J Mol Evol (1996) 43:551–562

JOURNAL OF MOLECULAR EVOLUTION

Molecular Phylogeny of the Free-Living Archezoan *Trepomonas agilis* and the Nature of the First Eukaryote

T. Cavalier-Smith, E.E. Chao

Evolutionary Biology Program, Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, V6T 1Z4, Canada

Received: 18 April 1996 / Accepted: 17 July 1996

Abstract. We have sequenced the small ribosomal subunit RNA gene of the diplozoan *Trepomonas agilis*. This provides the first molecular information on a free-living archezoan. We have performed a phylogenetic analysis by maximum likelihood, parsimony, and distance methods for all available nearly complete archezoan small subunit ribosomal RNA genes and for representatives of all major groups of more advanced eukaryotes (metakaryotes). These show Diplozoa as the earliest-diverging eukaryotic lineage, closely followed by microsporidia. *Trepomonas* proves to be much more closely related to *Hexamita,* and, to a lesser degree, to *Spironucleus,* than to *Giardia.* The close relationship between the free-living *Trepomonas* on our trees and the parasites *Hexamita inflata* and *Spironucleus* refutes the idea that the early divergence of the amitochondrial Archezoa is an artefact caused by parasitism. The deep molecular divergence between the three phagotrophic genera with two cytostomes (*Hexamita, Trepomonas, Spironucleus*) and the saprotrophic *Giardia* that lacks cytostomes is in keeping with the classical evidence for a fundamental difference in the symmetry of the cytoskeleton between the two groups. We accordingly separate the two groups as two orders: Distomatida for those with two cytostomes/cytopharynxes and Giardiida ord. nov. for *Giardia* and *Octomitus* that lack these, and divide each order into two families. We suggest that this fundamental divergence in manner of feeding and in the symmetry of the cytoskeleton evolved in a free-living diplozoan very early indeed in the evolution of the eukaryotic cell, possibly very soon after the origin of the diplokaryotic state (having two nuclei linked together firmly by the cytoskeleton) and before the evolution of parasitism by distomatids and giardiids, which may have colonized animal guts independently. We discuss the possible relationship between the two archezoan phyla (Metamonada and Microsporidia) and the nature of the first eukaryotic cell in the light of our results and other recent molecular data.

Key words: Diplozoa — Cytostomes — Diplokaryosis — *Trepomonas* — Ribosomal RNA phylogeny

Introduction

The three earliest-diverging eukaryotic phyla on most molecular trees are the amitochondrial Metamonada (Sogin et al. 1989a,b), Microsporidia (Vossbrinck et al. 1987), and Parabasala. It has been difficult to establish which of these three groups diverged first (Leipe et al. 1993; Cavalier-Smith 1993a; Philippe and Adoutte 1995; Galtier and Gouy 1995). The Metamonada and Microsporidia, classified together in the protozoan subkingdom Archezoa (Cavalier-Smith 1983, 1996a) because unlike Parabasala they lack hydrogenosomes and permanent Golgi dictyosomes, are widely considered to be the most primitive eukaryotes and to be primitively without mitochondria. This view has been criticized by Wolters (1991) and Siddall et al. (1992), who have suggested that the deep branching of these three phyla may be an arte-

Correspondence to: T. Cavalier-Smith

fact caused by exceptionally high rates of molecular evolution resulting from parasitism. Microsporidia are all obligate intracellular parasites. Parabasala (trichomonads and hypermastigotes) were traditionally thought to be all inhabitants of animal guts, but a few free-living trichomonads have recently been discovered (Farmer 1993). Metamonada comprise three classes: Oxymonadea, which are purely gut symbionts; Retortamonadea, which are also gut symbionts but may also have free-living members (Farmer 1993); and Trepomonadea, which include the enteromonads (all of which are gut symbionts) and the diplomonads, which are free-living as well as gut symbionts. Despite the fact that there is neither a clear theoretical basis nor empirical evidence for the idea that parasitism accelerates molecular evolution, it is clearly desirable to test the ideas of Wolters (1991) and Siddall et al. (1992) by obtaining molecular sequences from free-living metamonads.

