likely acquired an extracellular parasitic lifestyle independent from *Trypanoplasma-Cryptobia* clade, as the SSU rDNA phylogeny united *C. helicis* with free-living *Parabodo nitrophilus* and *P. caudaus* (MLBP of 96% and BPP of 1.0; Fig. 2). Altogether, I propose that parasitism emerged at least twice in Parabodonida (Fig. 3). The potential third parasitic lineage in Parabodonida is *Jarrellia atramenti* found in the mucus of the respiratory tract of the pygmy sperm whale (Poynton et al., 2001). As only morphological and no molecular data are available for this species, it is necessary to assess the phylogenetic relationship between *Jarrellia* and other parabodonids to better understanding the evolution of parasitism in this order.

Among the diversity of neobodonids, *Azumiobodo* is the sole parasitic member known to date (Hirose et al., 2012). This flagellate was found in the tunics of ascidians with soft tunic syndrome by histopathology (Kumagai et al., 2010). Intriguingly, *Azumiobodo* appeared to bear a phylogenetic affinity to a commensal in the ascidian intestine *Cruzella marina* (Frolov and Malysheva, 2002) in the SSU rDNA analyses (MLBP of 79% and BPP of 0.98; Fig. 2). The affinity between *Azumiobodo* and *Cruzella* prompts us to propose serial lifestyle changes in Neobodonida as follows—the common ancestor of *Azumiobodo* and *Cruzella* established a symbiotic relationship with ascidians, and, after separation of the two species, the former became a pathogenic extracellular

parasite causing soft tunic syndrome. Although the above scenario needs to be examined in future studies, I are certain that the transition from a free-living to a parasitic lifestyle occurred once after the divergence of the extant neobodonids (Fig. 3).

Independent from the lifestyle changes discussed above, the transition from a free-living to a parasitic lifestyle was proposed for Prokinetoplastida (Simpson et al., 2006). As all the members belonging to Prokinetoplastida known to date are parasitic, it is straightforward to assume that they were emerged from a parasitic ancestor (Fig. 3). In the SSU rDNA phylogeny (Fig. 2), prokinetoplastids were split into two subclades, one is of parasites that infect the skins, fins, and gills of fishes (e.g., *Ichthyobodo necator*) and the other is of intracellular parasites of amoebozoans (e.g., *Perkinsela* sp.). The conspicuous difference in parasitic mode between the two subclades in Prokinetoplastida—ectoparasitism and intracellular parasitism—demands a more complex scenario than that assuming a single transition from a free-living to a parasitic lifestyle prior to the divergence of prokinetoplastids (Fig. 3). Notably, the hosts of *Perkinsela* sp. and its relatives are amoebozoans, which parasitize marine animals (Munday et al., 2001; Young et al., 2007). Thus, it is attractive to hypothesize that the common ancestor of *Perkinsela* sp. and its relatives was an ectoparasite of a marine animal, then switched the host to an amoebozoan parasitizing marine animals, and finally invaded and settled in the cytoplasm of the amoebozoan host. The above scenario will be favored if future surveys find a novel ectoparasitic species that branches at the base of the clade of *Perkinsela* sp. and its relatives. It is also important to pursue the possibility of some free-living prokinetoplastids being overlooked in natural environments, as environmental sequence data hinted that the full diversity of Prokinetoplastida has yet to be unveiled (Heyden et al., 2004; Moreira et al., 2004). I certainly need to reevaluate how

parasitism was established in Prokinetoplastida after its organismal diversity is sufficiently depicted in the future. Finally I took a good case study for revealing an internal phylogenies and involved well estimations how evolutionary events had occurred by these phylogenetic analyses is a good case study.

2-4-3 Future perspectives

The phylogenetic relationship among the five orders in Kinetoplastea was found to be not changed largely before and after incorporating three parasitic kinetoplastids, *Azumiobodo*, *Trypanoplasma*, and *Perkinsela* into a phylogenomic alignment (Deschamps et al., 2011; this study). However, I still need to improve the taxon sampling in phylogenomic alignments to re-examine the global Kinetoplastea phylogeny in the future. For instance, the monophyly of Neobodonida received only weak statistical support in the SSU rDNA analyses (Fig. 2). Thus, future phylogenomic analyses considering additional neobodonids are required to strengthen their monophyly inferred from the SSU rDNA alignment. Another potential concern is the diversity of Eubodonida. It is not clear whether the eubodonids identified so far represent the true diversity of this order, as the SSU rDNA phylogeny (Fig. 2) implies that the diversity of Eubodonida is considerably lower than that of any other order in Kinetoplastea. Thus, future studies may identify novel kinetoplastid flagellates that bear phylogenetic affinities to the currently known eubodonids. If such kinetoplastid flagellates exist, it would be worth to incorporate them into phylogenomic analyses to represent the proper diversity of Eubodonida.

Because of the serious threats to public health posed by them, members of Trypanosomatida have attracted more research attention than other kinetoplastids. From an evolutionary biological perspective, trypanosomatids are regarded as one of the model

organisms for studying how and when kinetoplastid flagellates acquired a parasitic lifestyle and pathogenicity (Deschamps et al., 2011; Lukeš et al., 2014). Nevertheless, I also anticipate that Parabodonida and Neobodonida will provide insights into the mechanism involved in the transformation of a free-living species into a parasite. In Parabodonida, a snail parasite *Cryptobia helicis* showed a specific affinity to the free-living members, *Parabodo caudatus* and *Parabodo nitrophilus* (Fig. 2). This change in lifestyle took place after the divergence of parabodonids (Fig. 3), and I may have a chance to pinpoint the set of genes that played a pivotal role in the change by comparing the genomic and transcriptomic data of the parasite and its closest free-living relatives. Similarly, the comparison between two closely related neobodonids, *Azumiobodo* and *Cruzella*, a parasite and a commensal of ascidians, respectively, may provide insights into the origin of parasitism and pathogenicity.

