Molecular phylogeny and evolution of the amitochondriate protists

Nobuko Arisue

Doctor of Science

Department of Biosystems Science
School of Advanced Sciences
The Graduate University for Advanced Studies (Sokendai)
2001 (School year)

Abstract

Amitochondoriate protists are unicellular eukaryotes that lack mitochondria. Diplomonadida (including *Giardia*), Palabasala (including *Trichomonas*), Entamoebidae (including *Entamoeba*), Pelobionta (including *Mastigamoeba*), and Microsporidia (including *Encephalitozoon*) are well known as major lineages. Morphological evidence and findings on the ribosomes showing 'primitive eukaryotic' features for these lineages led a proposal of the 'Archezoa' hypothesis that several amitochondriate protist lineages (Archezoa) diverged preceding the endosymbiotic origin of proto-mitochondria, and thus have been living relics of the early phase of eukaryotic evolution.

Early studies on the eukaryotic phylogenies based on the small subunit (SSU) ribosomal RNA (rRNA) and on the translation elongation factors (EF) supported this hypothesis, placing three amitochondriate protists lineages, Microsporidia, Parabasala and Diplomonadida, at the basal position of the eukaryotic tree. However, phylogenies based on other genes, such as tubulins, mitochondorial-type heat shock protein 70 (HSP70mit) and the largest subunit of RNA polymerase II (RpoII), suggested that Microsporidia are not deep branching eukaryotes but are closely related to Fungi. Furthermore, phylogenies of various genes based on the recent accumulation of many sequence data from various protist lineages sometimes gave conflicting results with each other, indicating that the phylogenetic relationships among major eukaryotic lineages have still been an open problem.

On the other hand, mitochondrion-related genes that are coded in nuclear DNA and target for mitochondria were isolated from Entamoebidae, Microsporidia, Parabasala and Diplomonadida. The findings suggested that ancestors of these amitochondriate lineages once harbored mitochondria and lost them secondarily during their evolution.

On these backgrounds of the studies on early eukaryotic evolution, this work was intended to elucidate an evolutionary status of the amitochondriate protists. First, in order to establish a robust placement of the amitochondriate protist lineages in the eukaryotic tree, phylogenetic relationships among major eukaryotic lineages including amitochondriate

protists were statistically analyzed in detail by applying a combined maximum likelihood (ML) method to the sequence data of multiple genes. Next, in order to establish whether the ribosomal features of potentially early-branching lineages are 'primitive eukaryotic', the ribosomes of the amitochondriate protists, *Giardia intestinalis* (Diplomonadida) and *Trichomonas vaginalis* (Parabasala), were analyzed, and the components were compared to those of other eukaryotic organisms including amitochondriate protists.

Chapter 1 of this article described the analyses of the phylogenetic relationships among major eukaryotic lineages including amitochondriate protists, with the reports on 27 original sequence data of various genes mostly derived from amitochondriate protists.

At first, phylogenetic positions of Microsporidia and of stramenopiles were analyzed. In order to clearly settle a phylogenetic position of Microsporidia among major eukaryotic lineages, a combined ML analysis was performed using 6,391 positions from 10 genes for which data from Microsporidia were available. These genes were EF-1α, EF-2, valyl- and isoleucyl- tRNA synthetases (ValRS, IleRS), RpoII, Actin, α-tubulin, β-tubulin, SSUrRNA, and large subunit (LSU) rRNA. Although several phylogenies based on individual genes, such as EF-1α, IleRS, and SSUrRNA, did not support a close relationship between Microsporidia and Fungi, the combined analysis clearly demonstrated a relationship, (Metazoa, (Fungi, Microsporidia)) with a very high statistical support. In addition, another combined ML analysis was performed to examine a relationship between stramenopiles and Alveolata, using 5,423 positions from eight genes for which data from stramenopiles were available (EF-1α, EF-2, cytosolic-type HSP70 (HSP70c), non-catalytic 'B' subunit of vacuolar ATPase, Actin, β-tubulin, SSUrRNA, and LSUrRNA). The analysis demonstrated also with a very high statistical support that stramenopiles and Alveolata were the closest relatives with each other.

In the next, the phylogenetic position of the Pelobiont *Mastigamoeba balamuthi* was analyzed in relation to the position of *E. histolytica*. A combined ML analysis using 3,935 positions from four genes, SSUrRNA, LSUrRNA, EF-1α, and EF-2, suggested that *M. balamuthi* was the closest relative of *E. histolytica* and that Mycetozoa were placed at the

sistergroup to the common ancester of *M. balamuthi* and *E. histolytica*. These findings supported the notion, which had previously been proposed primarily on cytological evidence, that both *M. balamuthi* and *E. histolytica* are closely related to the Mycetozoa and that these three together represent a major eukaryotic lineage (Conosa).

Finally, on the basis of the findings as mentioned above and as currently reported in the literatures, 13 major eukaryotic lineages were divided into seven groups: 1. (Metazoa, (Fungi, Microsporidia)), 2. (Mycetozoa, (Pelobionta, Entamoebidae)) [Conosa], 3. (Viridiplantae, Rhodophyta), 4. (stramenopiles, Alveolata), 5. Euglenozoa, 6. Diplomonadida, and 7. Parabasala. Phylogenetic relationships among these groups with an outgroup were examined by a combined ML analysis of the genes, EF-1α, EF-2, ribosomal protein (Rp) S14, RpS15a, RpL5, RpL8, RpL10a, IleRS, ValRS, RpoII, chaperonin 60, HSP70mit, endoplasmic reticulum-type HSP70, HSP70c and cytosolic-type HSP90, chaperonin-containing testis complex polypeptide-1 subunit (CCT) α, CCTγ, CCTδ, CCTζ, Actin, α-tubulin, β-tubulin, SSUrRNA, and LSUrRNA. The combined ML analysis clearly supported with statistical confidence that Diplomonadida and Parabasala diverged earlier than other five groups in the eukaryotic tree, although the branching order between these two lineages were still open for further analysis. In addition, especially when among-site rate heterogeneity was taken into consideration, it was clearly supported that (Metazoa, (Fungi, Microsporidia)), Conosa, and (Viridiplantae, Rhodophyta) were the first, the second and the third earliest offshoots among the five groups excluding Diplomonadida and Parabasala. The analysis at the first time demonstrated robustly that Diplomonadida and Parabasala are the early branching eukaryotes, although presence of a potential artefact derived from a long branch attraction could not be ruled out entirely.

Chapter 2 of this article described the analyses of the ribosomes of potentially early branching amitochondriate protists, *G. intestinalis* and *T. vaginalis*.

Sedimentation analyses demonstrated that the sedimentation coefficients of these ribosomes were larger than that of *Escherichia coli* and smaller than that of *Saccharomyces*

cerevisiae or *Artemia salina*. Based on the radical free and highly reduced two dimensional polyacrylamide gel electrophoresis analysis, N-terminal sequencing analysis, and/or similarity search on the public database, the number of ribosomal proteins were estimated to be at least 74 for *G. intestinalis* and approximately 80 for *T. vaginalis*. These numbers were comparable with that of a 'typical' eukaryote (about 80) and larger than that of *E. coli* (about 55). The N-terminal sequences of the protein spots and alignment analyses of all the ribosomal proteins currently available revealed that the sequences of *G. intestinalis* and *T. vaginalis* are clearly of 'typical' eukayotic type with no exception.

On the other hand, sequence comparison analyses of rRNAs revealed that the SSU and LSUrRNAs of *G. intestinalis* and *T. vaginalis* were remarkably shorter in length than those of 'typical' eukaryotes. All the helices that belong to the universal core, however, were strictly conserved also in *G. intestinalis* and *T. vaginalis*. In contrast, variable regions of both rRNAs were reduced to be short in *G. intestinalis* and *T. vaginalis*.

As far as these results are concerned, the protein components and the essential parts of the rRNAs of the *G. intestinalis* and *T. vaginalis* ribosomes are clearly of 'typical' eukaryotic type. No 'primitive eukaryotic' features are found in the ribosomes of these amitochondriate protists. The smaller sedimentation coefficients of the ribosomes of *G. intestinalis* and *T. vaginalis* than those of 'typical' eukaryotes are due to the smaller size of rRNAs with shortened variable regions. These findings give additional evidence for fully developed eukaryotic nature of *G. intestinalis* and *T. vaginalis*. Probably Diplomonadida and Parabasala already had obtained major eukaryotic properties commonly found in the 'typical' eukaryotes.

Contents

Introduction

Among unicellular eukaryotic lineages (protists), there are some organisms that lack mitochondria. Major amitochondriate protists lineages, Diplomonadida (including *Giardia*), Parabasala (including *Trichomonas*), Entamoebidae (including *Entamoeba*), Pelobionta (including *Mastigamoeba*) and Microsporidia (including *Encephalitozoon*) are well-known.

Morphological studies up to 80's had revealed that several organisms of these lineages also lack peroxisomes, typical golgi dictyosomes and/or nucleoli (Cavalier-Smith 1987; Li 1999). In addition, ribosomes of Microsporidia and Parabasala had been described repeatedly as 'prokaryotic' or 'primitive eukaryotic' because of their shorter rRNAs than those of most eukaryotes (Vossbrinck et al. 1987; Chakrabarti et al. 1992), small (70S) sedimentation coefficient values (Ishihara and Hayashi 1968; Oka et al. 1973), and of lack of internal transcribed spacer region 2 in Microsporidia (Vossbrinck and Woose 1986). These lines of evidence showing 'primitive eukaryotic' features for these lineages led a proposal of the 'Archezoa' hypothesis that several amitochondriate protist lineages (Archezoa) diverged preceding the endosymbiotic origin of proto-mitochondria, and thus have been living relics of the early phase of eukaryotic evolution (Cavalier-Smith 1987).

Early studies on the eukaryotic phylogeny were carried out by small subunit ribosomal RNA (SSUrRNA) genes (Sogin et al. 1989; Sogin 1991; Leipe et al. 1993). According to the SSUrRNA tree by Leipe et al. (1993), among amitochondriate protists, *Vairimorpha necatrix* (Microsporidia), *Ttitrichomonas foetus* (Parabasala) and the common ancestor of *Giardia lamblia* (*G. intestinalis*) and *Hexamita inflata* (Diplomonadida) represented early offshoots among eukaryotic lineages. However, another amitochondriate protist *Entamoeba histolytica* (Entamoebidae) branched off after the divergences of Euglenozoa and other mitochondrion-containing protists. Although an alternative examination of the eukaryotic phylogeny by the use of elongation factors (EF) 1α and 2 was not necessarily comparable with the SSUrRNA tree in detail, it also supported the early emergence of the three amitochondriate protist lineages and the late emergence of *E. histolytica* (Hashimoto et al.

1994, 1995ab, 1997; Shirakura et al. 1994; Hashimoto and Hasegawa 1996; Nakamura et al. 1996ab; Kamaishi et al. 1996; Yamamoto et al. 1997). However, phylogenies based on other genes, such as tubulins, mitochondrial-type heat shock protein 70 (HSP70mit) and the largest subunit of RNA polymerase II (RpoII), suggested that Microsporidia are not deep branching eukaryotes but are closely related to Fungi. Furthermore, phylogenies of various genes based on the recent accumulation of many sequence data from various protist lineages sometimes gave conflicting results with each other, indicating that the phylogenetic relationships among major eukaryotic lineages have still been an open problem.

As the SSUrRNA tree suggested the possibility of secondary loss of mitochondria in *E. histolytica*, an approach to find mitochondrial traces was applied to this organism. In 1995, evidence for secondary loss of mitochondrial function in *E. histolytica* has been obtained by the detection of mitochondrion-related chaperonin (CPN60) genes that are coded in the nuclear DNA (Clark & Roger 1995). Followed by this finding, other amitochondriate lineages, Diplomonadida, Parabasala and Microsporidia were also reported to have mitochondrion-related genes and it was confirmed that all of these amitochondriate lineages were once harbored mitochondria and lost them secondarily in their evolutionary past. These findings are listed in Table I. The CPN60 and mitochondrial-type heat shock protein 70 (HSP70mit) were shown to be localized to the hydrogenosome in *Trichomonas vaginalis* (Bozner 1997), while in *E. histolytica*, CPN60 is localized in a recently identified double membrane bounded organelle, called mitosome (Tovar, Fischer and Clark 1999) or crypton (Mai et al. 1999). These lines of evidence suggest a common evolutionary origin of these organelles with the mitochondrion (Martin and Müller 1998; Dyall and Johnson 2000; Rotte et al. 2000).

On these backgrounds of the studies on early eukaryotic evolution, our work was intended to elucidate an evolutionary status of the amitochondriate protists. First, in order to establish a robust placement of the amitochondriate protist lineages in the eukaryotic tree, phylogenetic relationships among major eukaryotic lineages including amitochondriate

protists were statistically analyzed in detail by applying a combined maximum likelihood method to the sequence data of multiple genes. Next, in order to establish whether the ribosomal features of potentially early-branching lineages are 'primitive eukaryotic', the ribosomes of amitochondriate protists, *G. intestinalis* and *T. vaginalis*, were characterized, and the components were compared to those of other eukaryotic organisms including amitochondriate protists.

Chapter 1 Phylogenetic analyses among major eukaryotic lineages including amitochondriate protists

1.1 Introduction

Until recently phylogenetic analysis of the eukaryotic tree was performed mainly by SSUrRNA sequences. According to the SSUrRNA trees previously published, amitochondriate protists were robustly located at early branching positions among eukaryotic lineages (Sogin et al. 1989, Sogin 1991; Leipe et al. 1993). However, the SSUrRNA phylogeny may sometimes be unreliable due to the drastic difference of G+C compositions among species (Hasegawa, Hashimoto and Adachi 1992; Hasegawa and Hashimoto 1993). Although, some methods to overcome the base composition problem have been developed and reliable findings may have been obtained on the deep branching eukaryotes of SSUrRNA phylogeny (Galtier and Gouy 1995; Swofford et al. 1996), the result inferred only by a single gene was shown to be still fragile (Cao et al. 1994). Since alternative phylogenetic analyses based on other molecular data were highly desired, re-examination of the phylogenetic relationships among eukayotic lineages were performed by using amino acid sequence data of translation elongation factors 1α (EF-1α) and 2 (EF-2) (Hashimoto et al. 1994, 1995ab, 1997; Shirakura et al. 1994; Hashimoto and Hasegawa 1996; Nakamura et al. 1996ab; Kamaishi et al. 1996; Yamamoto et al. 1997). Although the EF phylogenies were not necessarily consistent with the SSUrRNA phylogeny and did not give a good resolution among early branching eukaryotic lineages. In addition, a different view of the phylogenetic placement of eukaryotic lineages was suggested. In contrast to the basal eukaryotic position of Microsporidia in the SSUrRNA and EF trees, phylogenies of β-tubulin (Edlind et al. 1996) and α-tubulin (Keeling and Doolittle 1996) placed Microsporidia within the fungal lineages. A phylogeny of the largest subunit of RNA polymerase II (RpoII) also supported the Fungal relationship of Microsporidia (Hirt et al. 1999). Recently many molecular data have been reported and stored in public database. Using these data, many researchers have made effort to resolve the evolutionary relationships of eukaryotes, however, up to date, robust results have not been

obtained.

To contribute to more clearly resolve the eukaryotic phylogeny, relationships among major eukaryotic lineages including amitochondriate protists were explored using multiple molecular sequence data with combined maximum likelihood (ML) methods. The major eukaryotic lineages were schematically represented in Fig. 1.1. First, the close relationship among Metazoa, Fungi and Microsporidia, and that between stramenopiles and Alveolata were confirmed. Next, monophyly of Mycetozoa, Entamoebidae and Pelobionta with the latter two being the closest relatives were demonstrated. Finally, by constraining these relationships in advance, relationships among seven eukaryotic major lineages, 1) Metazoa, Fungi, and Microsporidia, 2) Mycetozoa, Entamoebidae, and Pelobionta, 3) Viridiplantae and Rhodophyta, 4) Euglenozoa, 5) stramenopiles and Alveolata, 6) Diplomonadida, and 7) Parabasala, were explored using an outgroup.

1.2 Phylogenetic position of Microsporidia and of stramenopiles

When the evolutionary relationships among major eukaryotic lineages are analyzed by an exhaustive topology search of the ML method, eight lineages (10,395 tree topologies) are the maximum numbers executable in personal computer. In this respect, it is necessary to confirm close relationships among major lineages of eukaryotes, and to make constraints in advance for these relationships, for reducing the number of lineages less than eight. For this purpose, a close relationship among Metazoa, Fungi and Microsporidia and, that between stramenopiles and Alveolata, were examined by combined ML methods using multiple molecular sequence data. Previous studies suggested a close relationship among Metazoa, Fungi and Microsporidia by individual analyses of different genes (Edlind et al. 1996; Keeling and Doolittle 1996; Germot, Phillipe and Le Guyader 1997; Peyretaillade et al. 1998b; Hirt et al. 1999; Fast, Logsdon and Doolittle 1999). Previous rRNA phylogenies (Van de Peer, Ben Ali and Meyer 2000; Van de Peer and De Wachter 1997; Van de Peer et al. 2000) and a phylogeny by concatenated protein data (Baldauf et al. 2000) suggested a close

relationship between stramenopiles and Alveolata.

Microsporidia are unicellular amitochondriate protists. In the phylum, 144 genera and over 1,000 species are contained (Bryan & Schwartz 1999). They are obligate intracellular parasites infecting many animal lineages such as human, fish and insects, and forming spore. Over 80 genera of Microsporidia have been reported in human disease (Weiss and Vossbrinck 1999). In recent years, Microsporidia have emerged as important opportunistic parasites in HIV infected patients. The genome size of many Microsporidia have been determined and vary from 2.3~19.5 Mb (Weiss and Vossbrinck 1999). The genus *Encephalitozoon* have a small genome size less than 3 Mb and *E. intestinalis* is the smallest genome (2.3 Mb) among eukaryotic genomes reported to date. The genome project of *E. cuniculi* has finished and all genes encoded by the genome were reported by Katinka et al. (2001). The 2.9 Mb genome consisting of the 11 chromosomes has reduced intergenetic spacers and shows shortness of most putative proteins relative to their eukaryotic orthologs. A Microsporidian, *Nosema bombycis* has 70S ribosome (Ishihara and Hayashi 1968). The rRNA genes of Microsporidians are shorter than those of other organisms including both eukaryotes and prokaryotes (Vossbrinck et al. 1987; Sogin et al. 1989; Chakrabarti et al. 1992; Peyretaillade et al. 1998b). The 5.8S rRNA region of Microsporidians are directly join to the large subunit (LSU) rRNA as seen in prokaryotes (Vossbrinck and Woose 1986). In addition, the early branching position of Microsporidia in eukaryotic lineages were suggested by SSUrRNA and EF phylogeny (Vossbrinck et al. 1987; Sogin et al.1989; Leipe et al. 1993; Hashimoto et al. 1997; Hashimoto and Hasegawa 1996; Kamaishi et al. 1996). From these lines of evidence, Microsporidia were previously considered to be primitive eukaryotes. However, early emergence of Microsporidia in the SSUrRNA and the EF phylogenies might have been affected by a long branch attraction (LBA) (Philippe and Laurent 1998) due to accelerated evolutionary rates of both genes of Microsporidia. On the other hand, placement of the Microsporidia with Fungi was proposed by phylogenies of α-tubulin (Keeling and Doolittle 1996), β-tubulin (Edlind et al. 1996), HSP70mit (Germot, Phillipe and Le Guyader 1997),

RNA polymerase II (RpoII) (Hirt et al. 1999) and TATA box binding protein (Fast, Logsdon and Doolittle 1999). The LSUrRNA phylogeny with a model of among-site rate heterogeneity (Peyretaillade et al. 1998b) also did not support the early branching position of Microsporidia in a eukaryotic tree. In this study, the phylogenetic position of Microsporidia was analyzed by multiple molecular data sets. For this purpose, ValRS, isoleucyl (Ile) RS and EF-2 sequences were newly determined from several protists. A combined analyses of multiple genes supported a close relationship between Microsporidia and Fungi with 100% bootstrap proportions (BP). In addition, the sister group of this clade was also clearly shown to be Metazoa with 100% BP.

In order to confirm the phylogenetic position of stramenopiles among eukaryotic lineages, cloning and sequencing analyses were performed for several genes from a human parasite, *Blastocystis hominis*, which has recently been allied with stramenopiles by the SSUrRNA phylogeny (Silberman et al. 1996).

Blastocystis hominis is an obligate anaerobic parasite of the human intestine. It has a number of unique biological features, such as a large membrane-bound central vacuole taking up as much as 95 % of the cell, multiple nuclei, and cytochrome-free mitochondria (e.g. Boreham and Stenzel 1993; Stenzel and Boreham 1996; Zierdt 1991). *Blastocystis hominis* shows a variety of cell forms during its life cycle, but no flagellated forms have ever been identified. There are indications that its anaerobic mitochondria are functional, although many mitochondrial enzymes demonstrate no activity. *Blastocystis hominis* thus is an organism of great interest from various biological points of view, and its phylogenetic position within eukaryotes has long been unclear.

Blastocystis hominis was first described in 1912 as a yeast (Brumpt 1912). Subsequent physiological and electron microscopic studies indicated that it more likely belons to the protozoa (Zierdt 1973). A recent molecular phylogenetic study based on SSUrRNA (Silberman et al. 1996) demonstrated that both *B. hominis* (Nand strain) and a *Blastocystis* isolate from guinea pig are clearly placed within a complex heterogeneous lineage,

'stramenopiles', that include unicellular and multicellular protists whose flagella bear tripartite tubular hairs (Patterson 1989). Both heterotrophs and autotrophs (photosynthetic) belong to this the lineage, including brown algae, golden-brown algae, diatoms, slime nets, and water molds. The stramenopiles are located within the terminal 'crown' groups of the SSUrRNA tree, together with Viridiplantae, Rhodophyta, Metazoa, and Fungi. In contrast, a protein phylogeny of EF-1α suggested that *B. hominis* very likely diverged earlier than terminal crown groups (Ho et al. 2000; Nakamura et al. 1996). To further confirm the phylogenetic position of *B. hominis* among eukaryotes using additional sequence information, SSUrRNA, cytosolic type heat shock protein 70 (HSP70c), EF-2, and the non-catalytic 'B' subunit of the vacuolar ATPase (v-ATPB) from *B. hominis* (HE87-1 strain) and EF-2 data from an Alveolata parasite, *Plasmodium falciparum* were determined and performed phylogenetic reconstruction. The HSP70c analysis clearly supported the SSUrRNA phylogeny in which *B. hominis* is positioned within stramenopiles. Although no clear resolution among the major eukaryotic lineages was obtained by the individual molecular phylogenies, a combined analysis of various genes clearly positioned *Blastocystis*/stramenopiles as a sister group of Alveolata.

1.2.1 Materials and methods

Genes from three protists cloned and sequenced for this study were listed in Table 1.1.

Entamoeba histolytica Extraction of genomic DNA of *E. histolytica* strain HM-1:IMSS (ATCC 30456) was according to the protocol for *Trypanosoma brucei* (Patnaik, Kulkarni and Cross1993). Gemomic and cDNA libraries of *E. histolytica* were kindly provided from Dr. L. Sánchez (The Rockefeller Univ., New York). Primers were synthesized based on highly conserved regions of the ValRS and IleRS sequences. Those were: ValRS: ATTRIETM (forward) 5'-GCNACNACNMGNATHGARACNATG-3' and FCNK(L/I)W(N/Q)A (reverse) 5'-GCNTKCCANADYTTRTTRCARAA-3' ; IleRS: AWTTTPWTL (forward) 5'-GCNTGGACNACNACNCCNUGGACNYT-3' and

DQTRGWFY (reverse) 5'-TARAACCANCCNCDNGTYTGRTC-3'. Corresponding fragments were amplified by polymerase chain reactions (PCR) with these primers and genomic DNA of *E. histolytica*. The fragments were purified by using a Gel Extraction Kit (QIAGEN) and cloned into pT7 vector (Novagen). The excised inserts labeled with a Random Primers DNA Labeling Systems (GIBCO/BRL) were used to probe and isolate clones containing ValRS and IleRS genes from libraries.

Plasmodium falciparum Genomic DNA of the *P. falciparum* FCR3 strain was kindly provided by Drs. K. Kita and A. Takeo (Univ. of Tokyo). Using degenerate primers synthesized based on the conserved amino acid residues in the alignment of ValRS, IleRS and EF-2 sequences and unique primers synthesized based on the partial sequencing result, major part of the *P. falciparum* ValRS, IleRS and EF-2 genes were amplified. The amino acid and the corresponding nucleotide sequences used for synthesizing the degenerate primers are: ValRS: PPPNVTG (forward) 5'-CCNCCNCCNAAYGTNACNGG-3', ATTRIETM (forward) 5'-GCNACNACNMGNATHGARACNATG-3', GTDALRF (reverse) 5'-AANCKNARNGCRTCNGTNCC-3', FCNK(L/I)W(N/Q)A (reverse) 5'-GCNTKCCANADYTTRTTRCARAA-3'; IleRS: FYDGPPFAT (forward) 5'-TTYTAYGAYGGNCCNCCNTTYGCNAC-3', AWTTTPWTL (forward) 5'-GCNTGGACNACNACNCCNUGGACNYT-3', DQTRGWFY (reverse) 5'-TARAACCANCCNCDNGTYTGRTC-3', WYIRLNR (reverse) 5'-CKRTTNARNCKDATRTACCA-3'; EF-2: GAGELHLE (forward) 5'-GGNGCNGGNGARYTNCAYYTNGA-3', FPQCVFDHW (reverse) 5'-CARTGRTCRAANACRCAYTGNGGRAA-3', GVCVQTETV (forward) 5'-GGNGTNTGYGTNCARACNGARACNGT-3', NMSVIAHVD (forward) 5'-AAYATGWSNGTNATHGCNCAYGTNGA-3'. Plasmid vectors pT7 (Novergen) or pCR2.1 (Invitrogen) were used for cloning the PCR amplified bands.

Blastocystis hominis *Blastocystis hominis* (Strain HE87-1) was maintained in culture according to the method of Yoshikawa, Kuwayama, and Enose (1995). A total of 2×10^7 cells

were processed by a standard ultracentrifuge technique for genomic DNA, and total and poly(A)$^+$ RNA extraction. Poly(A)$^+$ RNA was treated with murine reverse transcriptase and the resultant mRNA:cDNA duplex was used for a cDNA library construction in λZAP II vector (ZAP-cDNA synthesis kit, STRATAGENE). Either mRNA:cDNA duplex or genomic DNA was used as a template for PCR amplification of the partial gene fragment. A pair of sense (5'-TACTTGGTTGATCCTGCC-3') and anti-sense (5'-AACCTTGTTACGACTT-3') primers was synthesized to amplify almost the entire SSUrRNA gene of *B. hominis* HE87-1 strain: a 1.6-kb amplified product was purified and cloned into pT 7 plasmid. For each analysis of the protein coding sequences, a pair of sense and anti-sense primers, respectively degenerated for the highly conserved amino acid regions in the alignment of eukaryotic sequences was synthesized: 5'-GGNGCNGGNGARYTNCAYYTNGA-3' (GAGELHLE) and 5'-CARTGRTCRAANACRCAYTGNGGRAA-3' (FPQCVFDHW) for EF-2, 5'-GGNATHGAYYTNGGNACNAC-3' (GIDLGTT) and 5'-TCYTCNCCNCCNARRTGNGT-3' (THLGGED) for HSP70c, and 5'-TAYCCNGARGARATGATHCARACNGG-3' (YPEEMIQTG) and 5'-GGRTGNGTDATRTCRTCRTTNGGCAT-3' (MPNDDITHP) for v-ATPB. Amplified products were 0.75 kb (EF-2), 0.7 kb (HSP70c), and 0.6 kb (v-ATPB) in length, respectively, and these were cloned into pT 7 plasmids. After confirming that the nucleotide sequences of these fragments actually encoded the corresponding proteins, these fragments were labeled with digoxigenin (DIG, Boehringer Mannheim, Mannheim, Germany) by the randomly priming method according to manufacturer's manual and used as probes for screening the *B. hominis* cDNA library described above. Several reactive plaques were isolated, purified, and excised into pBluescript I plasmids.

Nucleotid sequences of clones were determined on both strands with the use of an automated DNA sequencer (310 Genetic Aanalyzer, ABI PRISM).

Sequence alignments For preliminary sequence alignments of the proteins, the SAM 2.1 program were used. That applies a linear hidden Markov model to facilitate recognition of

conserved subdomains within a protein family (Hughey and Krogh 1996). The alignments were further improved by manual editing. The secondary structure-based SSU and LSUrRNA alignment database constructed by Van de Peer et al. (http://rrna.uia.ac.be/)(Van de Peer and De Wachter 1997) was searched, and most of eukaryotic SSU and LSUrRNA sequences used for the phylogenetic analyses were obtained. Original *B. hominis* SSUrRNA sequences and few sequences which were not included in the database were aligned with the data set manually.

Phylogenetic analysis The maximum likelihood (ML) method was used to infer phylogenetic relationships among related sequences. The analysis was performed with the NUCML, PROTML, and TOTALML programs in the MOLPHY (version 2.3) program package (Adachi and Hasegawa 1996). The HKY85 model of nucleotide substitutions (Hasegawa, Kishino and Yano 1985) and the JTT-F model of amino acid substitutions (Jones, Taylor and Thornton 1992) were assumed in the analysis, because these models better approximated each data set than simpler models. Transition/transversion ratio for the HKY85 model and amino acid compositions for the JTT-F model were estimated from the data set under analysis. Since the number of operational taxonomic units (OTUs) to be analyzed was large, we searched for an optimal ML tree by the exact ML method only among the alternative trees that were produced by the local rearrangement option (-R) and the quick-add OTU search option (-q -n2000) of the NUCML or PROTML program. Several different trees including the neighbor joining (NJ) tree were used as initial trees for the -R option analysis. In addition, for only limited numbers of the tree topologies of interest, the ML analysis that takes among-site rate heterogeneity into consideration (Yang 1996) was carried out by assuming a discrete Γ-distribution with eight categories. The analysis was performed by the BASEML and CODEML programs in the PAML package (Yang 1997). Bootstrap proportion (BP) for each node of the best tree with the best log-likelihood score among alternatives was calculated by applying the resampling estimated log-likelihood (RELL) method (Kishino, Miyata, and Hasegawa 1990) with 10,000 resamplings to the alternative trees mentioned above. Statistical

confidence of the best tree and its alternatives was evaluated by *p*-values of the test by Kishino and Hasegawa (1989) for comparison between the best tree and each alternative tree and of the test by Shimodaira and Hasegawa (1999) that takes multiple comparison correction into consideration (Goldman, Anderson, and Rodrigo 2000). The program package CONSEL (Shimodaira and Hasegawa 2001) was used for the calculation of these *p*-values.

In order to evaluate the support for a particular tree, the results based on individual genes were combined to analyze the total information residing in the sequences studied. In this analysis the relationships of seven eukaryotic lineages (Microsporidia, Fungi, Metazoa, Mycetozoa, Viridiplantae, Euglenozoa, Alveolata) or (*Blastocystis*/stramenopiles, Alveolata, Metazoa, Fungi, Mycetozoa, Viridiplantae, Euglenozoa) were explored by exhaustive search of all possible 945 alternative trees. Genes used for these combined analyses were summarized in Table 1.2. First, the -R option of the NUCML or PROTML analysis was applied for each data set, and the relationship within each eukaryotic lineage was assumed in advance according to the result of the analysis. Next, an exhaustive search was performed on each data set by examining the 945 possible tree topologies. Finally, by the use of TOTALML program in the MOLPHY package, the results of individual analyses were combined by summing up the estimated log-likelihoods of all the positions for a given topology.

1.2.2 Results and discussion

1.2.2.1 Cloning and sequencing of genes

Using the PCR-based sequence analysis, partial EF-2 of *P. falciparum* were deduced from a nucleotide sequence of 2,336 bp (445 amino acid in length), which covers a major part of the putative EF-2 gene. The deduced amino acid sequence showed the highest similarity score with the EF-2 of an apicomplexan parasite, *Cryptosporidium parvum*, which belongs to Alveolata. The sequence was therefore identified as a *P. falciparum* counterpart of EF-2. Major part of ValRS and IleRS genes of *P. falciparum* were also cloned and sequenced using the PCR-based sequence analysis. Those coded 628 and 799 amino acids respectively for

ValRS and IleRS. With the use of the PCR based fragments for probes, ValRS and IleRS genes of *E. histolytica* were screened from genomic DNA and cDNA libraries. Isolated clones were sequenced and nucleotides sequences of the ValRS and IleRS genes were determined. Both of the clones for ValRS and IleRS did not contain a part of the gene putatively encoding N-terminal few residues. The trancated open reading frames (ORFs) encoded 1,004 and 1,036 amino acids for ValRS and IleRS, respectively. These sequences contained the conserved KMSK(S/R) and HIGH motives characteristic of all class I aminoacyl-tRNA synthetase (Moras 1992). Nuclear-coded ValRS genes of eukaryotes is regarded as a mitochondrial origin. In the region downstream of the KMSKS motif of ValRS sequences, a 37-residue insertion was shared by amitochondriate eukaryotes, and β- and γ-proteobacteria (Hashimoto et al. 1998). ValRS sequences of *E. histolytica* and *P. falciparum* contained this signature sequence. Presence of a mitochondriate eukaryotic type ValRS sequence in an amitochondtriate, *E. histolytica* supports the secondary loss of mitochondria in the evolutionary past of this protist. Large insertions with asparagines rich sequences were observed both ValRS (position 393~434) and IleRS (position 179~266) sequences of *P. falciparum*, as often seen in *P. falciparum* sequences (Zhang et al. 1999). Sequences of *P. falciparum* EF-2, ValRS and IleRS had striking bias for A or T at the third codon position, a codon usage bias typical of *P. falciparum* genes (Musto et al. 1999).

Using the PCR-based cloning method described above, two different SSUrRNA sequences were obtained for the *B. hominis* HE87-1 strain. These were tentatively named HE87-1A (AB023499) and HE87-1B (AB023578). Both sequences, except for the primer regions were 1,625 bp long with a 99.6% identity. The HE87-1A and HE87-1B sequences showed a 97.3% and 97.6% identity, respectively, with the SSUrRNA sequence of Nand strain (U51151) (Silberman et al. 1996), and showed a 96.2% and 94.7% identity, respectively, with that of *Blastocystis* sp. (U51152) (Silberman et al. 1996) (U51152). Although the sequences of HE87-1 and Nand strains were very similar to each other, a 35 bp region within the variable region, which corresponds to the E23 region of the *Homo sapiens* sequence (Van de

Peer et al. 2000), was significantly different between the two strains.

Several positive plaques for HSP70c, EF-2, and v-ATPB were obtained for each screening of the *B. hominis* cDNA. One of the positive clones for each gene was completely sequenced with both strands for each gene. The ORFs of the three genes revealed 623, 846, and 534 amino acid proteins, which were highly similar to typical eukaryotic HSP70c, EF-2 and v-ATPB sequences, respectively. Thus, the ORFs were identified as *B. hominis* homologs of the corresponding proteins. These three sequences and the EF-1α sequence (D64080) reported previously (Nakamura et al. 1996) were characterized by high G+C contents (72% through 88%) at the third codon positions, and showed similar codon usage bias (data not shown).

The 5'-noncoding regions of the above sequences consisted of 12, 5, and 35 nucleotides in the HSP70c, EF-2, and v-ATPB cDNAs, respectively. Unless these clones independently had depletions at the 5'-end due to technical artifact, the noncoding regions were very short compared with those of vertebrate cDNAs (Kozak 1987). Short leader sequences have been found in another protist, *G. intestinalis* (Adam 1991; Holberton and Marshall 1995). Binding of the ribosome to the mRNA and the initiation of translation may be more closely linked in *G. intestinalis* and also in *B. hominis*, than in vertebrates and other higher eukaryotes. A poly(A) tail was preceded by a very short 3'-noncoding sequence in each cDNA. The lengths were 28, 22, and 23 nucleotides in the cDNAs of HSP70c, EF-2, and v-ATPB, respectively. There was no stretch that was similar to a polyadenylation consensus sequence either in Metazoa, Fungi, or Vviridiplantae (AATAAA), or in *G. intestinalis* (AGTRAA) (Adam 1991; Peattie et al. 1989).

1.2.2.2 Phylogenetic position of Microsporidia

For genes shown in Table 1.2, at least one sequence could be available from each eukaryotic lineage were used for combined ML analysis (Table 1.2.). Combined data set consisted of 3,879 and 2,512 positions of protein and rRNA, respectively with total of 6,391

positions. Of the 945 trees in the combined analysis without Γ-distribution, 29 trees were selected as candidate trees (Table 1.3.). Those were, the ML trees, 20 trees that were not statistically different from the ML tree by the criterion of 3SE of log-likelihood difference (Kishino and Hasegawa 1989), and eight trees each of which was selected as the ML tree by the individual analysis, however were not included in the above trees. The ML analysis with Γ-distribution was carried out for these 29 trees. Same tree topology (Tree 1) was selected as the best tree in the analysis with Γ-distribution as in the analysis without Γ. As shown in Fig. 1.2, with over 99% BP support, the ML tree clearly indicated a close relationship between Microsporidia and Fungi, and this clade showed sister group relationship to Metazoa with 100% BP. Except for the analyses of α- and β-tubulins the ML-trees selected by individual analyses of the eight genes showed a log-likelihood difference from Tree 1 larger than its 3SE in the combined analysis. However, even when among-site rate heterogeneity was not taken into consideration (without Γ) analyses of the individual genes except for SSUrRNA and EF-1α did not reject Tree 1 in a statistically significant way from the corresponding ML tree. When analyses with Γ were performed, none of the difference between each ML tree and Tree 1 was statistically significant, suggesting that no individual analysis was in disagreement with Tree 1. In the criterion by Kishino and Hasegawa (1989), Tree 2 through Tree 8 (Table 1.3, without Γ) or Tree 2 and Tree 5 through Tree 8 (Table 1.3, with Γ) were not significantly different from the ML tree (p>0.05). All of these trees supported the ((Microsporidia, Fungi), Metazoa) relationship. However, if multiple comparison correction was taken into consideration by the method of Shimodaira and Hasegawa (1999), Tree 2 through 21 (Table 1.3, without Γ) or Tree 2 through 20 excluding Tree 17 were not significantly different from the ML tree (p>0.05). The ((Microsporidia, Fungi), Metazoa) relationship was not obtained in several of these trees but, at least monophyly of these three lineages was obtained in all of these trees.

