ELSEVIER

Gene 272 (2001) 131–139

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

The unusually long small subunit ribosomal RNA gene found in amitochondriate amoeboflagellate *Pelomyxa palustris*: its rRNA predicted secondary structure and phylogenetic implication

Irina A. Milyutina[a,*], Vladimir V. Aleshin[a], Kirill A. Mikrjukov[b], Olga S. Kedrova[a], Nikolai B. Petrov[a]

[a]*Section of Evolutionary Biochemistry, A.N. Belozersky Institute of Physicochemical Biology, Moscow State University, Moscow, 119899, Russia*
[b]*Biology Division, Moscow State University, Moscow, 119899, Russia*

Received 4 November 2000; received in revised form 24 February 2001; accepted 1 June 2001
Received by E. Sverdlov

Abstract

In order to ascertain a phylogenetic position of the freshwater amitochondriate amoeboflagellate *Pelomyxa palustris* its small subunit (SSU) rRNA gene was amplified and sequenced. It was shown to be 3502 bp long. The predicted secondary structure of its rRNA includes at least 16 separate expansion zones located in all the variable regions (V1–V9), as well as in some conservative gene regions. Most insertions are represented by sequences of low complexity that have presumably arisen by a slippage mechanism. Relatively conservative, uniformly positioned motifs contained in regions V4 and V7, as well as in some others, made it possible to perform folding. In maximum likelihood, maximum parsimony, and neighbor-joining trees, *P. palustris* tends to cluster with amitochondriate and secondary lost mitochondria amoebae and amoeboflagellates *Entamoeba*, *Endolimax nana*, and *Phreatamoeba balamuthi*, comprising together with them and aerobic lobose amoebae *Vannella*, *Acanthamoeba*, *Balamuthia*, and *Hartmannella* a monophyletic cluster. Another pelobiont, *Mastigamoeba invertens*, does not belong to this cluster. No specific similarity was discovered between the SSU rRNA of *P. palustris* and amitochondriate taxa of 'Archezoa': Diplomonada, Parabasalia, Microsporidia. *Pelomyxa palustris* SSU rRNA does not occupy a basal position in the phylogenetic trees and could be ascribed to the so-called eukaryotic 'crown' group if the composition of the latter were not so sensitive to the methods of tree building. Thus, molecular and morphological data suggest that *P. palustris* represents a secondarily modified eukaryotic lineage. © 2001 Published by Elsevier Science B.V. All rights reserved.

Keywords: Archamoebae; Pelobionts; V4 region; Expansion sequences; Slippage

1. Introduction

Pelomyxa palustris is a giant free-living amitochondriate amoeba with hundreds of nuclei and numerous immobile flagella. It is a common inhabitant of microoxic decaying organic sediment in stagnant backwater lakes and ponds. Traditionally, *P. palustris* was classified in the phylum Rhizopoda, along with Lobosea and other amoeboid organisms (Page, 1987). However, the phylogenetic position of this species is highly controversial (Griffin, 1988; Brugerolle, 1991; Goodkov and Seravin, 1991; Simpson et al., 1997). *Pelomyxa* lacks mitochondria or any obvious typically stacked Golgi dictyosomes, a contractile vacuole or most membrane-bound organelles, and is characterized by a very simple mastigote system (Seravin and Goodkov, 1987a,b; Griffin, 1988). These peculiarities as well as the existence of various archebacterial and eubacterial endosymbionts in its cytoplasm underlie the intriguing hypothesis that *P. palustris* is not related to other lobose amoebae but derives from some primitive eukaryotic lineage diverging before the earliest mitochondrial protists. According to this idea, the monotypic phylum Caryoblastea Margulis, 1974 was established for *P. palustris*. In recent classifications, this species is included in the phylum Archamoebae within the most primitive kingdom of eukaryotes, Archezoa (Cavalier-Smith, 1983). Kingdom Archezoa also includes some other free-living and parasitic amoeboid protists, the pelobionts, that inhabit anoxic environments and like *P. palustris* lack mitochondria and Golgi dictyosomes.

Abbreviations: ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor-joining; rDNA, DNA coding for rRNA; rRNA, ribosomal RNA; SDS, sodium dodecyl sulfate; SSU, small subunit

* Corresponding author. Tel.: +7-095-939-1440; fax: +7-095-939-3181.
E-mail address: petr@bioevol.genebee.msu.su (I.A. Milyutina).

0378-1119/01/$ - see front matter © 2001 Published by Elsevier Science B.V. All rights reserved.
PII: S0378-1119(01)00556-X

Until now molecular data have existed only for two pelobionts: *Mastigamoeba invertens* and *Phreatamoeba balamuthi* (considered by some authors as *Mastigamoeba*). The SSU rDNAs of both species have different structure. Whereas the sequence of *M. invertens* has a typical size (Stiller and Hall, 1999), the sequence of *P. balamuthi* is among the longest eukaryotic SSU rRNA genes (Hinkle et al., 1994). According to the phylogenetic analyses both sequences branch separately in the middle part of the SSU tree (Stiller and Hall, 1999). This is contrary, however, to the analysis of the *M. invertens* RNA polymerase II gene, which shows it as one of the earliest lineages of extant protists.

In order to resolve the controversy concerning the origin of pelobionts we have amplified and sequenced the SSU rRNA gene of *P. palustris*. Although the partial large subunit of the rRNA sequence of this species has been reported by Morin and Mignot (1995), this sequence has never been published. Therefore, the sequence data presented here are the first data on the SSU rRNA gene of *P. palustris*.