2-5 Figures and Tables

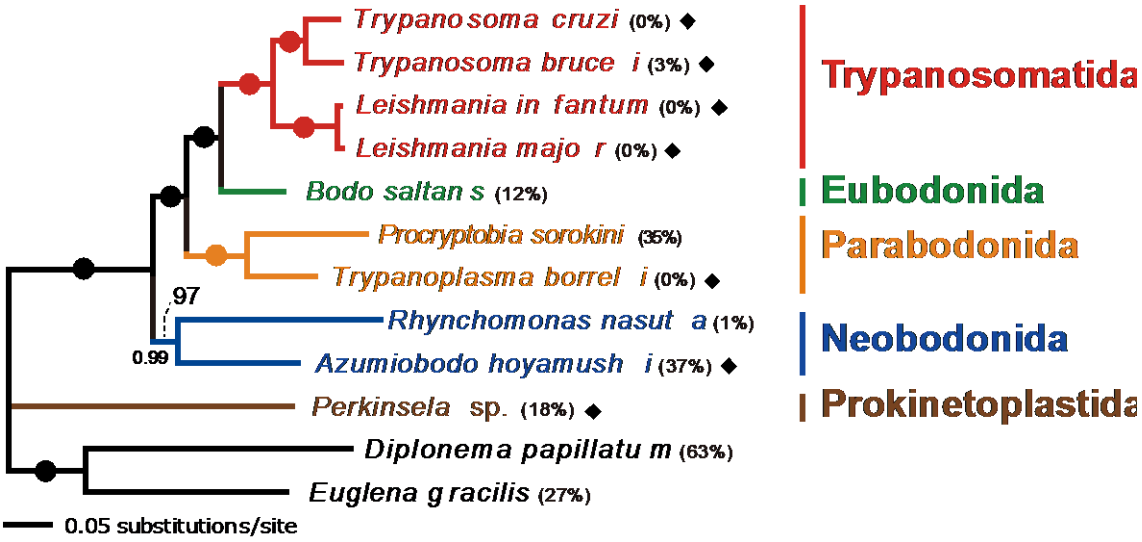

Fig. 1. Global Kinetoplastea phylogeny inferred from an alignment comprising 43 genes encoding highly conserved proteins. The tree topology was inferred from the maximum-likelihood (ML) method. The nodes marked by dots were supported by ML bootstrap values of 100% and Bayesian posterior probabilities of 1.0. For each taxon, the percentage of missing data is presented in parentheses. Parasites are marked by diamonds.

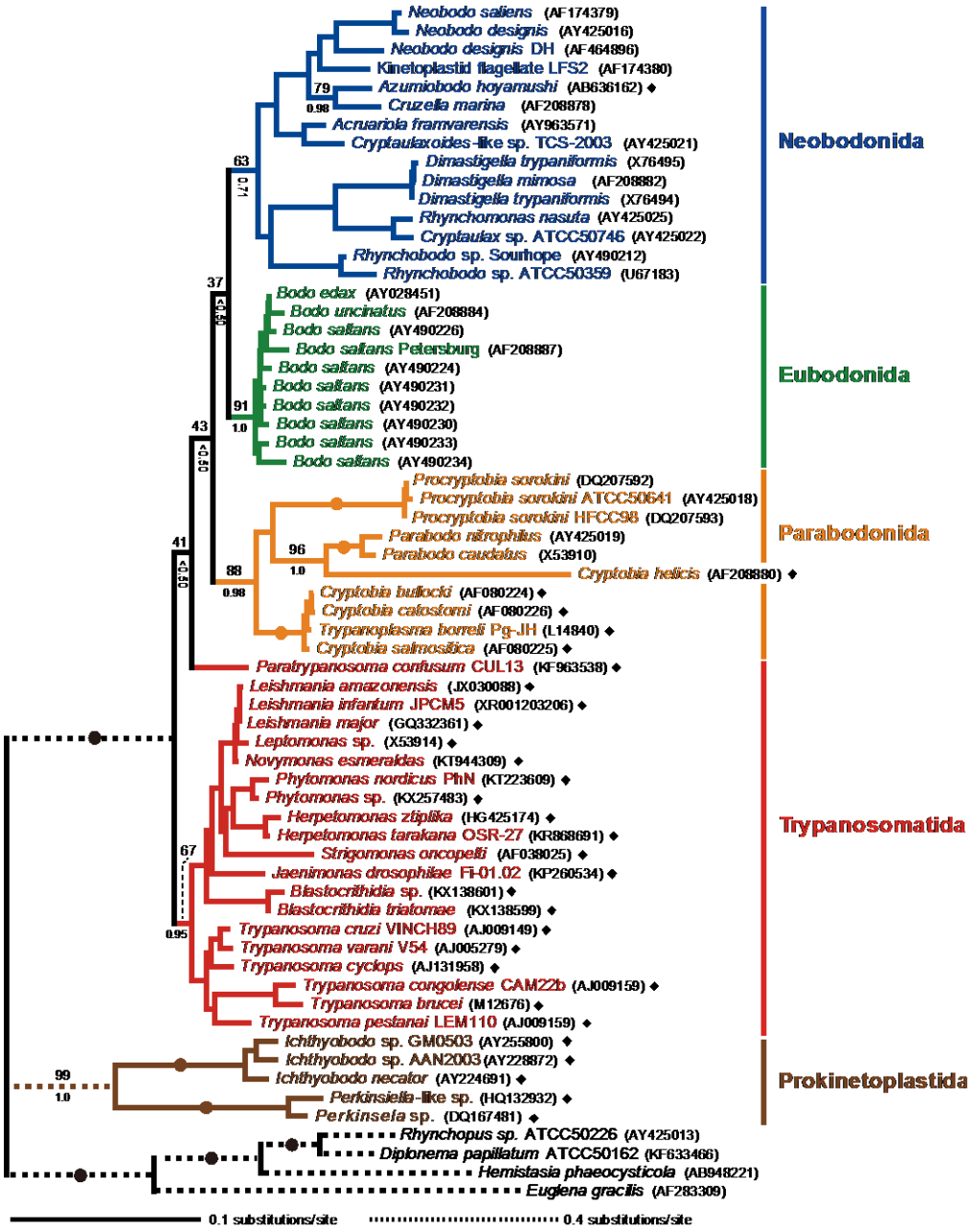

Fig. 2. Global Kinetoplastea phylogeny inferred from an alignment of small subunit ribosomal DNA sequences. The tree topology was inferred from the maximum-likelihood (ML) method. The nodes marked by dots were supported by ML bootstrap values (MLBPs) of 100% and Bayesian posterior probabilities (BPPs) of 1.0. MLBPs and BPPs are presented only for the nodes that are critical for discussing the evolution of lifestyle in Kinetoplastea. Parasites are marked by diamonds.