Although individual phylogenies did not always indicate the close phylogenetic position of Microsporidia with Fungi as the ML tree, the combined ML analysis using 10 molecular

data sets clearly supported the ((Microsporidia, Fungi), Metazoa) relationship with extremely high BPs. Phylogenetic position of other eukaryotic lineages were not clearly resolved in this analysis. However, the BPs for the close relationship between Alveolata and Euglenozoa, and for sister group relationship of this clade to Viridiplantae in the analysis with Γ-distribution were higher than those of the analysis without Γ-distribution.

1.2.2.3 Phylogenetic position of stramenopiles

Phylogenetic position of *Blastocystis* The individual analysis of SSUrRNA and HSP70c genes clearly demonstrated that *B. hominis* HE87-1 strain is positioned within stramenopiles (data not shown), being comparable with the previous SSUrRNA phylogeny by Silberman et al. (1996), including *B. hominis* Nand strain and *Blastocystis* sp. isolated from a guinea pig. Since the phylogenetic analyses based on the two data sets, SSUrRNA and HSP70c, independently reached the same conclusion, the close relationship of *Blastocystis* with stramenopiles is corroborated. The stramenopiles, including brown algae, golden-brown algae, diatoms, labyrinthulida, bicosoecids, oomycetes, and a number of other lineages, are one of the most diverse assemblages of eukaryotes. It includes a wide range of unicellular and multicellular, as well as heterotrophic and autotrophic organisms. The stramenopiles have been described as a monophyletic taxon that comprises all organisms with tripartite tubular hairs on their flagella and cell surface (Patterson 1989; Cavalier-Smith 1986; Grell, Heini, and Schéller 1990). The homology of the tubular hairs has been asserted based on their ultrastructure, location, and biogenesis. However, *Blastocystis* has neither flagella nor tubular hairs (Zierdt 1991), in spite of its clear phylogenetic placement within the stramenopiles. Since *Proteromonas lacertae*, the closest relative of *Blastocystis* in the SSUrRNA tree described above and the trees reported previously (Silberman et al. 1996; Sogin and Silberman 1998), has both, the lack of these features in *Blastocystis* might be due to their secondary loss on the line leading to the common ancestor of *Blastocystis*.

Phylogenetic position of stramenopiles among major eukaryotic lineages based on

the combined analysis Of the 945 trees in the combined analysis without Γ-distribution, 18 trees were selected as candidate trees (Table 1.4): the ML tree; 14 trees that were not statistically different from the ML tree by the criterion of 3SE of log-likelihood difference (Kishino and Hasegawa, 1989); and 3 trees, each of which was selected as the ML tree by the individual analysis, but were not included in the above 15 trees. The ML analysis with Γ-distribution was carried out for these 18 trees. With extremely high BP supports, the ML tree clearly demonstrates a close relationship between Metazoa and Fungi (99.51%), and between Alveolata and stramenopiles (99.95%) (Fig. 1.3). The position of Mycetozoa as an outgroup to the Metazoa-Fungi clade is also supported, and is congruent with the results of previous reports (Baldauf 1999; Baldauf and Doolittle 1997; Kuma et al. 1995). If the criterion by Kishino and Hasegawa (1989) is used, most of the alternative trees except either for Tree 2 and Tree 3 (Table 1.4, without Γ) or for Tree 2 (Table 1.4, with Γ) were significantly different from the ML tree ($p< 0.05$). However, if multiple comparison correction is taken into consideration by the method of Shimodaira and Hasegawa (1999), Tree 2 through Tree 13 (Table 1.4, without Γ) or Tree 2 through Tree 10, Tree 12, and Tree 13 (Table 1.4, with Γ) were not significantly different from the ML tree ($p>0.05$). Most of these trees except for Tree 13 place stramenopiles as the closest relative of Alveolates, confirming the sister group relationship of these two lineages as shown by the high BP (Fig. 1.3). Although the ML trees for the individual data sets were different, none of them showed statistically significant differences from the ML tree of the combined analysis (Tree 1) in each individual analysis either with or without Γ-distribution (data not shown), demonstrating that none of the ML trees of the individual genes rejects the combined ML tree (Fig. 1.3.). It has to be stressed, thus, that as far as the seven lineages examined are concerned, sister group relationship between Alveolata and stramenopiles is robustly reconstructed by the total evidence of the eight genes. However, the possibility that the root is in either of these two lineages cannot be ruled out by the present analysis.

Overall, the individual results of the phylogenies in this analysis are congruent with

each other in recovering the monophyly of major eukaryotic lineages, and the combined analysis clearly demonstrates the Alveolata-stramenopile relationship, which was suggested also by the previous combined analysis by Baldauf et al. (2000). However, the relationship of the Alveolata-stramenopile clade to the other eukaryotic major lineages still remains to be established by further analyses.

1.3 Phylogenetic position of the Pelobionta *Mastigamoeba balamuthi*

Because of their medical importance, most of the research efforts on amitochondriate protists were directed toward human parasites, including the Diplomonadida *G. intestinalis*, the Entamoebidae *E. histolytica*, the Parabasala *Trichomonas vaginalis*, and several microsporidian species. Free-living amitochondriate species were studied much less in spite of their intrinsic significance as organisms that possibly did not experience the selective pressures by any host organism. Pelobionta, which are free-living amoeboflagellates, lack mitochondria and are found in anoxic and hypoxic environments (Brugerolle 1991, 1993). While members of this lineages (*Mastigamoeba, Mastigella, Pelomyxa*) are good candidates for exploring the amitochondriate condition, most of them have received relatively little attention until very recently, probably with the exception of the symbiont-bearing giant amoeba, *Pelomyxa palustris* (Andresen, Chapman-Andresen, and Nilsson 1968; Griffin 1988; Whatley, John, and Whatley 1979). A symbiont free species, *Mastigamoeba balamuthi* (described as *Phreatamoeba balamuthi*) (Chavez, Balamuth, and Gong 1986) has been available in axenic cultures for several years, but its potential has so far been poorly exploited. Pelobionta are characterized by a cone of microtubules that connect the basal body of the flagellum to the nucleus (Brugerolle 1993; Cavalier-Smith 1998; Simpson et al. 1997; Walker et al. 2001). They have been regarded to form an independent protist lineage and were placed as Archamoebae into the lineage Archezoa, a lineage erected for the primitively amitochondriate eukaryotes (Cavalier-Smith 1987; Cavalier-Smith 1991; Cavalier-Smith 1993). The ancestral status of Archezoa has been questioned recently (Cavalier-Smith 1998;

Roger 1999). In a revised proposal for the overall classification of living organisms, Cavalier-Smith (1998) removed Archamoebae from the redefined Archezoa, and placed them together with the amitochondriate Entamoebidae and the mitochondriate Mycetozoa (slime molds) in the lineage Conosa of the kingdom protozoa. A cladistic analysis of ultrastructural data did not provide decisive evidence for or against the grouping together the Pelobionta and Mycetozoa (Walker et al. 2001). This study did not include the Entamoebidae because they lack a number of characters. While not using the term Conosa, this hypothesis has been accepted by at least one contemporary textbook on phylogenetic classification (Lecointre and Le Guyader 2001).

The study of SSUrRNA gene sequences from protists has played a major role in the assessment of the phylogeny of amitochondriate protists in general (Sogin and Silberman 1998; Van de Peer and De Wachter 1997; Van de Peer et al. 2000). One major result was that diverse amitochondriate lineages were found to separately diverge from the "main" line of eukaryotes, representing independent clades. A second major conclusion was that some "amitochondriate" lineages emerged early (Sogin et al. 1989) and could be assumed to probably lack mitochondria ancestrally while others acquired their amitochondriate phenotype by a secondary loss of mitochondria. The subsequent analysis of a number of protist protein genes provided more material for phylogenetic analysis, but often the results were not congruent with the SSUrRNA findings. The majority of the studies were able to confirm the monophyly of most protist lineages that have already been recognized by their morphology and biology.

Although both Pelobionta and Entamoebidae were regarded initially as early, premitochondrial lineages (Bakker-Grunwald and Wöstmann 1993; Cavalier-Smith 1987; Cavalier-Smith 1991; Griffin 1988; Reeves 1984; Whatley, John, and Whatley 1979), neither the Pelobionta nor Entamoebidae represent early branches on SSUrRNA trees (Hinkle et al. 1994; Milyutina et al. 2001; Silberman et al. 1999). The SSUrRNA data, however, suggest a close relationship of these two lineages (Milyutina et al. 2001; Silberman et al. 1999). Their

late emergence on the SSUrRNA trees suggested that their amitochondriate phenotype was acquired secondarily by loss of an ancestral mitochondrion. This conclusion is supported by the presence in *E. histolytica* of nuclear genes of probable mitochondrial origin (Clark and Roger 1995) and of a minute organelle without mitochondrial energy conserving function but again of probable mitochondrial origin (Mai et al. 1999; Tovar, Fischer, and Clark 1999). Small organelles bounded by two membranes and with ultrastructural similarities to hydrogenosomes are present in Pelobionta as well (Andresen, Chapman-Andresen, and Nilsson 1968; Chavez, Balamuth, and Gong 1986; Griffin 1988; Seravin and Goodkov 1987; Simpson et al. 1997; Walker et al. 2001). Whether these are homologous to the organelles detected in *E. histolytica* or to hydrogenosomes (Müller 1993) will have to be elucidated by biochemical studies. Phylogenetic studies of other sequences gave contradictory results for another *Mastigamoeba* species (*M. invertens*) (Stiller and Hall 1994; Stiller, Duffield, and Hall 1998) and for *E. histolytica* (Hashimoto and Hasegawa 1996).

To contribute to the clarification of the evolutionary relationships of the Pelobionta, and in particular to test the hypothesis that this lineage is part of a clade that also contains Mycetozoa and Entamoebidae, sequences of large subunit (LSU) rDNA and the elongation factors, EF-1α and EF-2 were obtained from *M. balamuthi*. Phylogenetic analysis of these data together with the available SSU rDNA sequence support the notion that Pelobionta, Entamoebidae and Mycetozoa form a natural lineage.

1.3.1 Materials and methods

Strain and culture An axenic culture of *Mastigamoeba balamuthi* (ATCC 30984) was received from Dr. Frederick Schuster (Brooklyn College, City University of New York, Brooklyn) and cultured in PYGC medium (1% proteose-peptone, 1% yeast extract, 56mM glucose, 6mM L-cysteine HCl, 86mM NaCl, 10mM K_2HPO_4, 10mM KH_2PO_4, 0.0001% resazurin) without agar at 37 °C. Cells were lysed in a buffer containing 0.5% Sarkosyl, 10mM EDTA (pH 8.0), 10mM Tris-HCl (pH 7.5), 100mM NaCl, and 0.5 mg ml^{-1} proteinase K for 15

min at 55 °C, and treated with RNase A (50 mg ml^{-1}) for 1 h at 37 °C. DNA was extracted with phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol and precipitated with sodium acetate and ethanol (Sambrook, Fritch, and Maniatis 1989). The samples were centrifuged at 60,000×*g* for 30 min at 4 °C and the supernatant was used as template for PCR. Total RNA was extracted with the use of Tri Reagent LS - RNA/DNA/PROTEIN Isolation Reagent (Molecular Research Center), according to the manufacturer's instructions. Total RNA was treated with SuperScript™II RNase H⁻ Reverse Transcriptase (RT) (GIBCO BRL, Life Technologies, Rockville, MD), and the resultant RNA:cDNA duplex was used as template for RT-PCR analysis.

Sequence The sequence of the SSUrRNA was retrieved from GenBank (L23779) (Hinkle et al. 1994). rDNA sequence containing an LSUrRNA unit was obtained by PCR with the use of three primer sets: (A) LSU-F: 5'-ACCCGCYGAAYTTAAGCATAT-3' and LSU-R: 5'-GGATTCTRCTTAGAGGCGTTCAG-3', (B) SSU-3'F: 5'-GGTGAACCTGCAGAAGGATC-3' and LSU-5'R: 5'-GTTACTAAGGGAATCACGGTTG-3', and (C) LSU-3'F: 5'- GGTTGGTGATGGAAAGCTAA-3' and SSU-5'R: 5'-GACTACTGGCAGGATCAACC-3'. The primers, LSU-F and LSU-R, were designed based on the highly conserved 5'- and 3'-regions of the LSUrRNA alignment described in the next section. The unique primers, LSU-5'R and LSU-3'F, and those in the SSU rDNA region, were designed from the sequence of the fragment obtained with the primer set (A), and from the published SSU rRNA sequence, respectively. With *M. balamuthi* genomic DNA as template in standard PCR, the primer sets (A), (B), and (C) gave 4.2-kb, 1.4-kb, and 0.8-kb products, respectively, which were cloned into plasmid vectors, pT7 (Novagen) or pCR2.1 (Invitrogen), and sequenced on both strands by primer walking using an automated DNA sequencer (310 Genetic Analyzer, ABI PRISM). Based on the sequence information obtained, an additional primer set, (D) LSU-F2: 5'-ACCTCATCTCGAGTGGGAATACCC-3' and LSU-R2: 5'-GACCTTGCGATCATAATAAGGC-3', expected to cover almost completely the *M. balamuthi* LSUrRNA region, was synthesized and used for the RT-PCR analysis of LSU

rRNA.

The EF-1α sequence was the consensus sequence of four cDNA sequences deposited in the GenBank dbEST database (BE636534, BE636560, BE636596, BE636646) by us.

The EF-2 sequence was reconstructed from a cDNA sequence (BE636642) containing the C-terminal 227 amino acids and the 3'-untranslated region and several partial sequences obtained by PCR. Degenerate primers were synthesized based on the highly conserved regions of the eukaryotic EF-2 sequences (Hashimoto et al. 1995a), A1: 5'-GGNGCNGGNGARYTNCAYYTNGA-3' (forward), A2: 5'-CARTGRTCRAANACRCAYTGNGGRAA-3' (reverse), B1: 5'-GGNGTNTGYGTNCARACNGARACNGT-3' (forward) and C1: 5'-AAYATGWSNGTNATHGCNCAYGTNGA-3' (forward). Unique primers B2: 5'-TCGCGGAACGACACAACAGG-3' and C2: 5'-CCATCATCTTGGCCTTATCT-3' were synthesized based on the sequences obtained in the course of the cloning steps. Using genomic DNA as template in standard PCR, the three primer sets, (A) A1 and A2, (B) B1 and B2, and (C) C1 and C2, produced 0.8-kb, 1.3-kb, and 0.7-kb products, respectively. The PCR products were cloned and sequenced as mentioned above.

Sequence alignments. Alignments of SSU and LSUrRNA sequences are available at the rRNA database (http://rrna.uia.ac.be/) (Van de Peer and De Wachter 1997). For this study, 32 eukaryotic SSUrRNA sequences were selected from the database. These included the *M. balamuthi* sequence and the sequences of various representative organisms, for which LSUrRNA sequences were also present in the database. The *M. balamuthi* LSUrRNA sequence was aligned with the LSUrRNA alignment manually. The deduced EF-1α and EF-2 amino acid sequences of *M. balamuthi* were also aligned manually with their eukaryotic homologs available from the GenBank database.

Phylogenetic analysis. Same methods described in section 1.2.1. were used for phylogenetic inference. In this analysis, relationships among seven eukaryotic including *M. balamuthi*, *E. histolytica*, Mycetozoa, Metazoa/Fungi, Viridiplantae/Rhodophyta,

Alveolata/stramenopiles and Euglenozoa were analyzed by a combined analysis in the same way described in section 1.2.1.

1.3.2 Results and discussion

1.3.2.1 Cloning and sequencing of the genes

LSUrRNA sequence The three PCR-derived clones for the rRNA, corresponding to the primer sets (A), (B), and (C), contained sequences with a length of 4,150 bp, 1,447 bp, and 753 bp, respectively. These partial sequences were concatenated and aligned with 31 other eukaryotic LSUrRNA sequences. The alignment identified an approximately 4,380-bp region as the LSUrRNA gene of *M. balamuthi*, although the 5' and 3' ends could not be clearly determined from the alignment. The LSUrRNA sequence was longer than those of *E. histolytica*, Mycetozoa, and other protists. Several large insertions exceeding 100 bp were found in the *M. balamuthi* sequence: 478 bp at position 1,326 of the *Homo sapiens* sequence (U13369), 122 bp at position 2,450, 128 bp at position 2,543, 149 bp at position 2,649, and 234 bp at position 4,072. The alignment showed that these insertions were located within variable regions of the LSUrRNA. They were longer than those observed in any of the other 31 eukaryotes, except for the first insertion in *Homo sapiens* LSUrRNA (865 bp).

EF-1α sequence The EF-1α sequence, which was reconstructed from four cDNA sequences, revealed a 1,365 nucleotide-long ORF coding for a 454 amino acid putative protein. The deduced EF-1α sequence was aligned with 42 eukaryotic homologs. The *M. balamuthi* EF-1α had all sequence features typical of eukaryotic EF-1αs.

EF-2 sequence The PCR-derived clones of the EF-2 gene obtained with primer sets (A), (B), and (C) contained 811 bp, 1,295 bp, and 727 bp fragments, A, B, and C, respectively, putatively encoding 253, 409, and 118 amino acid sequences. A possible intron sequence (75 bp long) was detected in fragment C, located in the first of the four highly conserved motifs of the GTPase superfamily (Bourne, Sanders, and McCormick 1991). The cDNA sequence encoded the C-terminal 227 amino acid region of the EF-2 gene and contained a stop codon.

Fragment A and the cDNA clone had a 173 amino acid overlap with identical nucleotide sequences. Although primer B2 was synthesized based on the sequence of fragment A, sequences of the overlapping region of the two fragments A and B were slightly different, suggesting the presence of multiple copies of EF-2 genes in *M. balamuthi*. The fragments B and C were combined (fragment C+B), because the overlapping sequences between the two were identical. Since we could not obtain an upstream region of fragment A, the N-terminal (fragments C+B) and the C-terminal (fragment A) portions of *M. balamuthi* EF-2 were treated as separate sequences.

rRNA genes Gel electrophoresis of the total RNA of *M. balamuthi* revealed that the intact LSU and SSUrRNAs are approximately 4.5 kb and 2.7 kb in size. These values are comparable with the putative lengths of the LSUrDNA determined by us and of the SSUrDNA reported previously (Hinkle et al. 1994). Using genomic DNA or RT-treated RNA:cDNA duplex as the template with the primer set (D), PCR analysis was carried out under standard conditions. An approximately 4.5-kb product, corresponding to the 4,380 bp of the putative LSUrRNA sequence, was obtained in the RT-PCR analysis as well as in the analysis using genomic DNA as template, suggesting that there is no intron within this region. Neither of the two group I introns in the LSUrRNA commonly found in two myxogastrid Mycetozoa, *Didymium iridis* and *Physarum polycephalum* (Johansen, Johansen, and Haugli 1992), were detected in the *M. balamuthi* LSUrDNA

1.3.2.2 Phylogeny of each gene

rRNA phylogeny Phylogenetic reconstruction was performed first separately for each of the SSU and LSUrRNA data sets including 32 eukaryotic sequences (data not shown). Although monophylies of Metazoa, Fungi, Viridiplantae, Alveolata with stramenopiles, Euglenozoa, and Diplomonadida were clearly reconstructed in each analysis, we were unable to resolve the branching order for any of the major eukaryotic lineages. In the best tree for the SSUrRNA gene, *M. balamuthi* was located next to *E. histolytica* with a significantly high BP

support (98%). The common ancestor of *M. balamuthi* and *E. histolytica* was placed as a sister group to the so-called "Crown group". Mycetozoa were placed outside (as an outgroup) of these lineages. Two myxogastrid species, *Didymium nigripes* and *Physarum polycephalum*, were clearly grouped together (100%), but the cellular slime mold *Dictyostelium discoideum* was not monophyletic with the myxogastrids. The close relationship of *M balamuthi* with *E. histolytica* in our analysis is comparable with the SSUrRNA trees by Silberman et al. (1999) and by Milyutina et al. (2001), but is incongruent with a distance matrix-based tree by Hinkle et al. (1994), in which *M. balamuthi* does not group with any amoeba or any other eukaryotes specifically. The incongruity probably resulted from the differences in positions selected, in species sampling, and in the phylogenetic reconstruction methods. However, since statistical confidence for their tree was not evaluated, the significance of the incongruity is uncertain.

The best tree of LSUrRNA placed *M. balamuthi* as the sister group of the myxogastrids with a low BP support. The *M. balamuthi*/myxogastrids clade was the sister group of the 'Crown group'. *Entamoeba histolytica* and *D. discoideum* were the closest relatives of each other and their common ancestor was positioned within the 'Crown group'. A small number of alternative trees in which Mycetozoa, *M. balamuthi*, and *E. histolytica* were closely related could not be ruled out either in the SSU or in the LSU analysis.

To see whether the use of a larger number of positions gives a more robust result for the phylogenetic analysis, the concatenated data set of the SSU and LSU data for each species was analyzed in the same way. The best tree placed *M. balamuthi* together with *E. histolytica* with 74% bootstrap support (Fig. 1.4), and the common ancestor of *M. balamuthi* and *E. histolytica* as a sister group of the 'Crown group', similarly to the best tree obtained by the SSU analysis. *Dictyostelium discoideum* was grouped with the myxogastrids and monophyly for the Mycetozoa was observed, although it was not supported by a high BP. The alternative tree, in which Mycetozoa were linked with the *M. balamuthi*/*E. histolytica* clade, showed a smaller log-likelihood value than the best tree in Fig. 1.4 by 14.6±10.6 (± denotes 1SE). When the Γ-distribution was taken into consideration, the difference diminished to 2.6±2.7.

These log-likelihood difference values show no clear difference between the alternative tree and the tree in Fig. 1.4. However, a log-likelihood difference for relocation of the Mycetozoa/*M. balamuthi*/*E. histolytica* clade in the alternative tree to the Metazoa/Fungi clade was 49.8±13.8 without Γ (9.1±6.7 with Γ), showing that the relocation is very unlikely in the analysis without Γ-distribution. Compared to those within the "Crown group", branch lengths of the other lineages, including *M. balamuthi*, *E. histolytica*, and Mycetozoa, are long, while those for Parabasala and Diplomonadida are extremely long. Therefore, long branch attraction (LBA) (Philippe and Laurent 1998) is very likely affecting the positioning of *M. balamuthi*, *E. histolytica*, and Mycetozoa.

EF-1α phylogeny Phylogenetic reconstruction of EF-1α clearly placed *M. balamuthi* as the sister group to the Mycetozoa with high BP support (96%) (Fig. 1.5). The clade including mycetozoa, *M. balamuthi*, and *Porphyra purpurea,* which was unexpectedly placed here (no sufficient BP support), was located near the Metazoa/Fungi clade. *Entamoeba histolytica* was very distant from the Mycetozoa/*M. balamuthi* clade, and was linked with *Euplotes crassus* (Ciliophora). These formed the sister group to the Apicomplexa. The Ciliophora did not form a monophyletic clade, in agreement with previous analyses (Moreira, Le Guyader and Philippe 1999; Roger et al. 1999). Log-likelihood difference between the best tree in Fig. 3 and the alternative tree, in which *E. histolytica* is moved to the *M. balamuthi* position, was 53.9±21.8 (34.0±15.5 with Γ), making the alternative possibility unlikely. In contrast to the SSU+LSU tree in Fig. 2, long branches are not found in Mycetozoa or Euglenozoa but are found for *E. histolytica* and for almost all the species of Alveolata, including Ciliophora and Apicomplexa. As in the SSU+LSU tree, Diplomonadida and Parabasala show extremely long branches in the EF-1α tree. Thus, the tree suggests again that the position of *E. histolytica* within Alveolata may very likely be due to the LBA phenomenon.

EF-2 phylogeny The best trees for the N- and C-terminal portions of the EF-2 gene show the monophyly of the major eukaryotic lineages, Metazoa, Fungi, Viridiplantae and

Rhodophyta, Euglenozoa, Apicomplexa, and Ciliophora (Fig. 1.5). In the tree of the N-terminal portion (Fig. 1.5A), monophyly of the Alveolata was also clearly supported. The tree of the N-terminal portion placed *M. balamuthi* as the sister group of the Metazoa/Fungi clade. The outgroup to the (*M. balamuthi*, Metazoa/Fungi) clade was shared by the common ancestor of *D. discoideum* and *E. histolytica*. In contrast, in the tree of the C-terminal portion (Fig. 1.5B), *M. balamuthi* was grouped with *E. histolytica* and their common ancestor was close to the Euglenozoa. However, the position of *M. balamuthi*, *E. histolytica*, and *D. discoideum* in relation to the major eukaryotic lineages could not clearly be resolved in either analysis, probably because of insufficient phylogenetic information in the EF-2 data set.

1.3.2.3 Combined analysis of four different genes

In order to obtain a more robust inference on the phylogenetic positions of *M. balamuthi*, *E. histolytica*, and Mycetozoa, a combined ML analysis was carried out. First, an exact ML analysis was performed separately on each of the three data sets, concatenated SSU and LSUrRNAs, EF-1α , and concatenated N- and C-terminal portions of EF-2, by exhaustively examining the 945 alternative trees possible among the seven eukaryotic lineages, *M. balamuthi*, *E. histolytica*, Mycetozoa, Metazoa/Fungi, Viridiplantae/Rhodophyta, Alveolata/stramenopiles, and Euglenozoa. Relationships within each lineage consisting of more than three organisms were assumed in advance according to the results of the preliminary analysis. Next, the results obtained with the individual data sets were combined by summing up the estimated log-likelihoods of all the positions for a given tree topology. Finally, the total log-likelihood values and the corresponding BP values were compared among 945 alternative trees. The best three trees of the combined analysis were compared with the three ML trees obtained by the individual analyses (Table 1.5). For these six trees only, an ML analysis with Γ- distribution was also performed. Tree 1 (Fig. 1.7) is the ML tree among 945 alternatives in the analysis without Γ-distribution. When among-site rate heterogeneity is taken into consideration by Γ-distribution, Tree 2 (Fig. 1.7) becomes the best

among the six trees. An outgroup to the ((*M. balamuthi*, *E. histolytica*), Mycetozoa) clade is different between the two trees. Tree 2 (Fig. 1.7) is congruent with a previously proposed placing of Mycetozoa as a sister group to Metazoa/Fungi (Baldauf and Doolittle 1997; Baldauf et al. 2000; Kuma et al. 1995).

In the analysis without Γ-distribution, four alternative trees (Trees 2 through 5) are not significantly different from the ML tree (Tree 1) (p>0.05) by a statistical test by Shimodaira and Hasegawa (1999). In the analysis with Γ-distribution, none of the alternative trees are significantly different from Tree 2 (p>0.05). Although the ML trees for the individual data sets are different, none of them show statistically significant difference from Trees 1 through 3 (p>0.05), demonstrating that the phylogenetic relationship common to all these trees, ((*M. balamuthi*, *E. histolytica*), Mycetozoa), is not rejected by any of the individual analyses. The close relationship between *M. balamuthi* and *E. histolytica* and the monophyly of Mycetozoa with the *M. balamuthi/E. histolytica* clade is supported by 87% and 79% BP values, respectively, in the combined analysis without Γ-distribution (Fig. 1.7).

In addition, separating the rRNA and the EF-2 data sets into two parts, (i.e. SSU and LSU, and N- and C-terminal portions respectively), an alternative combined analysis for five individual data sets was also performed. The result consistently suggested a close relationship of Mycetozoa with the *M. balamuthi/E. histolytica* clade (data not shown).

Phylogenetic reconstruction of the individual sequences analyzed or reanalyzed in this study revealed robust sister group relationships in different combinations of two of the three OTUs tested, *M. balamuthi, E. histolytica* and Mycetozoa. Only the combined evaluation of all genes tested resulted in a tree in which the ML value was the highest for the monophyly of the three lineages (Fig. 1.7), although the BP support was not very strong (79%). We would like to emphasize, however, that no tree for any of the individual genes rejected such a monophyly in a statistically significant way (Table 1.5). These results suggested the validity of the hypothesis that these organisms are related and belong to a single, monophyletic lineage, as proposed earlier, primarily based on morphological and biological considerations (Cavalier-Smith 1998; Walker et

al. 2001). Recently, a close relationship between *E. histolytica* and *D. discoideum* was supported by a LogDet/paralinear distance tree of CPN60 (Horner and Embley 2001). They also mentioned that α-tubulin, SSUrRNA, and EF-2 trees have depicted this topology.

The monophyly of Mycetozoa (slime molds) themselves has remained an open question for a long time. The major lineages that belong to the slime molds (cellular and myxogastrid slime molds) are phenotypically quite divergent and in a number of SSUrRNA studies they branched independently from each other from the eukaryotic trunk (Hinkle et al. 1994; Leipe et al. 1993; Zettler, Sogin, and Caron 1997). Only recent work using EF-1α and the combination of several molecular markers (Baldauf and Doolittle 1997; Baldauf et al. 2000) demonstrated a robust monophyly of this lineage. EF-1α analysis of this study with the inclusion of *M. balamuthi* is congruent with this conclusion (Fig. 1.5). In spite of its low BP support, the best tree of our SSU+LSU analysis also revealed a monophyly of Mycetozoa (Fig. 1.4).

If the monophyly of the three lineages, Pelobionta, Entamoebidae and Mycetozoa as discussed here is real, it expands the list of eukaryotic lineages that consist of organisms with either mitochondriate or amitochondriate energy metabolism. Mycetozoa are aerobic organisms and have typical mitochondria (Dykstra 1977; Ogawa et al. 2000). While little is known about the metabolism of *M. balamuthi* and other Pelobionta, morphological data clearly show the absence of typical mitochondria. Double membrane bounded organelles of about 500-1000 nm in diameter are present and often resemble hydrogenosomes (Andresen, Chapman-Andresen and Nilsson 1968; Brugerolle 1993; Griffin 1988; Seravin and Goodkov 1987; Simpson et al. 1997; Walker et al. 2001). These organelles might have derived from ancestral mitochondria and could be homologous to the small organelles of *E. histolytica* (Tovar, Fischer and Clark et al. 1999; Mai et al. 1999) or to genuine hydrogenosomes (Müller 1993). Without biochemical studies their biological nature has to remain conjectural. *Entamoeba histolytica* is a characteristic type I amitochondriate organism with fermentative metabolism (Müller 1988; Müller 1998; Reeves 1984). The presence of mitochondriate and amitochondriate groups in the same lineage allows us to strongly argue for a mitochondriate

nature of their common ancestor and for the development of the amitochondriate phenotype by secondary loss of the role of the mitochondrion in energy metabolism and a drastic reduction of the size of the remnant organelle (Martin and Müller 1998; Roger 1999; Rotte et al. 2000). These considerations have been generally accepted for *E. histolytica* (Tovar, Fischer, and Clark 1999) but the status of Pelobionta, including *M. balamuthi* has remained an open question. Such coexistence of mitochondriate and amitochondriate organisms is known in other monophyletic lineages as well (e.g., in ciliates (Embley et al. 1995) and chytrid fungi (Hackstein et al. 1999)). In these lineages the mitochondriate phenotype is regarded ancestral and the amitochondriate one secondary. The processes involved in the conversion of mitochondriate phenotype into amitochondriate ones are little understood. The coexistence of mitochondriate, free-living and parasitic amitochondriates in a single lineage might provide a good basis for the exploration of these processes.

1.4 Phylogenetic relationships among seven major eukaryotic lineages

Diplomonadida and Parabasala have been considered as candidates for early branching organisms in the eukaryotic tree because of their simple cellular structures. Diplomonadida are amitochondriate flagellates. Most of them have identical two nuclei and two sets of flagella. They lack some cellular apparatus such as nucleolus (Friend 1966; Ferguson 1979; Desser, Hong and Siddall 1993), peroxisomes and Golgi dictyosomes (Cavalier-Smith 1987). Nine genera are known and most of them are parasites, while *Hexamita* and *Trepomonas* are free-living genera. *Giardia intestinalis* is a well known organism in diplomonadida and its genome project is going on. The genome size of *G. intestinalis* is 12 Mb. The genome is polyploid with at least four copies of each of five choromosomes.

Parabasala are mostly flagellated protozoa. Trichomonads are well known among them. Almost all of them are parasites and little is known for free-living groups. Trichomonads have an organelle, the hydrogenosome, which metabolizes pyruvate and malate to produce ATP, H_2, CO_2 and acetate (Müller 1993). The endosymbiotic origin for this membrane bounded

organelle was suggested to be bacteria such as *Clostridium* for similarities in aerobic metabolism (Whatley, John and Whatley 1979). Later, Cavalier-Smith (1987) proposed that those were highly modified mitochondria. Recently two homologs of the mitochondrion targeting proteins in mitochondriate eukaryotes, CPN60 and HSP70mit, were shown to be localize to the hydrogenosome of *T. vaginalis* (Bozner 1997). Phylogenetic analyses supported that these genes were clearly located within the corresponding eukaryotic mitochondrial clade, suggesting a common evolutionary origin of the hydrogenosome with the mitochondrion.

As mentioned in Introduction, Diplomonadida and Parabasala were suggested to be early branching eukaryotic lineages. However, theie branching positions were not supported by all of phylogenies without exceptions. In α-tubulin (Keeling and Doolittle 1996) and β-tubulin (Edlind et al. 1996) phylogenies, *E. histolytica* located at the bottom of the eukaryotic tree, while in ValRS tree (Hashimoto et al. 1998) both *G. intestinalis* and *T. vaginalis* were positioned at the sister group of *Arabidopsis*. In endplasmic reticulum-type HSP70 (Germot and Philippe 1999) and HSP70mit (Arisue et al. 2002) phylogenies, the earliest and the second earliest branching eukaryotes were *Trypanosoma* (Euglenozoa) and *Eimeria* (Alveolata), respectively. Although, these phylogenetic relationships were not strongly supported by bootstrap probabilities, and these genes may not have correctly reflect the relationships among eukaryotes. In addition the branching order of Diplomonadida and Parabasala were not clearly be resolved. At present, rRNAs (Leipe at al. 1993; Hinkle et al 1994; Ohkuma et al 1998; Philippe and Germot 2000), β-tubulin (Edlind et al. 1996) and the cyclin-dependent kinase (Riley et al. 1995) trees showed earlier branching position of Parabasala than Diplomonadida, while trees of EF-1α and 2 (Yamamoto et al. 1997), translation initiation factor eIF-2γ (Keeling, Fast and McFadden 1998), RNA polymerase II (Hirt et al. 1999) and HSP70c (Germot and Philippe 1999) represented earlier branching position of Diplomonadida than Parabasala.

To address this question, phylogenetic relationships among major eukaryotic lineages, a

combined ML analysis using 24 genes data was examined. According to preliminary results (§ 1.1. and § 1.2 this article; Moreira, Guyader and Philippe 2000; Bapteste et al. 2002) constraints were assumed in advance to some nodes, and relationships among 8 lineages including outgroups (Fig 1.8) were explored by examining possible 10,395 alternative trees.

For this purpose, some of new sequences were determined (Table 1.6) .Among them, cytosolic type heat shock protein 90 (HSP90c) is an ATP dependent molecular chaperon involved in the folding and activation of a number of substrate proteins. RNA polymerase are conserved among all living organism. Eukaryotic RpoII is involved in the pre-mRNA synthesis. It is a large enzyme comprising at least 12 distinct subunits and possessing the molecular mass over 500 kDa. The HSP90c and RpoII sequences are highly conserved among all extant organisms and are useful for phylogenetic analyses.

1.4.1 Materials and methods

Genes cloned and sequenced for this study were listed in Table 1.6.

Genomic DNA was extracted from *T. vaginalis* strain C-1:NIH (ATCC 30001) by Blood & Culture kit (QIAGEN) according to the manufacture's protocol. *Trichomonas vaginalis* genomic DNA library was kindly given by Dr. M. Müller (The Rockefeller Univ., New York). Two and one sets of degenerate primers were synthesized based on highly conserved regions of the IleRS and HSP90c sequences, respectively: IleRS: FYDGPPFAT (forward) 5'-TTYTAYGAYGGNCCNCCNTTYGCNAC-3'and AWTTTPWTL (reverse) 5'-GCNTGGACNACNACNCCNUGGACNYT-3', DQTRGWFY (forward) 5'-CTRGTYTGNDCNCCNACCAARAT-3' and WYIRLNR (reverse) 5'-CKRTTNARNCKDATRTACCA-3'; HSP90c: ETF(A/T)FQAEI (forward) (5'-GARACNTTYRCNTTYCARGCNGARAT-3') and QQNKI(L/I)KV (reverse) (5'-ACYTTNADDATYTTRTTYTGYTG-3'). Corresponding fragments were amplified by PCR, purified by using a Gel Extraction Kit (QIAGEN), and cloned into pT7 vector (Novagen). The excised inserts labeled with a Random Primers DNA Labeling Systems

(GIBCO/BRL) were used to probe and isolated clones containing IleRS or HSP90c genes from *T. vaginalis* genomic DNA library.

For strains and methods for genomic DNA extraction of *E. histolytia*, see section § 1.1.1. For *G. intestinalis*, strain WB, clone 6 (ATCC 300957) was used. DNA and total RNA of *G. intestinalis* were extracted with the use of Blood & Culture kit (QIAGEN) and Rneasy Mini Kit (QIAGEN), respectively, according to the manufacturer's instructions. Total RNA was treated with SuperScript™ II Rnase H⁻Reverse Transcriptase (RT) (GIBCO BRL), and the resultant RNA :cDNA duplex was used as template for RT-PCR analysis, in order to isolate cDNA encoding HSP90c. Unique primers for RpoII and HSP90c of *G. intestinalis* and *E. histolytica* were designed and synthesized based on the partial sequences of these genes found in the genome sequencing project databases of *G. intestinalis* (http://evol3.mbl.edu/Giardia-HTML/) and *E. histolytica* (http://www.tigr.org/tdb/e2k1/eha1/). Those were *G. intestinalis*: HSP90c: 5'-GGATGCCCGCTGAAGTCTTCGAG-3' and 5'-GTTCACAGTCTCAGTCAACTTCGTC-3'; RpoII: 5'-CCTGAAGTCGAGAAAGGCACTTCATGG-3' and 5'-GCGACCAATCAATATAGTGAGCACGC-3', 5'-CAGCTGATAGCGGATATGTTGAAAG-3' and 5'-GAATGTTCTGCAGAATATTGTCGTCCG-3'; *H. histolytica*: HSP90c: 5'- GAAATGACTGAAACATATCAATTCC-3' and MEEVD* (reverse) 5'-CYYARTCNACYTCYTCCAT-3'; RpoII: 5'-GAACTTGCCGATCTTTGTATGGGCCC-3' and 5'-GCTCCTGGATAATCATTCATTCCAGC-3', 5'-CAATGGATGAAGTTGGAGTAC-3' and 5'-CTTGTGGAGATAATCCCATCACATAAC-3', 5'-GAACCTTACCACATTATTGGAAAGAAG-3' and 5'-CAAATTCATTTATCTTGCGGTTTAGGG-3'. Degenerated primer was designed and synthesized for C-terminal portion of the *E. histolytica* HSP90c.

