(1983).
33. None of these considerations affects the function of glutamine as substrate for cytosolic, glutamine-dependent CPS II in pyrimidine synthesis.
34. N. M. Ritter, D. D. Smith, J. W. Campbell, *J. Exp. Zool.* **243**, 181 (1987).
35. T. P. Mommsen *et al.*, *Can. J. Zool.* **61**, 1835 (1983); P. J. Walsh, C. Bedolla, T. P. Mommsen, *J. Exp. Zool.* **241**, 133 (1987).
36. R. A. Shankar and P. M. Anderson, *Arch. Biochem. Biophys.* **239**, 248 (1985).
37. T. Mommsen and T. Moon, unpublished data.
38. T. Mommsen, C. Moyes, P. Walsh, unpublished data.
39. G. W. Brown, Jr., and S. G. Brown, in *Evolutionary Biology of Primitive Fishes*, R. E. Foreman *et al.*, Eds. (Plenum, New York, 1985), pp. 321–337.
40. C. A. Casey and P. M. Anderson, *J. Biol. Chem.* **258**, 8273 (1983).
41. F. Vellas and A. Serfaty, *J. Physiol. (Paris)* **68**, 592 (1974).
42. J. E. Vorhaben and J. W. Campbell, *J. Cell Biol.* **73**, 300 (1977); J. W. Campbell, D. D. Smith, Jr., J. E. Vorhaben, *Science* **228**, 349 (1985).
43. S. Grisolia *et al.*, Eds., *The Urea Cycle* (Wiley, New York, 1976); M. J. Jackson, A. L. Beaudet, W. E. O'Brien, *Annu. Rev. Genet.* **20**, 431 (1986); D. D Smith and J. W. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 160 (1988).
44. We thank C. D. Moyes for help with the preparation of mitochondria and C. J. Lusty, T. W. Moon, C. D. Moyes, J. S. Nelson, J. Nener, and P. A. Wright for helpful discussions. The following sources for experimental animals are gratefully acknowledged: J. McCosker (California Academy of Sciences) for liver pieces of a *Latimeria* specimen (CAS Catalog no. 33111); C. V. Sullivan and J. Musick, the Explorers Club, and the Society for the Protection of Old Fishes for liver pieces of a recently caught *Latimeria* (Virginia Institute of Marine Science, Catalog no. 8118); D. R. Jones for two *Protopterus aethiopicus*; The Fish Store (Bellevue, WA) for the *Protepterus annectens*; and the Jimbo for *O. beta* specimens. We thank K. Mopper, D. Kieber, and G. Vaughn for amino acid analyses. Supported through grants from the Natural Sciences and Engineering Research Council of Canada to T.P.M. (A2898) and the National Science Foundation to P.J.W. (DCB-8608728).

22 April 1988; accepted 13 October 1988

Phylogenetic Meaning of the Kingdom Concept: An Unusual Ribosomal RNA from *Giardia lamblia*

MITCHELL L. SOGIN, JOHN H. GUNDERSON, HILLE J. ELWOOD, ROGELIO A. ALONSO, DEBRA A. PEATTIE

An analysis of the small subunit ribosomal RNA (16*S*-like rRNA) from the protozoan *Giardia lamblia* provided a new perspective on the evolution of nucleated cells. Evolutionary distances estimated from sequence comparisons between the 16*S*-like rRNAs of *Giardia lamblia* and other eukaryotes exceed similar estimates of evolutionary diversity between archaebacteria and eubacteria and challenge the phylogenetic significance of multiple eukaryotic kingdoms. The *Giardia lamblia* 16*S*-like rRNA has retained many of the features that may have been present in the common ancestor of eukaryotes and prokaryotes.

THE TAXONOMIC SEGREGATION OF organisms into two or more kingdoms is a legacy from early systematic biologists who relied on morphological variation at the macroscopic level to differentiate plants from animals. With the discovery of the microbial world and the development of analytical tools for defining subcellular features, the number of proposed kingdoms has increased, and the debate about evolutionary relationships between the major groups of eukaryotic organisms has intensified (*1*). Controversies over conflicting taxonomic schemes are usually due to a lack of consensus about which characteristics are most useful for inferring phylogenies. As an alternative to traditional methods, comparisons of gene sequences that share a common ancestry can be used to infer objective phylogenetic frameworks (*2*). The 16*S*-like rRNAs have proved to be particularly well suited for estimating relationships between even the most divergent taxa (*3*). Here we describe extensive differences between the 16*S*-like rRNAs of *Giardia lamblia* and other eukaryotes, including diverse protozoans.

The protozoan parasite *G. lamblia* is a diplomonad, which can be propagated in vitro but normally lives attached to the intestinal mucosa of its host. *Giardia lamblia* has two nuclei and eight flagella but lacks mitochondria and normal endoplasmic reticulum (ER) or Golgi (*4*). Either these features were introduced into other eukaryotes after the divergence of diplomonads or they were lost in the *G. lamblia* evolutionary lineage. To identify its phylogenetic placement, we determined the sequence of the *G. lamblia* 16*S*-like rRNA coding region and compared it to the rRNAs of divergent taxa.