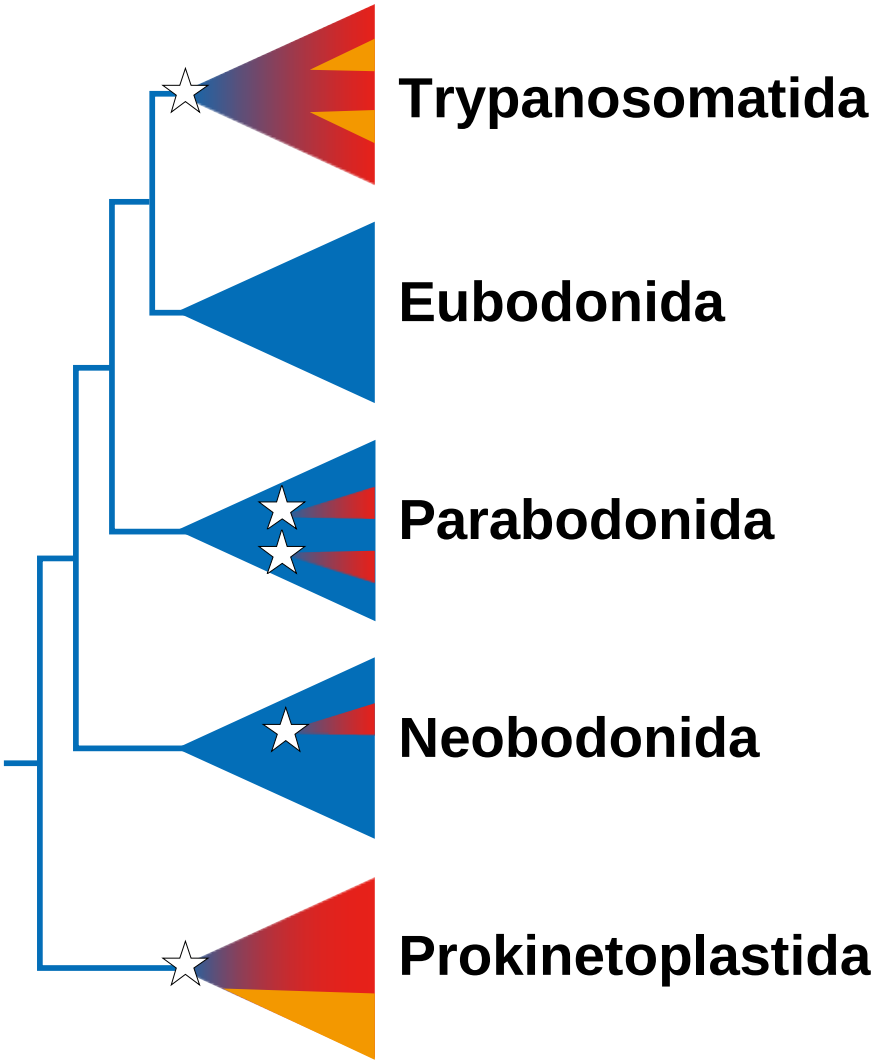

Fig. 3. Transition from a free-living to a parasitic lifestyle in Kinetoplastea. The branching order among Trypanosomatida, Eubodonida, Parabodonida, Neobodonida, and Prokinetoplastida is based on Fig. 1. Blue branches/clades indicate free-living species, while red clades indicate parasitic species. The putative changes in lifestyle are marked by stars. Due to the lack of a precise phylogenetic position, a parasitic parabodonid *Jarrellia atramenti* is omitted from this figure. Orange triangles indicate the lineages that acquired the ability to invade host cytoplasm, namely *Leishmania* spp. and *Trypanosoma cruzi* in Trypanosomatida and the intracellular parasites of marine amoebas in Prokinetoplastida.

Table 1. List of the 43 proteins used for inferring the global Kinetoplastea phylogeny.

	Trypanosoma cruzi	*Trypanosoma brucei*	*Leishmania major*	*Leishmania infantum*	*Bodo saltans*	*Tryanoplasma borreli*	*Procryptobia sorokini*	*Rhynchomonas nasuta*	*Azumiobodo hoyamushi*	*Perkinsela* sp.	*Euglena gracilis*	*Diplonema papillatum*
eukaryotic initiation factor 5a	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	✓
fibrillarin	✓	–	✓	✓	✓	✓	–	✓	✓	✓	✓	–
histone H3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
histone H4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
ribosomal protein L1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein L11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein L12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein L15	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein L2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
ribosomal protein L22	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
ribosomal protein L23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal Protein L25	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
ribosomal protein L27	✓	✓	✓	✓	✓	✓	–	–	✓	–	✓	–
ribosomal protein L30	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
ribosomal protein L32	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–	✓
ribosomal protein L35	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
ribosomal protein L37	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–
ribosomal protein L38	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
ribosomal protein L39	✓	✓	✓	✓	✓	✓	–	–	✓	✓	–	✓
ribosomal protein L4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein L5	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	–
ribosomal protein L6	✓	✓	✓	✓	✓	✓	✓	–	✓	✓	–	–
ribosomal protein L7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein L9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein S14	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein S15	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	–
ribosomal protein S16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S17	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	–	✓
ribosomal protein S18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein S20	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein S23	✓	✓	✓	✓	✓	✓	✓	–	✓	✓	✓	–
ribosomal protein S25	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓
ribosomal protein S26	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S27	✓	✓	✓	✓	✓	✓	✓	–	✓	✓	–	✓
ribosomal protein S3	✓	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	✓
ribosomal protein S4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ribosomal protein S5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
ribosomal protein S9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tubulin, alpha	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tubulin, beta	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
Total alignment positions available	6839	6613	6816	6825	5052	6838	4471	6785	4321	5621	4986	2526
Percentage of missing data (%)	0.04	3.35	0.38	0.25	11.55	0.06	34.65	0.83	36.85	17.85	27.13	63.08

Chapter 3 Strain PAP020 as an early branch in Metamonada.