We have accordingly sequenced the small subunit ribosomal RNA gene from *Trepomonas agilis,* a free-living diplozoan (the earlier name for diplomonads, which seems to be an unnecessary junior synonym for diplozoans) belonging to the class Trepomonadea of the phylum Metamonada. We have carried out a molecular phylogenetic analysis by three different methods, which all show that *Trepomonas* is more closely related to *Hexamita* than to other diplozoans. Our analysis clearly rules out the possibility that the early divergence of diplozoa is merely the result of parasitism. Although it is impossible to be sure that diplozoa are the earliest-diverging eukaryotes, as our results suggest, their cellular properties are of key importance for understanding the nature of the first eukaryotic cells. We attempt to integrate molecular and cell biological evidence for the early evolution of diplozoa and discuss recent molecular evidence (Soltys and Gupta 1994) that might perhaps be interpreted as evidence that diplozoans (and possibly even all other Archezoa) are secondarily amitochondrial, like *Entamoeba* (Clark and Roger 1996) and several other eukaryotes.

Materials and Methods

Trepomonas agilis (ATCC 50336) was grown in TYGM-9 medium (medium 1171 of Nerad 1993). Cells were harvested during the logarithmic growth phase by gentle centrifugation, and the pellets were immediately treated with 2% CTAB for the extraction of genomic DNA (Lichtenstein and Draper 1985). Their small subunit rRNA genes were amplified by the polymerase chain reaction (PCR) using conserved primers (Medlin et al. 1988) and a lower-than-standard annealing temperature (i.e., 55°C), cloned in both orientations in M13 (mp18 and 19) phage, and sequenced fully on both strands using 12 conserved primers as described by Bhattacharya et al. (1990); and ABI Dyedeoxy sequencing kit; a Cetus 480 thermal cycler for cycle sequencing; and an ABI 373A automated sequencer. For the more conserved primers, the ABI protocol was followed; for the less conserved ones, a lower temperature cycle was necessary (Cavalier-Smith and Chao 1995) in order to obtain clear results. The Genbank accession number is U53120.

The sequencer traces were edited using the trace editor TED (Gleeson and Hillier 1991), and the sequences were aligned manually with over 500 eukaryotic and 50 bacterial small ribosomal RNA genes using the Genetic Data Environment software (Smith et al. 1994): The other sequences were obtained from Genbank or the EMBL database, except for those originating in our own laboratory. Phylogenetic trees were calculated using Phylip v. 3.5 (Felsenstein 1992) or fastDNAml v. 1.0.6 (Olsen et al. 1994). All trees were calculated with a random order of addition of taxa, using the jumble options, using a Sun Sparc 10 workstation with four 80-MHz processors. For maximum likelihood and parsimony several independent calculations were done using different random orders of addition in order to increase the probability of finding the shortest or most likely tree.

Results

Since maximum likelihood has been shown to be superior to either distance or parsimony methods for molecular phylogenetics over a wide range of parameters, and less vulnerable to violations of its basic assumptions (Nei 1991; Kuhner and Felsenstein 1994), and since it has not previously been used to examine the question of the nature of the earliest-diverging eukaryotes, we chose it as our primary method of analysis, despite its central processing unit (CPU)-intensive character. Figure 1 shows the relationships calculated for the 1,260 most-conserved and best-aligned nucleotide positions for the small subunit ribosomal RNA genes of a larger number of eukaryote species than have previously been included in a maximum likelihood tree; it took over 2 weeks and used over 95% of the CPU time to calculate. Because the incorrect rooting of the tree would be misleading as to the order of early divergences, and as rooting is more reliable if several representatives of the outgroup rather than a single one are included, we used four phylogenetically diverse archaebacteria simultaneously as the outgroup. There is now a general consensus that archaebacteria, not eubacteria, are the sister group to eukaryotes (Cavalier-Smith 1987a; Iwabe et al. 1989; Gogarten et al. 1989; Doolittle and Brown 1994). Although certain eukaryotic proteins seem to be more similar to eubacterial than to archaebacterial ones (Golding and Gupta 1995), this is not true for ribosomal RNA, for which archaebacteria are undoubtedly the more suitable outgroup (Woese 1987; Olsen and Woese 1994).

The maximum likelihood tree clearly shows that *Trepomonas* is related to the parasitic diplozoans and is much more closely related to *Hexamita* and *Spironucleus* than to *Giardia.* It is significantly closer to *Hexamita* than to *Spironucleus.* The tree also shows diplozoans rather than metamonads or parabasalans as the deepest eukaryotic branch. This is in accordance with our earlier neighbor-joining distance analysis using six archaebacteria and 20 eubacteria as outgroups, and 150 eukaryotes and virtually the entire sequence alignment (3,400 positions: Cavalier-Smith 1993a), as well as with the neighbor-joining analysis of Van De Peer et al. (1993) using a single archaebacterium and only more conserved parts of