2. Materials and methods

2.1. Biological material and DNA manipulations

The specimens of *P. palustris* were collected in a Sphagnum swamp in the Solnechnogorsk district of the Moscow Region. The specimens were extracted from sapropel and each specimen was washed with sterile desaeriated distilled water under microscopic control. Collected cells were placed in solution containing 10 mM Tris–acetate (pH 8.2) and 50 mM EDTA and lysed by adding 0.5% SDS. DNA was isolated by phenol extraction and precipitated by ethanol. The entire SSU rRNA gene was amplified by polymerase chain reaction using universal eukaryotic primers (A, 5′-TCTGGTTGATCCTGCCAGT-3′ and B, 5′-TGATCCTTCTGCAGGTTCACCTAC-3′) for nuclear 16S-like rRNA coding regions and internal rDNA primers (direct d3, 5′-TGGAGGGCAAGTCTGGTG-3′ and d5, 5′-AAACTTAAAGGAATTGACG-3′ (corresponding to positions 939–956 and 2285–2303 in the sequence of the SSU rRNA gene of *P. palustris*), as well as reverse r5, 5′-TGGTGCCCTTCCGTCAAT-3′ and r7, 5′-GGGCGGT-GTGTACAAA-3′ (complementary to positions 2297–2314 and 3292–3307, respectively)). Products of amplification from three highly overlapping fragments (A, r5; d3, r7; d5, B) were purified in agarose gel and cloned in pGEM-T Vector (Promega). The clones obtained proved to be identical based on restriction maps obtained after cleaving with *Csp*6I and *Taq*I. One to three clones for each of the amplified fragments were sequenced using the *fmol* DNA sequencing system (Promega) on both strands. Polymorphic sites were designated as ambiguous. The sequence has been deposited in GenBank under Accession number AF320348.

2.2. Structural and phylogenetic analysis

The secondary structure was inferred on the basis of the universal model of eukaryotic SSU rRNA (Van de Peer et al., 2000). The secondary structure of the variable and inserted regions was predicted with the *mfold* program version 3.0 (Zuker et al., 1999). The choice between the suboptimal structures was based on the presence of short conserved marker strings as indicated below. Graphic output was made with RnaViz (De Rijk and De Wachter, 1997).

For phylogenetic analysis the SSU rRNA sequence of *P. palustris* and several missing sequences were aligned manually to the set of protist SSU rRNA sequences (Pawlowski et al., 1999). The alignment used is accessible by request (e-mail: petr@bioevol.genebee.msu.su). Hypervariable regions were rejected from the phylogenetic analysis and the final matrix included 1209 unambiguously aligned nucleotides. Maximum likelihood (ML), maximum parsimony (MP) and neighbor-joining (NJ) trees were inferred with fastDNAml version 1.1.1a (Olsen et al., 1994), Dnapars from the PHYLIP 3.572 package (Felsenstein, 1993), PAUP version 4.0.0d55 (Wisconsin Package Version 9.0, GCG), and Neighbor from the PHYLIP package, respectively, under various parameters. MP and NJ analyses were performed with 1000 bootstrap replicates and ML ones with 20 replicates. In MP and ML analyses, correction for the rate heterogeneity over sites was performed using results of preliminary PUZZLE analysis under the model of Tamura and Nei (Strimmer and von Haeseler, 1996). The set of suboptimal trees obtained in the ML analysis was compared by the Kishino–Hasegawa test.

3. Results

3.1. Structural features of the P. palustris SSU rRNA gene

The *P. palustris* SSU rRNA gene (3502 nucleotides in length) is among the longest SSU rRNA genes described so far. Eight nucleotides positions were described as polymorphous through clones. The RNA predicted secondary structure (Fig. 1) suggests that it contains all the elements typical of eukaryotes except for probably helix 4. Additional nucleotides responsible for the extraordinary length of *Pelomyxa* SSU are distributed among at least 16 insertions, most of them located in variable zones of the gene. All the variable zones V1–V9 including zone V6 variable in prokaryotes and relatively conservative helices 8, 16, and 24 were targets for expansion. As a rule, the insertions have a low complexity, and extensive purine and pyrimidine blocks are typical of these segments. Most insertions are incomplete palindromes and form predicted helices that are added to loops of the universal hairpins. As a result, hairpins 6, 16, 17, E23_7, 37, and 46 look as composite. The basal parts of these helices are represented by sequences with a relatively