Amplified corresponding fragments were cloned into pCR2.1 plasmid (Invitrogen).

Nucleotide sequences of clones were determined on both strands with the use of an automated DNA sequencer (310 Genetic Aanalyzer, ABI PRISM).

Ribosomal protein genes of *G. intestinalis* (S14, S15a, L5, L8 and L10a) and *T. vaginalis* (S14, S15a, L5 and L8) were also cloned and sequenced. Those were described in § 2.2.

Sequence alignment The deduced amino acid sequences were aligned by manual with their eukaryotic homologs available from public databases.

Phylogenetic analysis Same methods described in § 1.1.1. were used for phylogenetic inference. In this study, eight lineages, outgroup, Diplomonadida, Parabasala, Euglenozoa, Viridiplantae/Rhodophyta, Alveolata/stramenopiles, Conosa, and Metazoa/Fungi/Microsporidia were explored. Genes used for combined ML analysis were listed in Table 1.7. As outgroups, prokaryotic homologs or duplicated genes were used.

1.4.2 Results and discussion

1.4.2.1 Cloning and sequencing of the genes

HSP90c By the screening of the *T. vaginalis* libraries, a positive clone was obtained and was completely sequenced with both strands. Although a few N-terminal residues were lacking, an almost entire part of a putative HSP90c ORF with 710 amino acid (2,139 bp) was deduced. Using the PCR-based cloning methods, complete putative ORFs 706 and 704 amino acids (2,118 bp and 2,112 bp) were obtained from *E. histolytica* and *G. intestinalis*, respectively. Those HSP90c sequences were well conserved among. An upstream portion of *T. vaginalis* hsp90c gene was slightly extended compared with those of other organisms. The HSP90c genes of *T. vaginalis* and *E. histolytica* HSP90c were not interrupted by intron like sequences. However, comparison of a cDNA sequence of the *G. intestinalis* HSP90c with its genomic DNA sequence in the database revealed that the gene has a intron (Fig. 1.9). It is located at the middle part of the of the HSP90c gene. The complete sequence of the intron has not been found in the database, but its length was estimated to be over 1.5 kb by an alignment

of fragment data including a part of the intron sequence. The sequence 'GTATGT...AG' in *G. intestinalis* HSP90c gene (Fig. 1.9) is very similar to a consensus sequence 'GT(A/G)AGT...AG' spliceosomal introns. None of the *G. intestinalis* sequences reported to date has introns (Adam 2000). This is the first indication of the possible presence of spliceosomal intron in *G. intestinalis*. Up to date, some of small nuclear RNAs (snRNAs) were found in *G. intestinalis*, but these were related to pre-rRNA processing mRNAs (Niu et al. 1994; Gohsh et al. 2001). Spliceosomal snRNAs involved in pre-mRNA processing have not been detected by northern hybridization analysis by using a complementary sequence to a snRNA consensus sequence as a probe, or by immunological techniques. Because spliciosomal snRNAs have not been detected, Diplomonadida including *G. intestinalis* have been considered to have neither introns nor spliceosomal machinery for splicing pre-mRNAs (Li 1999). However, sequences which have similarity with splicing related snRNPs or proteins were found in the genome project database of *G. intestinalis*. The finding of the spliceosomal intron-like sequence in the HSP90c genes also confirme the notion that *G. intestinalis* has spliceosomal machinery. Among major five amitochondriate lineages analyzed in this work, no Parabasala genes were reported to have introns, while Entamoebidae, Pelobionta and Micropsorideia have some intron-containing genes: protein kinase $p34^{cdc2}$ (Lohia and Samuelson 1993), serin/threonin protein kinase (Urban et all. 1996) and 39 kDa antigen of *E. histolytica* (Plaimauer et al. 1994); ribosomal protein L27a of *E. histolytica* and *E. disper* (Wilihoeft et al. 2001); translation elongation factor 2 of *M. balamuthi* (this work, see § 1.3.2.1); 11 ribosomal proteins (S8, S17, S24, S26, L5, L19, L27a, L37, L37a and L39) and two tRNA of *E. cuniculi* (Biderre, Méténier and Vivarès 1998; Kathinka et al. 2001). However the size of these spliceosomal introns were only 46~115 bp in *Entamoeba* and 23~52 bp in *E. cuniculi*. Compared to these values, the intron of the *G. intestinalis* HSP90c gene with over 1.5 kb in length is extremely long.

RpoII The putative ORF of *G. intestinalis* RpoII coded for a 2,076 amino acid (6,228 bp). two large insertions that shared by none of other eukaryotes were found at the position of

572~680 and 1,100~1,188 in the amino acid sequence. Cloned and sequenced E. histolytica RpoII contained a major part of the coding region of with 1,589 amino acid (4,767 bp), although it lacked approximately 50 N-terminal residues. C-terminal portion of the *E. histolytica* RpoII was reduced, compared with other eukaryotic consensus sequences. As well as in *T. vaginalis* RpoII sequence (Quon, Delgadillo and Johnson 1996), the 'YSPTSP(S/A)' repeat on the C-terminal domain which is conserved among Metazoa, Fungi, and Viridiplantae was lacking both in *G. intestinalis* and in *E. histolytica*. Among amitochondriate protists, *Nosema locustae*, *Vairimorpha necatrix* (Microsporidia) and *Mastigamoeba invertens* (Pelobionta) have this repeat on the C-terminal domains.

IleRS A major part of the IleRS gene of *T. vaginalis* was cloned and sequenced using the PCR-based method. The sequence coded for 676 amino acids in length. This contained characteristic motives, HYGH and KMSKS, commonly found in all class I aminoacyl-tRNA synthetases (Moras 1992), although these motives were slightly different from the consensus sequences. The sequence was well aligned with other eukaryotic orthologs with no remarkable gaps.

Ribosomal proteins Described in details in § 2.3.4.

1.4.2.2 Combined analysis of 24 different genes

For further estimation of evolutionary relationships among major eukaryotic lineages, a combined ML analysis were performed using 24 genes (Table 1.7). Individual ML analysis was carried out on each of the 24 genes by exhaustively examining the 10,395 alternative trees possible for the seven eukaryotic lineages (Metazoa/Fungi/Microsporidia, Conosa, Viridiplantae/Rhodophyta, Euglenozoa, Alveolata/stramenopiles, Palabasala and Diplomonadida)with an outgroup. A relationship within each lineage consisting of more than three organisms was assumed in advance according to the results of the preliminary analysis. The results obtained by the individual data sets were combined by summing up the estimated log-likelihoods of all the positions for a given tree topology. Finally, the total log-likelihood

values were compared among 10,395 alternative trees. Of the 10,395 trees in the combined analysis without Γ-distribution, 160 trees were selected as candidate trees: the ML tree; 143 trees that were not statistically different from the ML tree by the criterion of 4SE of log-likelihood difference (Kishino-Hasegawa, 1989); and 16 trees each of which was selected as the ML tree by the individual analysis, but were not included in the above 143 trees. The ML analysis with Γ-distribution was examined for these 160 trees. The RELL bootstrap analysis (Kishino et al. 1990) was carried out for these 160 alternative trees in the analyses both without and with Γ-distribution. Approximately unbiased (AU) test (Shimodaira 2002) was used for selecting candidate alternative trees by using a significance level at p=0.05 revealed by the analyses (Table 1.8). The best three trees without and with Γ-distribution analyses were schematically represented with BPs in Fig. 1.10. The best three trees in the analysis without Γ-distribution located Diplomonadida and Parabasala at the earliest and the second earliest branching position respectively in the eukaryotic tree, while in the analysis with Γ-distribution, the branching order of Diplomonadida and Parabasala was different among the best three trees. The best tree of the analysis (Fig. 1.10B, Tree 3) showed the earliest branching position of Diplomonadida than Parabasala. The common ancestor of Diplomonadida and Parabasala branched at the basal position in the second best tree (Fig. 1.10B, Tree 9), while Parabasala branched earlier than Diplomonadida in the third best tree (Fig. 1.10B, Tree 11). If we use the AU test for excluding the alternative trees that are significantly different from the ML tree (Tree 1), five trees (Tree 2 through Tree 5, and Tree 15) were not excluded in the analysis without Γ-distribution as shown in Table 1.8 (p>0.05). All these trees indicate that Diplomonadida and Parabasala were the earliest and the second earlist offshoots of the eukaryotic tree. In contrast, in the analysis with Γ-distribution, 16 trees except for Tree 15 in Table 1.8 were not significantly different (p>0.05) from the ML tree (Tree 3). Among these trees, various branching positions for Diplomonadida and Parabasala were shown.

Phylogenetic relationships among rest five lineages, Metazoa/Fungi/Microsporidia,

Conosa, Viridiplantae/Rhodophyta and Alveolata/stramenopiles were not clearly resolved in the analysis without Γ-distribution. The relashionships were dfferent among the best three trees and only low BPs (26.5 ~ 63.2%) were given to the internal nodes. On the other hand, the relationship among the five lineages was consistently reconstructed in the best three trees in the analysis with Γ-distribution. The relationship ((((Euglenozoa, Alveolata/stramenopiles), Viridiplantae/Rhodophyta), Conosa), Metazoa/Fungi/Microsporidia) was supported with high BPs for the corresponding nodes (85.5~91.0%). Comparison of the estimated log-likelihood value between the analyses without and with Γ-distribution revealed that the value of the latter analysis (-279603.2) was far higher than that of the former one (-295739.9), demonstrating that the model with among-site rate heterogeneity taken into consideration by the use of Γ-distribution for better approximated the data set, and thus that the result derived from the analysis with Γ-distribution would be more reliable than that without Γ-distribution .

In this study, Diplomonadida and Parabasala were demonstrated to be early branching lineages in the eukaryotic tree, although the branching order between the two were not well resolved. The basal positioning of these two lineages in the eukaryotic tree was robustly demonstrated at the first time by the present combined analysis. However, we cannot rule out entirely the possibility that the best tree (Tree 3 in Fig. 1.10B) has still been affected by the LBA artefact (Philippe and Laurent 1998). Recently, by the ML analysis of 25,000 amino acid positions concatenated from over 100 genes, Bapteste et al. (2002), demonstrated a eukaryotic tree, in which Diplomonadida and Kinetoplastida (Euglenozoa) were the earliest and the second earliest offshoots of the Eukaryota. The earliest branching position of Diplomonadida was consistent with our best tree, although they mentioned that the positions of Diplomonadida and Kinetoplastida were caused by LBA. Since the data sets and the methods for phylogenetic inference are different between their analysis and our present analysis, and since Parabasala were not included in their analysis, we cannot simply compare our best tree with their tree. However, the position of Euglenozoa in our best tree is significantly different from that of their tree. In our tree, the closest relative of Euglenozoa is the

Alveolata/stramenopiles clade, and Viridiplantae/Rhodophyta is located at the sistergroup position of the ancestor of these these two groups. The monophyly of these three groups, Euglenozoa, Alveolata/stramenopiles and Viridiplantae/Rhodophyta was also reconstructed by an analysis with concatenated amino acid sequence data of four proteins (Baldauf et al. 2000), although they had no root in their tree and they could not clearly settle the positions of Diplomonadida and Parabasala. The monophyly of these three groups is clearly supported also by the other evidence that the fused dihydrofolate reductase and thymidilate synthetase genes are present only in Alveolata, Viridiplantae, and Euglenozoa (Philippe et al. 2000). In Metazoa and Fungi, the genes for these enzymes are independent as well as in prokaryotes.

1.5 Summary and conclusions

Phylogenetic relationships among major eukaryotic lineages including amitochondriate ones were studied by combined ML analyses of various genes, with or without among-site rate heterogeneity taken into consideration by the use of Γ-distribution.

At first, phylogenetic positions of Microsporidia and of stramenopiles were analyzed. Sequences of the genes encoding SSUrRNA, v-ATPB, and HSP70c from *Blastocystis hominis*; EF-2 from *B. hominis* and *Plasmodium falciparum* (Alveolata); and IleRS and ValRS from *P. falciparum* and *Entamoeba histolytica* (Entamoebidae) were obtained, and phylogenies based on these genes were individually examined by the ML method. The ValRS phylogeny suggested that Microsporidia were closely related to Fungi, while the IleRS phylogeny gave no clear resolution on the relationships among eukaryotic lineages. The SSUrRNA and HSP70c phylogenies suggested that *B. hominis* was grouped with stramenopiles and that stramenopiles and Alveolata were closely related with each other.

In order to further settle a phylogenetic position of Microsporidia among major eukaryotic lineages, a combined ML analysis was performed using 6,391 positions from 10 genes for which data from Microsporidia were available. These genes were EF-1α, EF-2, IleRS, ValRS, RpoII, Actin, α-tubulin, β-tubulin, SSUrRNA, and LSUrRNA. Although

several phylogenies based on individual genes, such as EF-1α, IleRS, and SSUrRNA, did not support a close relationship between Microsporidia and Fungi, the combined analysis clearly demonstrated a relationship, (Metazoa, (Fungi, Microsporidia)) with a very high statistical support. In addition, an another combined ML analysis was performed to examine a relationship between stramenopiles and Alveolata, using 5,423 positions from eight genes for which data from stramenopiles were available (EF-1α, EF-2, HSP70c, v-ATPB, Actin, β-tubulin, SSUrRNA, and LSUrRNA). The analysis demonstrated also with a very high statistical support that stramenopiles and Alveolata were the closest relatives with each other.

In the next, the phylogenetic position of the Pelobiont *Mastigamoeba balamuthi* was analyzed in relation to the position of *E. histolytica* (Entamoebidae). Sequences of the genes encoding LSUrRNA, EF-1α, and EF-2 from *M. balamuthi* were obtained and phylogenies based on these genes as well as on the SSUrRNA gene were individually examined by the ML method. The SSUrRNA, LSUrRNA and EF-1α trees suggested that *M. balamuthi* had a most recent common ancester either with Mycetozoa or with *E. histolytica*. A clade formed by *M. balamuthi*, *E. histolytica*, and Mycetozoa was not rejected statistically for any of the individual genes examined. A combined ML analysis using 3,935 positions from four genes suggested that *M. balamuthi* was the closest relative of *E. histolytica* and that Mycetozoa were placed at the sistergroup to the common ancestor of *M. balamuthi* and *E. histolytica*. These findings supported the notion, which had previously been proposed primarily on cytological evidence, that both *M. balamuthi* (Pelobionta) and *E. histolytica* (Entamoebidae) are closely related to the Mycetozoa and that these three together represent a major eukaryotic lineage (Conosa).

Finally, on the basis of the findings as mentioned above and as currently reported in the literatures, 13 major eukaryotic lineages were divided into seven groups: 1. (Metazoa, (Fungi, Microsporidia)), 2.(Mycetozoa, (Pelobionta, Entamoebidae)) [Conosa], 3. (Viridiplantae, Rhodophyta), 4. (stramenopiles, Alveolata), 5. Euglenozoa, 6. Diplomonadida, and 7. Parabasala. Then, phylogenetic relationships among these groups

with an outgroup were examined by a combined ML analysis of the genes, EF-1α, EF-2, RpS14, RpS15a, RpL5, RpL8, RpL10a, IleRS, ValRS, RpoII, CPN60, HSP70c, HSP70mit, HSP70er, HSP90c, CCTα, CCTγ, CCTδ, CCTζ. Actin, α-tubulin, β-tubulin, SSUrRNA, and LSUrRNA. To perform this analysis with these 24 individual genes, sequences of the several genes that are absent in three amitochondriate protists were obtained: RpS14, RpS15a, RpL5, RpL8, RpL10a, RpoII, and HSP90c from *Giardia intestinalis* (Diplomonadida); RpS14, RpS15a, RpL5, RpL8, IleRS and HSP90c from *Trichomonas vaginalis* (Parabasala); and RpoII and HSP90c from *E. histolytica* (Entamoebidae).

The phylogenies based on the 24 individual genes supported different trees as the best among 10,395 alternatives possible for eight groups. The conclusions obtained were sometimes in disagreement with each other among genes examined. However, a combined ML analysis using 10,910 positions for these 24 genes could reduce alternative possibilities to a very few trees. The analysis clearly supported that Diplomonadida and Parabasala diverged earlier than other five groups in the eukaryotic tree, although the branching order between these two lineages were still open for further analysis. In addition, especially when among-site rate heterogeneity was taken into consideration, it was clearly supported that (Metazoa, (Fungi, Microsporidia)), Conosa, and (Viridiplantae, Rhodophyta) were the first, the second and the third earliest offshoots among the five groups excluding Diplomonadida and Parabasala. The analysis at the first time demonstrated robustly that Diplomonadida and Parabasala are the early branching eukaryotes, although presence of a potential artifact derived from a long branch attraction could not be ruled out entirely.

In conclusion, on the basis of the combined ML analyses examined comprehensively in this work, five amitochondriate protist lineages were phylogenetically clearly positioned within a eukaryotic tree. Microsporidia were closely related to Fungi. Pelobionta and Entamoebidae were the closest relatives with each other, and these two lineages were linked with Mycetozoa to form a group, Conosa. Diplomonadida and Parabasala diverged earlier than other eukaryotic lineages.

Chapter 2

Characterization of the ribosomes of amitochondriate protists

2.1 Introduction

The ribosomes of three amitochondriate protists, Diplomonadida (e.g. *Giardia intestinalis*) Parabasala (e.g. *Trichomonas vaginalis*) and Microsporidia (e.g. *Encephalitozoon cuniculi*), were reported to exhibit certain features that made them 'prokaryotic', 'intermediate' or 'primitive eukaryotic'. These features are 1) shorter small subunit (SSU) and large subunit (LSU) ribosomal RNAs (rRNAs), than those found in most eukaryotes (Vossbrinck et al. 1987; Sogin et al. 1989; Chakrabarti et al. 1992; van Keulen et al. 1992; De Rijk, Gatehouse and De Wachter 1998; Peyretaillade et al. 1998b), 2) the prokaryotic 70S sedimentation coefficient of microsporidian (*Nosema bombycis*) (Ishihara and Hayashi 1968) and trichomonad ribosomes (Oka et al. 1973; Champney, Chittum and Samuels 1992), 3) the presence of fewer proteins (40~56) in four trichomonad ribosomes (Champney, Chittum and Samuels 1992) and 4) absence in Microsporidia of the internal transcribed spacer region 2 (ITS2) present in large subunit (LSU) rRNA region of all other eukaryotes (Vossbrinck and Woose 1986). The 5.8S rRNA region of Microsporidia are directly join to the LSUrRNA as seen in prokaryotes. Unfortunately, the characterization of ribosomes does not encompass all four properties for any of these groups thus it remains to be seen whether these really separate these groups from the rest of eukaryotes.

The possibility that Diplomonadida, Parabasala and Microsporidia represent 'primitive eukaryotes' or 'intermediate' forms, received seeming support from two independent lines of evidence. First, none of these groups contain typical mitochondria and are functionally 'amitochondriate' (Müller 1988; Müller1998), a finding that led to the proposal that these lineages branched off from the truck of the eukaryotic tree preceding the mitochondrial endosymbiotic event (Cavalier-Smith 1987; Cavalier-Smith 1993). Second, earlier phylogenetic inference based on SSUrRNA (Leipe et al. 1993; Sogin and Silberman 1998) and on translation elongation factors (Hashimoto et al. 1997) placed these amitochondriate

lineages as the earliest branches of the eukaryotic tree. More recent work increasingly indicates that neither line of evidence provides a compelling argument for the 'primitive' or 'intermediate' status of these three lineages (Roger 1999; Katz 1998; Philippe and Adotte 1998; Philippe and Laurent 1998; Embley and Hirt 1998). In chapter 1 of this article, the branching position of Microsporidia was clearly supported not to be early in eukaryotic lineages but close to fungi. This suggests Microsporidia are highly evolved eukaryotes and their small rRNAs, absence of the ITS2, and the prokaryotic-like 70S sedimentation coefficients are derived features obtained on the line leading to Microsporidia. In contrast to Microsporidia, Diplomonadida and Parabasala were demonstrated to be early in the eukaryotic tree as mentioned in chapter 1 (§ 1.4.2.2). If basal positions for these lineages are really correct, there still remain a possibility that ribosomes from these lineages are 'intermediate' or 'primitive eukaryotic'.

This situation prompted us to study in more depth the ribosomes in the organisms that belong to Diplomonadida and Parabasala. In this chapter we report the analyses of the components of ribosomes from *Giardia intestinalis* (Diplomonadida) and the *Trichomonas vaginalis* (Parabasala), trying to find the differences of the ribosomal components that separate between these organisms and 'higher' eukaryotic ones.

2.2 Materials and methods

Strain and culture *Trichomonas vaginalis*, strain C-1:NIH (ATCC 30001) was cultured in TYM diamond medium (Diamond 1957). *Giardia intestinalis*, strain WB, clone 6 (ATCC 30957) was cultured in TYI-S-33 medium (Keister 1983). *Mastigamoeba balamuthi* (ATCC 30984) was cultured in PYGC medium (see, § 1.3.1). All cultures were performed at 37℃. Trophozoites in log-phase were harvested by centrifugation at 800 × *g* at 4℃. Because the trophozoites of *G. intestinalis* grow on the wall of the vial, cultures of *G. intestinalis* were chilled on ice for 10 min to detach them before harvest. Precipitates were washed twice in phosphate buffered saline (pH7.4) and stored at -80℃ until use.

Ribosome sedimentation analysis The trophozoites, (1×10^6) of *G. intestinalis* or *T. vaginalis* were resuspended in 0.5ml of Buffer A (100mM NH_4Ac, 15mM $MgAc_2$, 20mM Tris-HCl, pH7.6) and homogenized with 0.5g of 0.3mm ϕ glass beads by the use of micro tube mixer (VORTEX-GENIE 2, Scientific Industries) at 4℃, 5 min in maximum speed. Immediately after homogenization, the material was centrifuged at $10,000 \times g$ for 15 min at 4℃. The supernatant which contains ribosome was replaced into the new tube and measured optical density (OD) at 260nm. The extracted ribosome of $3OD_{260}$ units was loaded onto a 5-20% or 10-40% (w/v) linear sucrose gradient, prepared in Buffer A and centrifuged (L90, Beckman) at 4℃, at 40,000 rpm ($202,000 \times g$) for 2hr (5-20% (w/v) sucrose), or at 35,000 rpm ($155,000 \times g$) for 3hr (10-40% (w/v) sucrose) with the use of swinging bucket rotor (SW40Ti, Beckman). The distribution of RNA in the gradients was monitored by the absorbance at 260nm. As controls, *Escherichia coli* (strain W3110), *Saccharomyces cerevisiae* (strain A364A) and *Artemia salina* (brine shrimp) ribosomes were also sedimented at the same experiment. Preparation of ribosomes from *E. coli* and *S. cerevisiae* were performed according to Horie, Wada and Fukutome (1981) (*E. coli*) and Guthrie and Fink (1991) (*S. cerevisiae*). The ribosome of *A. salina* was prepared from dried eggs of a resting stage according to Mizumoto et al. (1974) and Zasloff and Ochoa (1971).

For the analyses of dissociate subunits, Buffer B (100mM NH_4Ac, 1mM $MgAc_2$ and 20mM Tris-HCl (pH7.6)) was used for *E. coli*, and Buffer C (500mM KCl, 16mM $MgCl_2$, 20mM Tris-HCl (pH7.8), 20mM 2-mercaptoethanol and 0.2mM EDTA) was used for *E. coli* and *S. cerevisiae* (Guthrie & Fink 1991).

Preparation and purification of ribosomes The trophozoites of *G. intestinalis* and *T. vaginalis* were homogenized in Buffer A containing 6 mM 2-mercaptpethanol with 0.3mm ϕ glass beads by the use of micro tube mixer (TMW-4836, IWAKI) for 10 min at 4℃. The mixture was centrifuged at $10,000 \times g$ for 15 min, and the supernatant was saved into new tube. The pellet was suspended with Buffer A, centrifuged again under the same condition and the pellet was removed completely. Ammonium acetate (NH_4Ac) was added to the

combined supernatants to a final concentration of 1M, incubated on ice for 30 min, and centrifuged at 10,000×*g* for 10 min. The supernatant of this high salt washed ribosome suspension was loaded onto a 10-40%(w/v) linear sucrose density gradient in the same high-ionic strength buffer (Buffer D: 1M NH_4Ac, 15mM $MgCl_2$, 20mM Tris-HCl, pH7.6) with 6mM 2-mercaptoehanol, centrifuged at 40,000 rpm (202,000×*g*) for 2 hr in a swing rotor (SW40Ti, Beckman) at 4℃ for *G. intestinalis*, or at 40,000 rpm (140,000×*g*) for 3 hr in an angle rotor (45Ti, Beckman) at 4℃ for *T. vaginalis*. The supernatant was fractionated and the absorbance at 260nm was recorded. The fractions containing intact ribosome (*G. intestinalis*) or small or large subunit of ribosome were recollected separately, overlaying each one onto a 40% (w/v) sucrose cushion in Buffer D. The ribosomes were finally pelleted by centrifugation at 30,000 rpm (55,800×g) for 12 hr at 4℃ for *G. intestinalis*, or at 40,000rpm (99,000×*g*) for 12 hr at 4℃ for *T. vaginalis* in an angle rotor (90Ti, Beckman).

Extraction and electrophoresis of ribosomal proteins Protein extracts were prepared from the ribosomal fractions by the acetic acid method (Hardy et al. 1969) and were dialyzed over-night in 2%(v/v) acetic acid with membrane (MW=1000, Spectra/Por). The proteins were lyophilized (FZ-1, LABCONCO) and stored at -80℃. The ribosomal protein fraction was separated by radical-free and highly reduced method of two-dimensional polyacrylamide gel electrophoresis (RFHR 2D-PAGE) (Wada 1986a,b) except 8M urea was employed instead of 6M urea. After electrophoresis, the gels were stained with amido black10B in 1%(v/v) acetic acid and destained with same concentration of acetic acid. Gel images were scanned with a densitometer (PD110, Molecular Dynamics) and the number of protein spots were counted on each gel.

N-terminal sequencing of the proteins The ribosomal proteins on the gels were electroblotted from the gels to a polyvinylidene difluoride (PVDF) membrane. Protein spots on the membrane were cut out and applied for N-terminal amino acid sequencing by a protein sequencer (G1005A, Hewlett packard). Up to 20 sequencing cycles were performed on each spot.

Sequence similarity analysis and database search N-terminal amino acid sequence of each protein was compared with all protein sequences in the public database using the BLAST (http://www.ncbi.nlh.gov/BLAST/) and FASTA (http://fasta.genome.ad.jp/) programs to determine the corresponding homologous sequences present in other organisms.

Cloning and sequencing of ribosomal proteins Genomic DNA was extracted from *G. intestinalis* and *T. vaginalis* with the use of a Blood and Culture DNA kit (QIAGEN) according to the manufacture's protocol. The sequences for the S14, S15a, L5, L8 and L10a ribosomal protein genes were searched for the *G. intestinalis* genome project database (http://evol3.mbl.edu/Giardia-HTML/). Based on the identified sequences a set of oligonucleotide primers were designed for upstream (forward) and downstream (reverse) regions of the ORFs of these ribosomal proteins: S14;
5'-GTACGAGTTTGCCTTTGCATTAAAG-3' (forward),
5'-CAAGCGCCGTGTTATGTAAACATAG-3' (reverse);
S15a: 5'-GCTGAGCGCACATCGCGTGCCGGCG-3' (forward),
5'-CGTCGCCGCGTTTACTCACAGGTCC-3' (reverse);
L5: 5'-GGAGCCCTCTGGTGAAATACAAACC-3' (forward),
5'-CTGTTCAGGCACAACCTGGATCCCG-3' (reverse);
L8: 5'-CGCCCACTTGATCCGCTGTGTGATA- 3' (forward),
5'-GCAGGGAAAAGGTCTCCGTCATAGC-3' (reverse);
L10a: 5'-CAATATTAATTTAGAATCTGAGTC-3' (forward),
5'-GGTGATTTACTCGGCAGAGCCTCCC-3' (reverse). Degenerated forward primers for *T. vaginalis* ribosomal proteins S14, S15a, L5 and L8 were designed based on N-terminal amino acid sequences, while reverse primers degenerated for a highly conserved region in the C-terminal portion of these ribosomal proteins were also designed: S14; GVAHIYA (forward) (5'-GGNGTNGCNCAYATHTAYGC-3'), GRIEDVT (reverse) (5'-GGNMGNATHGARGAYGTNAC-3'); S15a: DALKAITN (forward) (5'-GAYGCNYTNAARGCNATHACNAA-3'), GRLNKCGV (reverse)

(5'-GGNMGNYTNAAYAARTGYGGNGT-3'); L8: MGHIVRAQ(forward) (5'-ATGGGNCAYATHGTNMGNGCNCA-3'), GVAMNPV(D/E) (reverse) (5'-TCNACNGGRTTCATNGCNACNCC-3'). An uneven PCR method was used to obtain N- or C- terminal portion of each gene (Chen & Wu 1997). Specific primers for the uneven PCR method were synthesized based on the sequence information already obtained, while arbitary primers were based on the sequences described in Chen and Wu (1997). Plasmid vectors, pT7 (Novergen) or pCR2.1 (Invitrogen), were used for cloning the PCR amplified bands. Nucleotide sequences were determined on both strands with the use of an automated DNA sequencer (Model 310 Genetic Analyzer, ABI).

Cloning and sequencing of LSUrRNAs For *M. balamuthi* LSUrRNA , see § 1.2.1. in detail. Genomic DNA of *A. salina* was extracted from the embryo by Blood and Culture DNA kit (QIAGEN) according to the manufacture's protocol. As partial LSUrRNA sequences of *A. salina* were found in the database, three sets of primers were designed based on these sequences: A1: 5'-CAAACTTGATCCTTTAGAGGAAG-3' (forward),
A2: 5'-CATCGACCTACGACTCGCACACAAG-3' (reverse),
B1: 5'-GATCCCACGGAAAGGAGTTGGTTG-3' (forward),
B2: 5'-CCTCCCATTTTAAGGGCTAGTTGC-3' (reverse)
C1: 5'-GTACCGATAACCGCATCAGGTCC-3' (forward) and
C2: 5'-GATTCTACTTAGAGGCGTTCAG-3' (reverse). The PCR products using these sets of primers were cloned and sequenced as mentioned above. By concatenating the sequences for these three PCR-based fragments, an almost entire part of the *A. salina* LSUrRNA sequence was obtained.

RNA extraction Total RNAs were extracted from *T. vaginalis*, *G. intestinalis*, *S. cerevisiae* and *E. coli* with the use of Rneasy Mini kit (QIAGEN) and from *M. balamuthi* with Tri Reagent LS- RNA /DNA /PROTEIN Isolation Reagent (Molecular Research Center). Extracted RNAs were electrophoresed on a 1%(w/v) agarose gel containing 2.2 M formaldehyde for comparing the size of the SSU- and LSUrRNAs among different organisms.

Estimation of secondary structure model of rRNA GENETYX-WIN (ver.5) was used for the estimation of secondary structure model of *M. balamuthi* LSUrRNA. This program is constructed based on the Zuker's method (Zuker 1989).

2.3 Results and discussion

2.3.1 Sedimentation profiles of ribosomes

The sedimentation profiles from various organisms were analyzed by sucrose density gradient centrifugation (Fig. 2.1). The peaks of the *G. intestinalis* and *T. vaginalis* intact ribosomes were located between the peaks corresponding to *S. cerevisiae* and *E. coli*, indicating that the sedimentation coefficients of *G. intestinalis* and *T. vaginalis* are larger than that of *E. coli* and smaller than that of *S. cerevisiae*. The sedimentation coefficient of *A. salina* intact ribosome was shown to be 81S by Nieuwenhuysen & Clauwaert (1981). The *S. cerevisiae* peak located at slightly left side of the *A. salina* peak, indicating that the *S. cerevisiae* ribosome has smaller sedimentation coefficient than that of *A. salina*. The sedimentation profile of *T. vaginalis* showed three peaks. These correspond respectively left to right, to small subunit, large subunit and to intact ribosomes of *T. vaginalis*. In general, the ribosomes of well studied organisms show a profile with a monosome peak in the high Mg^{2+} concentration used in Buffer A (100mM NH_4Ac, 15mM $MgAc_2$, 20mM Tris-HCl, pH7.6) (Iwabuchi and Ochiai 1969; Cook et al. 1971; Zasloff and Ochoa 1971; Chaires and Kegeles 1977), but in the same condition, the ribosome of *T. vaginalis* showed a profile with dissociated subunits. The affinity that associates each subunit of *T. vaginalis* ribosome might be weak and thus the intact ribosome might be easily broken into subunit forms by centrifugal force on the sedimentation analysis, otherwise, factors needed for monosome formation in *T. vaginalis* might be lost during preparation of the ribosome extract. It was reported by Oka et al. (1973), that the ribosome of *Tritrichomonas foetus*, one of the Trichomonad species is very sensitive to the stress of cell destruction, and therefore that the freeze-throw method is

effective to obtain an intact ribosome.

The sedimentation profile of dissociated ribosomal subunits were compared (Fig. 2.2). Because ionic conditions suitable for obtaining a profile with dissociated subunits are different in *T. vaginalis* and *S. cerevisiae*, different buffers were used. In these analyses, the sedimentation profile of the *E. coli* dissociated subunits was used as a control to compare the sedimentation coefficients between *T. vaginalis* and *S. cerevisiae*. The peaks for small and large subunits of *T. vaginalis* and of *S. cerevisiae* were located to the right, compared with those of *E. coli*, indicating that sedimentation coefficients of dissociated subunits for these organisms are larger than those of *E. coli*. In addition, the sedimentation coefficients for *T. vaginalis* showed slightly smaller than those of *S. cerevisiae*, being comparable with the profile shown in Fig. 2.1, although a peak obtained for the *T. vaginalis* intact ribosome was very low.

2.3.2 Isolation of ribosomal proteins and 2D RFHR-PAGE analyses

Ribosomal proteins of *G. intestinalis* and *T. vaginalis* were extracted from the high salt washed ribosomal preparations and resolved by the RFHR 2D-PAGE.

Giardia intestinalis At least 74 protein spots were identified on the gel (Fig. 2.3A), and these were schematically represented in Fig. 2.3B. Proteins of high salt washed ribosomes of rat (*Rattus norvegicus*) and *E. coli* were also analyzed in the same way, giving the profiles shown in Fig. 2.3C and D. The number of the separated ribosomal protein spots of *E. coli* (~57) was much lower than that of *G. intestinalis* or rat (~80). The 2D profiles of *G. intestinalis* and rat were more similar to each other than to *E. coli*, indicating that the physicochemical nature of the individual ribosomal protein is conserved between *G. intestinalis* and rat . This clearly suggests the typical eukayotic nature of the protein components of the *G. intestinalis* ribosome.

Trichomonas vaginalis Forty-two and 55 protein spots were identified respectively on the gels for the small and large subunits (Fig. 2.4A, C), and these were schematically

represented in Fig. 2.4B and D. Each of the ribosomal proteins, S26, L21 and L37a, was identified by two different spots on the gel with a same N-terminal sequence (Table 2.2). Although the cause are actually unknown, different location for these proteins on the gel may have resulted from post-translational modification, presence of the products from different copy genes, and/or differentiation of the movement on the gel due to the formation of disulfide linkage within a molecule. Same sequence was found from three spots out of 18 spots examined (17%) in the present analysis. If the percentage was assumed also to be applicable to other spots, roughly 17% of all the spots would correspond to the ones that were derived from same ribosomal proteins. The total number of different ribosomal proteins on these gels could thus be regarded as about 80.

Ribosomal proteins have already been extensively analyzed by 2D gel electrophoresis for various eukaryotic organisms, including rat (*Rattus norvegicus*) (Wool 1979; Wool, Chan and Gluck 1995), *S. cerevisiae* (Mager et al. 1997; Michel, Traut and Lee 1983), *Dictyostelium discoideum* (Ramagopal and Ennis 1980), *Tetrahymena thermophila* (Petridou et al. 1983), *Podospora anserina*(Dequard-Chablat et al. 1986) and *Chlamydomonas reinhardtii* (Fleming, Boynton and Gillham 1987). The number of ribosomal proteins of these organisms ranges from 71 to 93. The wide range is probably due to the difference of the criteria used for the identification of the protein spots on the 2D gel. From our RFHR 2D-PAGE analysis, *T. vaginalis* has as many ribosomal proteins as 'typical' eukaryotes, indicating that no prokaryotic feature is observed at least on the number of ribosomal proteins. The fewer number of ribosomal proteins of four trichomonad (40~56) were reported by Champney, Chittum and Samuels (1992). Since *T. vaginalis* and four trichomonad species are phylogenetically very closely related with each other, it is very unlikely that the number of ribosomal proteins on the 2D gel varies from 40 through 56, and, in contrast, that *T. vaginalis* has about 80 ribosomal proteins in our analysis.

Using low Mg^{2+} composition buffer (350 mM KCl, 7.5mM $MgAc_2$, 50mM Tris-HCl pH 7.6) for sedimentation analyses of trichomonads, they separated ribosomal particles from the

fractions of the sedimentated gradient and analyzed ribosomal proteins by 2D gel electrophoresis or by high pressure liquid chromatograph, and then, they counted the number of proteins. In their buffer condition, however, the *T. vaginalis* ribosomes clearly show a profile with dissociate subunits in our system. Since the medium (Diamond 1957) and the centrifugation condition were almost the same between their analysis and our analysis, there is a possibility that they fractionated only large subunit ribosomes from the four trichomonads. In 'typical' eukaryotes, the number of ribosomal proteins are about 30 and 50 respectively on small and large subunits. If they had counted the number of ribosomal proteins only on the large subunit trichomonad ribosomes, the numbers 40~56 would be appropriate, although the variation of the numbers is still too large.