3-1 Abstract

A novel protist, strain PAP020, was isolated from mangrove sediments sampled in the Republic of Palau in November 5, 2011. The laboratory culture of strain PAP020 has been maintained under the anaerobic condition with prey bacteria. PAP020 is oval-shaped cell with two flagella and showed no clear morphological characteristic with other previously described eukaryotes. I explored the position of this protist using the maximum-likelihood (ML) phylogenetic analysis of small subunit ribosomal DNA (SSU rDNA) sequences. In the SSU rDNA tree, PAP020 showed no strong affinity to other eukaryotes. As neither microscopic observation nor SSU rDNA phylogeny provided any clues for the phylogenetic affiliation of PAP020, I suspected that PAP020 belongs to an as-yet-to be recognized lineage. To determine the precise phylogenetic position of PAP020, I prepared and analyzed an alignment comprising 148 genes extracted from the transcriptomic data of PAP020. The ML tree inferred from the 148-protein alignment covering 83 phylogenetically diverse eukaryotes (including PAP020) reconstructed a clade comprising PAP020, parabasalids and fornicates with high statistical support. Within this clade, PAP020 branched at the base of the Fornicata clade including *Giardia intestinalis* and its free-living relatives (i.e. *Carpediemonas*-like organisms or CLOs) with a full BP value. Due to the early branching nature of PAP020 in the Metamonada clade, I anticipate that this protist may hold keys to predict the anaerobic metabolism of the common ancestor of Metamonada.

Chapter 4 Estimation of the phylogenetic possession of a novel protist, strain SRT308 by phylogenomic analysis

4-1 Abstract

A novel unicellular flagellate, strain SRT308, was isolated from a marine sediment sample collected from the Republic of Palau in 2013, and has been maintained in the laboratory. I firstly explored the position of this flagellate using the maximum-likelihood (ML) phylogenetic analysis of small subunit ribosomal DNA (SSU rDNA) sequences. In the SSU rDNA tree, SRT308 showed no strong affinity to any eukaryotes or eukaryotic lineages, suggesting that this flagellate represents an unprecedented eukaryotic lineage. Therefore, to determine the accurate phylogenetic position strain SRT308, I conducted a ML and Bayesian analysis based on 153 nucleus-encoded gene sequences, which included a part of the transcriptomic data of this flagellate. The ML and Bayesian tree inferred from the 153-protein alignment reconstructed a clade comprising SRT308 as base of Euglenozoa, that clade composed Kinetoplastea, Euglenida, and Diplonemea, with a MLBPs of 100% and BPPs of 1.0. The clade of SRT308 + Euglenozoa further grouped with Heterolobosa, Jakobida, and *Tsukubamonas globosa*, forming the Discoba clade. Moreover, to confirm that SRT308 has no phylogenetic affinity with other Discoba, I conducted a ML and Bayesian analysis for Discoba taxon rich dataset using SSU rDNA sequences. SRT308 placed at base of Euglenozoa by Discoba taxon rich phylogenetic analyses. Finally, I here propose that SRT308 branches at the base of the clade of Euglenozoa and this microeukaryote may hold keys to predict mitochondrial diversity in Euglenozoa

Chapter 4. General discussion

In chapter 2, I reexamined the internal phylogenetic relationship of Kinetoplastea by combining phylogenomic analysis and SSU rDNA phylogenetic analysis, and verify how many parasitic acquisitions occurred in Kinetoplastea. As a result I advocate a hypothesis that acquisitions of parasitic capability occurred more than five times and transitions of parasitic styles occurred three or more times in each kinetoplastida linages. This study has become a good case study, which clarified internal phylogenetic relationships of main eukaryotic linages, which is Discoba of Excavata in this study, and speculated evolutionary events (Fig. 12).

In chapter 3 and 4, I succeeded in robustly estimating the each phylogenetic positions of two novel eukaryotes discovered in the Republic of Palau by large-scale molecular phylogenetic analysis. As a result, it was shown that PAP020 branches from the base of Fornicata, SRT 308 branches the base of Euglenozoa. Therefore, Excavata has been shown to have greater diversity than previously know and it is provided the possibility that both of the novel eukaryotes are the key organisms for understanding the early evolution in each sister linages, Fornicata and Euglenozoa. This study has become a good case study which a novel microeukaryote have contributed to understanding a part of the true diversity of main eukaryotic linage, which is Excavata in this study (Fig. 12).

My three studies described that one was based on internal linage of Excavata, others on the whole Excavata as a result of large scale multi genes phylogeny. In other words, all my studies contributed to the understanding of the true diversity inside Excavata and unfortunately not provide highly impact for the understanding of the true diversity of whole eukaryotic relationship, also revealing the root of eukaryote.

There are many internal linages that are not robustly grasped by many lineage

relationships, and along with this, details of evolutionary events have not been disclosed in linage independently (Hampl et al. 2009; Wakeman et al. 2014; Pánek et al. 2016). Such elucidation of internal linage relationship is expected to be clarified by large-scale molecular phylogenetic analysis and single genetic analysis enriched in taxon sampling as I used in verification of Kinetoplastea internal phylogenetic relationship. On the other hand, mentioned above, it has remained the problem, describing true eukaryotic relationship. Moreover, there are many types of missing links which is possibility to fill with overlooked microeukaryotes, large and small in the eukaryotic relationship like as (Yubuki et al. 2010; Yabuki, Ishida, and Cavalier-Smith 2013; Kamikawa et al. 2014; Yabuki et al. 2014; Yabuki and Tame 2015; Burki et al. 2016). As I demonstrated in chapters 3 and 4, novel eukaryotic organisms are discovered from the environment and determined the systematic position by large-scale molecular phylogenetic analysis, thereby filling the missing links one by one in whole eukaryote. Moreover, determining the phylogenetic position for unknown eukaryotes, that have not been studied much, also contribute to the understanding of eukaryotic diversity. Finally, the whole of the eukaryotic lineage relationship unveils when almost all missing links in eukaryotic phylogeny will filled up.

I will contribute to the elucidation of the whole true eukaryotic relationship continue by focusing on undiscovered or unexamined microeukaryote, “overlooked eukaryote”, and demonstrating large-scale molecular phylogenetic analysis.