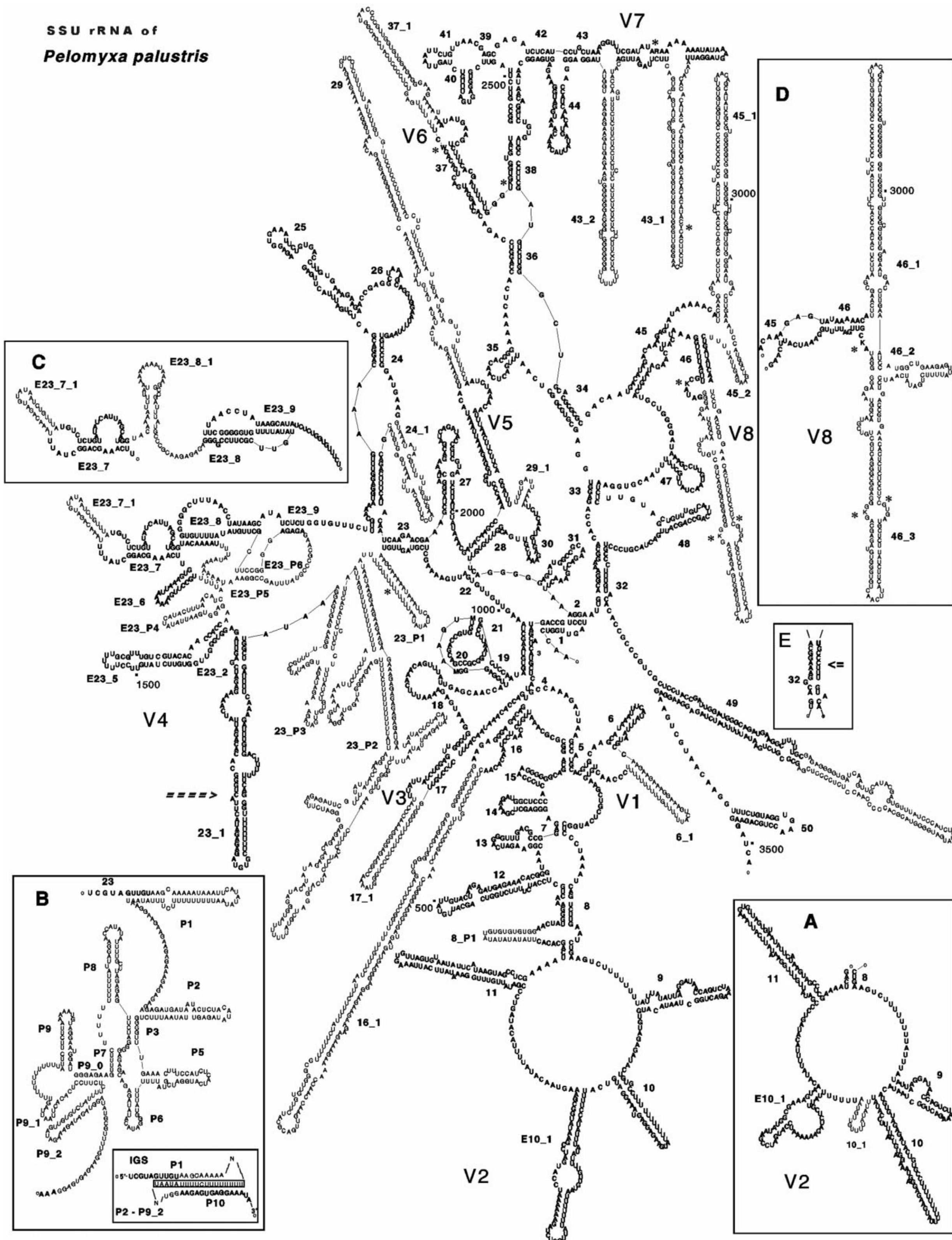

Fig. 1. Predicted secondary structure of *P. palustris* SSU rRNA. The helices are numbered in order of occurrence from the 5′ terminus. Those common to all eukaryotes sequence regions are in bold, and those specific of *Pelomyxa* are in standard characters. Eukaryotic and *Pelomyxa* specific helices are prefixed with an ‘E’ and ‘P’, respectively. Insets (A–E) represent the alternative variants of folding for some helices. The small inset in (B) shows the possible structure of helix P10 comprising the internal guide sequence. Asterisks indicate polymorphic positions among clones. The short arrow in (E) indicates the non-canonical pair AC, and the long punctuated arrow indicates the conservative motif ACUG in the 3′ branch of helix 23_1.

conservative primary structure in both branches, and their extremely long apical parts are formed by unique sequences of low complexity.

In the V1 region, the inserted helix containing blocks of 15 purine and 14 pyrimidine residues branches off from the internal loop of helix 6. A small insertion $(UA)_5(UG)_5$ is a side branch from the internal loop of helix 8. The preceding conservative motive UCA that is a part of helix 8 allows this insertion to be located accurately on the secondary structure model. This motif flanks the 3′ end of the hypervariable region V2. The 5′ end of the region V2 is flanked with a conservative trinucleotide (GGCUAAUAC)AUG which is included in helix 9 rather than bordering it, as it was proposed for rRNA of *Saccharomyces cerevisiae* (Van de Peer et al., 2000) and pelobiont *P. balamuthi* (Hinkle et al., 1994). Each of the three helices 10, E10_1, and 11 seems moderately elongated if folding occurs in accordance to the max $-\Delta G$. Helix E10_1 is short and imperfect (Fig. 1A) if its folding is done in accordance with a canonical distribution of typical nucleotide motifs (UUAUU and AGUAA as basal parts of 5′ and 3′ branches, respectively, A_6 in the internal loop, unmatched WU and CAU which are adjacent to the helix from the 5′ end and 3′ end, respectively).

Four expansion zones are distinguished within the hypervariable region V4 of *P. palustris*. For their interpretation, it is necessary to distinguish hairpin 23_1, which is a stable element of the secondary structure in the V4 region though it is hypervariable in terms of its primary structure. Predicted hairpin 23_1 in the central part of the V4 consists of a long imperfect helix with two internal loops and the conservative motif ACUG in the 3′ branch (Fig. 1, long arrow). In the same place, the motif ACUG is often found in E23_1 of eukaryotes as well as in P23_1 of archebacteria and eubacteria, which testifies to their homology and explains why this helix is numbered 23_1 without indices E or P. An area of 354 nucleotide residues, typical of *P. palustris*, with a biased composition where U accounts for 40% and C for just 9%, precedes the 23_1 helix. Fig. 1 shows one of several options of the folding of this fragment as separate hairpin 23_P1 and branching hairpins 23_P2 and 23_P3. This area can be folded in another fashion that gives a group I intron-like structure with pseudoknot and helices P1–P9 (Fig. 1B).