2.3.3 N-terminal sequencing of *Giardia intestinalis* and *Trichomonas vaginalis* ribosomal protein

Ribosomal protein spots randomly selected from the 2D gels were sequenced. The number of sequenced proteins were 15 from *G. intestinalis* (Fig. 2.3B) and 18 from *T. vaginalis* (12 from small and 6 from large subunit gel, Fig.2.4B,D). These were compared with all protein sequences in the public databases using the FASTA programs. Significant similarity to known ribosomal proteins was found (Table 2.1). Each of these sequences, except L17 of *G. intestinalis* and S12 of *T. vaginalis*, showed the highest similarity score to a 'typical' eukaryotic ribosomal protein among the ribosomal proteins identified in each FASTA analysis. The ribosomal protein L17 of *G. intestinalis* and S12 of *T. vaginalis* showed the highest FASTA score with those of archaebacterial homologs, although they also have similarities with corresponding homologs of eukayotes (data not shown). None of the ribosomal proteins that exist only in bacteria (b) or archaebacteria (a) were found in *G. intesinalis* and *T. vaginalis*.

2.3.4 Comparison of amino acid sequences for five ribosomal proteins

Complete amino acid sequences of ribosomal proteins S14, S15a, L5, L8 and L10a deduced from nucleotide sequences of the corresponding genes of *G. intestinalis* and *T. vaginalis* were aligned with those of archaebacterial, eubacterial and eukaryotic homologs (Fig. 2.5) and pairwise amino acid similarities were compared among them. The putative open reading frames (ORF's) of *G. intestinalis* and *T. vaginalis* ribosomal proteins cloned and sequenced in this study were not interrupted by intron-like sequences.

S14 As shown in the alignment (Fig. 2.5) N-terminal potion of S14 ribosomal proteins was not conserved among organisms. Two gaps were found in the alignment. One is located at the positions 105-112 and the other is at 144-146. The former is shared only by eubacteria and the latter is shared by eubacteria and an archaebacterium, *Haloarcula marismortui*. Eukaryotes including *G. intestinalis* and *T. vaginalis* and an archaebacterium *Aeropyrum pernix*, do not share these gaps. Pairwise similarities of *A. pernix* sequence to mitochondriate eukaryotes (55.33~61.31%) are higher than that of *A. pernix* to *H. marismortui* (archaebacteria) (54.96%). Pairwise similarities between *G. intestinalis* or *T. vaginalis* and mitochondriate eukaryote (45.28~61.01%) are lower than those of *A. pernix* and mitochondriate eukaryotes (47.59~61.31%). These values do not necessarily indicate the prokaryotic feature of *G. intestinalis* or *T. vaginalis* S14 protein. Phylogenetic analysis clearly supported the monophyly of these organisms with amitochondriate eukaryotes (data not shown).

It has been reported that ribosomal protein S14 of *S. cerevisiae* regulates its expression by binding to ribosomal protein S14B pre-mRNA and to SSUrRNA (Fewell and Woolford 1999). Its eubacterial homologs, S11, is located on the platform region that faces the LSU and is believed to be involved with codon-anticodon recognition (Oakes et al. 1985).

S15a S15a ribosomal proteins were highly conserved among three primary kingdoms in size and sequence (Fig. 2.5). Especially these were well conserved among eukaryotes. Of all the 130 residues, 75 are identical among eukaryotes in the alignment. The *M. balamuthi*

S15a sequence found in the EST project database indicates high pairwise similarity values (64.62~72.09%) with other eukaryotes. The pairwise similarity between *G. intestinalis* and *T. vaginalis* is 56.92%, and it is the lowest among any pairs of the eukaryotes.

Rat S15a protein was deduced from the nucleotide sequences of two recombinant cDNAs and 10 to15 copies of the S15a gene have been found in the nucleus of rat (Chan et al. 1994). Its eubacterial counterpart, S8, is centrally located in the SSU adjacent to proteins S2, S4, S5, S12, S15 and S17 (all of these are in eubacterial nomenclature). S8 directly interacts with S5 and has crucial role in folding the central domain of SSUrRNA. S8 is also an important regulatory protein controlling the expression of 10 ribosomal protein genes in the spc operon by binding specifically to a region of the polycistronic mRNA (White et al. 2000).

L5 Sequence lengths of L5 are variable among three kingdoms (Fig. 2.5). The eubacterial L5 homologs (L18) are reduced in length compared with eukaryotic homologs. Ribosomal protein L5 (L18 in eubacterial nomenclature) binds to 5S rRNA and has been implicated in the intracellular transport of 5S rRNA. By the study of *Homo sapiens* L5, following sequence features were reported (Rosorius et al. 2000): interaction of L5 with 5S rRNA is mediated by the L5 N- and C-terminal (positions 1-37 and 251-296 in the alignment); portions 20-36 mediate nuclear import and nuclear localization of L5; C-terminal region of the positions 254-264, rich in basic amino acid residues, mediates nuclear targeting; the domain at positions 100-110 serve as a nuclear export signal. The C-terminal portions involved in an interaction with 5S rRNA reported in *H. sapiens* were not conserved in eubacteria. It is suggested that the location of the sequence involved in an interaction with 5S rRNA are different in eubacteria. Even within the eukaryotic lineages, sequence similarities are not high (42.53~52.82%). The similarities between *G. intestinalis* or *T. vaginalis* and other eukaryotes are 42.53~51.52%, those are higher than those between *G. intestinalis* or *T. vaginalis* and prokaryotic homologs. There are insertions at position 111 (LSVGIKE) of *G. intestinalis* and at position 190 (GKDRKDVPAEKIVKGN) of *T. vaginalis*. These are not within the 5S rRNA binding domain or within the nuclear export signal. The N-terminal

sequences of 5S binding domains are relatively conserved among eukaryotes including *G. intestinalis* and *T. vaginalis*, but those of C-terminal portions are less conserved, although C-terminal regions are commonly rich in basic amino acid residues.

L8 Two gaps are found to be shared by eukaryotes including *G. intestinalis* and *T. vaginalis* at positions 12 and 70-74 of the alignment (Fig. 2.5). These are the eukaryotic sequence signatures. *G. intestinalis* and *T. vaginalis* show high similarity values with other eukaryotes. Archaebacterial sequences are more similar to eukaryotic homologs than to eubacterial homologs. Sequence similarities among eukaryotes are higher than those between eukaryotes and archaebacteria. At position 143-144, *T. vaginalis* shared the gap with prokaryote, although the rest of the *T. vaginalis* sequence clearly shows a eukaryotic type feature.

Although, L8 has been less studied, the eubacterial homolog, L2, has been well studied on its biochemical properties. A recent study identified two ribosomal proteins, L2 and L3, as essential protein components of the peptidyltransferase center (Khaitovich et al. 1999). In addition, L2 is known to be a primary LSUrRNA binding protein. It binds directly and specifically to highly conserved stem-loop structure encompassing nucleotides from 1794 to 1825 in *E. coli* LSUrRNA (Egebjerg, Christiansen and Garrett 1991).

L10a L10a of *T. vaginalis* has already been reported to be typical eukaryotic type (Wu and Hashiomoto 1999). The *G. intestinalis* L10a sequence also clearly showed eukaryotic features. Although the sequence similarities between *T. vaginalis* or *G. intestinalis* and other eukaryotes were less than those among other eukaryotes, *G. intestinalis* and *T. vaginalis* shared 82 identical positions with *H. sapiens*, *S. serevisiae* and *A. thaliana* and shown less similarities with the sequence of prokaryotic organisms. The eukaryotic L10a is a homolog of L1 in eubacteria. The L1 locates on lateral protuberance opposite the L7/L12 stalk of the large ribosomal subunit (Stoffler-Meilicke and Stoffler 1990). The *E. coli* L1 associates independently, specifically and strongly with the LSUrRNA (Draper 1990). The L1 protein also can regulate gene expression by binding to its own mRNA, acting as a translational

repressor (Gourse, Sharrock and Nomura 1986). Because of profound interest in the dual function of L1, i.e. rRNA binding and mRNA binding, this protein has been extensively studied from various organism (Petitjean, Bonneaud and Lacrote 1995; Olvera and Wool 1996; Kohre et al. 1998; Nevskaya et al. 2000). Positions 31, 34, 129-132, 209 and 210 in the alignment are suggested to be involved in interaction with RNA (Urlaub et al. 1995). Positions 31, 34, 209 and 210 are highly conserved between eukaryotes and prokaryotes, although sequence motif of positions 129-132 are not conserved. It is considered that RNA binding domains in eukaryotes is slightly different from those of eubacteria. Eukaryotic L1 has well conserved R(L/I)(Y/F) motif in the C-terminal portion. *G. intestinalis* and *T. vaginalis* also have this motif.

2.3.5 Comparison of rRNAs from various organisms

The ribosomal proteins of *G. intestinalis* and *T. vaginalis* have eukaryotic characters in numbers and sequences, but the sedimentation coefficients of their ribosomes are smaller than that of *S. cerevisiae* and *A. salina* ('typical' eukaryotes). What makes their ribosomes smaller? Probably it is the length of rRNAs. Figure 2.6 shows rRNAs extracted from various organisms, and lengths of rRNAs are summarized in Table 2.3. The rRNAs of *G. intestinalis* are shorter not only than those of *S. cerevisiae* but also than those of *E. coli*. Ribosomal RNAs of *T. vaginalis* are slightly longer than *G. intestinalis* but shorter than those of *S. cerevisiae* and *E. coli*. As the components of the ribosomal proteins of the eukaryotic ribosome are different from those of prokaryotic ribosome, the relationships between rRNA length and sedimentation coefficient could not be compared between eukaryotic and prokaryotic organisms. However, among the four eukaryotic organisms examined the order of sedimentation coefficients (*A. salina* > *S. cerevisiae* > *T. vaginalis* > *G. intestinalis*) is in agreement with that of the nucleotide lengths of rRNAs. The Microsporidia, *Nosema bombycis*, was reported to have a sedimentation coefficient resemble to *E. coli* (70S, 50S and 30S, intact, LSU and SSU respectively) (Ishihara and Hayashi 1968). The length of *N.*

bombycis rRNAs are not clear, but from a partial sequence in database and an alignment in the rRNA WWW server (http://rrna.uia.ac.be/), nearly 1,230 bp of SSUrRNA are presumed. This is not in disagreement with the hypothesis described above. Ribosomal RNAs comprise 50 to 70% of the mass of the ribosomal particle (Spirin 1999). Generally, the prokaryotic ribosome is about two third RNA and one third protein, while in eukaryotes, the ratio is approximately fifty-fifty. Shown in Table 2.3, the rRNAs of eukaryotes are variable in length, the largest one is twice as large as the smallest one. If a half of the eukaryotic ribosomal composition is shared by RNA, such a remarkable difference in length must influence on the ribosome volume or sedimentation coefficient.

2.3.6 Secondary structure model of rRNAs

In this section, on the basis of the estimated secondary model of rRNA structures, rRNA sequences were compared among eukaryotes. Figure 2.7 shows the secondary structure model of *S. cerevisiae* SSUrRNA sequence. The universal core consists of the sequence regions which are conserved among eubacteria, archaebacteria, eukaryotes and mitochondria (Gutell 1993). Even in the reduced SSUrRNAs of Microsporidia these regions are strictly conserved. A single number is attribute to 50 universal helices, which are present in all hitherto known (~1993) SSUrRNAs of eubacteria, archaebacteria and plastids. A microsporidian, *E. cuniculi*, lacks the helices 11, 18, 43 and 46. Another microsporidian, *Vairimorpha necatrix*, also lacks the helix 11 as well as *E. cuniculi* (Neefs et al. 1991).

Helices specific to the eukaryotic model are numbered with 'E'(e. g. E10_1, E23) and shaded in green. These helices are less conserved and have large variability among eukaryotic organisms. Table 2.4 clearly shows the variation of the sequence size in the helices E10_1 and E23. In E10_1 area, no helix-like structure is confirmed in *G. intestinalis*, *T. vaginalis* and *E. cuniculi*. In contrast, *M. balamuthi* has a very extended sequence. The E10_1 area of *M. balamuthi* seems to be involved in the formation of plural number of helices. The helices in E23 area show a most complicated structure among SSUrRNAs. As an example, in the *H.*

sapiens SSUrRNA, the helices make multifurcation structures bearing two more helices (Van de Peer et al. 2000). On the problem of estimation, the secondary structure of SSUrRNA of *S. cerevisiae* E_23 area is simply drawn, but the exact form is supposed to be similar to that of *H. sapiens* (Neefs et al. 1993). We could not estimate the secondary structure model of the *M. balamuthi* SSUrRNA E23 area. However, the nucleotide number of this area is three times longer than those of *H. sapiens* or *S. cerevisiae*. Formation of the helices is speculated to be very complicated. As *M. balamuthi* has very elongated SSUrRNA, additional helices exist as shown in Fig. 2.7. Several eukaryotic organisms, such as *Euplotes aediculatus* (Ciliophora), *Euglena gracilis* (Euglenozoa) and *Physarum polycepharum* (Mycetozoa), have additional helices (E8_1) around the helix 8. Although the total length of SSUrRNA is larger than those of these organisms, no helix corresponding to the helix E8_1 was observed in *M. balamuthi*. Around the helix 45, an additional helix was observed in *M. balamuthi*. The helix is absent in *H. sapiens* and in *S. cerevisiae*, but is present in many other organisms belonging to Alveolata, Euglenozoa and amoebae (*Acanthamoeba* and *Naegleria*). Recently, a very extended SSUrRNA with 3,502 bp was sequenced from *Pelomyxa palustris* and its predicted secondary structure model was reported (Milytina et al 2001). *Pelomyxa palustris* is a free-living and amitochondriate amoeba. It is classified in Pelobionta with *M. balamuthi*. On the predicted SSUrRNA secondary structure model of *P. palustris*, the extended sequence pattern and location of additional helices is similar to *M. balamuthi*, suggesting the close phylogenetic relationship between these two organisms that belong to Pelobionta. However, *P. palustris* has more additional helices than *M. balamuthi* and these are located within the helices 6, 8, 16, 17, 24, 29,and 37, while in the E10_1 area, the nucleotide length of *P. palustris* is shorter than that of *M. balamuthi*.

The secondary structure model estimated from the *S. cerevisiae* LSUrRNA sequence is shown in Fig. 2.8. Helices numbered with under bar and shaded in green are not universal among organisms. In the reduced LSUrRNA sequence of a microsporidian, *E. cuniculi*, the universal core parts are conserved, but the helices B7, B8 D3, D5 and E15 are missing

(Peyretaillade et al. 1998b). Compared to the highly extended SSUrRNA sequence of *M. balamuthi*, its LSUrRNA sequence is not so elongated (4,485 bp), but it is longer than those of other protists with additional helices observed around E12, E14, E15, and H1. Many eukaryotic organism have helices E20_1-2, but these are not present in *M. balamuthi*. The structural variability of the helices C1 and G5 were shown in Fig. 2.9. The sequence length of *M. balamuthi* in the C1 area is as twice as longer than that of *S. cerevisiae* and the predicted secondary structure is very complicated with many helices. On the other hand, in the LSUrRNAs of *G. intestinalis*, *T. vaginalis* and *E. cuniculi*, only one small helix is formed in the area. The structural features of the C1 helices of *G. intestinalis*, *T. vaginalis* and *E. cuniculi* are similar, while the sequences in the area are not resemble with each other. In the *G. intestinalis* sequence with high GC content, stems are mainly composed of G-C pairs while in *T. vaginalis* with low GC content, A-U pairs are used mainly in the stem formation. Same as the C1 area, structural variations are also seen in the G5 area (Fig. 2.9B), but the G5 helix is not present in the LSUrRNA sequence of *E. cuniculi*. Figure 2.10 shows the predicted secondary structure model of *M. balamuthi* LSUrRNA which was cloned and sequenced in this study. Based on the multiple alignment analysis, the 5' and 3' portions are presumed. The universal core and helices are well conserved. However, compared to the predicted LSUrRNA structure of *S. serevisiae*, larger number of helices are observed.

In this section, by comparing secondary structure models of rRNAs among various eukaryotic organisms, it was clearly demonstrated that the universal core regions and universal helices are conserved even in the reduced rRNAs of *G. intestinalis* and *T. vaginalis*. The smaller sedimentation coefficients of *G. intestinalis* and *T. vaginalis* ribosomes than those of 'typical' eukaryotic organisms may be resulted from reduced rRNAs with shortened variable regions.

2.3.7 Database search of the *Giardia intestinalis* ribosomal proteins

Up to December 2001, 73 ribosomal protein sequences were available in the *Giardia*

lamblia (*G. intestinalis*) genome project database (http://www.mbl.edu/Giardia/index2.html). The BLAST analysis of each ribosomal protein sequences against the sequences in the public databases showed for all 73 *G. intestinalis* sequences the highest similarity score to homologs of eukaryotic protein sequences (data not shown). A homologs of ribosomal protein L24 has not been found in the *G. intestinalis* genome database, but it was identified from N-terminal sequencing analysis in this study. The ribosomal proteins derived from genome database search and N-terminal sequencing analysis allowed to identify 74 ribosomal protein species, 32 and 42 for the small and large subunits, respectively in *G. intestinalis*. Among these 74 ribosomal proteins, 34, 26 and 13 belonged respectively to the classes, eab, ea and e (Fig. 2.11). Comparison of the ribosomal proteins of *G. intestinalis* with those of *S. cerevisiae* revealed that L31, L41 (class ea), L6, L38 and 25(class e) have not been identified in *G. intestinalis*. As the genome project is now on progress, some of these ribosomal proteins may be identified in the near future. As shown in Fig. 2.11, the numbers of ribosomal proteins are different among eukaryotes, archaebacteria and eubacteria. Since 80 proteins in round number are identified both in *R. norvegicus* and *S. cerevisiae*, the round number of the ribosomal proteins of 'typical' eukaryotes are assumed to be 80. At least 74 ribosomal proteins found in *G. intestinalis* are comparable with this number. In addition, sequence alignment analyses of the ribosomal proteins showed that most of them can be easily aligned with their corresponding eukaryotic counterparts without remarkable gaps, suggesting again that the protein components of the *G. intestinalis* ribosome are very likely to be 'typical' eukaryotic type. Recently, the genome project of the Microsporidia, *E. cuniculi* have finished and the whole picture of this organism was reported (Kathinka et al. 2001). As shown in Fig. 2.11, in addition to its shortened rRNAs, *E. cuniculi* dose not have 8 ribosomal proteins. Of course, *E. cuniculi* synthesizes its proteins using its own ribosome. The reduced ribosome could work as same as the ribosomes of other eukaryotic organisms. The proteins being lost might not important for protein synthesis. As well as the situation in *E. cuniculi*, there is a possibility that several ribosomal proteins are missing in the *G. intestinalis* ribosome. species, however

larger number of ribosomal proteins than prokaryotes have identified and the sequence similarity each of protein to 'typical' eukaryote was confirmed, the ribosome of *G. intestinalis* must be eukaryotic and not indicating any prokaryotic features.

2.4 Summary and conclusions

The ribosomes of amitochondriate protists *Giardia intestinalis* and *Trichomonas vaginalis* were analyzed. The sedimentation coefficients of these ribosomes were larger than that of *Escherichia coli* and smaller than that of *Saccharomyces cerevisiae* or *Artemia salina*.

The RFHR-2D-PAGE analysis of *G. intestinalis* ribosomal proteins resolved 74 protein spots, a number close to that of a 'typical' eukaryote (about 80) and larger than that of *E. coli* (about 55). The analysis of *T. vaginalis* ribosomal proteins resolved 97 protein spots (42 from small and 55 from large subunits). Since same proteins were occasionally identified from two different spots in the N-terminal sequencing analysis, the number of different ribosomal proteins of *T. vaginalis* was roughly estimated to be about 80, based on the frequency of duplicated protein spots. The N-terminal sequences of 15 and 18 randomly selected protein spots, respectively from *G. intestinalis* and *T. vaginalis* showed high similarity scores with 'typical' eukaryotic ribosomal proteins. The N-terminal sequencing analysis and the database search of the *G. intestinalis* genome sequencing project up to December 2001 already identified 74 *G. intestinalis* ribosomal proteins, a number almost comparable with the number for 'typical' eukaryotes. Alignment analyses of the ribosomal proteins including the ones, of which complete sequences were determined from *G. intestinalis* and *T. vaginalis* in this work, revealed that the sequences of *G. intestinalis* and *T. vaginalis* are clearly of 'typical' eukayotic type with no exception.

On the other hand, sequence comparison analyses revealed that the SSU and LSUrRNAs of *G. intestinalis* and *T. vaginalis* were remarkably shorter in length than those of 'typical' eukaryotes. All the helices that belong to the universal core, however, were strictly conserved also in *G. intestinalis* and *T. vaginalis*. In contrast, variable regions of both rRNAs were

reduced to be short in *G. intestinalis* and *T. vaginalis*.

As far as these results are concerned, the protein components and the essential parts of the rRNAs of the *G. intestinalis* and *T. vaginalis* ribosomes are clearly of 'typical' eukaryotic type. No 'prokaryotic' or 'intermediate' features are found in the ribosomes of these amitochondriate protists. The smaller sedimentation coefficients of the ribosomes of *G. intestinalis* and *T. vaginalis* than those of 'typical' eukaryotes are due to the smaller size of rRNAs with shortened variable regions. These findings give additional evidence for fully developed eukaryotic nature of *G. intestinalis* and *T. vaginalis*. Probably Diplomonadida, Parabasala and other eukaryotes already had obtained major eukaryotic properties commonly found in the 'typical' eukaryotes.

Conclusion: Evolutionary status of the amitochondoriate protists

According to the 'Archezoa' hypothesis, several amitochondoriate protist lineages were used to be considered as primitive eukaryotes that diverged preceding the endosymbiotic origin of proto-mitochondoria.

Five amitochondoriate protist lineages that had previously been regarded as Archezoa were phylogenetically clearly positioned within a eukaryotic tree on the basis of the combined ML analyses examined comprehensively in this work: Microsporidia were closely related to Fungi; Pelobionta and Entamoebidae were the closest relatives with each other, and these two lineages were linked with Mycetozoa to form a group, Conosa; Diplomonadida and Parabasala diverged earlier than other eukaryotic lineages.

Early branching positions of the two amitochondriate lineages, Diplomonadida and Parabasala were clearly supported in this work, and thus these lineages could remain as candidates for Archezoa based only on this evidence. However, many lines of other evidence confirmed recently that all of the five amitochondriate lineages including these two once harbored mitochondria in their evolutionary past. Since it is likely that several other amitochondriate protist lineages also lost their mitochondria secondarily, the mitochondrial endosymbiosis suggested to have taken place prior to the divergence of all extant eukaryotes as shown in Fig. I. Studies on the morphology, biochemistry, and molecular biology of Diplomonadida and Parabasala have revealed that many 'typical' eukaryotic features had already obtained by the time when these two lineages diverged.

In this work, 'typical' eukaryotic features of the ribosomes in *Giardia intestinalis* (Diplomonadida) and *Trichomonas vaginalis* (Parabasala) were clearly elucidated. In addition, presence of a spliceosomal intron was shown at the first time in the *G. intestinalis* HSP90c gene. These findings give additional evidence for fully developed eukaryotic nature of *G. intestinalis* and *T. vaginalis*. According to Roger (1999), Fig. I shows a selection of structural features, processes, molecules, that are found in both well studied 'typical' eukaryotic lineages (e. g., Metazoa, Fungi, Plantae) and Diplomonadida and/or Parabasala, but are absent

in archaebacteria. It is clearly suggested that the last common ancestor of Diplomonadida, Parabasala and other eukaryotic lineages was not a primitive but a relatively fully developed eukaryote.

Acknowledgments

I would like to give my best thanks to my supervisor Dr. T. Hashimoto, for his continual direction and discussion throughout the course of this study.

I wish to thank Dr. M. Hasegawa for helpful discussion. I would like to express my thanks to Drs. A. Wada, H. Yoshida and Y. Maki (Osaka Medical College) for technical support and discussion for ribosomal works. I also thank Drs. M. Müller, L. B. Sánchez (The Rockefeller Univ.), L. M. Weiss (Albert Einstein College of Medecine), H. Yoshikawa (Nara Women's Univ.), K. Kita, A. Takeo (Univ. of Tokyo) and T. Uchiumi (Shinsyu Univ.) for providing experimental samples and discussion. I am grateful for technical assistants in Hashimoto Lab. (Institute of Statistical Mathematics), without their help I could not accomplish all of my experiments.

Finally, I would like to thank my family who have always been supported me.

Literature cited

Adachi, J. & Hasegawa, M. 1996. Computer Science Monographs, No. 28, MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. Tokyo: The Institute of Statistical Mathematics.

Adam, R. D. 1991. The biology of *Giardia* spp. *Microbiol. Rev.*, **55**:706--732.

Adam, R. D. The *Giardia lamblia* genome. 2000. *Int. J. Parasitol.*, **30**:475--484.

Arisue, N., Sánches, L. B., Weiss, L. M., Müller, M. & Hashimoto, T. 2002. Mitochondorial-type hsp70 genes of the amitochondoriate protests, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Parasitol. Int.*, in press.

Andresen, N., Chapman-Andresen, C. & Nilsson, J. R. 1968. The fine structure of *Pelomyxa palustris*. *Compt. Rend. Trav. Lab. Carlsberg*, **36**:285--320.

Bakatselou, C., Kidgell, C. & Clark C. G. 2000. A mitochondrial-type hsp70 gene of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.*, **110**:177--182.

Bakker-Grunwald, T. & Wöstmann, C. 1993. *Entamoeba histolytica* as a model for the primitive eukaryotic cell. *Parasitol. Today*, **9**:27--31.

Baldauf, S. L. 1999. A search for the origins of animals and fungi: comparing and combining molecular data. *Am. Natr.*, **154**:S178--S188.

Baldauf, S. L. & Palmer, J. D. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA*, **90**:11558--11562.

Baldauf, S. L. & Doolittle, R. F. 1997. Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. USA*, **94**:12007--12012.

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*, **290**:972--977.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M. & Philippe, H. 2002. The analysis of one hundred genes supports the grouping of three highly divergent amoebae, *Dictyostelium*, *Entamoeba* and *Mastigamoeba*. *Proc. Nature. Acad. Sci. USA*, in press.

Ben Ali, A., Wuyts, J., De Wachter, R., Meyer, A. & Van de Peer Y. 1999. Construction of a variability map for eukaryotic large subunit ribosomal RNA. *Nucleic Acids Res.*, **27**:2825--2831.

Biderre, C., Méténier, G. & Vivarès, C. P. 1998. A small spliceosomal-type intron occurs in a

ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.*, **94**:283--286.

Boreham, P. F. L. & Stenzel, D. J. 1993. *Blastocystis* in humans and animals: morphology, biology, and epidemiology. *Adv. Parasitol.*, **32**:1--70.

Bourne, H. R., Sanders, D. A. & McCormick, F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**:117--127.

Bozner, P. 1997. Immunological detection and subcellular localization of HSP70 and HSP60 homologs in *Trichomonas vaginalis*. *J. Parasitol.*, **83**:224-229.

Brugerolle, G. 1991. Cell organization in free-living amitochondriate heterotrophic flagellates. *In*: Patterson, D. J. & J. Larsen, J (ed.) The Biology of Free-living Heterotrophic Flagellates. Clarendon Press, Oxford, Pp. 133--148

Brugerolle, G. 1993. Evolution and diversity of amitochondrial zooflagellates. *J. Eukaryot. Microbiol.*, **40**:616--618.

Brumpt, E. 1912. *Blastocystis hominis* n. sp. et formes voisines. *Bull. Soc. Pathol. Exot.*, **5**:725--730.

Bryan, R. T. & Schwartz, D. A. 1999. Epidemiology of microsporidiosis. *In*: Wittner, M. & Weiss, L. M. (ed) The Microsporidia and Microsporidiosis. ASM press, Washington DC, Pp. 502--516.

Bui, E. T., Bradley, P. J. & Johnson, P. J. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl. Acad. Sci. USA*, **93**:9651--9656.

Cao, Y., Adachi, J., Janke, A., Pääbo, S. & Hasegawa, M. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J. Mol. Evol.*, **39**:519--527.

Cavalier-Smith, T. 1986. The kingdom Chromista: origin and systematics. *In*: Round, F.E. & Chapman, D.J. (ed.), Progress in phycological research. BioPress, Bristol, Pp. 309--347.

Cavalier-Smith, T. 1987. Eukaryotes with no mitochondria. *Nature*, **326**:332--333.

Cavalier-Smith, T. 1991. Archamoebae: the ancestral eukaryotes? *BioSystems*, **25**:25--38.

Cavalier-Smith, T. 1993. Kingdom protozoa and its 18 phyla. *Microbiol. Rev.*, **57**:953--994.

Cavalier-Smith, T. 1998. A revised six-kingdom system of life. *Biol. Rev.*, **73**:203--266.

Chaires, J. B. & Kegeles, G. 1977. Sucrose density gradient sedimentation of *E. coli* ribosomes. *Biophys. Chem.*, 7:173--178.

Chakrabarti, D., Dame, J. B., Gutell, R. R. & Yowell, C. A. 1992. Characterization of the rDNA unit and sequence analysis of the small subunit rRNA and 5.8S rRNA genes from Tritrichomonas foetus. *Mol. Biochem. Parasitol.*, **52**:75--83.

Champney, W. S., Chittum, H. S. & Samuels, R. 1992. Ribosomes from trichomonad protozoa have prokaryotic characteristics. *Int. J. Biochem.*, **24**:1125--1133.

Chan, Y. L., Olvera, J., Paz, V. & Wool, I. G. 1994. The primary structure of rat ribosomal protein S15a. *Biochem. Biophys. Res. Commun.*, **200**:1498--1504.

Chatton, B., Walter, P., Ebel, J. P., Lacroute, F. & Fasiolo, F. 1988. The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. *J. Biol. Chem.*, **263**:52--57.

Chavez, L. A., Balamuth, W. & Gong, T. 1986. A light and electron microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n.g., n.sp. *J. Protozool.*, **33**:397--404.

Chen, X. & Wu, R. 1997. Direct amplification of unknown genes and fragments by uneven polymerase chein reaction. *Gene*, **185**:195--199.

Clark, C. G. & Roger, A. J. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. U S A.*, **92**:6518--6521.

Cook, R. T., Rock, R. C., Aikawa, M. & Fournier, M. J. Jr. 1971. Ribosomes of the malarial parasite, *Plasmodium knowlesi*. I. Isolation, activity and sedimentation velocity. *Comp. Biochem. Physiol. B*, **39**:897--911.

Dequard-Chablat, M., Coppin-Raynal, E., Picard-Bennoun, M. & Madjar, J. J. 1986. At least seven ribosomal proteins are involved in the control of translational accuracy in a eukaryotic organism. *J Mol. Biol.*, **190**:167--175.

De Rijk, P., Gatehouse, H. S. & De Wachter, R. 1998. The secondary structure of Nosema apis large subunit ribosomal RNA. *Biochim. Biophys. Acta.*, **1442**:326--328.

De Rijk, P., Robbrecht, E., De Hoog, S., Caers, A., Van de Peer, Y. & De Wachter R.1999. Database on the structure of large subunit ribosomal RNA. *Nucleic Acids Res.*, **27**:174--178.

Desser, S. S., Hong, H. & Siddall, M. E. 1993. An ultrastructural study of *Brugerolleida algonquinensis* gen. nov. sp. nov., Diplomonadina; Diplomonadida., a flagellate parasite in the blood of frogs from Ontario, Canada. *Eur. J. Protistol.*, **29**:72--80.

Diamond, L. S. 1957. The establishment of various Trichomonads of animals and man in axenic cultures. *J. Parasitol.*, **43**:488--490.

Diamond, L. S., Harlow, D. R. & Cunnick, C. C. 1978. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans. R. Soc. Trop. Med. Hyg.*, **72**:431--432.

Draper, D. E. 1990. Structure and function of ribosomal protein-RNA complexes: thermodynamic studies. *In*: Hill, W. E., Dahlberg, A., Garrett, R. A., Moor, P. B. Schlessinger, D. & Warner, J. R. (ed) The ribosome: structure, function, and evolution. ASM press, Washington DC, Pp. 160--167.

Dyall, S. D. & Johnson, P. J. 2000. Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr. Opin. Microbiol.* **3**:404--411.

Dyall, S. D., Koehler, C. M., Delgadillo-Correa, M. G., Bradley, P. J., Plumper, E., Leuenberger, D., Turck, C. W. & Johnson, P. J. 2000. Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane -targeting pathways between hydrogenosomes and mitochondria. *Mol. Cell. Biol.*, **20**:2488--2497.

Dykstra, M. J. 1977. The possible phylogenetic significance of mitochondrial configurations in the acrasid cellular slime molds with reference members of the eumycetozoa and fungi. *Mycologia*, **69**:579--591.

Edlind, T. D., Li, J., Visvesvara, G. S., Vodkin, M. H., McLaughlin, G. L. & Katiyar, S. K. 1996. Phylogenetic analysis of β-tubulin sequences from amitochondrial protozoa. *Mol. Phylogenet. Evol.*, **5**:359--367.

Egebjerg, J., Christiansen, J. & Garrett, R. A. 1991. Attachment sites of primary binding proteins L1, L2 and L23 on 23 S ribosomal RNA of *Escherichia coli*. *J. Mo.l Biol.*, **222**:251--264.

Embley, T. M. & Hirt, R. P. 1998. Early branching eukaryotes? *Curr. Opin. Genet. Dev.*, **8**:624--629.

Embley, T. M., Finlay, B. J., Dyal, P. L., Hirt, R. P., Wilkinson, M. & Williams, A. G. 1995. Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. *Proc. R. Soc. London, B*, **262**:87--93.

Fast, N. M., Logsdon, J. M. Jr. & Doolittle, W. F. 1999. Phylogenetic analysis of the TATA box binding protein (TBP) gene from *Nosema locustae*: evidence for a microsporidia-fungi relationship and spliceosomal intron loss. *Mol. Biol. Evol.*, **16**:1415--1419.

Ferguson, H. W. 1979. Scanning and transmission electron microscopical observation on *Hexamita salmonis* related to mortalities in rainbow trout fry *Salmon gairdneri*. *J. Fish Diseases*, **2**:57--67.

Fewell, S. W. & Woolford, J. L. Jr. 1999. Ribosomal protein S14 of *Saccharomyces cerevisiae* regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Mol. Cell. Biol.*, **19**:826--834.

Fleming, G. H., Boynton, J. E. & Gillham, N. W. 1987. Cytoplasmic ribosomal proteins from *Chlamydomonas reinhardtii*: characterization and immunological comparisons. *Mol. Gen. Genet.*, **206**:226--237.

Frend, D. S. 1966. The fine structure of *Giardia muris*. *J. Cell. Biol.*, **29**:317--332.

Galtier, N. & Gouy, M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. USA*, **92**:11317--11321.

Germot, A. & Philippe, H. 1999. Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *J. Eukaryot. Microbiol.*, **46**:116--124.

Germot, A., Philippe, H. & Le Guyader, H. 1997. Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol.*, **87**:159--168.

Germot, A., Philippe, H. & Le Guyader, H. 1996. Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. *Proc. Natl. Acad. Sci. USA*,. **93**:14614--14617.

Goldman, N., Anderson, J. P. & Rodrigo, A. G. 2000. Likelihood based tests of topologies in phylogenetics. *Syst. Biol.*, **49**:652--670.

Ghosh, S. Ghosh, R., Das, P. & Chattopadhyay, D. 2001. Expression and purification of recombinant *Giardia* fibrillarin and its interaction with small nuclear RNAs. *Protein Expr. Purif.*, **21**:40--48.

Gouse, R. L.,Sharrock, R. A. & Nomura, M. 1986. Control of ribosomal synthesis in *Escherichia coli*. *In*: Hardesty, B. & Kramer, G.(ed) Structure, function, and genetics of ribosomes. Springer-Verlag, New York, Pp. 766--788.

Grell, K.G., Heini, A. & Schéller, S. 1990. The ultrastructure of *Reticulosphaera socialis* Grell (Heterokontophyta). *Eur. J. Protistol.*, **26**:37--54.

Griffin, J. L. 1988. Fine structure and taxonomic position of the giant amoeboid flagellate *Pelomyxa palustris*. *J. Protozool.*, **35**:300--315.

Gutell, R. R.1993. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Res.*, **21**:3051--3054.

Gutell, R. R., Gray, M. W. & Schnare, M. N. 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. *Nucleic Acids Res.*, **21**:3055--3074.

Guthrie, C. & Fink, G. R. (ed) 1991.Methods in Enzymology vol. 194: Guide to yeast genetics and molecular biology. Academic Press, San Diego, Pp. 454.

Hackstein, J. H. P., Akhmanova, A., Boxma, B., Harhangi, H. R. & Voncken, F. G. J. 1999. Hydrogenosomes: eukaryotic adaptations to anaerobic environments. *Trends Microbiol.*, 7:441--447.

Hardy, S. J. S., Kurland, C. G, Voynow, P. & Mora, G. 1969. Extraction of ribosomal proteins with acetic acid. *Biochemstry*, **8**:2897--2905.

Hasegawa, M. & Hashimoto, T. 1993. Ribosomal RNA trees misleading? *Nature*, **361**:23.

Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22**:160--174.

Hasegawa, M., Hashimoto, T. & Adachi, J. 1992. Origin and fvolution of eukaryotes as inferred from protein sequence data. In: Haryman, H. & Matsuno, K. (ed) The Origin and Evolution of the Cell, Singapore, Pp. 107--130.

Hashimoto, T. & Hasegawa, M. 1996. Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1α/Tu and 2/G. *Adv. Biophys.*, **32**:73--120.

Hashimoto, T., Nakamura, Y,. Kamaishi, T. & Hasegawa, M. 1997. Early evolution of eukaryotes inferred from protein phylogenies of translation elongation factors 1α and 2. *Arch. Protistenkd.*, **48**:287--295.

Hashimoto, T., Sanchéz, L. B., Shirakura, T., Müller, M. & Hasegawa, M. 1998. Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc. Natl. Acad. Sci. U S A*, **95**:6860--6865.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K. & Hasegawa, M. 1995a. Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences from elongation factor 2. *Mol. Biol. Evol.*, **12**:782--793.

Hashimoto, T., Nakamura, Y. Kamaishi, T, Adachi, J., Nakamura, F., Okamoto, K. & Hasegawa, M. 1995b. Phylogenetic place of kinetoplastid protozoa inferred from a protein phylogeny of elongation factor 1α. *Mol. Biochem. Parasitol.*, **70**:181--185.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. & Hasegawa, M. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.*, **11**:65--71.