The *G. lamblia* 16*S*-like rRNA sequence (*5, 6*) is unusually rich in G+C content (75%) and has only 1453 nucleotide positions, a size more typical of prokaryotes than of eukaryotes. A collection of 41 eukaryotic (*7–9*), 6 archaebacterial (*10*), and 7 eubacterial 16*S*-like rRNAs (*11*) were aligned with the *G. lamblia* sequence by a computer-assisted procedure that considers the conservation of both primary and secondary structure features. Evolutionary distances (*6*) were used in the distance matrix methods (*12*) to infer the phylogenetic tree shown in Fig. 1. In this phylogenetic framework, *G. lamblia* represents the earliest diverging lineage in the eukaryotic line of descent. The *G. lamblia* branching is followed by the microsporidian *Vairimorpha necatrix* and then by euglenoids (*Euglena gracilis*) and kinetoplastids (*Trypanosoma brucei*). Late in the evolution of eukaryotes there was a nearly simultaneous splitting of animals, fungi, chlorophytes plus plants, chromophyte algae, and ciliates plus dinoflagellates. The precise branching order for these lineages is statistically uncertain, since it spans a distance of fewer than one nucleotide change per 100 positions. Yet the general branching pattern is nearly constant in similar phylogenetic trees that include different representatives of these major eukaryotic lineages. Similar tree topologies are observed by using the parsimony methods implemented by Swofford's computer program, "Phylogenetic analysis using parsimony" (PAUP) (*13*), with the significant difference that *V. necatrix* branches before *G. lamblia*. Unequal rates of change in one or more lineages (which are represented by long segments in distance matrix trees) sometimes produce anomalously deep branching patterns or generate different tree topologies when parsimony rather than distance methods are used (*14, 15*). The segment connecting *G. lamblia* to the eukaryotic subtree is not abnormally long and therefore its early divergence is not due to an unusually high mutation rate in its rRNA genes. In contrast, *V. necatrix* seems to have evolved more rapidly than the rRNAs of other eukaryotes, which may explain the alternate branching orders in distance and parsimony analyses.

The tree geometry in Fig. 1 could be biased by the high G/C content in the 16*S*-like rRNAs of *G. lamblia* (75% G/C) and *Sulfolobus solfataricus* (67% G/C). If G/C-rich 16*S*-like rRNAs are used to represent the eubacteria and archaebacteria, then convergence toward G or C at a number of sites might influence the phylogenetic position of

M. L. Sogin and H. J. Elwood, National Jewish Center for Immunology and Respiratory Medicine, Department of Pediatrics, Denver, CO 80206, and Department of Microbiology, University of Colorado Health Sciences Center, Denver, CO 80262.
J. H. Gunderson, Department of General Biology, Vanderbilt University, Nashville, TN 37235.
R. A. Alonso and D. A. Peattie, Department of Tropical Public Health, Harvard School of Public Health, Boston, MA 02115.

G. lamblia. Since there is no accepted or standard method of compensating for heavily biased nucleotide compositions, we constructed phylogenies that excluded prokaryotes with a high G/C content. In these trees *G. lamblia* still represented the deepest branching lineage in the eukaryotic line of descent (*9*). If *G. lamblia* diverged early in the evolutionary history of eukaryotes, its 16*S*-like rRNA may have retained many of the features that were present in the common ancestor or ancestors of eukaryotes and prokaryotes. We explored this possibility by examining the *G. lamblia* sequence for features that are diagnostic for archaebacterial and eubacterial 16*S*-like rRNAs.

Nucleotide usage patterns were identified in absolute consensus sequences for our collection of archaebacterial, eubacterial, or archaebacterial plus eubacterial 16*S*-like rRNAs. (Invariant positions in our total 16*S*-like rRNA database were not included in the consensus sequences.) Table 1 displays the number of sites that are identical in comparisons of representative eukaryotic 16*S*-like rRNAs with archaebacterial- or eubacterial-modified consensus sequences. These sites are enumerated in four categories: archaebacteria, eubacteria, archaebacteria or eubacteria, and archaebacteria plus eubacteria. In most eukaryotic 16*S*-like rRNAs the nucleotide compositions at 130 to 169 sites are identical to the archaebacterial- or eubacterial-modified consensus sequences (Table 1, columns 1 and 2, respectively). Similarly, the "or" comparisons (Table 1, column 3) show that 170 to 194 positions are identical to either the archaebacterial- or the eubacterial-modified consensus sequences. The values for *V. necatrix* are unusually low. This reflects the relatively short length of the *V. necatrix* 16*S*-like rRNA and its accelerated rate of evolutionary change. Conversely, the values for *G. lamblia* are unusually high. The *G. lamblia* 16*S*-like rRNA coincides with the archaebacterial- and eubacterial-modified consensus sequences at 296 and 248 sites, respectively. In the archaebacterial- plus eubacterial-modified consensus sequence comparison, 149 positions are identical to the *G. lamblia* rRNA (Table 1, column 4), and 395 positions are either eubacterial or archaebacterial in character (Table 1, column 3). These 395 positions are distributed throughout the proposed secondary structure shown in Fig. 2. *Giardia lamblia* appears to have retained many of the features found in prokaryotic 16*S*-like rRNAs, including the Shine-Dalgarno (*16*) binding site for bacterial messenger RNAs. These characteristics were probably present in ancestral 16*S*-like rRNA sequences common to the eubacterial, archaebacterial, and eukaryotic lines of descent.