In *P. palustris* there are helices E23_2 and E23_5 recognized by the motif CUUNNNUG that is typical of the apical part of E23_5. Moreover, its SSU rRNA has an additional unique helix, which precedes E23_6 and is designated as E23_P4.

The secondary structure of the pseudoknot in V4 may be represented by two different alternatives. The preferable alternative was received to make sure that the conservative sequence AUUUUGU was included in a double-helix site as a 3′ branch of E23_8 since a similar motif is often found in the 3′ branch of the helix E23_8 of other organisms. It ensures enrichment with residues G typical of the loop E23_9. Additional looped-out nucleotide residues form a unique E23_P6 helix in the pseudoknot. The other alternative (Fig. 1C) implies the inclusion of the additional residues in the helix that precedes the pseudoknot, just as was predicted for strepsipters (Choe et al., 1999b).

There are numerous substitutions of the conservative residues and an extensive insertion in the area of helix 24 of *P. palustris*. This insertion is located in the internal loop of the helix forming a peculiar hairpin that is numbered 24_1. Its predicted structure was chosen from several suboptimal structures because it retained the relatively conservative sequence ANGNA in the internal loop of the 3′ branch of helix 24. In the V6 region, which is variable for prokaryotes, relatively conservative nucleotide residues unambiguously mark the localization of the additional nucleotides that form the helix attached to loop 37. On the other hand, for helix 29, which is three times longer than the usual helix, it is difficult to find the boundary of the added segment. Because elongated helix 29 was predicted to be non-branched, we can only suppose that most residues were added to the loop of the initial helix.

In V7 of *P. palustris*, two additional hairpins branch off the helix 43 3′ branch. A conservative unpaired dinucleotide AU in its 5′ and trinucleotide UAG in its 3′ branch were predicted between these additional hairpins. These motifs are similar to motifs AU and UM in the internal loop of helix 43 in other organisms. Thus, hairpins 43_1 and 43_2, flanking the central conservative internal loop, appear to be the result of two independent insertion events. Branching of hairpin 43 in region V7 was also described for several species from various taxa. For example, in *P. balamuthi* these additional hairpins branch off near the apical loop of hairpin 43 (Hinkle et al., 1994), whereas in Strepsiptera (Choe et al., 1999b) and Myxozoa additional hairpins branch off the 5′ branch of helix 43. Earlier, an observation was made that the size and the base composition of the hypervariable region V7 correlate with those of the hypervariable region V4 (Crease and Colbourne, 1998), but not in all the cases studied so far (Choe et al., 1999a). In the rRNA of *P. palustris*, there are several long insertions in V4 and V7 regions, but their nucleotide composition is significantly different: it is 52% G + C for two additional helices of region V7 and 28.5% G + C for several additional helices located between helices 23 and 23_1, i.e. for helices including the bulk of the additional nucleotides in V4.

The hypervariable region V8 of *P. palustris* rRNA is 250 nucleotides longer than the same region of the typical molecule. Undoubtedly, all these nucleotides precede the 3′ base of helix 46, as demonstrated by the retaining of segment GGAAAUC, where the central part corresponds to the relatively conservative motif of the single-stranded bridge between helices 46 and 45. Two major patterns can be found among the suboptimal structures inferred by *mfold* for the region V8. According to one of them, the additional nucleotides form helix 45_1, whereas to another this helix is not formed. The first predicted pattern resembles long helix 45_1 in *P. balamuthi*, species of genera *Entamoeba* and

Acanthamoeba, haemosporidiae (Van de Peer et al., 2000) as well as foraminifers (Pawlowski et al., 1999). Short helix 45_1 is characteristic of rRNA of *Endolimax*, kinetoplastides, and acrasiales, while *Vannella anglica* has a single-stranded loop in the same region. This helix is absent in rRNA of a pelobiont *M. invertens*. Despite the lack of obvious motifs in the sequences of the region E45_1, a significant bias of the base composition similar in *Pelomyxa*, *Phreatamoeba* and *Entamoeba* is observed in this region. In all these organisms, the $5'$ branch of helix E45_1 is markedly enriched with pyrimidine residues, and the $3'$ branch with purine residues. Ratios R/Y in the $5'$ and $3'$ branches of helices E45_1 are 0.42 and 2.00, 0.21 and 4.67, 0.38 and 5.00, and 1.17 ± 0.28 and 1.41 ± 0.28 for *Pelomyxa*, *Phreatamoeba*, *Endolimax*, and eight species of *Entamoeba*, respectively, whereas in unrelated *Acanthamoeba* such a bias of the base composition is not observed. For ten selected species of *Acanthamoeba*, the ratios are 1.19 ± 0.11 and 0.81 ± 0.11. One can suppose that the unrelated occurrences of expansion into this region underlie the differences in purine/pyrimidine ratios of region E45_1 between Archamoebae on the one hand and *Acanthamoeba* on the other. Thus, in this pattern hairpin 46 in *Pelomyxa* appears to be unprecedentedly long.

The major peculiarity of another predicted pattern of V8 in *P. palustris* is an inclusion of an alternative segment of RNA as a $5'$ branch in helix 46, as well as redistribution of some relatively conservative residues between double-stranded and single-stranded segments in the distal part of helix 45. Additional nucleotides are also distributed among three adjacent helices that branch off the hairpin 46 loop. Thus, in this model hairpin 46 appears typical (Fig. 1D) if three additional helices are not taken into account.