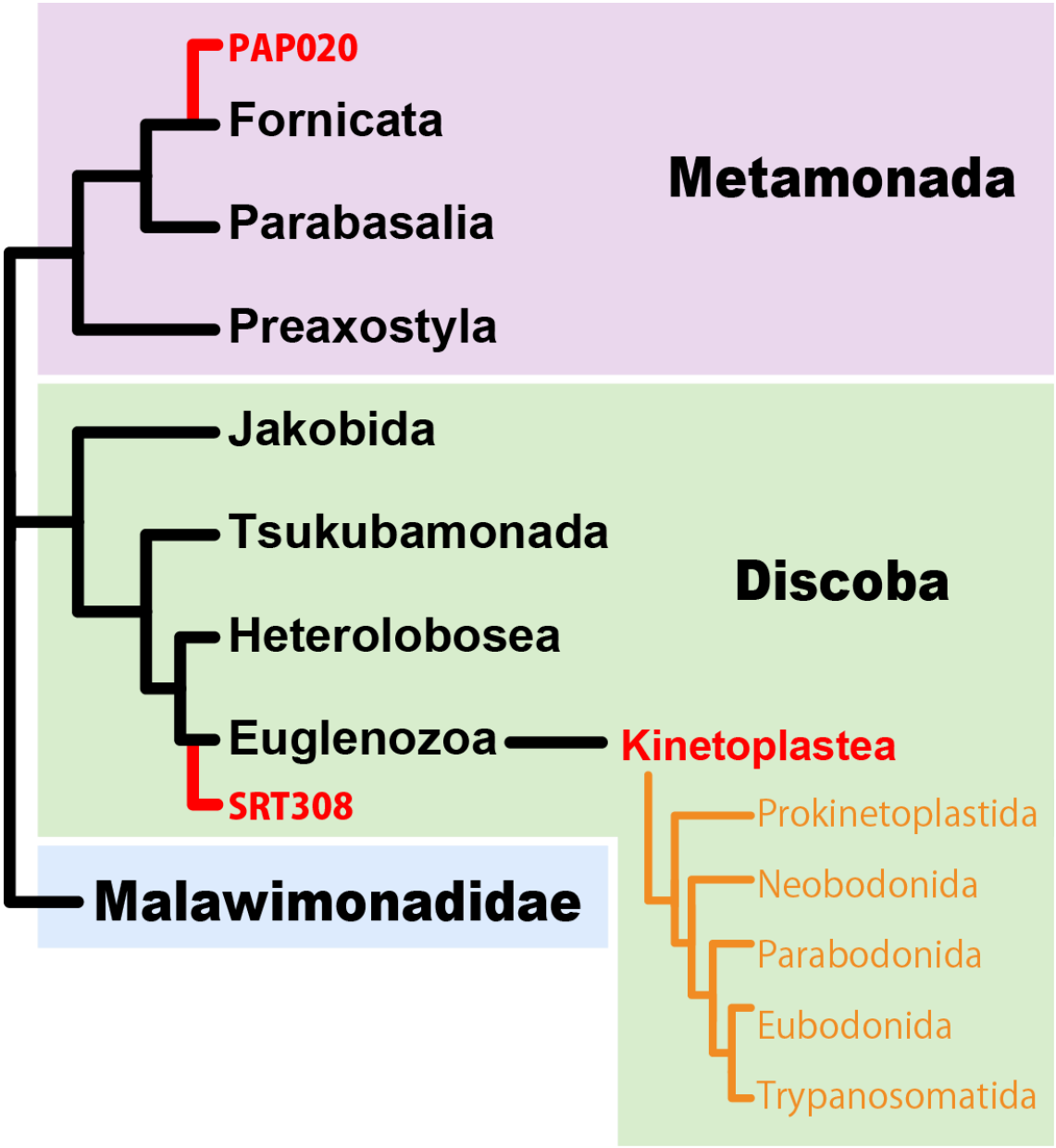

Fig. 12. The phylogeny described a detailed Excavata relathonship.

Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Yuji Inagaki (University of Tsukuba) for their helpful discussions and encouragements during the course of my work. I also thank to Prof. Tetsuo Hashimoto, Ph.D. Sohta A. Ishikawa (University of Tokyo), Ph.D. Goro Tanifuji (National Museum of Nature and Science), Assistant Prof. Takuro Nakayama (Tohoku University) and Mr. Keitaro Kume (University of Tsukuba) for technical advises and useful discussions. I am also grateful to Akira Kumagai (Miyagi Prefecture Fisheries Technology Institute) and Takashi Kamaishi (National Research Institute of Aquaculture) for providing Azumiobodo hoyamushi and useful discussions in chapter 2, phylogeny for Kinetoplasitda. I thank Takashi Shiratori and Prof. Ken-ichiro. Ishida (University of Tsukuba) for providing Strain PAP020 and SRT308 used in chapter 3 and 4 and useful discussions. Next Generation sequences of *Carpediemonas*-like organisms used in chapter 3 were provided by Assistant Prof. Ryoma. Kamikawa (Kyoto University). Finally, I thank all members of the laboratory of Molecular Evolution of Microbes, and Plant Phylogeny and Systematics (University of Tsukuba) for their help.

I was supported by Sasagawa Science Fundation (27-501) from April 2015 to January 2016.

Reference

Adl S. M., Simpson A. G. B., Farmer M. A., Andersen R. A., Anderson O. R., Barta J. R., Bowser S. S., Brugerolle G., Fensome R. A., Fredericq S., et al. 2005. The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *J. Eukaryot. Microbiol.*, **52**:399–451.

Adl S. M., Simpson A. G. B., Lane C. E., Lukeš J., Bass D., Bowser S. S., Brown M. W., Burki F., Dunthorn M., Hampl V., et al. 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.*, **59**:429–93.

Archibald J. M. 2009. The Puzzle of Plastid Evolution. *Curr. Biol.*, **19**:R81–R88.

Baldauf S. L., Roger A. J., Wenk-Siefert I. & Doolittle W. F. 2000. A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data. *Science (80-.).*, **290**.

Bork P., Bowler C., Vargas C. de, Gorsky G., Karsenti E. & Wincker P. 2015. Tara Oceans studies plankton at planetary scale. *Science (80-.).*, **348**:873.

Breglia S. A., Slamovits C. H. & Leander B. S. 2007. Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from Hsp90 gene sequences. *J. Eukaryot. Microbiol.*, **54**:86–92.

Brown J. R., Douady C. J., Italia M. J., Marshall W. E. & Stanhope M. J. 2001. Universal trees based on large combined protein sequence data sets. *Nat. Genet.*, **28**:281–285.

Brown M. W., Kolisko M., Silberman J. D. & Roger A. J. 2012. Aggregative Multicellularity Evolved Independently in the Eukaryotic Supergroup Rhizaria. *Curr. Biol.*, **22**:1123–1127.