Hinkle, G., Leipe, D. D., Nerad, T. A. & Sogin, M. L. 1994. The unusually long small subunit

ribosomal RNA of *Phreatamoeba balamuthi*. *Nucl. Acids Res.*, **22**:465--469.

Hirt, R. P., Logsdon, J. M. Jr., Healy, B., Dorey, M. W., Doolittle, W. F. & Embley, T. M. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. USA*, **96**:580--585.

Hirt, R. P., Healy, B., Vossbrinck, C. R., Canning, E.U. & Embley, T. M. 1997. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr. Biol.*, 7:995--998.

Ho, L. C., Armiugam, A., Jeyaseelan, K., Yap, E. H. & Singh, M. 2000. Blastocystis elongation factor-1[alpha]: genomic organization, taxonomy and phylogenetic relationships. *Parasitology*, **121**:135--144.

Holberton, D. V. & Marshall, J. 1995. Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Res.*, **23**:2945--2953.

Horie, K., Wada, A. & Fukutome, H. 1981. Conformational studies of *Escherichia coli* ribosomes with the use of acridine orange as a probe. *J. Biochem.*, **90**:449--461.

Horner, D. S. & Embley, T. M. 2001. Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes. *Mol. Biol. Evol.*, **18**:1970--1975.

Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. 1996. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. London, B*, **263**:1053--1059.

Hughey, R. & Krogh, A. 1996. Hidden Markov models for sequence analysis: extension and analysis of the basic method. *Comput. Appl. Biosci.*, **12**:95--107.

Ishihara, R. & Hayashi, Y. 1968. Some propaties of ribosomes from the sporoplasm of *Nosema bombycis*. *J. Invertebr. Pathol.*, **11**:377--385.

Iwabuchi, M. & Ochiai, H. 1969. Sedimentation properties of ribosomal particles in *Dictyostelium discoideum*. *Biochim. Biophys. Acta.*, **190**:211--213.

Johansen, S, Johansen, T. & Haugli, F. 1992. Structure and evolution of myxomycete nuclear group I introns: a model for horizontal transfer by intron homolog. *Curr. Genet.*, **22**:297--304.

Jones, D. T., Taylor, W. R. & Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **57**:94--97.

Kamaishi, T., Hashimoto, T., Nakamura, Y., Nakamura, F., Murata, S., Okada, N., Okamoto, K., Shimizu, M. & Hasegawa, M. 1996. Protein phylogeny of translation elongation

factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. *J. Mol. Evol.*, **42**:257-- 263.

Katinka, M. D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J. & Vivarès, C. P. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**:450--453.

Katz, L. A. 1998. Changing perspectives on the origin of eukaryotes. *Trends Ecol. Evol.*, **13**:493--497.

Keeling, P. J. & Doolittle, W. F. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.*, **13**:1297--1305.

Keeling, P. J., Fast, N. M. & McFadden, G. I. 1998. Evolutionary relationship between translation initiation factor eIF-2γ and selenocysteine-specific elongation factor SELB: change of function in translation factors. *J. Mol. Evol.*, **47**:649--655.

Keeling, P. J., Luker, M. A. & Palmer, J. D. 2000. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.*, **17**:23--31.

Keister D. B. 1983. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans. R Soc. Trop. Med. Hyg.*, 77:487--488.

Khaitovich, P., Mankin, A. S., Green, R., Lancaster, L. & Noller, H. F. 1999. Characterization of functionally active subribosomal particles from *Thermus aquaticus*. *Proc. Natl. Acad. Sci. USA*, **96**:85--90.

Kishino, H. & Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.*, **29**: 170--179.

Kishino, H., Miyata, T. & Hasegawa, M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.*, **30**:151--160.

Kohre, C., Mayer, C., Neumair, O., Grobner, P. & Piendl, W. 1998. Interaction of ribosomal L1 proteins from mesophilic and thermophilic Archaea and Bacteria with specific L1-binding sites on 23S rRNA and mRNA. *Eur. J. Biochem.*, **256**:97--105.

Kozak, M. 1987. An analysis of 5 -noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**:8125--8148.

Kubelik, A. R., Turcq, B. & Lambowitz, A. M. 1991. The Neurospora crassa cyt-20 gene encodes cytosolic and mitochondrial valyl-tRNA synthetases and may have a second function in addition to protein synthesis. *Mol. Cell. Biol.*, **11**:4022--4235.

Kuma, K., Nikoh, N., Iwabe, N. & Miyata, T. 1995. Phylogenetic position of *Dictyostelium* inferred from multiple protein data sets. *J. Mol. Evol.*, **41**:238--246.

Lecointre, G. & Le Guyader, H. 2001. Classification phylogenétique du vivant. Paris: Belin.

Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. 1993. Small subunit ribosomal RNA of *Hexamita inflata* and the quest of the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.*, **59**:41--48.

Li, J-Y. 1999. The primitive nucleus model and the origin of the cell nucleus. *Endocytobiosys. Cell. Res.*, **13**:1--86.

Lohia, A. & Samuelson, J. 1993. Cloning of the Eh cdc2 gene from *Entamoeba histolytica* encoding a protein kinase $p34^{cdc2}$ homologue. *Gene*, **127**:203--207.

Mager, W. H., Planta, R. J., Ballesta, J. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R. & Woolford, J. 1997. A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **25**:4872--4875.

Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R. & Samuelson, J. 1999. Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell. Biol.*, **19**:2198--2205.

Martin, W. F. & Müller, M. 1998. The hydrogen hypothesis of the first eukaryote. *Nature*, **392**:37--41.

Michel, S, Traut, R. R. & Lee, J. C. 1983. Yeast ribosomal proteins: electrophreeetic analysis in four two-dimensional gel system---correlation of nomenclatures. *Mol. Gen. Genet.*, **191**:251--256.

Milyutina, I. A., Aleshin, V. V., Mikrjukov, K. A., Kedrova, O. S. & Petrov, N. B. 2001. The unusually long small subunit ribosomal RNA gene found in amitochondriate amoeboflagellate *Pelomyxa palustris*: its rRNA predicted secondary structure and phylogenetic implication. *Gene*, **272**:131--139.

Mizumoto, K., Iwasaki, K., Tanaka, M. & Kaziro, Y. 1974. Studies on polypeptide elongation factor 2 from pig liver. I. Purification and properties. J. Biochem., **75**:104--1056.

Moras, D. 1992. Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.*, **17**:159--64.

Moreira, D., Le Guyader, H. & Philippe, H. 1999. Unusually high evolutionary rate of the elongation factor 1α genes from the ciliophora and its impact on the phylogeny of eukaryotes. *Mol. Biol. Evol.*, **16**:234--245.

Moreira, D., Le Guyader, H. & Philippe, H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* **405**:69--72.

Morrison, H. G., Roger, A. J., Nystul, T. G., Gillin, F. D. & Sogin, M. L. 2001. *Giardia lamblia* expresses a proteobacterial-like DnaK homolog. *Mol. Biol. Evol.*, **18**:530--541.

Müller, M. 1988. Energy metabolism of protozoa without mitochondoria. *Ann. Rev. Microbiol.*, **42**:465--488

Müller, M. 1993. The hydrogenosome. *J. Gen. Microbiol.*, **139**:2879--2889.

Müller, M. 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists -- a special case in eukaryotic evolution? *In*: Coombs, G. H., Vickerman, K., Sleigh, M. A. & Warren, A. (ed.), Evolutionary relationships among protozoa. Kluwer, Dordrecht. Pp. 109--131.

Musto, H., Romero, H., Zavala, A., Jabbari, K & Bernardi, G. 1999. Synonymous codon choices in the extremely GC-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. *J. Mol. Evol.*, **49**:27--35.

Nakamura, Y., Hashimoto, T., Kamaishi, T., Adachi, J., Nakamura, F., Okamoto, K. & Hasegawa, M. 1996a. Phylogenetic position of kinetoplastid protozoa inferred from the protein phylogenies of elongation factors 1alpha and 2. *J. Biochem.*, **119**:70--79.

Nakamura, Y., Hashimoto, T., Yoshikawa, H., Kamaishi, T., Nakamura, F., Okamoto, K. & Hasegawa, M. 1996b. Phylogenetic position of *Blastocystis hominis* that contains cytochrome-free mitochondria, inferred from the protein phylogeny of elongation factor 1α. *Mol. Biochem. Parasitol.*, **77**:241--245.

Neefs, J. M., Van de Peer, Y., De Rijk, P., Goris, A. & De Wachter, R. 1991. Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.*, **25**: 1987--2015.

Neefs, J. M., Van de Peer, Y., De Rijk, P., Chapelle, S. & De Wachter, R. 1993. Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res.*, **21**:3025--3049.

Nevskaya, N., Tischenko, S., Fedorov, R., Al-Karadaghi, S., Liljas, A., Kraft, A., Piendl W., Garber, M. & Nikonov, S. 2000. Archaeal ribosomal protein L1: the structure provides new insights into RNA binding of the L1 protein family., *Structure Fold. Des.*, **8**:363--371.

Nieuwenhuysen, P. & Clauwaert, J. 1981. Physicochemical characterization of ribosomal particles from the eukaryote *Artemia*. *J. Biol. Chem.*, **256**:9626--9632.

Niu, X. H., Hartshorne, T. & He, X. Y. & Agabian, N. 1994. Characterization of putative small nuclear RNAs from *Giardia lamblia*. *Mol. Biochem. Parasitol.*, **66**:49--57.

Oakes, M., Henderson, E., Scheinman, A., Clarke, M. & Lake J. A. 1985. Ribosome, structure, function, and evolution: mapping ribosomal RNA, proteins, and functional sites in three dimensions. *In*: Hardesty, B. & Kramer, G.(ed) Structure, function, and genetics of ribosomes., Springer-Verlag, New York, Pp. 47--67.

Ogawa, S., Yoshino, R., Angata, K., Iwamoto, M., Pi, M., Kuroe, K., Matsuo, K., Morio, T., Urushihara, H., Yanagisawa, K. & Tanaka, Y. 2000. The mitochondrial DNA of *Dictyostelium discoideum*: complete sequence, gene content and genome organization. *Mol. Gen. Genet.*, **263**:514--519.

Ohkuma, M., Ohtoko, K., Grunau, C., Moriya, S. & Kudo, T. 1998 Phylogenetic identification of the symbiotic hypermastigote *Trichonympha agilis* in the hindgut of the termite Reticulitermes speratus based on small-subunit rRNA sequence. *J. Eukaryot. Microbiol.*, **45**:439--444.

Oka, Y., Hayashi, H., Hayashi, N., Furuya, M., Ito, Y., Yagyu, M. & Osaki, H. 1973. Comparative studies on the isolation of ribosomes in *Trichomonas foetus*. *Jpn. J. Parasitol.*, **22**:193--198.

Olvera, J. & Wool, I. G. 1996. The primary structure of rat ribosomal protein L10a., *Biochem. Biophys. Res. Commun.* **220**:954--957.

Patnaik, P. K., Kulkarni, S. K. & Cross, G. A. 1993. Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. *EMBO J.*, **12**:2529--2538.

Patterson, D.J. 1989. Stramenopiles: chromophytes from a protistan perspective. *In*: Green, J.C., Leadbeater, B.S.C. & Diver, W.L. (ed.), The chromophyte algae: problems and perspectives. Clarendon Press, Oxford, Pp. 357--379.

Peattie, D. A., Alonso, R. A., Heim, A. & Caulfield, J. P. 1989. Ultrastructural localization of giardins to the edges of disk microribbons of *Giardia lamblia* and the nucleotide and deduced protein sequence of alpha giardin. *J. Cell Biol.*, **109**:2323--2335.

Petitjean, A., Bonneaud, N. & Lacroute, F. 1995. The duplicated *Saccharomyces cerevisiae* gene SSM1 encodes a eucaryotic homolog of the eubacterial and archaebacterial L1 ribosomal proteins., *Mol. Cell. Biol.* **15**:5071--5781.

Petridou, B., Cuny, M., Guerin, M. F. & Hayes, F. 1983. Ribosomal subunits and ribosomal proteins of *Tetrahymena thermophila*. Effect of the presence of iodoacetamide during ribosome extraction on the properties of the subunits. *Eur. J. Biochem.*, **135**:425--434.

Peyretaillade, E., Broussolle, V., Peyret, P., Metenier, G., Gouy, M. & Vivares, C. P. 1998a. Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Mol. Biol. Evol.*, **15**:683--689.

Peyretaillade, E., Biderre, C., Peyret, P., Duffieux, F., Metenier, G., Gouy, M., Michot, B. & Vivares, C. P. 1998b. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucl. Acids Res.*, **26**:3513--3520.

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

Philippe, H. & Adoutte, A. 1998. The molecular phylogeny of eukaryota: solid facts and uncertainties. *In*: Coombs,G., Vickerman, K., Sleigh, M. & Warren, A.(ed) Evolutionary relationships among protozoa. Chapman & Hall, London, Pp. 25--56.

Philippe, H., & Germot, A. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.*, **175**:830--834.

Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M. and Le Guyader, H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. B Biol. Sci.*, **267**(1449):1213--21.

Plaimauer, B., Ortner, S., Wiedermann, G., Scheiner, O. & Duchene, M. 1994. An intron-containing gene coding for a novel 39-kilodalton antigen of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.*, **66**:181--185.

Quon, D. V., Delgadillo, M. G. & Johnson, P. J. 1996. Transcription in the early diverging eukaryote *Trichomonas vaginalis*: an unusual RNA polymerase II and α-amanitin-resistant transcription of protein-coding genes. *J. Mol. Evol.*, **43**:253--262.

Ramagopal, S. & Ennis, H. L. 1980. Studies on ribosomal proteins in the cellular slime mold *Dictyostelium discoideum*. Resolution, nomenclature and molecular weights of proteins in the 40-S and 60-S ribosomal subunits. *Nucleic Acids Res.*, **25**:4872--4875.

Reeves, R. E. 1984. Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv. Parasitol.*, **23**:105--142.

Riley, D. E. & Krieger, J. N. 1995. Molecular and phylogenetic analysis of PCR-amplified cyclin-dependent kinase (CDK) family sequences from representatives of the earliest available lineages of eukaryotes. *J. Mol. Evol.*, **41**:407--413.

Roger, A. J. 1999. Reconstructing early events in eukaryotic evolution. *Am. Natr.*, **154**:S146--S163.

Roger, A. J., Sandblom, O., Doolittle, W. F. & Philippe, H. 1999. An evaluation of elongation

factor 1α as a phylogenetic marker for eukaryotes. *Mol. Biol. Evol.*, **16**:218--233.

Roger, A. J., Svard, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. & Sogin, M. L. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. U S A*, **95**:229--234.

Roger, A. J., Clark, C. G. & Doolittle, W. F. 1996. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc. Natl. Acad. Sci. U S A*, **93**:14618--14622.

Rosorius, O., Fries, B., Stauber, R. H., Hirschmann, N., Bevec, D. & Hauber, J. 2000. Human ribosomal protein L5 contains defined nuclear localization and export signals. *Biol. Chem.*, **275**:12061--12068.

Rotte, C., Henze, K., Müller, M. & Martin, W. 2000. Origins of hydrogenosomes and mitochondria. *Curr. Op. Microbiol.*, **3**:481--486.

Sambrook, J. Fritch, E. F. & Maniatis, T. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Seravin, L. N. & Goodkov, A. V. 1987. Microbody-like cytoplasmic granules in the amoeba *Pelomyxa palustris*. *Tsitologiia*, **29**:600--603.

Shimodaira, H. 2002. Assessing the uncertainty of the cluster analysis using the bootstrap resampling. *Proc. Inst. Stat. Math.* **50**, in press (in Japanese with English summary).

Shimodaira, H. & Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.*, **16**:1114--1116.

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

Shirakura, T., Hashimoto, T., Nakamura, Y., Kamaishi, T., Cao, Y., Adachi, J., Hasegawa, M., Yamamoto, A. & Goto, N. 1994 Phylogenetic place of a mitochondria-lacking protozoan, *Entamoeba histolytica*, inferred from amino acid sequences of elongation factor 2. *Jpn. J. Genet.*, **69**:119--135.

Silberman, J. D., Sogin, M. L., Leipe, D. D. & Clark, C. G. 1996. Human parasite fines taxonomic home. *Nature*, **380**:398.

Silberman, J. D., Clark, C. G., Diamond, L. S. & Sogin, M. L. 1999. Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small-subunit ribosomal RNA sequences. *Mol. Biol. Evol.*, **16**:1740--1751.

Simpson, A. G. B., Bernard, C., Fenchel, T. & Patterson, D. J. 1997. The organisation of

Mastigamoeba schizophrenia n.sp.: More evidence of ultrastructural idiosyncrasy and simplicity of pelobiont protists. *Europ. J. Protistol.*, **33**:87--98.

Spirin A. S. 1999. Ribosomes. Kluwer Academic/Plenum Publishers, New York, Pp. 75.

Sogin, M. L. 1991. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.*, **1**:457--463.

Sogin, M. L. & Silberman, J. D. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int. J. Parasitol.*, **28**:11--20.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**:75--77.

Stenzel, D. J. & Boreham, P. F. L. 1996. *Blastocystis hominis* revisited. *Clin. Microbiol. Rev.*, **9**:563--584.

Stiller, J. W. & Hall, B. D. 1994. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Eur. J. Biochem.*, **226**:853--859.

Stiller, J. W., Duffield, E. C. & Hall B. D. 1998. Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc. Natl. Acad. Sci. U S A*, **95**:11769--11774.

Stoffler-Meilicke, M. & Stoffler, G. 1990. Topography of the ribosomal proteins from *Escherichia coli* within the intact subunits as determined by immunoelectron microscopy and protein-protein crosslinking. *In*: Hill, W. E., Dahlberg, A., Garrett, R. A., Moor, P. B. Schlessinger, D. & Warner, J. R. (ed) The ribosome: structure, function, and evolution. ASM press, Washington DC, Pp.123--133.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. 1996. Phylogenetic inference. *In*: Hillis, D., Moritz, C. & Mable, B. (ed), Molecular Systematics, Sinaur, Pp. 407--514.

Tachezy, J., Sánchez, L. B. & Müller, M. 2001. Mitochondrial Type Iron-Sulfur Cluster Assembly in the Amitochondriate Eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as Indicated by the Phylogeny of *IscS*. *Mol. Biol. Evol.*, **18**:1919--1928.

Tovar, J., Fischer, A. & Clark, C. G. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.*, **32**:1013--1021.

Urban, B., Blasig, C., Forster, B., Hamelmann, C. & Horstmann, R. D. 1996. Putative serine/threonine protein kinase expressed in complement-resistant forms of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.*, **80**:171--178.

Urlaub, H., Kruft, V., Bischof, O., Muller, E.C. & Wittmann-Liebold, B. 1995. Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.*, **14**:4578--4588.

Van de Peer, Y. & De Wachter, R. 1997. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J. Mol. Evol.*, **45**:619--630.

Van de Peer, Y., Chapelle, S. & De Wachter, R. 1996. The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. *J. Mol. Evol.*, **42**:201--210.

Van de Peer, Y., Ben Ali, A. & Meyer, A. 2000. Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene*, **246**:1--8.

Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. 2000. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J. Mol. Evol.*, **51**:565--576.

van Keulen, H., Gutell, R. R., Campbell, S. R., Erlandsen, S. L. & Jarroll, E. L. 1992. The nucleotide sequence of the entire ribosomal DNA operon and the structure of the large subunit rRNA of *Giardia muris*. *J. Mol. Evol.*, **35**:318--328.

Vossbrinck, C. R. & Woose, C. R. 1986. Eukaryotic ribosomes that lack a 5.8S RNA. *Nature*, **320**:287--288.

Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. 1987. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature*, **326**:411--414.

Wada, A. 1986a. Analysis of *Eschelichia coli* ribosomal proteins by an improved two dimensional gel electrophoresis. I. Detection of four new proteins. *J. Biochem.*, **100**:1583--1594.

Wada, A. 1986b. Analysis of *Eschelichia coli* ribosomal proteins by an improved two dimensional gel electrophoresis. II. Detection of four new proteins. *J. Biochem.*, **100**:1595--1605.

Walker, G., Simpson, A. G. B., Edgcomb, V., Sogin, M. L. & Patterson, D. J. 2001. Ultrastructural identities of *Mastigamoeba punctachora, Mastigamoeba simplex* and *Mastigella commutans* and assessment of hypotheses of relatedness of the pelobionts (Protista). *Europ. J. Protistol.*, **37**:25--49.

Weiss, L. M. & Vossbrinck, C. R. 1999. Molecular biology, molecular phylogeny, and

molecular diagnostic approaches to the microsporidia. *In*: Wittner, M. & Weiss, L. M. (ed) The Microsporidia and Microsporidiosis. ASM press, Washington DC.

Whatley, J. M., John, P. & Whatley, F. R. 1979. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *Proc. R. Soc. London, B*, **204**:165--187.

White, S. W., Clemons-Jr, W. M., Davis, C., Ramakrishnan, V. & Wimberly, B. T. 2000. Structures of the architecture and mechanism of the ribosome. *In*: Garrett, R. A., Douthwaite, S. R., Liljas, A., Matheson, A. T., Moor, P. B. & Noller H. F. (ed) The ribosome--structure, function, antibiotics, and cellular interactions. ASM press, Washington DC, Pp. 73--83.

Wilihoeft, U., Campos-Gongora, E., Touzni, S., Bruchhaus, I. & Tannich, E. 2001. Introns of *Entamoeba histolytica* and *Entamoeba dispar*. *Protist*, **152**:149--156.

Wool, I. G. 1979. The structure and function of eukaryotic ribosomes. *Ann. Rev. Biochoem*, **48**:719--754.

Wool, I. G., Chan, Y. L. & Gluck, A. 1995. Structure and evolution of mammalian ribosomal proteins. *Biochem Cell. Biol.*, **73**:933--947.

Wu, G. & Hashimoto, T. 1999. Sequence analysis of genes encoding ribosomal proteins of amitochondriate protists: L1 of *Trichomonas vaginalis* and L29 of *Giardia lamblia*. *Parasitol. Int.*, **48**:135--134.

Yamamoto, A. Hashimoto, T., Asaga, E, Hasegawa, M. & Goto, N. 1997. Phylogenetic position of mitochondorion-lacking protozoan, *Trichomonas tenax*, based on amino acid sequences of elongation factors 1α and 2. *J. Mol. Evol.*, **44**:98-105.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.*, **11**:367--372.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **13**:555—556.

Yoshikawa, H., Kuwayama, N. & Enose, Y. 1995. Histochemical detection of carbohydrates of *Blastocystis hominis*. *J. Eukaryot. Microbiol.*, **42**:70--74.

Zasloff, M. & Ochoa, S. 1971. A supernatant factor involved in initiation complex formation with eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U S A*, **65**:3059--3063.

Zettler, L. A., Sogin, M. L. & Caron, D. A. 1997. Phylogenetic relationships between the Acantharea and the Polycystinea: a molecular perspective on Haeckel's Radiolaria. *Proc. Natl. Acad. Sci. USA*, **94**:11411--11416.

Zhang, K., Fujioka, H., Lobo, C. A., Kitayaporn, D., Aikawa, M. & Kumar, N. 1999. Cloning and characterization of a new asparagine-rich protein in *Plasmodium falciparum*. *Parasitol. Res.*, **85**:956--963.

Zierdt, C. H. 1973. Studies of *Blastocystis hominis*. *J. Protozool.*, **20**:114--121.

Zierdt, C. H. 1991. *Blastocystis hominis* ----- past and future. *Clin. Microbiol. Rev.*, **4**:61--79.

Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecules. *Science*, **244**:48--52.

Table I. Products of the mitochondorial symbiosis present in amitochondriate taxa.

Organisms	Mitochondorion-derived genes found to date	Reference
Entamoebidae		
Entamoeba histolytica	CPN60	Clark & Roger 1995
	HSP70mit	Bakatselou, Kidgell & Clark 2000 Arisue et al. 2002
Palabasala		
Trichomonas vaginalis	CPN60	Bui, Bradley & Johnson 1996 Roger, Clark & Doolittle 1996 Horner et al. 1996
	CPN10	Bui, Bradley & Johnson 1996
	HSP70mit	Germot, Philippe & Le Guyader 1996
	ValRS	Hashimoto et al. 1998
	iscS	Tachezy, Sánchez & Müller 2001
	AAC	Dyall et al. 2001
Diplomonadida		
Giardia intestinalis	CPN60	Roger et al. 1998
	ValRS	Hashimoto et al. 1998
	HSP70mit	Morrison et al. 2001, Arisue et al. 2002
	iscS	Tachezy, Sánchez & Müller 2001
Microsporidia		
Nosema locustae	HSP70mit	Germot, Philippe & Le Guyader 1997
Vairmorpha necatrix	HSP70mit	Hirt, Healy & Vossbrinck 1997
Encephalitozoon cuniculi	HSP70mit	Peyretaillade et al. 1998a
Encephalitozoon hellem	HSP70mit	Arisue et al. 2002
Glugea plecoglossi	HSP70mit	Arisue et al. 2002

Table 1.1. Genes cloned and sequenced for the analysis described in §1.1.

Organisms	Genes
Blastocystis hominis	SSUrRNA, EF-2, v-ATPB, HSP70c,
Plasmodium falciparum	EF-2, ValRS, IleRS
Entamoeba histolytica	ValRS, IleRS

Table 1.2. Genes used for the combined analysis described in §1.1.

Relationships examined	Genes
Microsporidia/Fungi	SSUrRNA, LSUrRNA, EF-1α, EF-2, RpoII, ValRS, IleRS, Actin, α-tubulin, β-tubulin
Stramenopiles/Alveolata	SSUrRNA, EF-1α, EF-2, HSP70c v-ATPB, Actin, β-tubulin

Table 1.3. Phylogenetic relationships among seven major eukaryotic lineages including Microsporidia inferred from the combined analysis of ten individual genes based on the maximum likelihood (ML) method (see Fig. 1.2).[a]

	without Γ[c]			with Γ[d]			The ML tree
Tree topology[b]	Δl_i[e]	KH[f]	SH[g]	Δl_i[e]	KH[f]	SH[g]	of
1 (Al,E,(P,(S,(A,(M,F)))))	(-92159.9)	-	-	(-87559.0)	-	-	
2 (Al,P,(E,(S,(A,(M,F)))))	27.8	0.148	0.821	15.1	0.214	0.859	α-, β-tubulin
3 (Al,E,(S,(P,(A,(M,F)))))	29.0	0.107	0.801	31.4	0.022	0.648	
4 (Al,P,((E,S),(A,(M,F))))	30.0	0.192	0.799	41.5	0.042	0.509	
5 (Al,(P,E),(S,((M,F),A)))	30.7	0.115	0.804	19.5	0.138	0.821	
6 (Al,E,((P,S),(A,(M,F))))	35.6	0.062	0.739	23.2	0.085	0.760	
7 (Al,S,((E,P),(A,(M,F))))	42.4	0.109	0.688	38.0	0.061	0.556	
8 (Al,(S,(P,E)),(A,(M,F)))	48.2	0.087	0.603	36.0	0.079	0.576	
9 (Al,(P,(E,S)),(A,(M,F)))	61.8	0.042	0.464	66.0	0.004	0.236	
10 (Al,P,(S,(E,(A,(M,F)))))	68.1	0.019	0.414	44.5	0.034	0.470	
11 (Al,(E,S),(P,((M,F),A)))	72.2	0.017	0.368	74.4	0.001	0.172	
12 (Al,(E,(P,S)),(A,(M,F)))	91.0	0.004	0.233	65.0	0.006	0.249	
13 (Al,S,(P,(E,(A,(M,F)))))	95.0	0.003	0.196	70.4	0.002	0.198	
14 (Al,S,(E,(P,(A,(M,F)))))	97.6	0.002	0.186	77.3	0.000	0.152	
15 (Al,(P,E),(S,(M,(F,A))))	107.8	0.001	0.141	76.2	0.002	0.163	
16 (Al,P,((E,S),(M,(A,F))))	112.6	0.005	0.122	101.4	0.001	0.056	
17 (Al,P,((E,S),(F,(A,M))))	113.3	0.004	0.127	104.9	0.000	0.049	
18 (Al,(S,(P,E)),(M,(F,A)))	120.2	0.003	0.099	92.4	0.002	0.089	
19 (Al,S,((E,P),(M,(A,F))))	122.9	0.002	0.083	97.2	0.001	0.065	
20 (Al,S,((E,P),(F,(A,M))))	127.3	0.001	0.074	102.1	0.000	0.054	
21 (Al,(P,(E,S)),(M,(F,A)))	130.1	0.002	0.074	120.5	0.000	0.022	
22 (Al,M,(P,(E,(S,(A,F)))))	203.8	0.000	0.002	193.4	0.000	0.000	EF1α
23 (Al,(P,E),(A,((M,F),S)))	229.7	0.000	0.001	203.3	0.000	0.000	EF2
24 (Al,(P,(E,M)),(S,(F,A)))	248.1	0.000	0.000	219.7	0.000	0.000	Actin
25 (Al,E,((M,F),(A,(P,S))))	262.4	0.000	0.000	233.6	0.000	0.000	ValRS
26 (Al,(P,(F,A)),(E,(M,S)))	271.3	0.000	0.000	254.8	0.000	0.000	SSUrRNA
27 (Al,E,(A,((P,S),(M,F))))	289.3	0.000	0.000	249.7	0.000	0.000	RpoII
28 (Al,M,(P,(S,(A,(E,F)))))	437.2	0.000	0.000	368.4	0.000	0.000	IleRS
29 (Al,(M,S),(F,(P,(E,A))))	527.6	0.000	0.000	446.7	0.000	0.000	LSUrRNA

[a] Of the 945 tree topologies possible for the seven lineages, only 29 are shown

[b] Al, Alveolata; E, Euglenozoa; P, Viridiplantae; S, Stramenopiles; A, Metazoa; M, Microsporidia; F, Fungi.

[c] ML analysis without consideration of among-site rate heterogeneity.

[d] ML analysis that takes among-site rate heterogeneity into consideration by the use of Γ distribution.

[e] Log-likelihood difference of alternative tree from that of the best tree. The log-likelihood value of the best tree is shown in parentheses.

[f] *p*-value for the test by Kishino and Hasegawa (1989).

[g] *p*-value for the test by Shimodaira and Hasegawa (1999) with multiple comparison correction.

Table 1.4. Phylogenetic relationships among seven major eukaryotic lineages inferred from the combined analysis of eight individual genes based on the maximum likelihood (ML) method (see Fig. 1.3).[a]

Tree topology[b]	without Γ[c]			with Γ[d]			The ML tree of
	Δl_i[e]	KH[f]	SH[g]	Δl_i[e]	KH[f]	SH[g]	
1. ((A,F),M,((Al,S),(P,E)))	(-76015.5)	-	-	(-70683.6)	-	-	EF-1α
2. ((A,F),M,(E,(P,(Al,S))))	-26.1	0.139	0.774	-18.0	0.144	0.774	β-tubulin
3. ((A,F),(M,E),(P,(Al,S)))	-27.7	0.196	0.752	-43.3	0.024	0.257	LSUrRNA
4. ((A,F),M,(P,(E,(Al,S))))	-49.5	0.010	0.439	-28.1	0.030	0.562	Actin
5. ((A,F),E,(M,(P,(Al,S))))	-55.0	0.046	0.347	-37.4	0.049	0.360	
6. ((A,F),(Al,S),(M,(P,E)))	-55.2	0.005	0.378	-32.8	0.010	0.470	
7. ((A,F),(P,E),(M,(Al,S)))	-56.6	0.003	0.361	-35.0	0.005	0.432	
8. (A,(F,M),((P,E),(Al,S)))	-68.1	0.007	0.271	-60.6	0.001	0.134	v-ATPB
9. ((A,F),(Al,S),(P,(M,E)))	-74.4	0.011	0.187	-67.3	0.001	0.068	
10. (A,(F,M),(E,(P,(Al,S))))	-86.6	0.007	0.142	-63.7	0.010	0.123	
11. ((A,F),P,((M,E),(Al,S)))	-94.8	0.003	0.078	-79.1	0.000	0.029	
12. ((A,F),E,((M,P),(Al,S)))	-101.2	0.002	0.064	-62.4	0.005	0.103	
13. ((A,F),M,(E,(S,(P,Al))))	-107.8	0.002	0.052	-69.0	0.003	0.074	
14. ((A,F),(M,E),(S,(P,Al)))	-111.9	0.004	0.043	-98.1	0.000	0.008	
15. ((A,F),(M,E),(Al,(P,S)))	-122.4	0.002	0.025	-101.2	0.000	0.008	
16. ((A,(F,E)),M,(P,(Al,S)))	-135.3	0.001	0.020	-86.3	0.002	0.029	HSP70c
17. ((A,F),P,(S,(Al,(M,E))))	-182.7	0.000	0.001	-140.1	0.000	0.000	SSUrRNA
18. ((A,(F,S)),M,(Al,(P,E)))	-311.5	0.000	0.000	-217.8	0.000	0.000	EF-2

[a] Of the 945 tree topologies possible for the seven lineages, only 18 are shown

[b] A, Metazoa; F, Fungi; M, Mycetozoa; Al, Alveolata; S, Stramenopiles; P, Viridiplantae; E, Euglenozoa.

[c] ML analysis without consideration of among-site rate heterogeneity.

[d] ML analysis that takes among-site rate heterogeneity into consideration by the use of Γ distribution.

[e] Log-likelihood difference of alternative tree from that of the best tree. The log-likelihood value of the best tree is shown in parentheses.

[f] *p*-value for the test by Kishino and Hasegawa (1989).

[g] *p*-value for the test by Shimodaira and Hasegawa (1999) with multiple comparison correction.

Table 1.5. Comparison of the alternative tree topologies for the phylogenetic relationships among seven eukaryotic lineages (see Fig. 1.7)[a].

	Total (3935 positions)				SSU+LSU (2852 positions)			
	without Γ[c]		with Γ[d]		without Γ		with Γ	
Tree topology[b]	Δl_i[e]	p-value[f]	Δl_i	p-value	Δl_i	p-value	Δl_i	p-value
1. (((Mb,Eh),M),E,(MF,(VR,AS)))	(-54631.3)	-	-11.7	0.535	-8.1	0.679	-1.5	0.810
2. (((Mb,Eh),M),MF,(AS,(VR,E)))	-8.5	0.790	(-50727.4)	-	-30.7	0.155	-2.7	0.733
3. (((Mb,Eh),M),E,(AS,(VR,MF)))	-15.1	0.618	-19.9	0.334	-13.5	0.493	-1.5	0.730
4. (((Mb,Eh),E),M,(MF,(VR,AS)))	-25.7	0.445	-26.5	0.216	(-33459.9)	-	(-30684.0)	-
5. ((Mb,M),MF,((Eh,AS),(E,VR)))	-71.5	0.071	-16.4	0.426	-103.4	0.000	-21.2	0.052
6. ((Mb,(Eh,MF)),M,(AS,(VR,E)))	-132.4	0.001	-34.0	0.123	-145.4	0.000	-28.6	0.018

	EF-1α (386 positions)				EF-2 (697 positions)			
	without Γ		with Γ		without Γ		with Γ	
Tree topology	Δl_i	p-value	Δl_i	p-value	Δl_i	p-value	Δl_i	p-value
1. (((Mb,Eh),M),E,(MF,(VR,AS)))	-31.2	0.105	-20.4	0.119	-14.9	0.397	-10.0	0.405
2. (((Mb,Eh),M),MF,(AS,(VR,E)))	-25.5	0.165	-15.3	0.223	-6.4	0.725	-2.3	0.811
3. (((Mb,Eh),M),E,(AS,(VR,MF)))	-38.9	0.053	-25.7	0.063	-16.9	0.359	-12.9	0.327
4. (((Mb,Eh),E),M,(MF,(VR,AS)))	-53.0	0.016	-33.4	0.025	-26.9	0.168	-13.4	0.310
5. ((Mb,M),MF,((Eh,AS),(E,VR)))	(-6226.1)	-	(-5817.4)	-	-22.2	0.263	-15.4	0.251
6. ((Mb,(Eh,MF)),M,(AS,(VR,E)))	-41.2	0.045	-25.6	0.055	(-14891.1)	-	(-14205.8)	-

[a] Of the 945 tree topologies possible for the seven lineages, only six are shown. For details also see text.

[b] Mb, *Mastigamoeba balamuthi*; Eh, *Entamoeba histolytica*; M, Mycetozoa; E, Euglenozoa; MF, Metazoa and Fungi; VR, Viridiplantae/Rhodophyta (Rhodophyta are included in the EF-2 data set); AS, Alveolata/stramenopiles (Stramenopiles are included in the SSU+LSU data set. Only apicomplexa are used in the EF-1α data set). The relationships within each lineage were assumed in advance according to the figures 1.4, 1.5, and 1.6.

[c] ML analysis without consideration of among-site rate heterogeneity.

[d] ML analysis that takes among-site rate heterogeneity into consideration by the use of Γ-distribution. The Γ shape parameters (α) estimated from data for these trees were approximately 0.37, 0.41, and 0.74 for the SSU+LSU, EF-1α, and EF-2 data sets, respectively.

[e] Log-likelihood difference of an alternative tree from that of the best tree. The log-likelihood value of the best tree is shown in parentheses.

[f] p-value for the test by Shimodaira and Hasegawa (1999) with multiple comparison correction.

Table 1.6. Genes cloned and sequenced for the analysis described in §1.3.

Organism	Genes
Giardia intestinalis	HSP90c, RpoII, RpS14, RpS15a, RpL5, RpL8, RpL10a
Trichomonas vaginalis	IleRS, HSP90c, RpS14, RpS15a, RpL5, RpL8
Entamoeba histolytica	HSP90c, RpoII

Table 1.7. Genes used for the combined analysis described in §1.3.