In *P. palustris* rRNA, a small peculiarity is observed in helix 32 at the border of domain III. The predicted secondary structure of helix 32 is usually highly conservative and characterized by two single residues that are looped from each of the branches at the base of the helix. In the case of *P. palustris*, such a canonical structure can be formed only if a non-Watson–Crick pair AC would be allowed close to the middle part of the helix (Fig. 1E).

3.2. Search of the closest phylogenetic relatives of *P. palustris*

NJ, MP and ML SSU rDNA sequence trees (Fig. 2) show that *P. palustris* tends preferably to cluster either with a group of species that corresponds closely to the Archamoebae established by Cavalier-Smith (1983, 1998) or, less strongly, with the Foraminifera. The cluster consisting of *Pelomyxa*, *Phreatamoeba*, *Entamoeba*, and *Endolimax* is present in 1357 of 1563 or in 6656 of 7123 suboptimal ML trees (inferred without or with correction for the rate heterogeneity over sites, respectively) that do not differ from one another by the Kishino–Hasegawa test. All species in this cluster are either amitochondriate amoebae or amoeboflagellates. The same cluster is supported by 68% bootstrap replicates in the weighted MP analysis and by only 25% in the NJ analysis though the latter is still the most frequent cluster formed by SSU rDNA *Pelomyxa*.

The second taxon whose SSU rDNA sequences tend to cluster with the SSU rDNA sequence of *Pelomyxa* is Foraminifera. The MP analysis using a PAUP heuristic search without an adjustment on the site-to-site rate heterogeneity unites *Pelomyxa* with Foraminifera in 32% bootstrap replicates compared to only 23% bootstrap replicates uniting *Pelomyxa* with Archamoebae. The NJ analysis links *Pelomyxa* to Foraminifera in only 18% bootstrap replicates.

In SSU rDNA phylogenetic trees, the lobose amoeba *V. anglica* appears to be a sister group to the cluster Archamoebae. It is noteworthy that *M. invertens* that was thought to be a representative of Archamoebae does not belong to this cluster, and very rarely, in less than 10% bootstrap replicates, joins it after *V. anglica*.

3.3. Phylogenetic position of Archamoebae as deduced from the rRNA analysis

The phylogenetic analysis of the selected data on the protists rDNA strongly supports the clades of stramenopiles, alveolates (without Haplosporidia), Cercozoa (cercomonadic amoeboflagellates and filose amoebae), and Lobosea, all of which are traditionally distinguished by the SSU rRNA analysis. Two other traditional clusters, Metazoa + Fungi and Euglenozoa (with Heterolobosea), have weaker support. However, traditional partition of the eukaryotic tree into the early branching lineages and so-called ‘crown’ is not observed in our analysis. Thus, in the ML analysis with the gamma adjustment on the site-to-site heterogeneity rate, Microsporidia and Diplomonada branch among the ‘crown’ taxa (Fig. 2B). Moreover, in our ML and NJ analyses as well as in Pawlowski et al. (1999), two classes of Radiolaria form a monophyletic cluster. This is in conflict with the results of the earlier original study by Amaral Zettler et al. (1997) where the first class, Acantharea, was shown to belong to the ‘crown’ group while the second one, Polycystinea, was declared to branch off before the radiation of the ‘crown’ group.

In ML SSU rDNA sequence trees, Archamoebae form a monophyletic cluster with lobose amoebae. The support of this cluster increases with increasing log likelihoods. For example, in ML analyses without a correction for the rates of heterogeneity over sites, it varies in the range of 64–86% under changes of the lower limits of lnL for suboptimal trees from -22883.07574 to -22838.95518. The rest of the amitochondriate protists (Microsporidia, Diplomonadida, and Parabasalia), initially included in Archezoa together with Archamoebae, branch far from *Pelomyxa*. When rooting the SSU rDNA tree with any prokaryotic SSU rRNA as an outgroup, this outgroup does not link with the branch leading to *Pelomyxa*, which should occur under the hypothesis on the early origin of *Pelomyxa*. Instead the bacterial