Brugerolle G. & Müller M. 2000. Amitochondriate flagellates. *In*: Green J. C. & Leadbeater B. S. C. (eds.), Flagellates: Unity, Diversity and Evolution. London, Taylor

& Francis Group., p. 166–169.

Burger G., Gray M. W., Forget L. & Lang B. F. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol. Evol.*, **5**:418–38.

Burki F., Kaplan M., Tikhonenkov D. V, Zlatogursky V., Minh B. Q., Radaykina L. V, Smirnov A., Mylnikov P., Keeling P. J. & Keeling P. J. 2016. Untangling the early diversification of eukaryotes : a phylogenomic study of the evolutionary origins of Centrohelida , Haptophyta and Cryptista.

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

Callahan H. A., Litaker R. W. & Noga E. J. 2002. Molecular Taxonomy of the Suborder Bodonina (Order Kinetoplastida), Including the Important Fish Parasite, Ichthyobodo necator. *J. Eukaryot. Microbiol.*, **49**:119–128.

Campbell D. A., Thomas S. & Sturm N. R. 2003. Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect.*, **5**:1231–1240.

Caraguel C. G. B., O'kelly C. J., Legendre P., Frasca S., Gast R. J., Desrés B. M., Cawthorn R. J. & Greenwood S. J. 2007. Microheterogeneity and Coevolution: An Examination of rDNA Sequence Characteristics in Neoparamoeba pemaquidensis and Its Prokinetoplastid Endosymbiont. *J. Eukaryot. Microbiol.*, **54**:418–426.

Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.*, **52**:297–354.

Cavalier-Smith T. 2007. A revised six-kingdom system of life. *Biol. Rev.*, **73**:203–266.

Cavalier-Smith T. 1987. The origin of Fungi and pseudofungi. *Evol. Biol. Fungi*, :339–353.

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. *Mol. Biol. Evol.*, **28**:53–58.

Doolittle W. F. 1999. Phylogenetic Classification and the Universal Tree. *Science (80-.).*, **284**.

Dyková I., Fiala I., Lom J. & Lukeš J. 2003. Perkinsiella amoebae-like endosymbionts of Neoparamoeba spp., relatives of the kinetoplastid Ichthyobodo. *Eur. J. Protistol.*, **39**:37–52.

Dyková I., Fiala I. & Pecková H. 2008. Neoparamoeba spp. and their eukaryotic endosymbionts similar to Perkinsela amoebae (Hollande, 1980): Coevolution demonstrated by SSU rRNA gene phylogenies. *Eur. J. Protistol.*, **44**:269–277.

Embley T. M. & Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature*, **440**:623–630.

Fernandes J. F. & Castellani O. 1966. Growth characteristics and chemical composition of Trypanosoma cruzi. *Exp. Parasitol.*, **18**:195–202.

Flegontov P., Gray M. W., Burger G. & Lukeš J. 2011. Gene fragmentation: A key to mitochondrial genome evolution in Euglenozoa? *Curr. Genet.*, **57**:225–232.

Flegontov P., Votýpka J., Skalický T., Logacheva M. D., Penin A. A., Tanifuji G., Onodera N. T., Kondrashov A. S., Volf P., Archibald J. M., et al. 2013. Paratrypanosoma is a novel early-branching trypanosomatid. *Curr. Biol.*, **23**:1787–93.

Forterre P. & Philippe H. 1999. Where is the root of the universal tree of life? *BioEssays*, **21**:871–879.

Frolov A. O. & Malysheva M. N. 2002. Ultrastructure of the flagellate Cruzella marina (Kinetoplastidea). *Tsitologiia*, **44**:477–84.

van der Giezen M., Tovar J., Abrahamsen M. S., Adams K. L., Palmer J., D., Ali V., Shigeta Y., Tokumoto U., Takahashi Y., et al. 2005. Degenerate mitochondria. *EMBO Rep.*, **6**:525–30.

Gould S. B., Waller R. F. & McFadden G. I. 2008. Plastid Evolution. *Annu. Rev. Plant Biol.*, **59**:491–517.

Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. a, Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**:644–652.

Gray M. W. 1999. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.*, **9**:678–687.

Gray M. W., Burger G. & Lang B. F. 2001. The origin and early evolution of mitochondria. *Genome Biol.*, **2**:REVIEWS1018.

Gribaldo S. & Philippe H. 2002. Ancient Phylogenetic Relationships. *Theor. Popul. Biol.*, **61**:391–408.

Gualdrón-López M., Brennand A., Hannaert V., Quiñones W., Cáceres A. J., Bringaud F., Concepción J. L. & Michels P. A. M. 2012. When, how and why glycolysis became compartmentalised in the Kinetoplastea. A new look at an ancient organelle. *Int. J. Parasitol.*, **42**:1–20.

Guy L. & Ettema T. J. G. 2011. The archaeal "TACK" superphylum and the origin of eukaryotes. *Trends Microbiol.*, **19**:580–587.

Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D., Li B., Lieber M., et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, **8**:1494–512.

Hampl V., Hug L., Leigh J. W., Dacks J. B., Lang B. F., Simpson A. G. B. & Roger A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc. Natl. Acad. Sci. U. S. A.*, **106**:3859–64.

Hasegawa M., Kishino H. & Yano T. 1989. Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea. *J. Hum. Evol.*, **18**:461–476.

Heyden S., Chao E. E., Vickerman K. & Cavalier-smith T. 2004. Ribosomal RNA Phylogeny of Bodonid and Diplonemid Flagellates and the Evolution of Euglenozoa. *J. Eukaryot. Microbiol.*, **51**:402–416.

Hirose E., Nozawa A., Kumagai A. & Kitamura S.-I. 2012. Azumiobodo hoyamushi gen. nov. et sp. nov. (Euglenozoa, Kinetoplastea, Neobodonida): a pathogenic kinetoplastid causing the soft tunic syndrome in ascidian aquaculture. *Dis. Aquat. Organ.*, **97**:227–35.

Huelsenbeck J. P. & Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**:754–755.

Huynen M. A. & Bork P. 1998. Measuring genome evolution. *Proc. Natl. Acad. Sci. U. S. A.*, **95**:5849–5856.