	Genes
Translation:	SSUrRNA, LSUrRNA, EF-1α, EF-2, RpS14, RpS15a, RpL5, RpL8, RpL10a, IleRS, ValRS
Transcription:	RpoII
Chaperon:	CPN60, HSP70c, HSP70mit, HSP70er HSP90c, CCTα, CCTγ, CCTδ, CCTζ
Cytoskelton:	Actin, α-tubulin, β-tubulin

Table 1.8. Phylogenetic relationships among seven major eukaryotic lineages and an outgroup inferred from the combined analysis of 24 individual genes based on the maximum likelihood (ML) method. (see Fig. 1.10).[a]

Tree Topology[b]	without Γ[c]				with Γ[d]			
	Δl_i[e]	KH[f]	SH[d]	AU[h]	Δl_i[e]	KH[f]	SH[d]	AU[h]
1. (O,(D,(P,(E,(AS,(VR,(C,MF))))))))	(-295739.9)	0.510	0.997	0.610	89.3	0.023	0.366	0.068
2. (O,(D,(P,(E,(C,(MF,(AS,VR))))))))	1.1	0.490	0.998	0.634	110.7	0.011	0.183	0.046
3. (O,(D,(P,(MF,(C,(VR,(AS,E))))))))	2.3	0.482	0.998	0.634	(-279603.2)	0.853	1.000	0.973
4. (O,(D,(P,(E,((AS,VR),(C,MF))))))	9.5	0.392	0.996	0.400	86.8	0.027	0.385	0.128
5. (O,(D,(P,(C,(MF,(VR,(AS,E))))))))	41.1	0.126	0.927	0.133	53.1	0.024	0.739	0.063
6. (O,((D,P),(E,(AS,(VR,(MF,C))))))	62.1	0.017	0.830	0.043	103.9	0.015	0.244	0.074
7. (O,(D,(P,(MF,(C,(E,(AS,VR))))))))	65.3	0.156	0.815	0.027	35.5	0.088	0.881	0.088
8. (O,((D,P),(E,((MF,C),(VR,AS)))))	74.5	0.045	0.773	0.030	100.3	0.021	0.263	0.125
9. (O,((D,P),(MF,(C,(VR,(AS,E))))))	84.7	0.106	0.672	0.020	22.0	0.147	0.951	0.095
10. (O,(P,(D,(E,((AS,VR),(MF,C))))))	90.8	0.018	0.650	0.017	107.8	0.016	0.207	0.065
11. (O,(P,(D,(MF,(C,(VR,(AS,E))))))))	97.2	0.075	0.581	0.003	25.8	0.109	0.921	0.259
12. (O,((D,P),(MF,(C,(E,(VR,AS))))))	146.9	0.019	0.263	0.012	56.1	0.051	0.679	0.142
13. (O,(P,(D,(MF,(C,(E,(AS,VR))))))))	158.7	0.012	0.213	0.015	59.9	0.037	0.654	0.084
14. (O,((D,P),(C,(MF,(E,(VR,AS))))))	208.0	0.001	0.055	0.000	119.5	0.003	0.145	0.057
15. (O,(D,(P,((MF,C),(E,(AS,VR))))))	238.4	0.000	0.019	0.054	126.2	0.002	0.115	0.033
16. (O,(E,(AS,(VR,(C,(MF,(D,P))))))))	255.3	0.000	0.016	0.006	73.8	0.041	0.520	0.101
17. (O,(P,(D,((C,E),(MF,(AS,VR))))))	255.5	0.000	0.008	0.000	195.6	0.000	0.003	0.055
18. (O,(E,((AS,(VR,(C,(MF,(D,P))))))))	271.3	0.000	0.008	0.005	87.5	0.003	0.385	0.121

[a] Of the 10,395 tree topologies possible for the seven lineages, only 18 are shown

[b] O, Outgroup; D, Diplomonadida; P, Parabasala; E, Euglenozoa; VR, Viridiplantae and Rhodophyta; AS, Alveolata and stramenopiles; C, Conosa; MF, Metazoa, Fungi and Microsporidia.

[c] ML analysis without consideration of among-site rate heterogeneity.

[d] ML analysis that takes among-site rate heterogeneity into consideration by the use of Γ distribution.

[e] Log-likelihood difference of alternative tree from that of the best tree. The log-likelihood value of the best tree is shown in parentheses.

[f] *p*-value for the test by Kishino and Hasegawa (1989).

[g] *p*-value for the test by Shimodaira and Hasegawa (1999) with multiple comparison correction.

[h] *p*-value for the approximately unbiased test (Shimodaira, 2002).

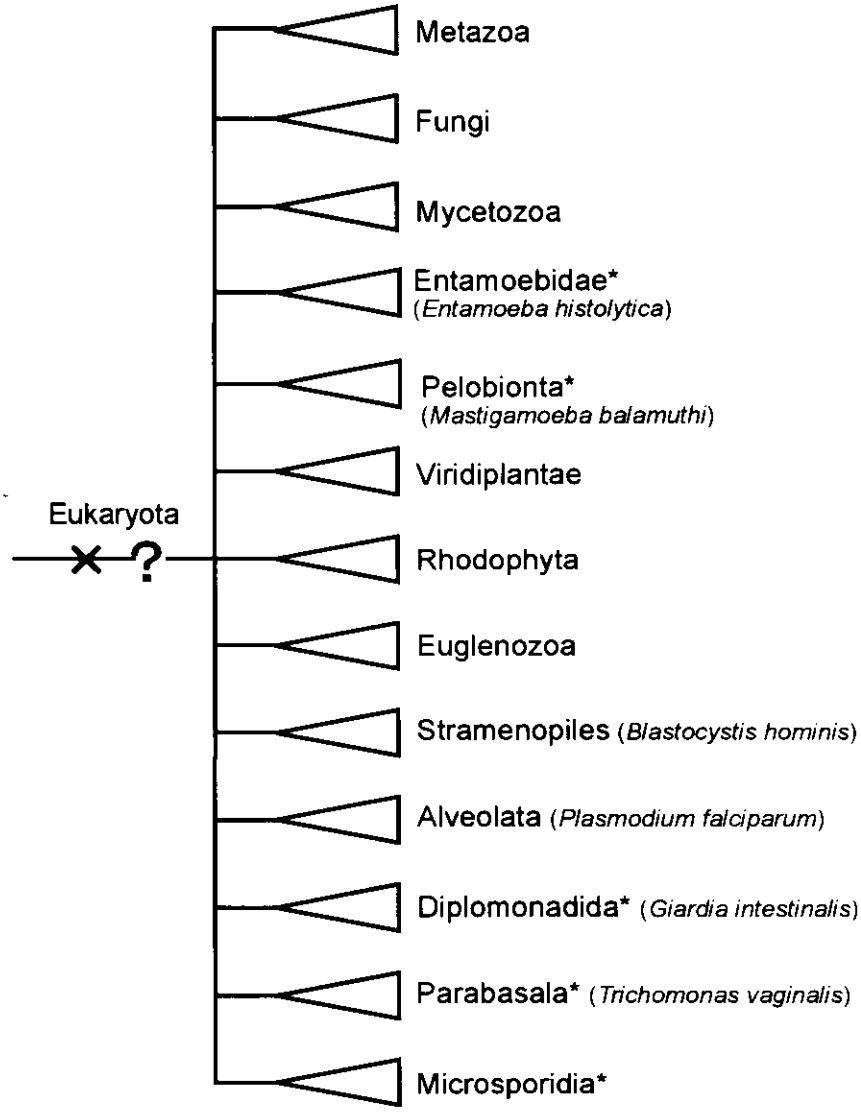


Fig. 1.1. Major eukaryotic lineages, the relationships among which were examined in this work. Organisms used for cloning and sequencing studies are shown in parentheses. Astarisk (*) denotes amitochondoriate protist groups.

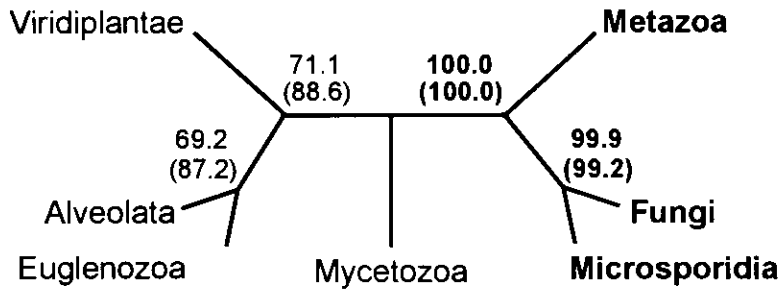


Fig. 1.2. Schematic representation of the phylogenetic relationships among seven eukaryotic lineages inferred from a combined analysis of individual genes. The best tree (Tree 1 in Table 1.3) is shown. Bootstrap proportions (BPs) calculated among 29 alternatives in Table 1.3 are attached to the internal branches for the analysis without Γ-distribution. BPs for the analysis with Γ- distribution are shown in parentheses.

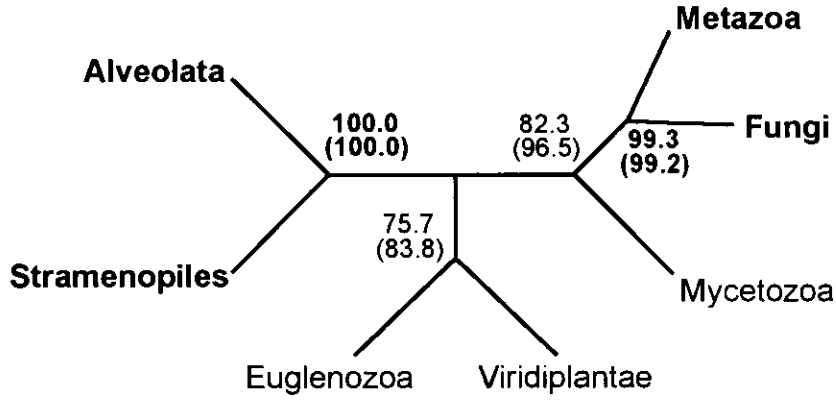


Fig. 1.3. Schematic representation of the phylogenetic relationships among seven eukaryotic lineages inferred from a combined analysis of individual genes. The best tree (Tree 1 in Table 1.4) is shown. Bootstrap proportions (BPs) calculated among 18 alternatives in Table 1.4 are attached to the internal branches for the analysis without Γ-distribution. BPs for the analysis with Γ-distribution are shown in parentheses.

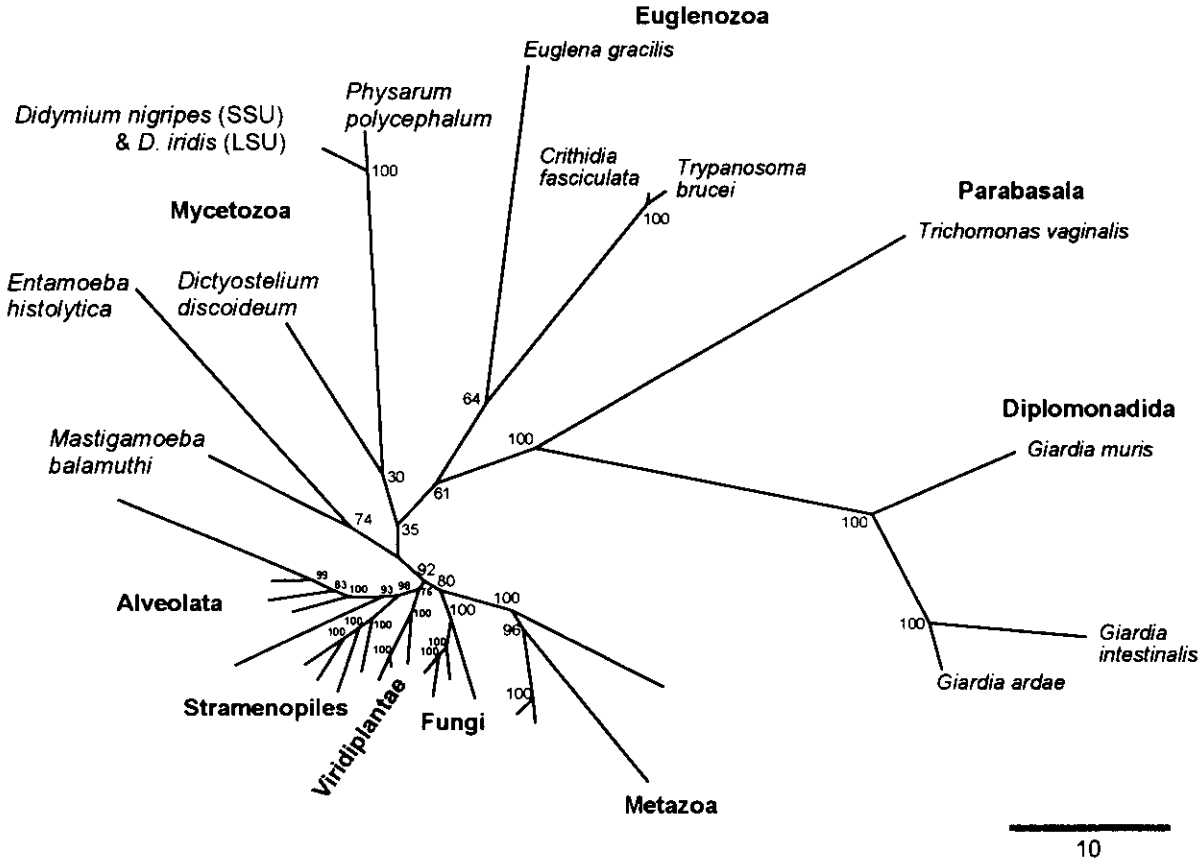


Fig. 1.4. Unrooted eukaryotic tree based on the concatenated small subunit (SSU) and large subunit (LSU) ribosomal RNAs (rRNAs). The best tree of the ML method is shown. Bootstrap proportions (BPs) are shown at the internal branches. The length of each branch is proportional to the estimated number of substitutions. Bar denotes 10% substitutions per site. For details of the phylogenetic reconstruction methods see text. Species names are shown only for taxa where space permits it. Species names of the other taxa and the accession numbers of the sequences used are: Metazoa, (*Drosophila melanogaster* M21017, (*Caenorhabditis elegans* X03680, (*Xenopus laevis* X02995, *Homo sapiens* U13369))); Fungi, (*Blastocladiella emersonii* X54264[SSU]/ X90410[LSU], (*Pneumocystis carinii* L27658/M86760, (*Schizosaccharomyces pombe* Z19578, *Saccharomyces cerevisiae* V01335))); Viridiplantae, (*Chlorella ellipsoidea* D13324/D17810, (*Oryza sativa* X00755/M16845, *Arabidopsis thaliana* X16077/X52320)); Stramenopiles, (*Phytophthora megasperma* X54265/X75632, (*Skeletonema pseudocostatum* X85394/Y11512, (*Tribonema aequale* M55286/Y07978, *Ochromonas danica* M32704/Y07976))); Alveolata, (*Prorocentrum micans* M14649, (*Cryptosporidium parvum* X64340/AF040725, (*Toxoplasma gondii* X75429, *Plasmodium falciparum* M19172/U21939))); *Mastigamoeba balamuthi* L23799/ AB066363-5 (this work); *Entamoeba histolytica* X65163; Mycetozoa, *Dictyostelium discoideum* K02641/ X00601, *Didymium nigripes* and *Didymium iridis* AF239230/X60210, *Physarum polycephalum* X13160/ V01159; Euglenozoa, *Euglena gracilis* M12677/X53361, *Crithidia fasciculata* Y00055, *Trypanosoma brucei* M12676/X14553; Parabasala, *Trichomonas vaginalis* U17510/AF202181; Diplomonadida, *Giardia muris* X65063, *Giardia intestinalis* X52949, *Giardia aldeae* Z17210/X58290. *Didymium nigripes* (SSU) and *Didymium iridis* (LSU) data were combined together, because each species clearly showed a close relationship to an another myxogastrid, *Physarum polycephalum* in the separate SSU and LSU analyses. Unambiguously aligned 2,852 (1,103 from SSU and 1,749 from LSU) nucleotide positions were used for the analysis.

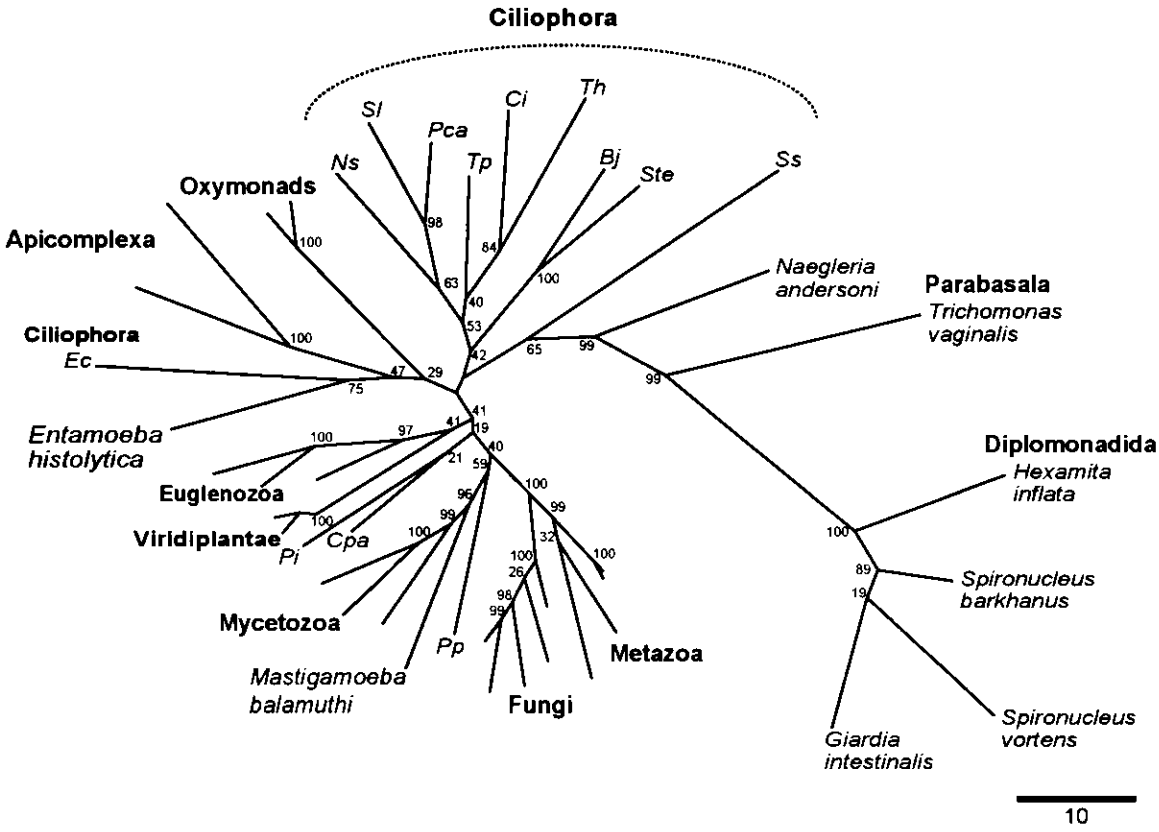


Fig. 1.5. Unrooted eukaryotic tree of EF-1α. The best tree of the ML method is shown. Bootstrap proportions (BPs) are shown at the internal branches. The length of each branch is proportional to the estimated number of substitutions. Bar denotes 10% substitutions per site. For details of the phylogenetic reconstruction methods see text. Species names are shown only for taxa where space permits it. Species names and abbreviations of the other taxa and the accession numbers of the sequences used are: Ss, *Spathidium* sp. AF056105; Ns, *Naxella* sp. AF056102; Sl, *Stylonychia lemnae* P25166; Pca, *Paranophrys carnivora* AF056103; Tp, *Tetrahymena pyriformis* Q04634; Ci, *Colpoda inflata* AF056098; Th, *Telotrochidium henneguyii* AF056109; Bj, *Blepharisma japonicum* AF056097; Sc, *Stentor coeruleus* AF056106; Ec, *Euplotes crassus* Q27139; Cpa, *Cyanophora paradoxa* AF092951; Pi, *Phytophthora infestans* AJ249839; Pp, *Porphyra purpurea* P50256; Metazoa, ((*Homo sapiens* P04720, *Xenopus laevis* P13549),(*Drosophila melanogaster* P08736, *Caenorhabditis elegans* P53013)); Fungi, (*Neurospora crassa* Q01372, (*Ajellomyces capsulatus* P40911, (*Schizosaccharomyces pombe* P50522, (*Saccharomyces cerevisiae* P02994, *Candida albicans* P16017)))); Mycetozoa, (*Planoprotostelium aurantium* AF016240, (*Physarum polycephalum* AF016243, *Dictyostelium dicoideum* P18624)); Viridiplantae, (*Oryza sativa* O64937, (*Arabidopsis thaliana* P13905, *Triticum aestivum* Q03033)); Euglenozoa, (*Euglena gracilis* P14963, (*Trypanosoma brucei* P41166, *Leishmania braziliensis* U72244)); Apicomplexa, (*Cryptosporidium parvum* P90519, *Plasmodium falciparum* Q00080); Oxymonadida, (Unidentified Oxymonadida A-14 AB007029, *Dinenympha exilis* AB007028); *Mastigamoeba balamuthi* BE636534/BE636560/BE636596/BE636646 (this work); *Entamoeba histolytica* P31018; *Naegleria andersoni* AF058283; Parabasala, *Trichomonas vaginalis* AF058282; Diplomonadida, *Hexamita inflata* U37081, *Spironucleus barkhanus* U29442, *Spironucleus vortens* U94406, *Giardia intestinalis* Q08046. Unambiguously aligned 386 amino acid positions were used for the analysis.

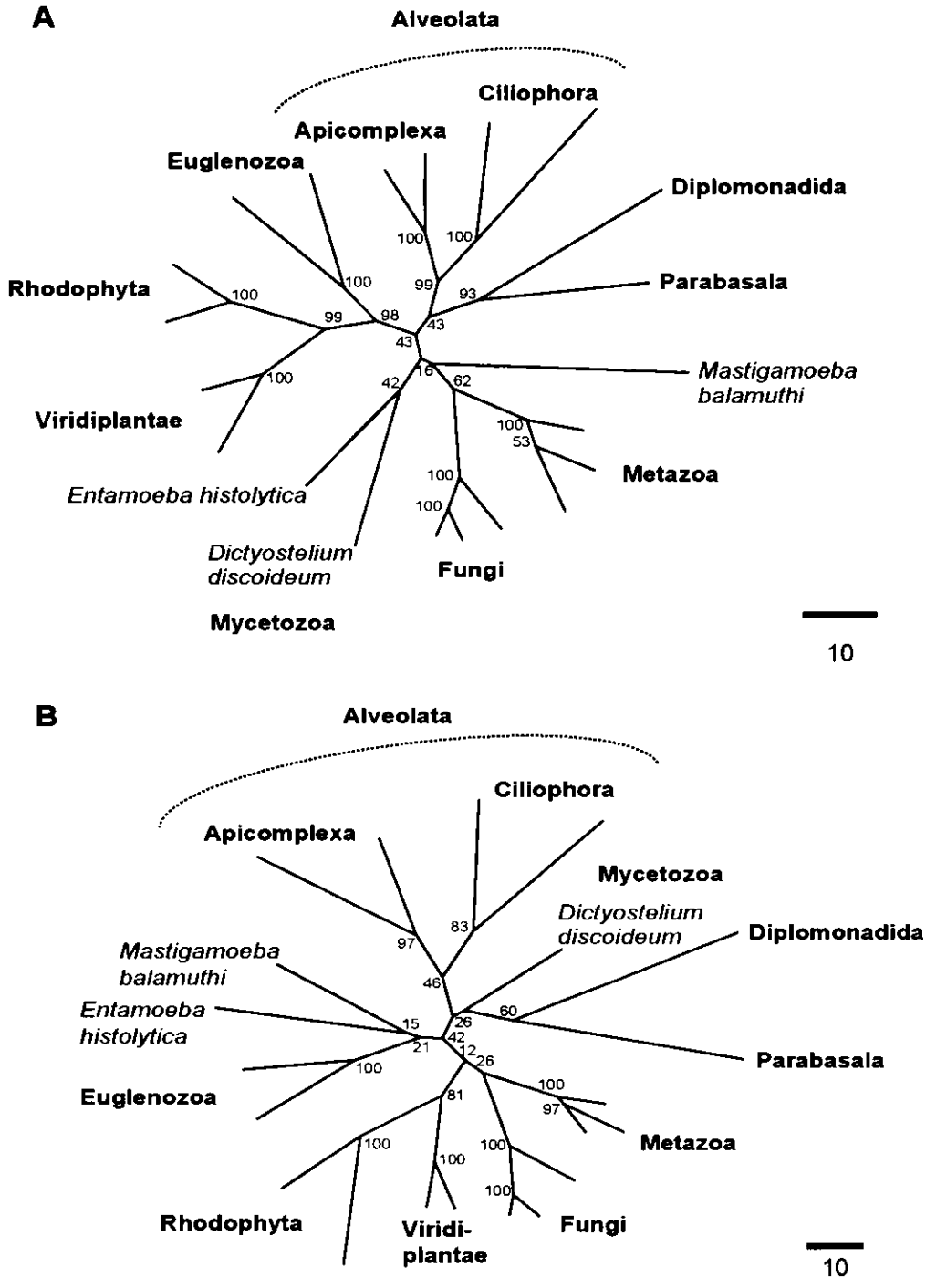


Fig. 1.6. Unrooted eukaryotic tree of EF-2. The best tree of the ML method by the use of the N-terminal (A) and C-terminal (B) portions of the EF-2 molecule are shown. Bootstrap proportions (BPs) are shown at the internal branches. The length of each branch is proportional to the estimated number of substitutions. Bar denotes10% substitutions per site. For details of the phylogenetic reconstruction methods see text. Species names are shown only for taxa where space permits it. Species names of the other taxa and the accession numbers of the sequences used are: Metazoa, (*Homo sapiens* P13639, (*Drosophila melanogaster* P13060, *Caenorhabditis elegans* P29691)); Fungi, (*Schizosaccharomyces pombe* O14460, (*Saccharomyces cerevisiae* P32324, *Candida albicans* O13430)); Viridiplantae (*Beta vulgaris* O23755, *Chlorella kessleri* P28996); Rhodophyta, (*Gelidium canariensis* AF213662, *Chondrus crispus* AF213661); Euglenozoa, (*Trypanosoma cruzi* D50806, *Euglena gracilis* AF213663); Apicomplexa, (*Plasmodium falciparum* AB023579, *Cryptosporidium parvum* Q23716); Ciliophora, (*Stylonychia mytilus* AF213664, *Tetrahymena pyriformis* AF213665); *Dictyostelium discoideum* P15112; *Entamoeba histolytica* Q06193; *Mastigamoeba balamuthi* AB066366/AB066367 (this work); Parabasala, *Trichomonas tenax* D78480; Diplomonadida, *Giardia intestinalis* D29835. Unambiguously aligned 467 and 230 amino acid positions, corresponding to the N- and C-terminal portions of the EF-2 molecule were used for the analyses.

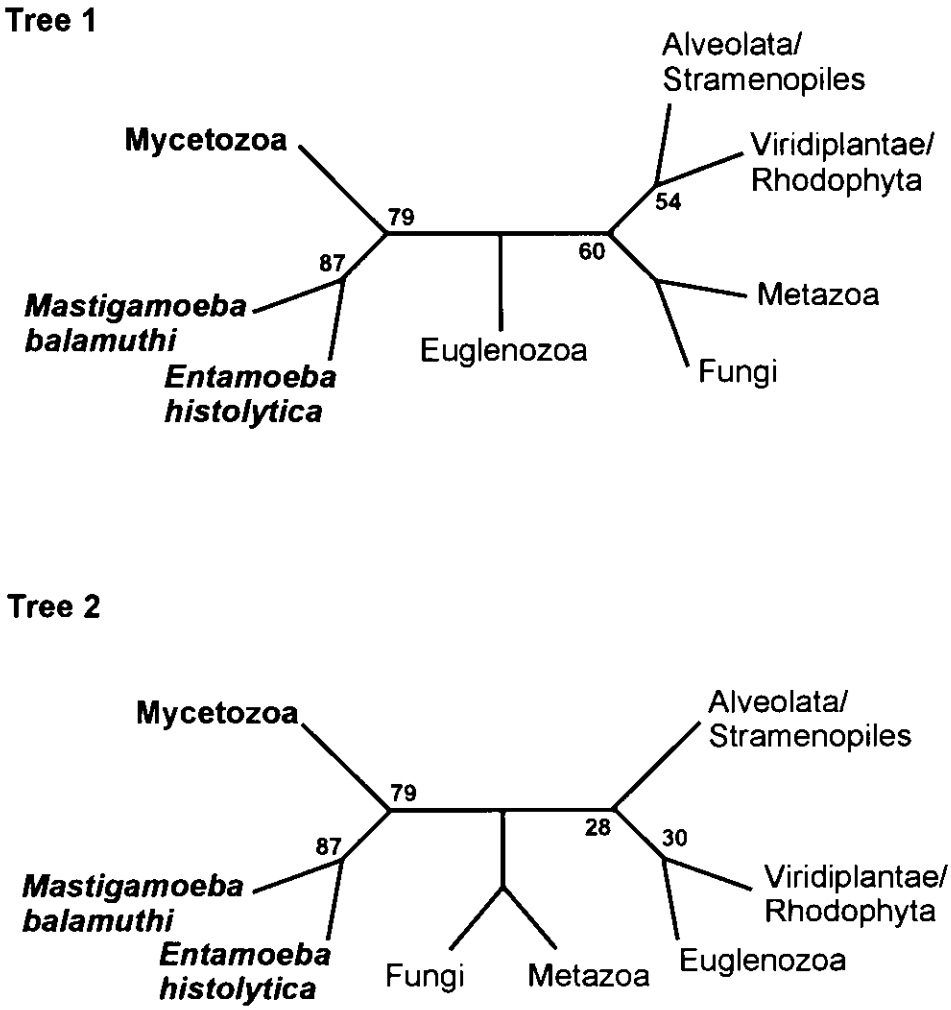


Fig. 1.7. Schematic representation of the phylogenetic relationships among seven eukaryotic lineages inferred from a combined analysis of individual genes. The best and the second best tree (Tree 1 and Tree 2 in Table 1.5) are shown. Bootstrap proportions calculated among 945 alternatives based on the analysis without Γ-distribution are attached to the internal branches.

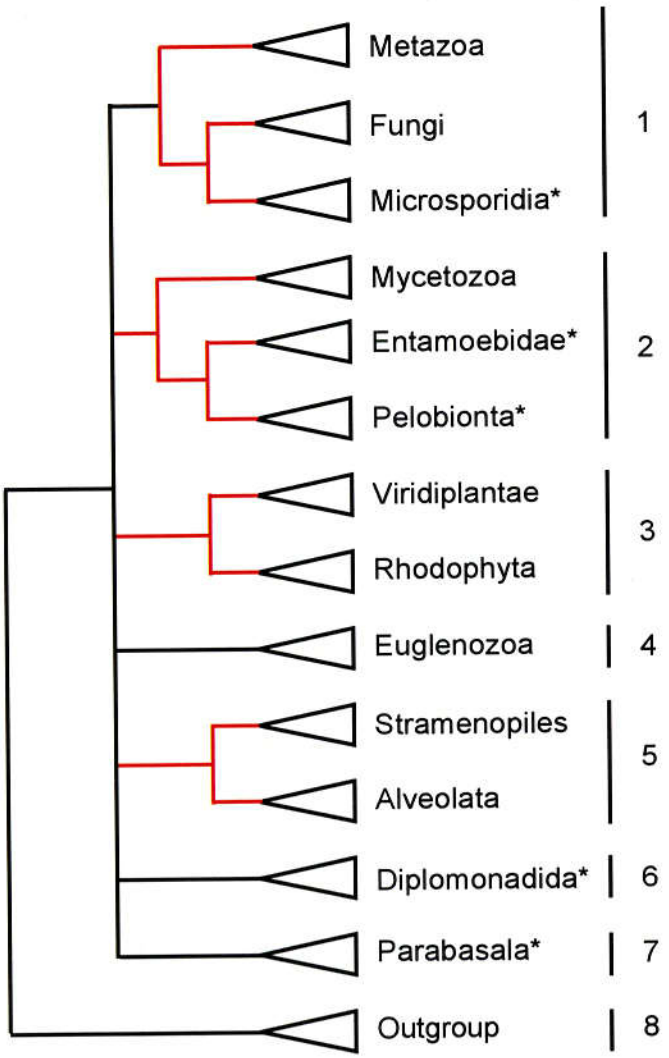


Fig. 1.8. Schematic representation of eight lineages used for a combined ML analysis. Constrained relationships are indicated in red. Astarisk (*) denotes amitochondriate protist groups.

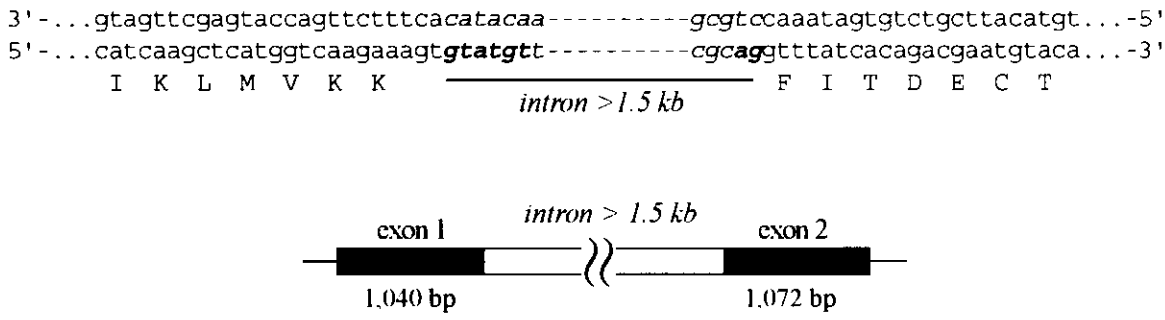


Fig. 1.9. Schematic representation of the *Giardia intestinalis* HSP90c gene. Signature sequences of spliceosomal intron are represented in bold.

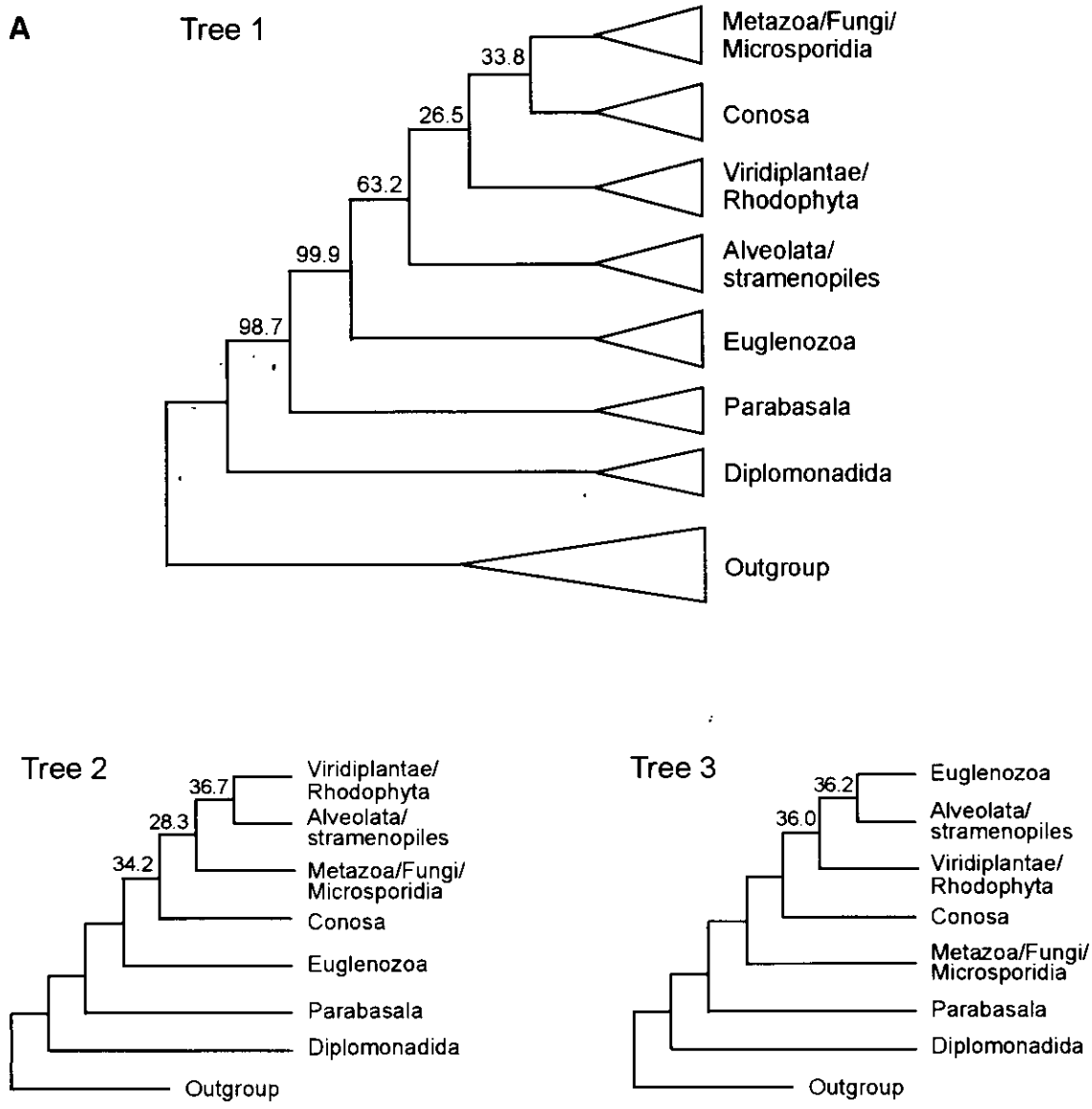
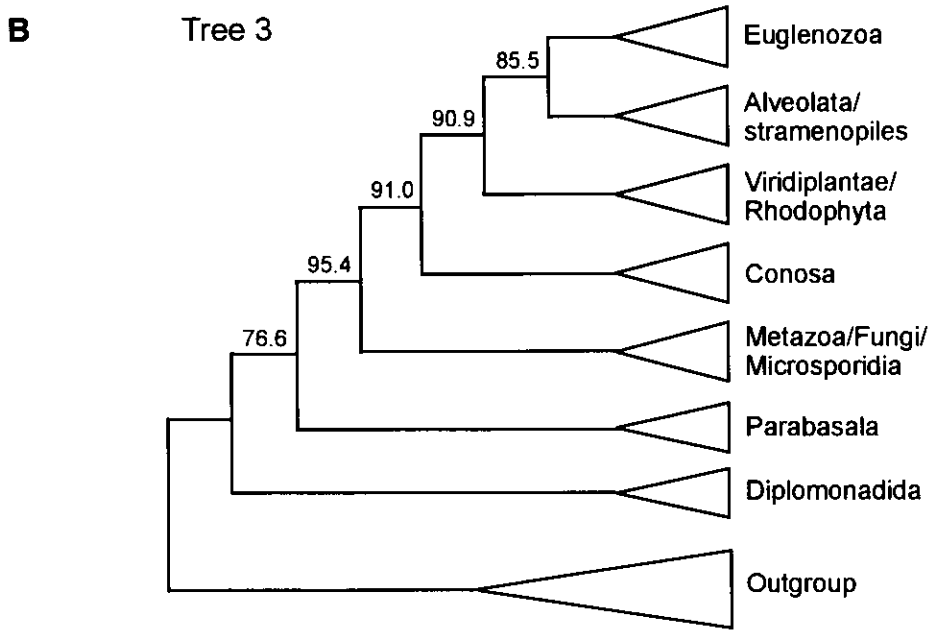
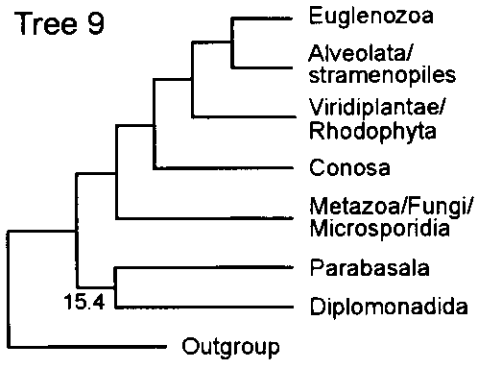


Fig. 1.10. Schematic representation of the phylogenetic relationships among seven eukaryotic lineages with an outgroup inferred from a combined analysis of 24 individual genes. Bootstrap proportions calculated among 160 alternatives were attached to the internal branches. A, The best three trees of the analyses without Γ-distribution; B, The best three trees of the analyses with Γ-distribution. Tree numbers are corresponding to Table 1.8.