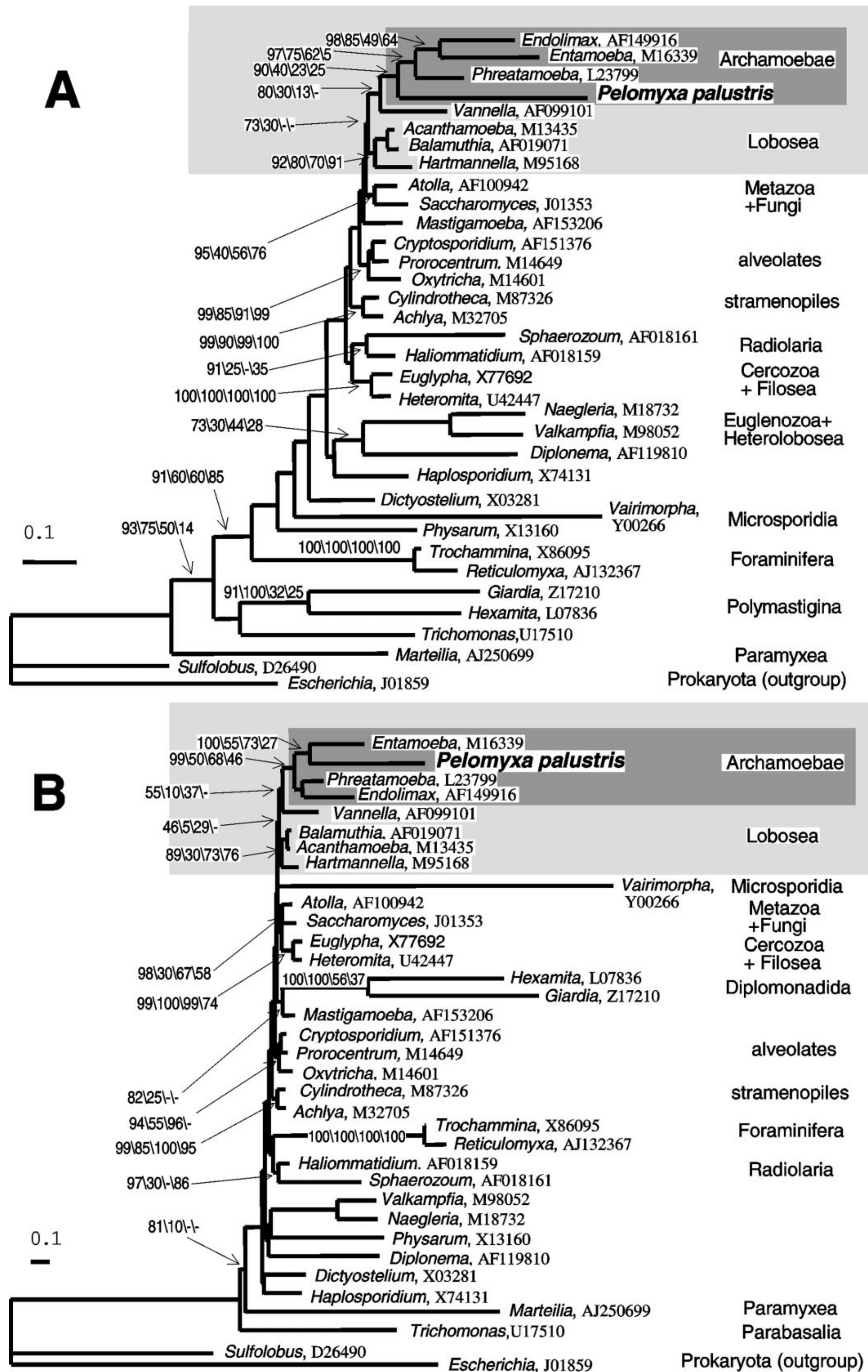

Fig. 2. Phylograms of eukaryotic SSU rRNA representing a consensus from 1000 suboptimal ML trees. (A) Trees were derived without any correction for the rate heterogeneity over sites. (B) Trees were derived using gamma correction for the rate heterogeneity over sites. Numbers at nodes represent the percentage support of major clusters derived from ML suboptimal trees granted by the Kishino–Hasegawa test and bootstrap support (%) in ML, MP, and NJ analyses (for A), or the percentage of quartet puzzling (for B) (the fourth number, if present). The horizontal scale is a distance scale for ML analysis.

sequences normally attract the long branches leading to *Giardia* and other Diplomonada, Parabasalia, and, especially often, to *Marteilia* (Paramyxea) whose SSU rDNA sequence was recently described to be one of the most diverged eukaryotic SSU rDNA sequences. Microsporidiae tested in our set of SSU rDNA sequences do not occupy the most basal position in the phylogenetic tree, which is consistent with the modern interpretation of their phylogenetic position (Hirt et al., 1999).

4. Discussion

4.1. Origin of expansion segments

Unusually long SSU rRNA genes were discovered earlier in a variety of unrelated species: foraminiferal protists (Pawlowski et al., 1996), strepsipter insects, crustacean isopods, some fungi and protist *Plasmodiophora brassica*. SSU rRNA genes much longer than 2000 bp were described in many taxa of protists and animals. Almost in all these cases, expansion segments that make the SSU rRNA gene longer are retained in the mature rRNA molecule. The only known exceptions are the Foraminifera whose mature SSU rRNA, as deduced from Northern blots, looses numerous expansion segments (Pawlowski et al., 1996) and various species of fungi, protists and algae whose SSU rRNA genes were shown to contain group I introns (Ivanova et al., 1998). Although the size of the mature SSU rRNA of *P. palustris* was not determined a low complexity of the insertions makes them more similar to the expansion segments that arise by the slippage mechanism (Tautz et al., 1988; Hancock and Vogler, 1998) than to the degenerate group I introns. The only exception is the insertion of 354 nucleotides preceding the 23_1 helix that comprises a complex of several helices. Although this insertion can be folded in a group I intron-like structure (Fig. 1B), it is energetically less profitable than its alternative shown in Fig. 1. Moreover, this insertion also has a low complexity, and we could not discover in it conserved canonical sequences P, Q, R, and S typical of group I introns.

In each of foraminiferal and some other taxa species, expansion segments occupy the same positions in rDNA, but differ strongly in terms of their primary structures. There is no doubt that in these cases they were inherited from the common ancestor, and sometimes paleontological records allow the time of their persistence to be established. For example, in rDNA of sternorhynchs (Insecta) and cladocerans (Crustacea), expansion segments originated in the Lower Permian when the common ancestor of these arthropods emerged, i.e. 350 million years ago, while rDNA of Foraminifera has had expansion segments since at least the Lower Cambrian, i.e. 540 million years or even since an earlier time. In lobose amoebae *Acanthamoeba* and archamoebae *Entamoeba*, expansion segments are widespread in many, though not all species. Although long SSU rRNA genes were described for all other archamoebae, *Pelomyxa*, *Phreatamoeba*, and *Endolimax*, there is no evidence that in these cases expansion segments originated from the common ancestor, since many of them are located in hypervariable regions representing the targets for independent expansions. It is more likely that massive expansion into SSU rRNA genes seems to be a common evolutionary tendency in these clades. In all of the presumable archamoebae, this tendency is not exhibited in *M. invertens*.