Janouškovec J., Gavelis G. S., Burki F., Dinh D., Bachvaroff T. R., Gornik S. G., Bright K. J., Imanian B., Strom S. L., Delwiche C. F., et al. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc. Natl. Acad. Sci. U. S. A.*, **114**:E171–E180.

Jedelský P. L., Doležal P., Rada P., Pyrih J., Šmíd O., Hrdý I., Šedinová M., Marcinčiková M., Voleman L., Perry A. J., et al. 2011. The Minimal Proteome in the

Reduced Mitochondrion of the Parasitic Protist Giardia intestinalis.Lightowlers B. (ed.). *PLoS One*, **6**:e17285.

Kamikawa R., Kolisko M., Nishimura Y., Yabuki A., Brown M. W., Ishikawa S. a, Ishida K., Roger A. J., Hashimoto T. & Inagaki Y. 2014. Gene content evolution in Discobid mitochondria deduced from the phylogenetic position and complete mitochondrial genome of Tsukubamonas globosa. *Genome Biol. Evol.*, **6**:306–15.

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

Katoh K. & Standley D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**:772–80.

Katz L. A. & Grant J. R. 2015. Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites. *Syst. Biol.*, **64**:406–415.

Keeling P. J. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **365**:729–48.

Kiethega G. N., Turcotte M. & Burger G. 2011. Evolutionarily conserved cox1 trans-splicing without cis-motifs. *Mol. Biol. Evol.*, **28**:2425–8.

Kolisko M., Silberman J. D., Cepicka I., Yubuki N., Takishita K., Yabuki A., Leander B. S., Inouye I., Inagaki Y., Roger A. J., et al. 2010. A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats. *Environ. Microbiol.*, **12**:2700–2710.

Kulda J. & Nohynkova E. 1978. Species., Flagellates of the human intestine and of intestines of other. *In*: Kreier J. P. (ed.), Parasitic Protozoa: Intestinal Flagellates,

Histomonads, Trichomonads, Amoeba, Opalinids, and Ciliates. New York, NY, USA:, Academic Press. p. 1–128.

Lartillot N., Brinkmann H., Philippe H., Philippe H., Delsuc F., Brinkmann H., Lartillot N., Kluge A., Rokas A., Williams B., et al. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.*, **7**:S4.

Lartillot N. & Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.*, **21**:1095–109.

Le S. Q. & Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.*, **25**:1307–20.

Leger M. M., Kolisko M., Kamikawa R., Stairs C. W., Kume K., Čepička I., Silberman J. D., Andersson J. O., Xu F., Yabuki A., et al. 2016. Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. *Nat. Ecol. Evol.*, **in press**.

Lopez P., Forterre P. & Philippe H. 1999. The Root of the Tree of Life in the Light of the Covarion Model. *J. Mol. Evol.*, **49**:496–508.

Lukeš J., Guilbride D. L., Votýpka J., Zíková A., Benne R. & Englund P. T. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell*, **1**:495–502.

Lukeš J., Hashimi H. & Zíková A. 2005. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr. Genet.*, **48**:277–299.

Lukeš J., Skalický T., Týč J., Votýpka J. & Yurchenko V. 2014. Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.*, **195**:115–122.

Makiuchi T. & Nozaki T. 2014. Highly divergent mitochondrion-related organelles in

anaerobic parasitic protozoa. *Biochimie*, **100**:3–17.

Marande W. & Burger G. 2007. Mitochondrial DNA as a Genomic Jigsaw Puzzle. *Science (80-.).*, **318**.

Marande W., López-García P. & Moreira D. 2009. Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. *Environ. Microbiol.*, **11**:3179–3188.

Michaeli S. 2011. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol.*, **6**:459–474.

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. *Int. J. Syst. Evol. Microbiol.*, **54**:1861–75.

Muller M., Mentel M., van Hellemond J. J., Henze K., Woehle C., Gould S. B., Yu R.-Y., van der Giezen M., Tielens a. G. M. & Martin W. F. 2012. Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiol. Mol. Biol. Rev.*, **76**:444–495.

Opperdoes F. R. & Borst P. 1977. Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: The glycosome. *FEBS Lett.*, **80**:360–364.

Palmer J. D. 2003. THE SYMBIOTIC BIRTH AND SPREAD OF PLASTIDS: HOW MANY TIMES AND WHODUNIT? *J. Phycol.*, **39**:4–12.

Pánek T., Zadrobílková E., Walker G., Brown M. W., Gentekaki E., Hroudová M., Kang S., Roger A. J., Tice A. K., Vlček Č., et al. 2016. First multigene analysis of Archamoebae (Amoebozoa: Conosa) robustly reveals its phylogeny and shows that Entamoebidae represents a deep lineage of the group. *Mol. Phylogenet. Evol.*, **98**:41–51.

Park J. S., Kolisko M., Heiss A. A. & Simpson A. G. B. 2009. Light microscopic observations, ultrastructure, and molecular phylogeny of Hicanonectes teleskopos n. g., n. sp., a deep-branching relative of diplomonads. *J. Eukaryot. Microbiol.*, **56**:373–384.

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

Philippe H. & Laurent J. 1998. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.*, **8**:616–623.

Philippe H., Lopez P., Brinkmann H., Budin K., Germot A., Laurent J., Moreira D., Müller M. & Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proceedings. Biol. Sci.*, **267**:1213–21.

Reyes-Prieto A., Weber A. P. M. & Bhattacharya D. 2007. The Origin and Establishment of the Plastid in Algae and Plants. *Annu. Rev. Genet.*, **41**:147–168.

Rodríguez F., Oliver, Jose L., Marin A. & Medina, J R. 1990. The General Stochastic Model o f Nucleotide Substitution. *J. Theor. Biol.*, **142**:485–501.

Roger A. J. 1999. Reconstructing Early Events in Eukaryotic Evolution. *Am. Nat.*, **154**:S146–S163.

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

Sakaguchi M., Nakayama T., Hashimoto T. & Inouye I. 2005. Phylogeny of the Centrohelida Inferred from SSU rRNA, Tubulins, and Actin Genes. *J. Mol. Evol.*, **61**:765–775.

Schneider R. E., Brown M. T., Shiflett A. M., Dyall S. D., Hayes R. D., Xie Y., Loo J. A. & Johnson P. J. 2011. The Trichomonas vaginalis hydrogenosome proteome is

highly reduced relative to mitochondria, yet complex compared with mitosomes. *Int. J. Parasitol.*, **41**:1421–34.