B


Tree 3
85.5
90.9
91.0
95.4
76.6
Euglenozoa
Alveolata/
stramenopiles
Viridiplantae/
Rhodophyta
Conosa
Metazoa/Fungi/
Microsporidia
Parabasala
Diplomonadida
Outgroup


Tree 9
Euglenozoa
Alveolata/
stramenopiles
Viridiplantae/
Rhodophyta
Conosa
Metazoa/Fungi/
Microsporidia
Parabasala
15.4
Diplomonadida
Outgroup

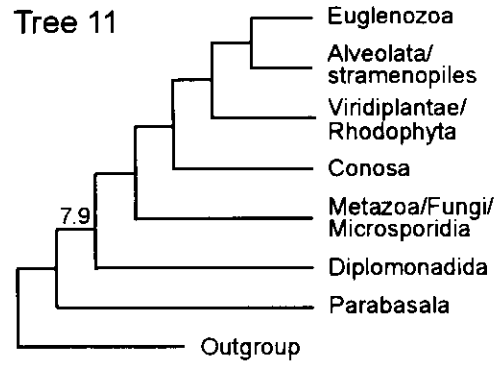

Tree 11
Euglenozoa
Alveolata/
stramenopiles
Viridiplantae/
Rhodophyta
Conosa
7.9
Metazoa/Fungi/
Microsporidia
Diplomonadida
Parabasala
Outgroup

Table 2.1. N-terminal amino acid sequences of the 15 *Giardia intestinalis* ribosomal proteins.

nomenclature		N-terminal sequence	Ribosomal protein that shows the highest FASTA score (accession number)	Fasta *E* value	class[a]
mammalian	*E. coli*				
S13	S15	GRMHSKGKGFXRSAFPYTRA	*Schizosaccharomyces pombe* S13 (P28189)	2.0×10^{-5}	cab
S17	-	GKVRTKTIKRAHRKI	*Arabidopsis thaliana* S17 (T48476)	0.0026	ca
S26	-	PVKRRNNGRSKYNEGSHNIFH	*Plasmodium falciparum* S26 (F71604)	0.0085	c
L8	L2	GRRIRVQRKGAGGVFRAHVA	*Homo sapiens* L8 (P25120)	3.0×10^{-5}	cab
L15	-	GAYKYVQEIQIRK	*Quercus suber* L15 (O82712)	0.0072	ca
L18a	-	MMQKYIVYGRRLPXEKFPE	*Saccharomyces cerevisia* e L20 (P47913)	0.0053	c
L22	-	PGKVFNLKKGGAVVRIVRKK	*Xenopus laevis* L22 (P50886)	3.9	c
L17	L22	GIVYEHTIRLNKAVRGKPXK	*Archaeoglobus fulgidus* L22P (AE000971)	2	cab
L24	-	VTTNQLSYXGRKILPGYGXR	*Arabidopsis thaliana* L24 (P38666)	6.7	ca
L26	L24	MYLNSAVTASRRKARXAYF	*Zea mays* L26 (AF093540)	0.0039	cab
L29	-	AKLKNHTSKNQNRKDHRNG	*Globodera rostochiensis* L29 (GRO133113)	5.0×10^{-8}	c
L37	-	SKGTASFGKRHTR	*Schistosoma mansoni* L37 (O44125)	0.0012	ca
L39	-	PANKTLSRKRRLAKKTKTNK	*Oryza sativa* L39 (P51426)	0.1	ca
L36a	-	VTYPAERRTFXKREGKHTV	*Homo sapiens* L44 (P09896)	0.019	ca
P2	L7/L12	MKHLAAYLLAKMGGKNEPAV	*Schizosaccharomyces pombe* A6 (T37490)	0.0013	cab

[a] The ribosomal protein is present in eukaryota (class e), archaebacteria (class a), and/or eubacteria (class b).

Table 2.2. N-terminal amino acid sequences of 15 *Trichomonas vaginalis* ribosomal proteins.

nomenclature		N-terminal sequence	Ribosomal protein that shows the highest FASTA score (accession number)	Fasta *E* value	class[a]
mammalian	*E. coli*				
S14	S11	PDGEIVMGVAHIYAXKNDGF	*Drosopila melanogaster* S14 (P14130)	4.4×10^{-5}	cab
S15a	S8	VRIPKTTKTRAARLVIERFY	*Homo sapiens* S15A (P39027)	0.0009	cab
S17		GRIPKTTKTRAARLVIERFY	*Schizosaccharomyces pombe* S17A (O42984)	0.24	ca
S23	S12	GKSKAXGRLAARKLRLAHKK	*Pyrococcus abyssi* S12P (F75182)	1.7	cab
S26	-	PVKRRNNGRGKMNRG	*Oxytricha nova* S26 (AJ305329)	0.072	c
L8	L2	GHIVRAQRRGNGSVFQAHTTHRVGP	*Arabidopsis thaliana* L8 (T04582)	2.5×10^{-7}	cab
L10	-	GRRPARXYRYIGNKPYPKARFXRGV	*Caenorhabditis elegans* L10 (Q09533)	6.8×10^{-9}	ca
L13	-	VAKNNQIPNDHLRKYWYHRVKTYFD	*Oryza sativa* L13 (AP001389)	0.0016	c
L15	-	GAFTYLRELYTKKQSEVMRFILRLRRWE	*Orconectes limosus* L15(Q9XYC2)	4.0×10^{-8}	ca
L18	-	GVDLDHRVPRKQRQEPKTADAQMKM	*Sus scrofa* L18 (Q95342)	0.66	ca
L21	-	GHTEGLRHATRYLFA	*Caenorhabditis elegans* L21 (P34334)	0.58	ca
L27a	L15	PTRFHKSRKLRGQVSMGFGRVGKHR	*Homo sapiens* L27a (AJ400879)	4.2×10^{-6}	cab
L37a	-	ARRTKKVGIRGKYGV	*Saccharomyces cerevisiae* L43 (P49631)	0.0018	ca
L39	-	AAHKTLRTKRILAKAMKANRPMPQF	*Mus musclus* L39 (AK005645)	0.00015	ca
L40	-	GKEAYEPSLRILAEKTNVKRSVXRKKYA	*Drosophila melanogaster* L40 (P18101)	0.0032	ca

[a] The ribosomal protein is present in eukaryota (class e), archaebacteria (class a), and/or eubacteria (class b).

Table 2.3. Sequence lengths of the rRNA from various organisms.

Organism (accession number)	SSUrRNA	LSUrRNA
Homo sapiens (U13369)	1,871	5,035
Artemia salina (X01723)	1,810	3,670[a]
Saccharomyces cerevisiae (U53879)	1,800	3,396
Trichomonas vaginalis (U17510/AF202181)	1,574	2,765
Giardia intestinalis (X52949)	1,454	2,684
Mastigamoeba balamuthi (L23779/AB066363-5)	2,741	4,485
Encechalitozoon cuniculi (AJ005581)	1,299	2,487
Escherichia coli (J01695)	1,542	2,904
Haloarcula marismortui (AF034620)	1,472	2,922

(bp)

[a] For LSUrRNA, AF169714, AS28S299 and data cloned and sequenced in this work are concatenated.

Table 2.4. Comparison of the sequence lengths of the SSUrRNA helices E10_1 and E23 among various organisms.

Organism (accession number)	E10_1	E23
Homo sapiens (U13369)	84	233
Saccharomyces cerevisiae (J01353)	50	225
Mastigamoeba balamuthi (L23779)	352	657
Trichomonas vaginalis (U17510)	-	118
Giardia intestinalis (X52949)	-	89
Encephalitozoon cuniculi (AJ005581)	-	46

(bp)

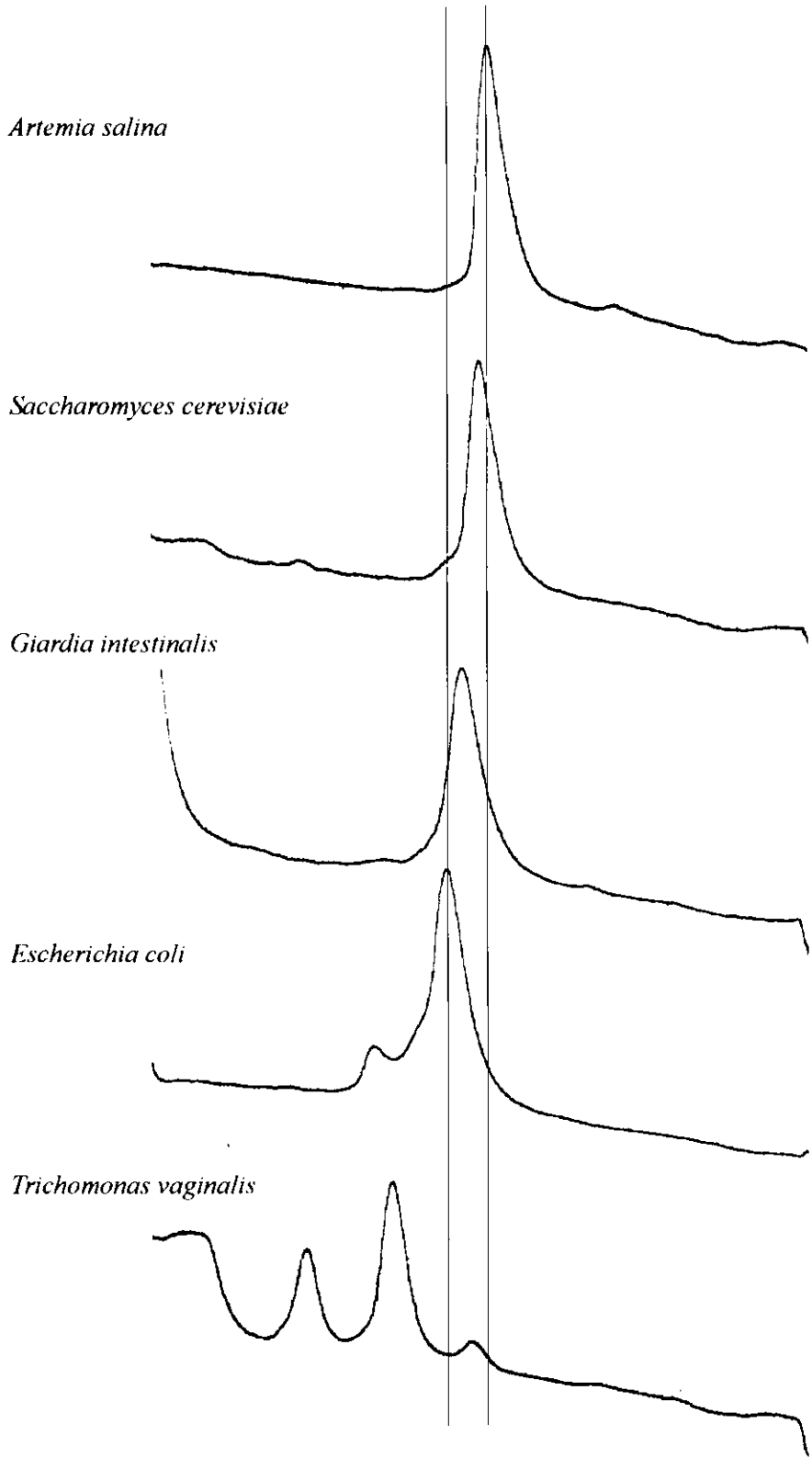


Fig. 2.1. Sedimentation profiles of the ribosomes from *Artemia salina*, *Saccharomyces cerevisiae*, *Giardia intestinalis*, *Eschelichia coli* and *Trichomonas vaginalis*. Buffer composition used in the analysis was 100mM NH_4Ac, 15mM $MgAc_2$, 20mM Tris-HCl pH7.6. Vertical and horizontal axes denote absorbance at 260nm and sedimentation gradient (top to bottom), respectively. Locations of the peaks for *E. coli* and *A. salina* are represented by vertical lines.

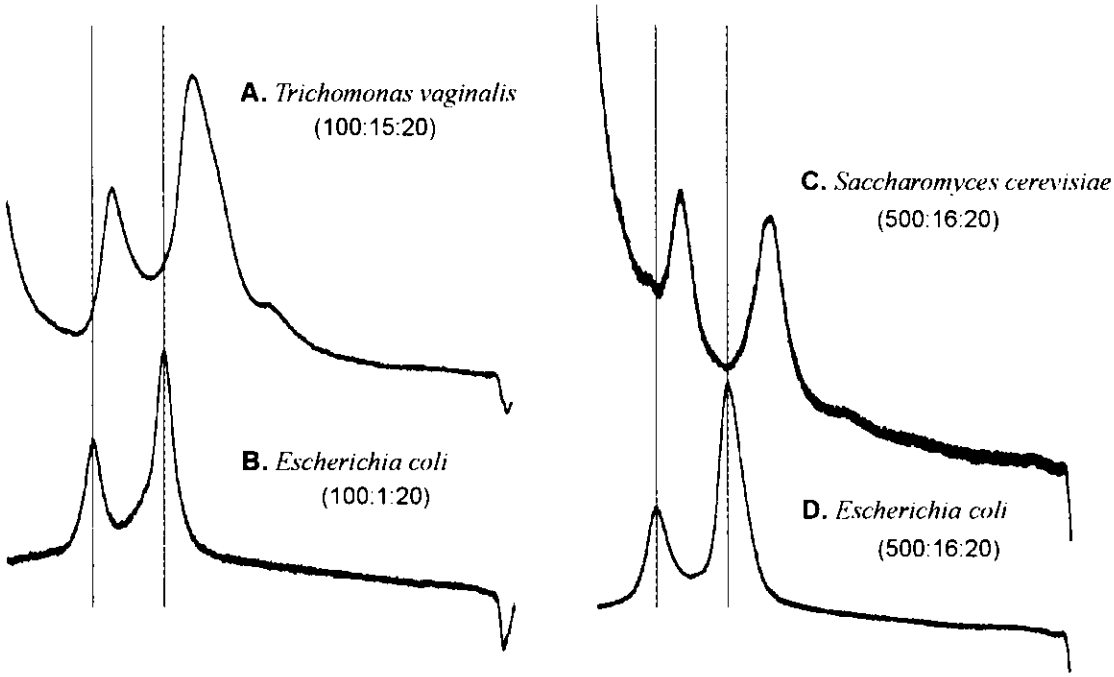


Fig. 2.2. Sedimentation profiles of the small and large subunit ribosomes of *Trichomonas vaginalis*, *Saccharomyces cerevisiae* and *Escherichia coli*. Millimolar concentrations of the buffer. (NH_4Ac: $MgAc_2$:Tris-HCl pH7.6) (A), (B) and (KCl: $MgCl_2$:Tris-HCl pH7.6) (C), (D), used in the analysis are shown in parentheses. 0.2 mM EDTA and 12mM 2-mercaptoethanol were also contained in the buffer. Vertical and horizontal axes denote absorbance at 260 nm and sedimentation gradient (top to bottom), respectively. Location of the peaks for *E. coli* small and large subunit ribosome are represented by vertical lines.

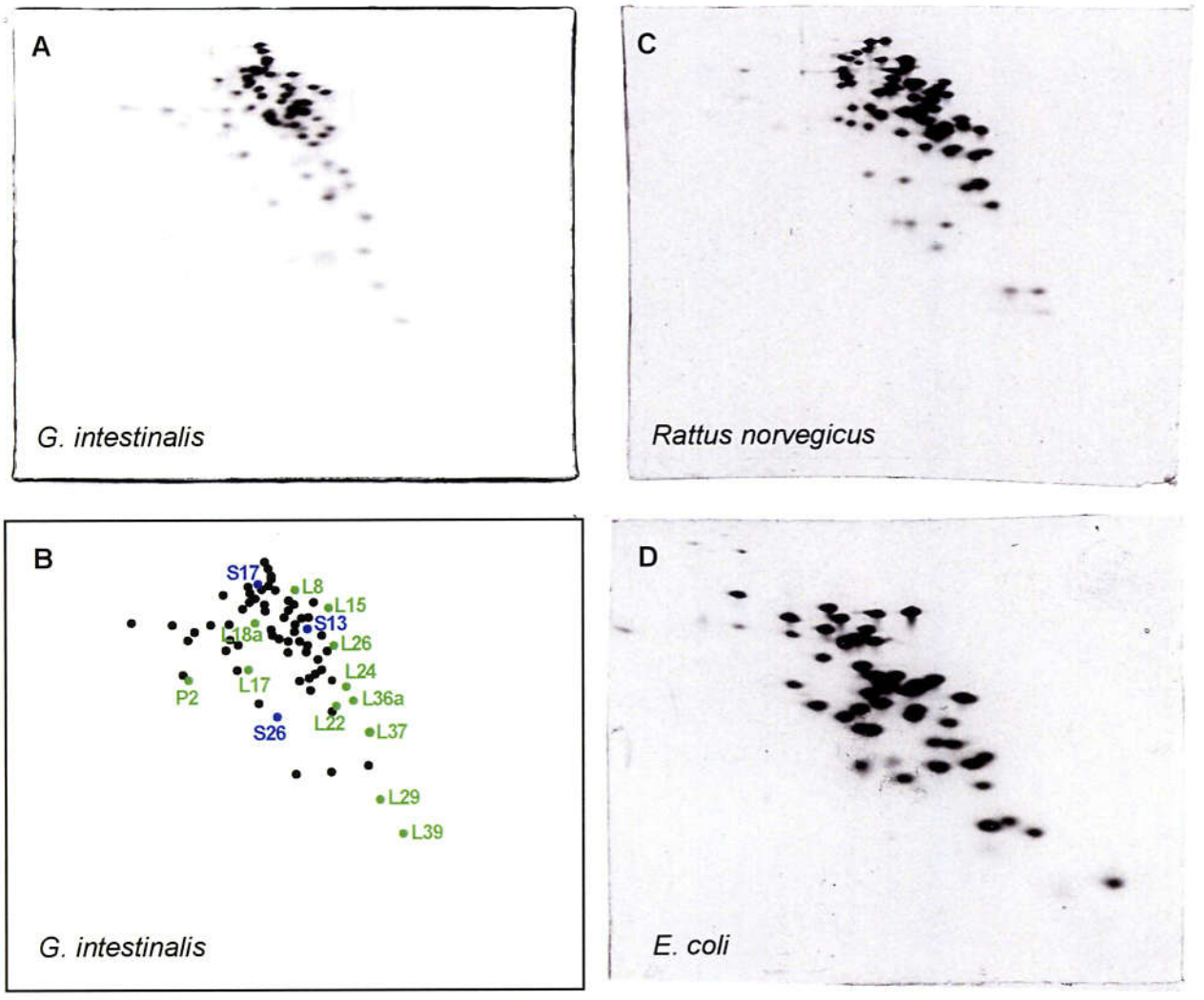


Fig. 2.3. Two-dimensional gel electrophoresis profiles of total proteins from ribosomes of three species. A, *Giardia intestinalis*; B, Schematically represented protein spots on the *G. lintestinalis* gel and ribosomal protein names identified by N-terminal amino acid sequencing analysis; C, *Rattus norvegicus*; D, *Eschelichia coli*.

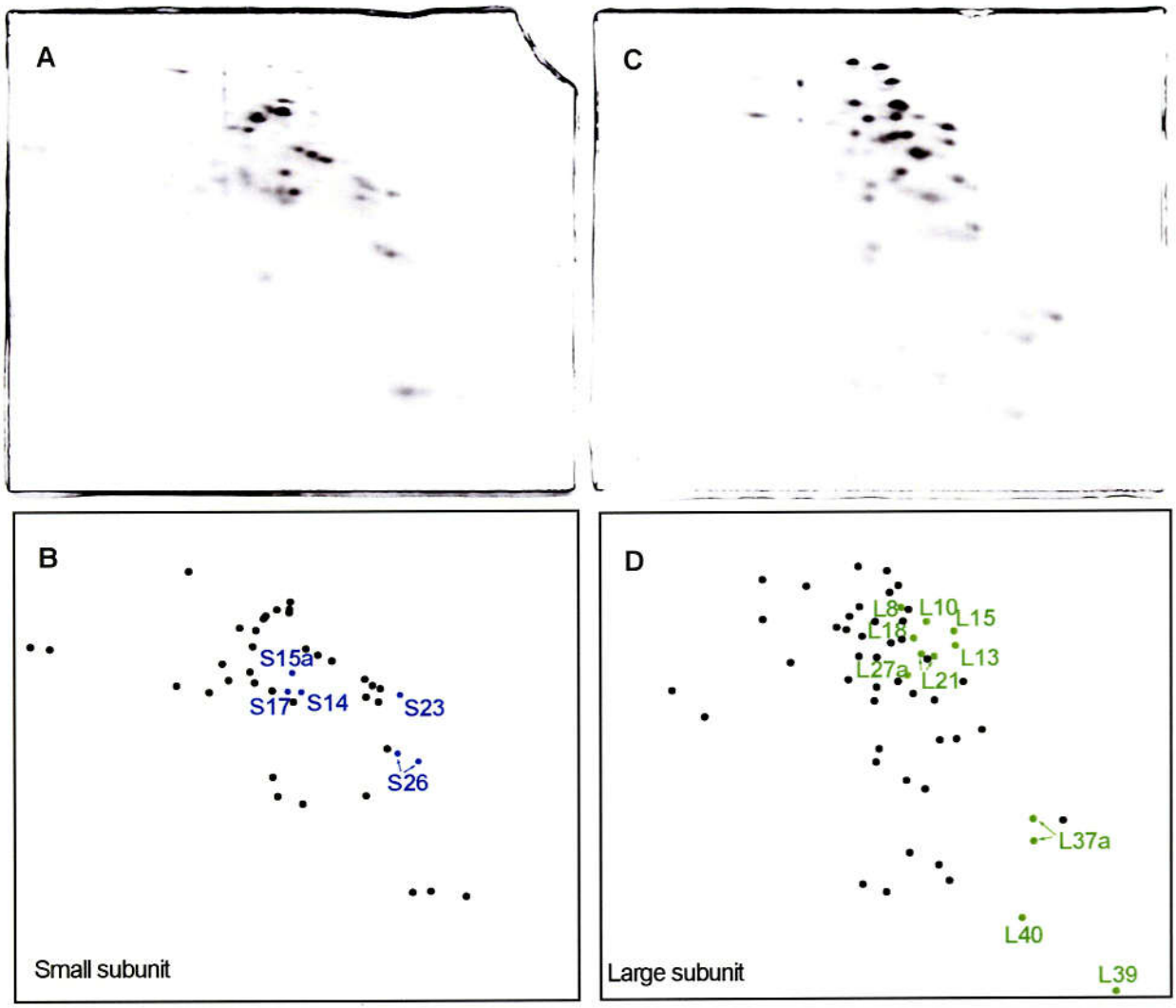


Fig. 2.4. Two-dimensional gel electrophoresis profiles of the total proteins from *Trichomonas vaginalis* ribosomal subunits. A, Large subunit; C, Small subunit; B,D, Protein spots on the gel of large (B) and small (D) subunit of *T. vaginalis* ribosome and ribosomal protein names identified by N-terminal amino acid sequencing analysis are shematically represented.

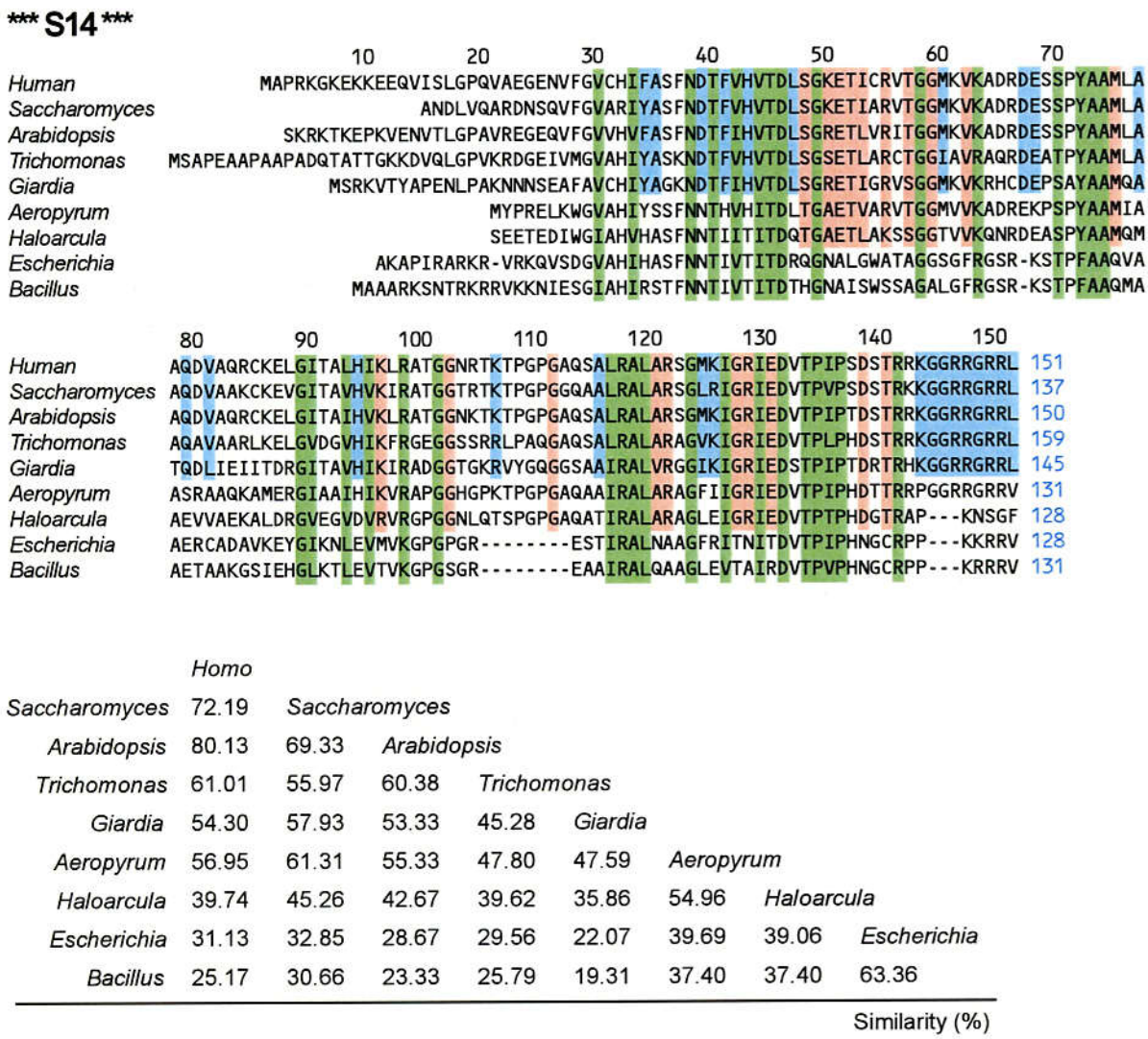
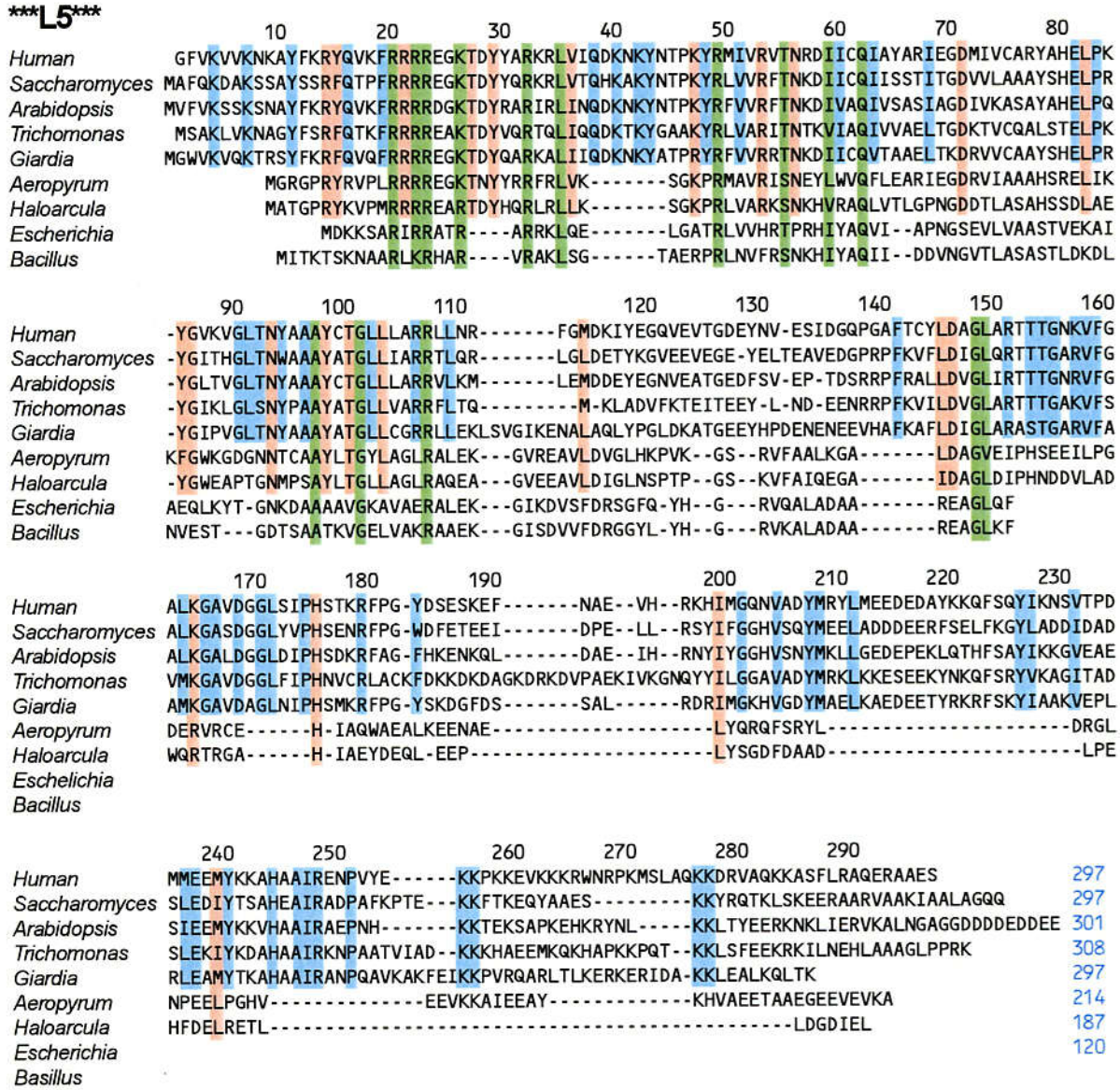


Fig. 2.5. Sequence alignments of five ribosomal proteins and pairwise amino acid identities between organisms. Residues identical among all sequences in the alignment are colored with green, residues identical among eukaryotes and archaebacteria with orange and residues identical in eukaryote with blue. Position numbers for the Homo sapiens sequence are shown above the alignment. The numbers at last column of each alignments are amino acid residues. Species and database accession numbers are: S14: *Homo sapiens* (P06366); *Saccharomyces cerevisiae* (P39516); *Arabidopsis thaliana* (P42036); *Trichomonas vaginalis* (this work); *Giardia intestinalis* (this work); *Aeropyrum pernix* (Q9YB55); *Haloarcula marismortui* (P10788); *Escherichia coli* (P02366); *Bacillus subtilis* (P04969); S15a: *H. sapiens* (P39027); *S. cerevisiae* (P04648); *A. thaliana* (P42798); *Mastigamoeba balamuthi* (BE636632); *T. vaginalis* (this work); *G. intestinalis* (this work); *A. pernix* (Q9YF89); *H. marismortui* (P12742); *E. coli* (P02361); *B. subtilis* (P12879); L5: *H. sapiens* (P46777); *S. cerevisiae* (P26321); *A. thaliana* (P49227); *T. vaginalis* (this work); *G. intestinalis* (this work); *A. pernix* (Q9YF94); *H. marismortui* (P14123); *E. coli* (P02419); *B. subtilis* (P46899); L8: *H. sapiens* (P25120); *S. cerevisiae* (P05736); *A. thaliana* (P46286); *T. vaginalis* (this work); *G. intestinalis* (this work); *A. pernix* (Q9YFN1); *H. marismortui* (P20276); *E. coli* (P02387); *B. subtilis* (P42919); L10a: *H. sapiens* (P53025); *S. cerevisiae* (P53030); *A. thaliana* (P08770); *T. vaginalis* (AF070993); *G. intestinalis* (this work); *A. pernix* (Q9Y9W6); *H. marismortui* (P12787); *E. coli* (P02384); *B. subtilis* (Q06797).