4.2. Is M. invertens a real archamoeba?

According to the modern taxonomic review, the genus of *Phreatamoeba* is a synonym of *Mastigamoeba* (Simpson et al., 1997). However in ML, MP, and NJ trees, not only do SSU rRNA genes of *M. invertens* and *P. balamuthi* not cluster with one another under any parameters of the analysis used but they also exhibit a significantly different affinity from respective sequences of other 'archamoebae': *Pelomyxa*, *Entamoeba*, and *Endolimax*. For this reason, we find it convenient to use the original name *P. balamuthi* from the description of the species by Chavez et al. (1986) until there is more reliable evidence of monophyly or polyphyly of the genus *Mastigamoeba*.

The trend of all the available SSU rDNA sequences of archamoebae (except for *M. invertens*) and aerobic amoeba *V. anglica* to cluster together in phylogenetic trees of SSU rDNA cannot be readily interpreted. All these sequences have diverged strongly from the SSU rRNA sequences of the eukaryotic core. This circumstance allows suspicion of their artificial clustering due to the effect of so-called long branch attraction. If this feature plays a major role in their clustering we must restrict ourselves to the trivial fact of the isolated taxonomic position of *Pelomyxa* and reject the hypothesis that brings it closer to any known form. However, the widespread opinion that long branches are clustered on the tree in an artificial manner is debatable. In the tested set of SSU rDNA sequences, there were many sequences forming long branches in the phylogenetic trees, but except for those of Archamoebae and Foraminifera, they do not exhibit a tendency to cluster with the SSU rRNA sequence of *Pelomyxa*. The clustering of *Pelomyxa* with Foraminifera is more likely to be artificial due to the long branch attraction because it is supported by more variable sites and gets reduced with the weight reduction of these sites in weighted MP analyses. In contrast, the obtained cluster of Archamoebae is relatively stable under any parameters and methods of tree building.

4.3. Archamoebae are monophyletic

Thus, the available data on SSU rDNA suggest monophyly of a group of archamoebae, whose composition is slightly different from that proposed by Cavalier-Smith (1998). In our analysis, Archamoebae include *Pelomyxa*, *Phreatamoeba*, *Endolimax*, and *Entamoeba*, but not *M. invertens*. Monophyly of these genera, except for *Pelomyxa*,

was supported by earlier molecular evidence (Silberman et al., 1999). Archamoebae of the new composition includes multiflagellate, uniflagellate and aflagellate forms, and the evolution of the mastigote system within it must be non-parsimonious. It is difficult to infer the organization of the group ancestor, and modern representatives of the Archamoebae share too few common features that could not be easily explained by convergence. So far, the molecular analysis has not revealed any features, such as conservative motifs or specific elements of the secondary structure, that could be regarded as obvious autapomorphy of Archamoebae.

Although *V. anglica* is not related to Lobosea according to earlier SSU rDNA analysis (Sims et al., 1999), it might be close to Archamoebae as follows from our ML analysis. These results support the hypothesis about monophyly of Amoebozoa (Cavalier-Smith, 1998) consisting of all aerobic and anaerobic lobose amoebae, Lobosea + *Vannella* + Archamoebae, which is viewed as the residue of the Sarcodina, but the support for this group is less reliable than of Archamoebae.

4.4. Pelomyxa is secondarily amitochondriate

Though the gene of SSU rRNA of *Pelomyxa* is nearly twice as large as the typical SSU rRNA gene, its unusual length does not seem to be a primitive characteristic. The *Pelomyxa* SSU rRNA gene appears to be derived from the ordinary eukaryotic gene, and does not exhibit any similarity with prokaryotic genes, which could be considered as simplesiomorphies. Moreover, amitochondriate Parabasalia and Metamonada, the only two groups that were left in the Archezoa group (Cavalier-Smith, 1998) do not exhibit any similarity of their rDNA with that of *Pelomyxa*. This could be regarded as a strong argument against the basal offshoot of archamoebae. However, if one admits that the basal place of Archezoa is nothing more than a result of the long branch attraction feature, then *Pelomyxa* as well as any other highly evolved taxon of the so-called 'crown' could be the earliest eukaryotic branch. The ancestor of the archamoebae can be characterized a little bit more accurately through its morphology. The clustering of archamoebae with aerobic *V. anglica* and other aerobic lobose amoebae would suggest that their ancestor was aerobic. It is well known that in *Entamoeba* the lack of mitochondrion is a derived and not ancestral state. The genome of *Entamoeba* contains the protein genes that are typical of mitochondrion, and a cryptic, presumably mitochondrion-derived organelle, mitosome, was recently described for the cells of *Entamoeba histolitica* (Mai et al., 1999; Tovar et al., 1999). Like Parabasalia and Metamonada, which are thought to be the earliest recent eukaryotic lineages developed on the basis of the typical eukaryotic cell (Embley and Hirt, 1998; Keeling and Palmer, 2000), the ancestor of *Pelomyxa* should be a typical aerobic eukaryote having mitochondria, Golgi dictyosomes, and an unreduced mastigote system (Griffin, 1988; Goodkov and Seravin, 1991).