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

Simpson A. G. B. 1997. The identity and composition of the Euglenozoa. *Arch. für Protistenkd.*, **148**:318–328.

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

Simpson A. G. B., Gill E. E., Callahan H. A., Litaker R. W. & Roger A. J. 2004. Early evolution within kinetoplastids (euglenozoa), and the late emergence of trypanosomatids. *Protist*, **155**:407–22.

Simpson A. G. B., Lukeš J. & Roger A. J. 2002. The Evolutionary History of Kinetoplastids and Their Kinetoplasts. *Mol. Biol. Evol.*, **19**:2071–2083.

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

Simpson A. G. B., Radek R., Dacks J. B. & O'kelly C. J. 2002. How Oxymonads Lost Their Groove: An Ultrastructural Comparison of Monocercomonoides and Excavate Taxa. *J. Eukaryot. Microbiol.*, **49**:239–248.

Simpson A. G. B., Roger A. J., Silberman J. D., Leipe D. D., Edgcomb V. P., Jermiin L. S., Patterson D. J. & Sogin M. L. 2002. Evolutionary history of "early-diverging" eukaryotes: the excavate taxon Carpediemonas is a close relative of Giardia. *Mol. Biol. Evol.*, **19**:1782–91.

Simpson A. G. B., Stevens J. R. & Lukeš J. 2006. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.,* **22**:168–74.

Spang A., Saw J. H., Jørgensen S. L., Zaremba-Niedzwiedzka K., Martijn J., Lind A. E., van Eijk R., Schleper C., Guy L. & Ettema T. J. G. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature,* **521**:173–179.

Spencer D. F. & Gray M. W. 2011. Ribosomal RNA genes in Euglena gracilis mitochondrial DNA: fragmented genes in a seemingly fragmented genome. *Mol. Genet. Genomics,* **285**:19–31.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* **30**:1312–3.

Stiller J. W. & Hall B. D. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.*, **16**:1270–9.

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

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

Tanifuji G., Kim E., Onodera N. T., Gibeault R., Dlutek M., Cawthorn R. J., Fiala I., Lukeš J., Greenwood S. J. & Archibald J. M. 2011. Genomic characterization of neoparamoeba pemaquidensis (Amoebozoa) and its kinetoplastid endosymbiont. *Eukaryot. Cell,* **10**:1143–1146.

Tessier L. H., van der Speck H., Gualberto J. M. & Grienenberger J. M. 1997. The cox1 gene from Euglena gracilis: a protist mitochondrial gene without introns and genetic

code modifications. *Curr. Genet.*, **31**:208–13.

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

Vlcek C., Marande W., Teijeiro S., Lukes J. & Burger G. 2011. Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Res.*, **39**:979–88.

Wakeman K. C., Reimer J. D., Jenke-Kodama H. & Leander B. S. 2014. Molecular Phylogeny and Ultrastructure of *Caliculium glossobalani* n. gen. et sp. (Apicomplexa) from a Pacific *Glossobalanus minutus* (Hemichordata) Confounds the Relationships Between Marine and Terrestrial Gregarines. *J. Eukaryot. Microbiol.*, **61**:343–353.

Woese C. R., Kandlert O. & Wheelis M. L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Evolution (N. Y).*, **87**:4576–4579.

Yabuki A., Chao E. E., Ishida K.-I. & Cavalier-Smith T. 2012. Microheliella maris (Microhelida ord. n.), an Ultrastructurally Highly Distinctive New Axopodial Protist Species and Genus, and the Unity of Phylum Heliozoa. *Protist*, **163**:356–388.

Yabuki A., Inagaki Y. & Ishida K. 2010. Palpitomonas bilix gen. et sp. nov.: A Novel Deep-branching Heterotroph Possibly Related to Archaeplastida or Hacrobia. *Protist*, **161**:523–538.

Yabuki A., Ishida K. I. & Cavalier-Smith T. 2013. Rigifila ramosa n. gen., n. sp., a Filose Apusozoan with a Distinctive Pellicle, is Related to Micronuclearia. *Protist*, **164**:75–88.

Yabuki A., Kamikawa R., Ishikawa S. a, Kolisko M., Kim E., Tanabe A. S., Kume K.,

Ishida K.-I. & Inagaki Y. 2014. Palpitomonas bilix represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Sci. Rep.*, **4**:4641.

Yabuki A., Nakayama T., Yubuki N., Hashimoto T., Ishida K. I. & Inagaki Y. 2011. Tsukubamonas globosa n. gen., n. sp., a novel excavate flagellate possibly holding a key for the early evolution in “discoba.” *J. Eukaryot. Microbiol.*, **58**:319–331.

Yabuki A. & Tame A. 2015. Phylogeny and Reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrächter & Schnepf, 1996. *J. Eukaryot. Microbiol.*, **62**:426–429.

Yasuhira S. & Simpson L. 1996. Guide RNAs and guide RNA genes in the cryptobiid kinetoplastid protozoan, Trypanoplasma borreli. *RNA*, **2**:1153–60.

Yubuki N., Céza V., Cepicka I., Yabuki A., Inagaki Y., Nakayama T., Inouye I. & Leander B. S. 2010. Cryptic Diversity of Free-Living Parabasalids, Pseudotrichomonas keilini and Lacusteria cypriaca n. g., n. sp., as Inferred from Small Subunit rDNA Sequences. *J. Eukaryot. Microbiol.*, **57**:554–561.

Yubuki N., Huang S. S. C. & Leander B. S. 2016. Comparative Ultrastructure of Fornicate Excavates, Including a Novel Free-living Relative of Diplomonads: Aduncisulcus paluster gen. et sp. nov. *Protist*, **167**:584–596.

Yubuki N., Inagaki Y., Nakayama T. & Inouye I. 2007. Ultrastructure and Ribosomal RNA Phylogeny of the Free-Living Heterotrophic Flagellate Dysnectes brevis n. gen., n. sp., a New Member of the Fornicata. *J. Eukaryot. Microbiol.*, **54**:191–200.

Zhao Y., Davis R. E. & Lee I.-M. 2005. Phylogenetic positions of “Candidatus Phytoplasma asteris” and Spiroplasma kunkelii as inferred from multiple sets of concatenated core housekeeping proteins. *Int. J. Syst. Evol. Microbiol.*, **55**:2131–2141.