***** S15a *****

	Homo								
Saccharomyces	76.74	*Saccharomyces*							
Arabidopsis	77.52	79.84	*Arabidopsis*						
Mastigamoeba	71.32	69.77	72.09	*Mastigamoeba*					
Trichomonas	70.77	67.69	68.46	63.85	*Trichomonas*				
Giardia	62.31	60.77	64.62	64.62	56.92	*Giardia*			
Aeropyrum	40.74	41.48	40.74	39.26	42.31	40.00	*Aeropyrum*		
Haloarcula	37.98	40.31	41.09	35.66	33.85	35.38	41.48	*Haloarcula*	
Escherichia	24.81	23.26	21.71	22.48	26.15	22.31	28.89	26.36	*Escherichia*
Bacillus	26.72	25.95	25.95	27.48	25.95	26.72	33.33	29.01	48.09

Similarity (%)



	Homo							
Saccharomyces	48.48	*Saccharomyces*						
Arabidopsis	52.82	49.50	*Arabidopsis*					
Trichomonas	44.48	42.21	42.53	*Trichomonas*				
Giardia	51.52	46.46	47.18	42.53	*Giardia*			
Aeropyrum	21.96	21.21	21.59	16.23	19.19	*Aeropyrum*		
Haloarcula	15.20	16.84	16.94	15.58	14.48	40.65	*Haloarcula*	
Escherichia	10.14	10.77	9.97	8.44	11.11	15.42	19.35	*Escherichia*
Bacillus	8.11	9.43	9.30	7.47	8.75	14.95	21.51	49.17

Similarity(%)

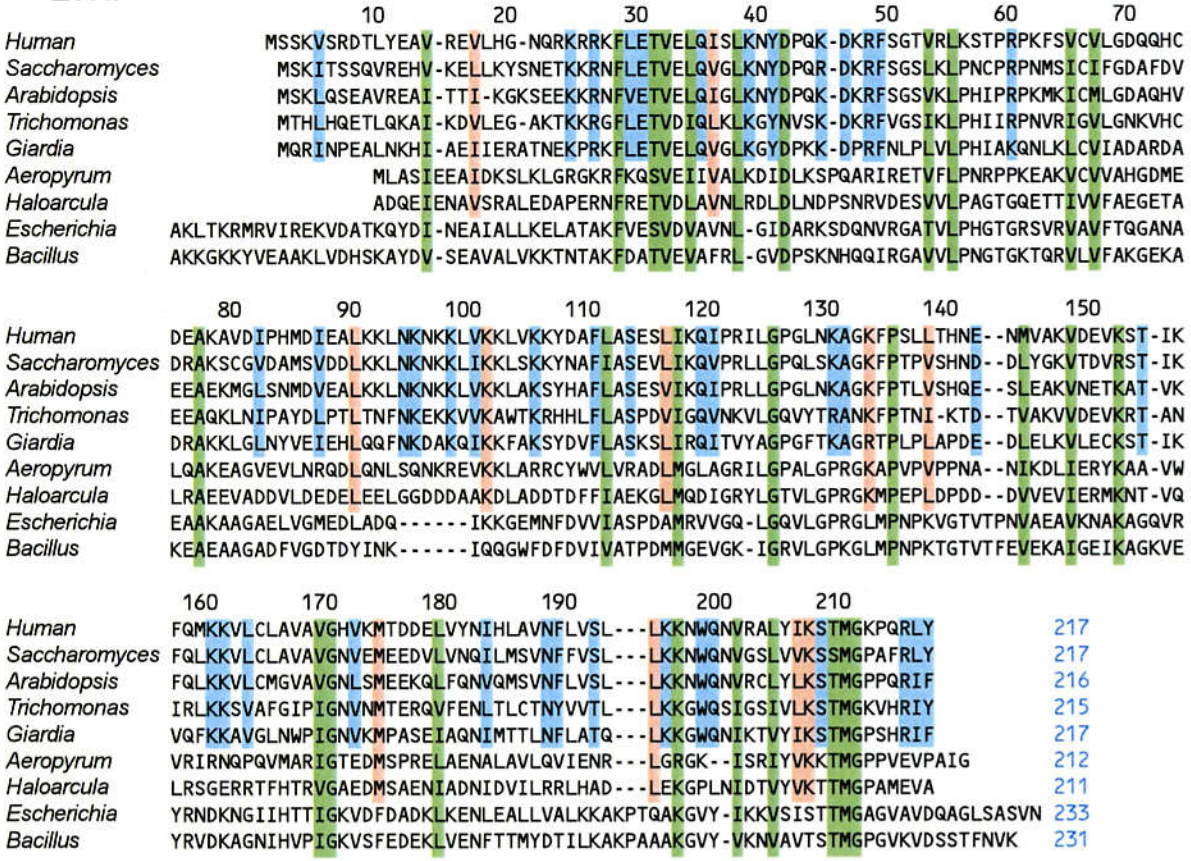
L8

```
                                                  10        20        30        40        50
Human                                     MGRVIRGQRKG-AGSVFRAHVKHRKG-AARLRAVDFAERHGYIKGIVKDIIHDPG
Saccharomyces                              GRVIRNQRKG-AGSIFTSHTRLRQG-AAKLRTLDYAERHGYIRGIVKQIVHDSG
Arabidopsis                               MGRVIRAQRKG-AGSVFKSHTHHRKG-PAKFRSLDFGERNGYLKGVVTEIIHDPG
Trichomonas                               MGHIVRAQRRG-NGSVFQAHTHHRVG-PAKFRALDAAERTSVISGMVKEIIHDPG
Giardia                                   MGRRIRVQRKG-AGSVFRAHVARRLG-AAKLRANDFAERTGSVRGVVKAILHDRG
Aeropyrum                                 MGKRLRQQRAGRGTPTYRSRAHIHPG-PAKYPPL----SGDTLRGKVVELIHDPG
Haloarcula                                 GRRIQGQRRGRGTSTFRAPSHRYKA-DLEHRKV---EDGDVIAGTVVDIEHDPA
Escherichia     AVVKCKPTSPGRRHVVKVVNPELHKGKPFAPLLEKNSKSGGRNNNGRITTRHIGGGHKQAYRIVDFKRNKDGIPAVVERLEYDPN
Bacillus        AIKKYKPSSNGRRGMTTSDFAEITTDKPEKSLLAPLHKKGGRNNQGKLTVRHQGGGHKRQYRVIDFKRDKDGIPGRVATVEYDPN

                      60        70        80        90       100       110       120       130
Human           RGAPLAKVVFRDPYRFKKRTELFIAAEGIHTGQFVYCGKKAQLNIGNVLPVGTMPEGTIVCCLEEKPGDRGKLARASGNYATVIS
Saccharomyces   RGAPLAKVVFRDPYKYRLREEIFIANEGVHTGQFIYAGKKASLNVGNVLPLGSVPEGTIVSNVEEKPGDRGALARASGNYVIIIG
Arabidopsis     RGAPLARVTFRHPFRFKKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLPLRSIPEGAVVCNVEHHVGDRGVLARASGDYAIVIA
Trichomonas     RGAPLAKLIYKKAEGQGFDSALVIAPEGIHTGQFIKCGAQADLHIGNILPLAQIPEGTEICNVEHRPGDGGRYGRCSGDSCRVIG
Giardia         RGAPLARVQFPNMRGEGRVNELFIAPEGMYTGQEVFSGNKATLHIGNILPVGNIPEGTYVCNVEEKPMDRGCLARASGTYCLVVA
Aeropyrum       RYVPLARVVREDGVEF-----LMPAAEGMYVGQIIEIGPAAKPEAGNILPLGKIPEGTEVFNVELRPGDGGKLARQAGSYALVVG
Haloarcula      RSAPVAAVEFEDGDRR-----LILAPEGVGVGDELQVGVDAEIAPGNTLPLAEIPEGVPVCNVESSPGDGGKFARASGVNAQLLT
Escherichia     RSANIALVLYKDGERR-----YILAPKGLKAGDQIQSGVDAAIKPGNTLPMRNIPVGSTVHNVEMKPGKGGQLARSAGTYVQIVA
Bacillus        RSANIALINYADGEKR-----YILAPKGIQVGTEIMSGPEADIKVGNALPLINIPVGTVVHNIELKPGKGGQLVRSAGTSAQVLG

                140       150       160       170       180       190       200       210       220
Human           HNPETKKTRVKLPSGSKKVISSANRAVVGVVAGGGRIDKPILKAGRAYHKYKAKRNCWPRVRGVAMNPVEHPFGGGNHQHI-GKP
Saccharomyces   HNPDENKTRVRLPSGAKKVISSDARGVIGVIAGGGRVDKPLLKAGRAFHKYRLKRNSWPKTRGVAMNPVDHPHGGGNHQHI-GKA
Arabidopsis     ENPDSDTTRIKLPSGSKKIVPSGSRAMIGQVAGGGRTEKPMLKAGNAYHKYRVKRNSWPKVRGVAMNPVEHPHGGGNHQHI-GHS
Trichomonas     HTEN--YTRIQLPSGRKALVSNICRATLGIVAGGGRPEKPLLKAGNVHYKYKAKRHTWPVVCGVAMNPVDHRHGGGSHQHM-GAP
Giardia         HNMETNKTRIRLPSGQKKLISSKARAMIGLVAGGGRTDKPLLKAGNAYHKYKAKRNCWPVVRGVAMNPVDHPHGGGNHQHI-GVS
Aeropyrum       RAGA--KTILRLPSGKDKEVPNDSRATIGIPAGAGRIEKPIMKAGFAYHKWKVKARKWPRVRGVAMNAVDHPFGGGRHQHK-GRP
Haloarcula      HDRN--VAVVKLPSGEMKRLDPQCRATIGVVGGGGRTDKPFVKAGNKHHKMKARGTKWPNVRGVAMNAVDHPFGGGGRQHP-GKP
Escherichia     RDGA--YVTLRLRSGEMRKVEADCRATLGEVGNAEHMLRVLGKAGAA--RWRGVR---PTVRGTAMNPVDHPHGGGEGRNF-GKH
Bacillus        KEGK--YVLVRLNSGEVRMILSACRASIGQVGNEQHELINIGKAGRS--RWKGIR---PTVRGSVMNPNDHPHGGGEGRAPIGRK

                      210       220       230
Human           STIRRDAPAGRKVGLIAARRTGRLRGTKTVQEKEN  257
Saccharomyces   STISRGAVSGQKAGLIAARRTGLLRGSQKTQD     253
Arabidopsis     STVRRDAPPGQKVGLIAARRTGRLRGQAAASAAKAD 258
Trichomonas     GTVARSARPGQKLGLIASRRTGHRRGTVNLA      251
Giardia         SCVPRNAVPGQKVGLIAARRTGCSGGKRR        251
Aeropyrum       STVARTAPPGRKVGHIAARRTGRRKR           238
Haloarcula      KSISRNAPPGRKVGDIASKRTGRGGNE          239
Escherichia     -PVTPWGVQTKGKKTRSNKRT-DKFIVRRRSK     272
Bacillus        SPMSPWGKPTLEFKTRKQKNKSDKFIVRRRKNK    276
```

	Homo							
Saccharomyces	70.43	*Saccharomyces*						
Arabidopsis	66.28	61.24	*Arabidopsis*					
Trichomonas	53.70	51.78	53.10	*Trichomonas*				
Giardia	66.15	63.24	60.85	57.37	*Giardia*			
Aeropyrum	47.86	44.66	48.45	46.61	48.61	*Aeropyrum*		
Haloarcula	47.47	41.50	42.64	43.43	45.82	48.54	*Haloarcula*	
Escherichia	25.00	26.84	26.47	27.21	29.04	30.15	32.72	*Escherichia*
Bacillus	25.72	25.36	23.91	26.45	26.81	27.90	30.07	57.97

Similarity (%)

***** L10a *****



	Homo							
Saccharomyces	59.45	*Saccharomyces*						
Arabidopsis	63.59	63.13	*Arabidopsis*					
Trichomonas	39.63	37.33	43.52	*Trichomonas*				
Giardia	43.32	41.47	45.62	36.41	*Giardia*			
Aeropyrum	23.04	22.12	24.07	18.60	23.50	*Aeropyrum*		
Haloarcula	23.04	23.96	23.61	20.47	29.03	34.43	*Haloarcula*	
Escherichia	18.88	22.32	17.60	21.89	24.03	25.32	27.47	*Escherichia*
Bacillus	18.61	17.75	18.18	20.35	20.78	22.51	23.38	49.36

Similarity (%)

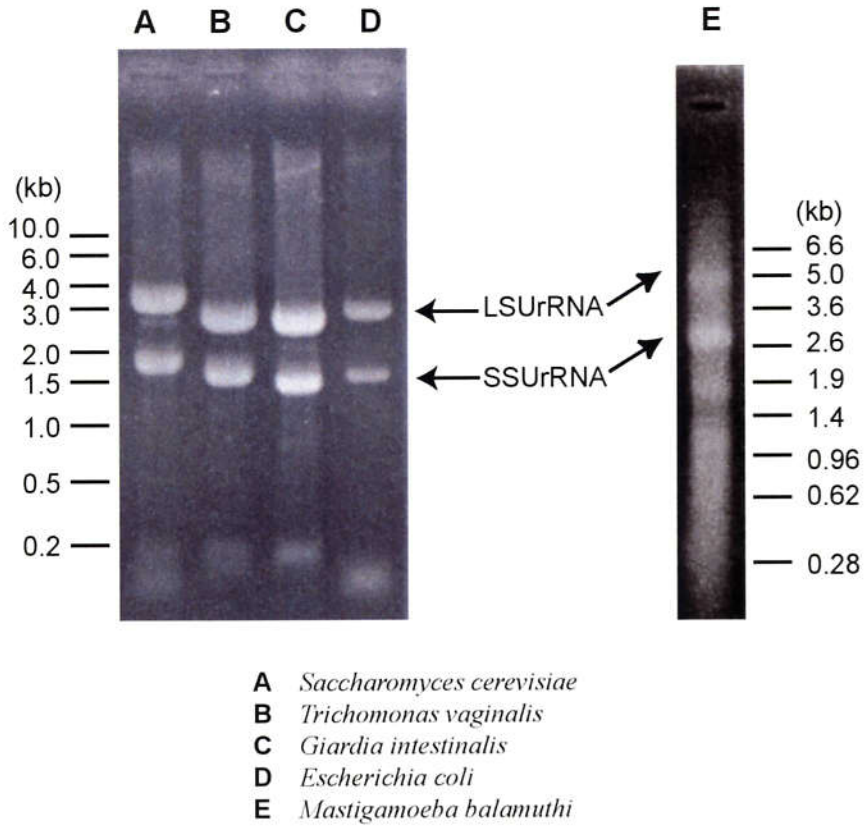


Fig. 2.6. Comparison of the length of rRNAs from various organisms.

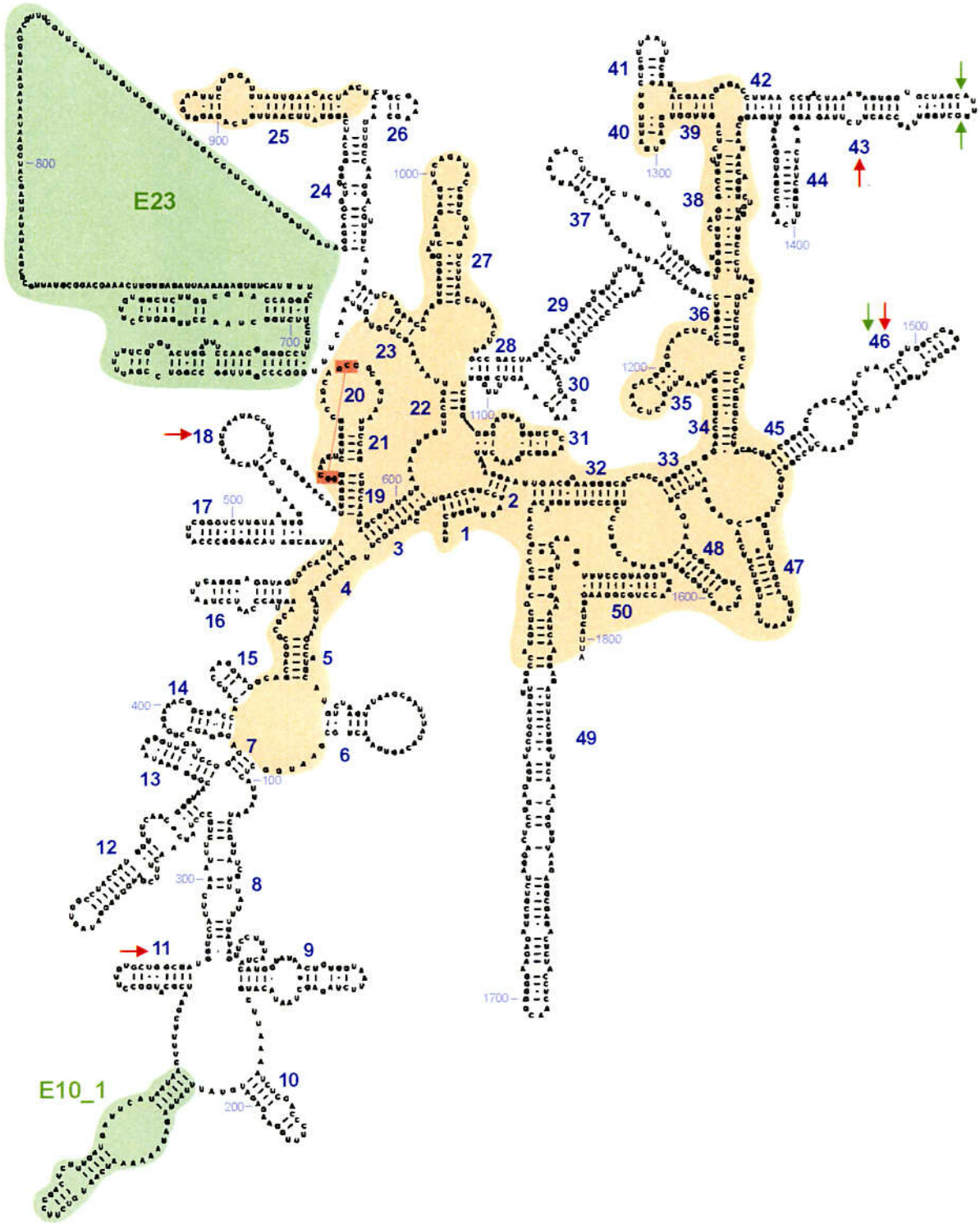


Fig. 2.7. Secondary structure model for *Saccharomyces cerevisiae* SSUrRNA. Quoted and modified from Gutell, R. R. (1993). The universal core sequences are shaded in orange, helices having variability among species are shaded in green. The sequence is written clockwise from 5' to 3', sites are numbered in blue every 100 nucleotides. Helix numbering is according to Neefs et al.(1993). The red arrow indicates the deletion of the helix in *Encephalitozoon cuniculi* SSUrRNA sequences and the green arrow indicates the insertion of helix in *Mastigamoeba balamuthi* SSUrRNA, which are observed in the corresponding regions.

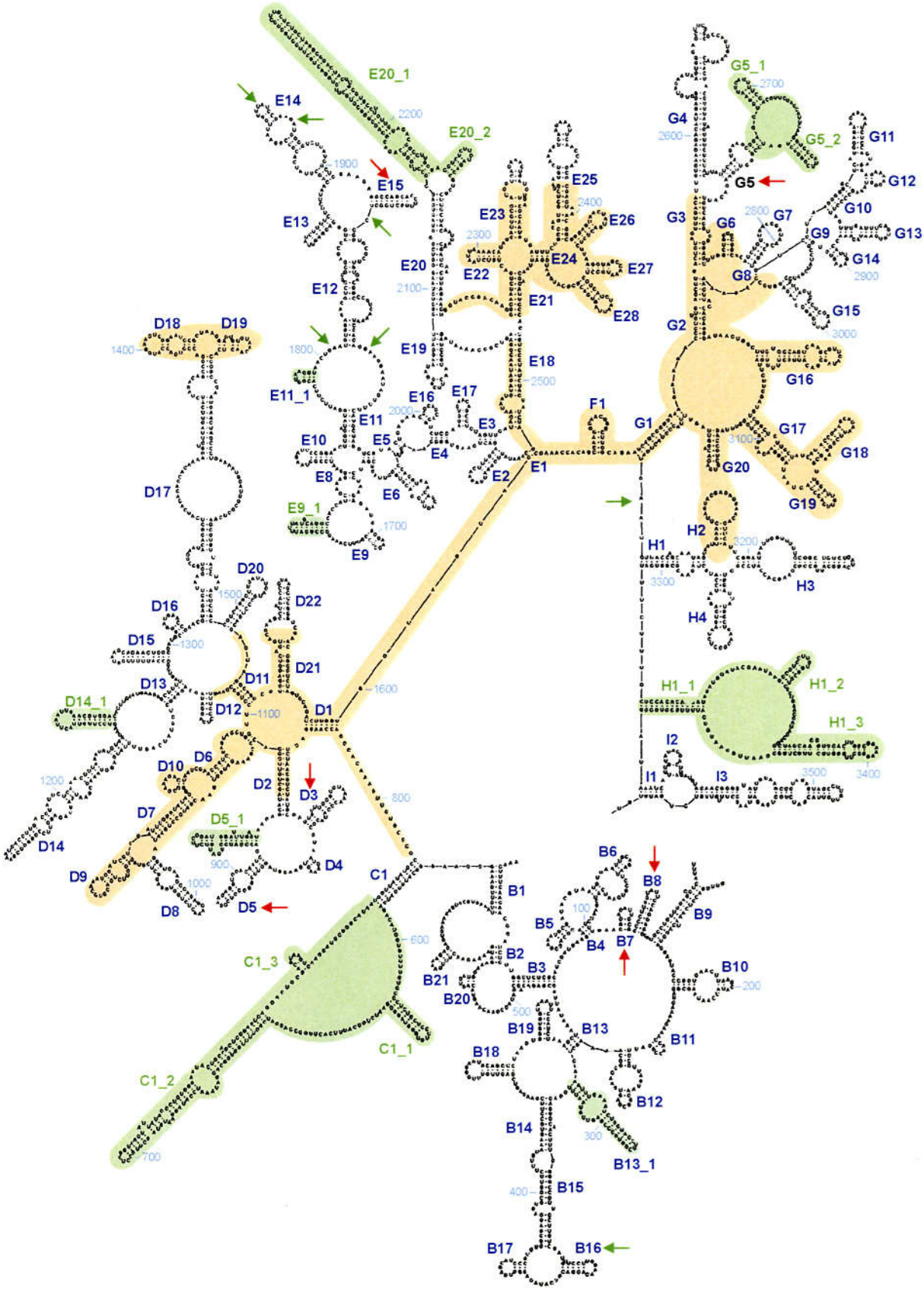


Fig. 2.8. Secondary structure model for *Saccharomyces cerevisiae* LSUrRNA. Quoted and modified from Ben Ali et al. (1999). The universal core sequences are shaded in orange, and helices which show variability among species are shaded in green. The sequence is written clockwise from 5' to 3', sites are numbered in blue every 100 nucleotides. Helix numbering is according to De Rijk et al.(1999). The red arrow indicate the deletion of helix in *Encephalitozoon cuniculi* SSUrRNAsequences and green arrows indicate the insertion of helix in *Mastigamoeba balamuthi* SSUrRNA sequence, which were observed in the corresponding regions.

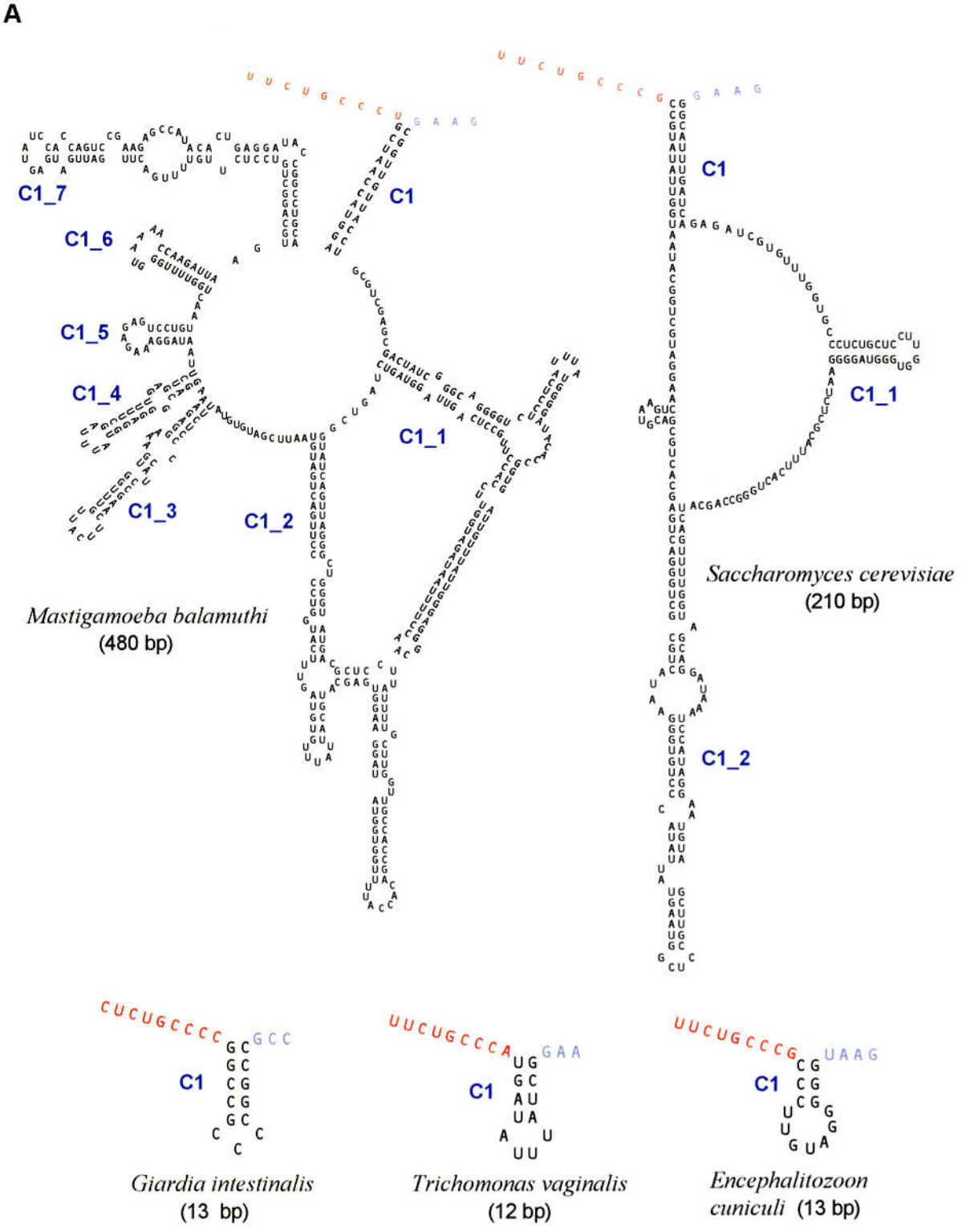
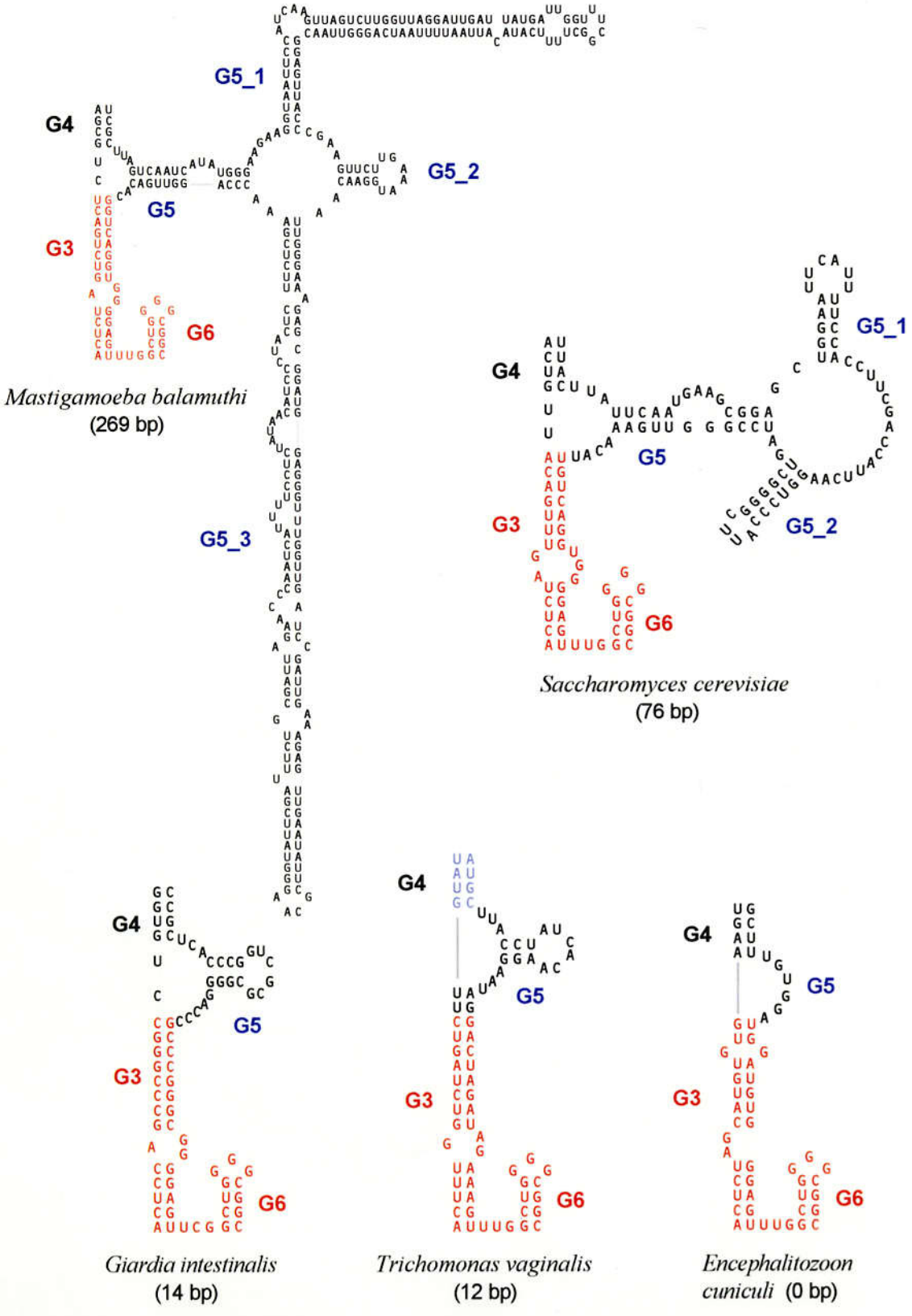


Fig. 2.9. Comparison of the estimated secondary structure models for helices C1 (A) and G5 (B) in LSUrRNA sequences from various organisms. Sequence length in C1 and G5 helices are shown in parentheses by bp. The universal core sequences are in red.

B


G4
G5_1
G5_2
G5
G3
G6
G5_3
Mastigamoeba balamuthi
(269 bp)
Saccharomyces cerevisiae
(76 bp)
Giardia intestinalis
(14 bp)
Trichomonas vaginalis
(12 bp)
Encephalitozoon cuniculi (0 bp)

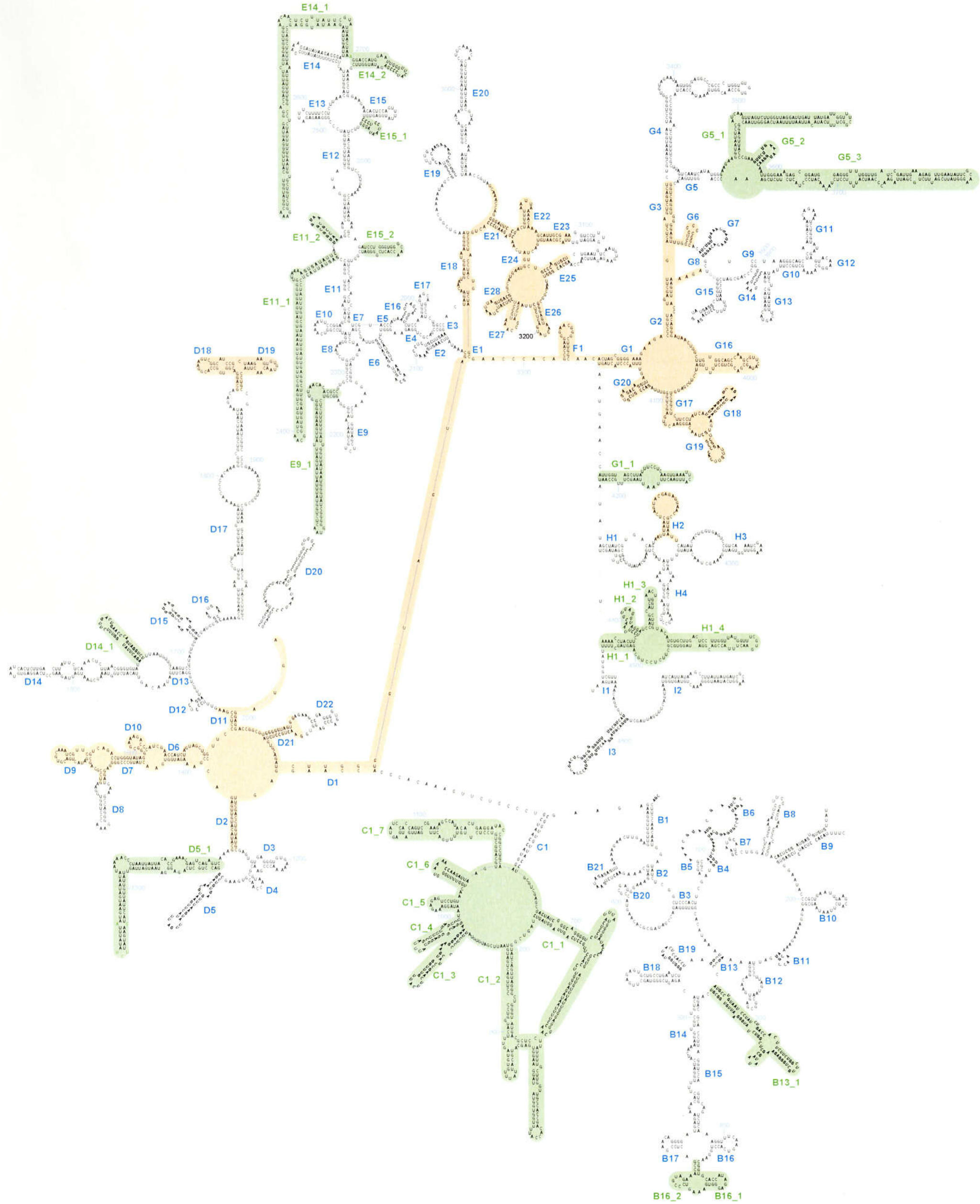


Fig. 2.10. Secondary structure model estimated from *Mastigamoeba balamuthi* LSUrRNA sequence. The universal core sequences are shaded in orange, and helices which show variability among species are shaded in green. The sequence is written clockwise from 5' to 3', sites are numbered in blue every 100 nucleotides. Helix numbering is according to De Rijk et al. (1999).

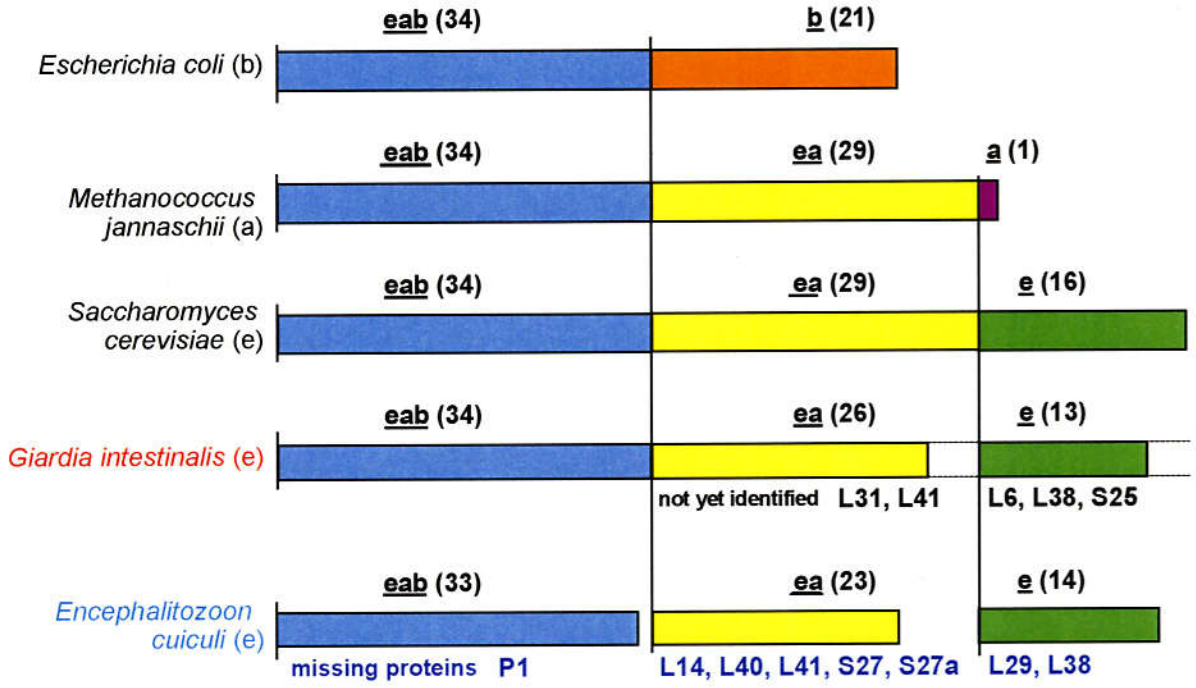


Fig. 2.11. Comparison of the protein components of the ribosome of *Giardia intestinalis* to the representative organisms from three primary kingdoms, *Eschelichia coli*, *Methanococcus jannaschii*, *Saccharomyces cerevisiae* and *Encephalitozoon cuniculi*. e: eukaryota; a: archaebacteria; b: eubacteria. Number of ribosomal proteins classfied into each class, such as eab, ea, e, a, b (see text), is shown in parentheses. On the *Giardia intestinalis* ribosome, N-terminal sequencing analyses of the proteins (Table 2.1), and a genome project database search have already identified 74 ribosomal proteins. On the *Encephalitozoon cuniculi* ribosome, ribosomal proteins are missing, compared to *S. cerevisiae* cytosolic ribosomal proteins. (Katinka et al. 2001).

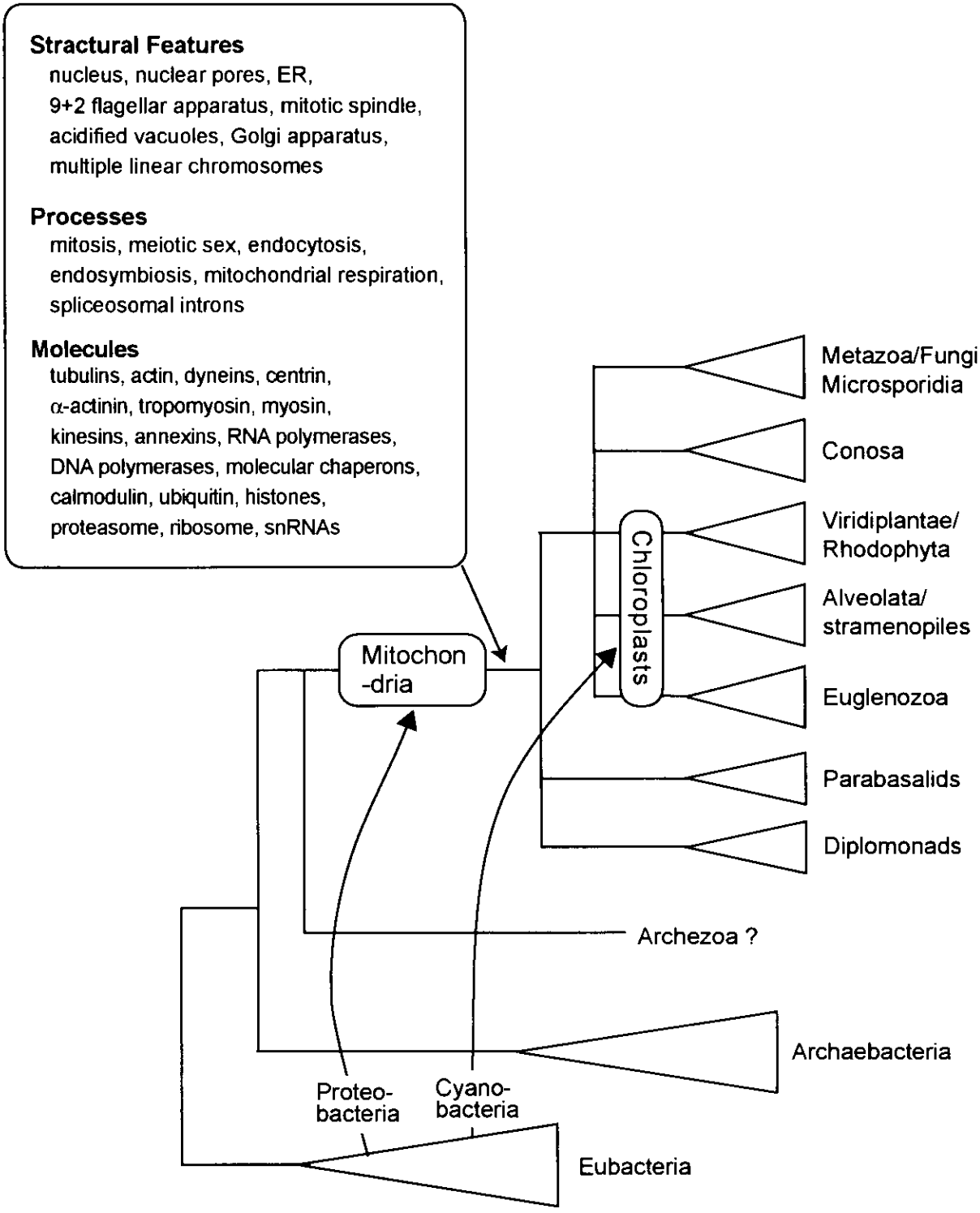


Fig. I. An updated view of early eukaryotic evolution