Acknowledgements

We thank Dr J. Pawlowski for kindly providing the set of aligned protist SSU rRNA sequences and for discussion, Dr N.I. Ivanova for valuable consultations on the group I introns, Drs A.V. Goodkov and A.S. Antonov for discussion, and two anonymous reviewers for helpful comments on the manuscript. This research was supported by RFBR grants 99-04-48840 and 00-15-97905, as well as grant 97-10-140 of the State Committee of Education of Russia.

References

Amaral Zettler, L., 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.

Brugerolle, G., 1991. Flagellar and cytoskeletal systems in amitochondrial flagellates – Archamoeba, Metamonada and Parabasalia. Protoplasma 164, 70–90.

Cavalier-Smith, T., 1983. A 6-kingdom classification and a unified phylogeny. In: Schwemmler, W., Schenk, H.E.A. (Eds.), Endocytobiology, II. De Gruyter, Berlin, pp. 1027–1034.

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

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, 397–406.

Choe, C.P., Hancock, J.M., Hwang, U.W., Kim, W., 1999a. Analysis of the primary sequence and secondary structure of the unusually long SSU rRNA of the soil bug, *Armadillidium vulgare*. J. Mol. Evol. 49, 798–805.

Choe, C.P., Hwang, U.W., Kim, W., 1999b. Putative secondary structures of unusually long strepsipteran SSU rRNAs and its phylogenetic implications. Mol. Cell 9, 191–199.

Crease, T.J., Colbourne, J.K., 1998. The unusually long small-subunit ribosomal RNA of the crustacean, *Daphnia pulex*: sequence and predicted secondary structure. J. Mol. Evol. 46, 307–313.

De Rijk, P., De Wachter, R., 1997. RnaViz, a program for the visualisation of RNA secondary structure. Nucleic Acids Res. 25, 4679–4684.

Embley, T.M., Hirt, R., 1998. Early branching eukaryotes? Curr. Opin. Genet. Dev. 8, 624–629.

Felsenstein, J., 1993. PHYLIP – Phylogeny Inference Package, version 3.5, University of Washington, Seattle, WA.

Goodkov, A.V., Seravin, L.N., 1991. New ideas on the nature of the giant amoeba *Pelomyxa palustris* – the position of this organism in the system of lower eukaryotes (Pelloflagelata – Classis n.) (in Russian). Zool. Zh. 70, 5–16.

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

Hancock, J.M., Vogler, A.P., 1998. Modelling the secondary structures of slippage-prone hypervariable RNA regions: the example of the tiger beetle 18S rRNA variable region V4. Nucleic Acids Res. 26, 1689–1699.

Hinkle, G., Leipe, D.D., Nerad, T.A., Sogin, M.L., 1994. The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. Nucleic Acids Res. 22, 465–469.

Hirt, R.P., Logsdon Jr., J.M., Healy, B., Dorey, M.W., Doolitle, W.F., Embly, 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.

Ivanova, N.V., DePriest, P.T., Bobrova, V.K., Troitsky, A.V., 1998. Group I introns in fungal 18S rDNA of the lichen family Umbilicariaceae. Dokl. Akad. Nauk 363, 400–404.

Keeling, P.J., Palmer, J.D., 2000. Parabasalian flagellates are ancient eukaryotes. Nature 405, 635–637.

Mai, Z., Ghosh, S., di Frisar, 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.

Morin, L., Mignot, J.-P., 1995. Are Archamoebae true Archezoa? The phylogenetic position of *Pelomyxa* sp., as inferred from large subunit ribosomal RNA sequencing. Eur. J. Protistol. 31, 402.

Olsen, G.J., Matsuda, H., Hagstrom, R., Overbeek, R., 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. 10, 41–48.

Page, F.C., 1987. The classification of the naked amoebae (Phylum Rhizopoda). Arch. Protist. 133, 199–217.

Pawlowski, J., Bolivar, I., Fahrni, J.F., Cavalier-Smith, T., Gouy, M., 1996. Early origin of Foraminifera suggested by SSU rRNA gene sequences. Mol. Biol. Evol. 13, 445–450.

Pawlowski, J., Bolivar, I., Fahrni, J.F., De Vargas, C., Bower, S.S., 1999. Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. J. Eucaryot. Microbiol. 46, 612–617.

Seravin, L.N., Goodkov, A.V., 1987a. Golgi apparatus in the amoeba *Pelomyxa palustris*. Dokl. Akad. Nauk + 296, 249–250.

Seravin, L.N., Goodkov, A.V., 1987b. The flagella of the fresh-water amoeba *Pelomyxa palustris*. Tsitologiya + 29, 721–724.

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 organization of *Mastigamoeba schizophrenia* n. sp.: more evidence of ultrastructural idiosyncrasy and simplicity in pelobiont protists. Eur. J. Protistol. 33, 87–98.

Sims, G.P., Rogerson, A., Aitken, R.J., 1999. Primary and secondary structure of the small-subunit ribosomal RNA of the naked, marine amoeba *Vannella anglica*: phylogenetic implications. J. Mol. Evol. 48, 740–749.

Stiller, J.W., Hall, B.D., 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. Mol. Biol. Evol. 16, 1270–1279.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., Dover, G.A., 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. 5, 366–376.

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.

Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., De Wachter, R., 2000. The European Small Subunit Ribosomal RNA database. Nucleic Acids Res. 28, 175–176.

Zuker, M., Mathews, D.H., Turner, D.H., 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski, J., Clark, B.F.C. (Eds.), RNA Biochemistry and Biotechnology, NATO ASI Series. Kluwer Academic, Dordrecht, pp. 11–43.