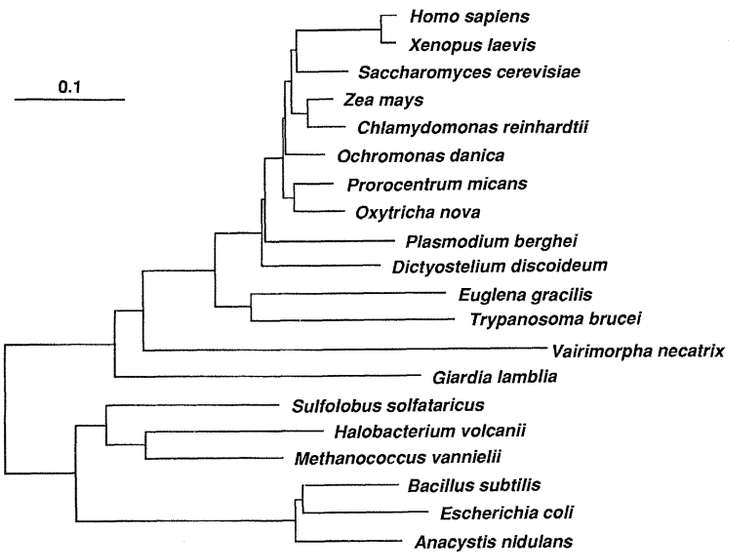

Fig. 1. Multikingdom tree inferred from 16*S*-like rRNAs. A computer-assisted method was used to align the 16*S*-like rRNA sequences from divergent representatives of the eubacteria, archaebacteria, and Eukaryota. The alignments were influenced by considering the evolutionary conservation of both primary and secondary structure features (*6*). The distance matrix methods (*12*) were used to infer an unrooted multikingdom tree in which the horizontal component of separation represents the evolutionary distance between organisms. The distance that corresponds to ten changes per 100 positions is indicated.

Table 1. Absolute consensus sequences were determined for 16*S*-like rRNAs from archaebacteria, eubacteria, or archaebacteria plus eubacteria. Invariant sites within the archaebacterial, eubacterial, and eukaryotic 16*S*-like rRNAs were deleted from the consensus, and these modified consensus sequences were compared to representative eukaryotic 16*S*-like rRNAs.

Organism	Number of sites identical to positions in modified 16*S* rRNA consensus sequences from			
	Archaebacteria (*10*)	Eubacteria (*11*)	Archaebacteria or eubacteria	Archaebacteria plus eubacteria
Homo sapiens	162	169	170	103
Xenopus laevis	140	145	186	99
Saccharomyces cervisiae	146	144	194	96
Zea mays	140	143	189	94
Dictyostelium discoideum	140	146	191	96
Euglena gracilis	144	142	194	93
Trypansoma brucei	130	135	180	86
Vairimorpha necatrix	106	103	146	63
Giardia lamblia	296	248	395	149

The retention of prokaryotic features in the *G. lamblia* rRNA is consistent with its early branching in the distance matrix analyses. This finding supports similar proposals based on the absence of organelles found in other eukaryotes, the prokaryotic-like organization of its rRNA coding region (*17*), the remarkably simple constellation of proteins in its cytoskeleton (*18*), and the evident lack of sexual life cycle stages. *Giardia lamblia* probably separated from other eukaryotes before the full development of subcellular features such as Golgi and ER, earlier than the endosymbiotic event or events that gave rise to mitochondria, and before the cytoskeleton had reached the level of complexity found in other eukaryotic microorganisms.

Our analyses show that the extent of genetic diversity in the eukaryotic subtree exceeds that seen within the entire prokaryotic world (*19*). Either the rate of evolutionary change in eukaryotic rRNA coding regions is more rapid than in prokaryotes or eukaryotes are as ancient as eubacteria and archaebacteria. Regardless of the correct scenario, the extreme genetic diversity represented by protists raises practical problems for taxonomists with respect to the number of eukaryotic kingdoms to be recognized and the characteristics to be used for differentiating major protistan groups. Both comparisons of phenotypic characters and measurements of genetic relatedness make it

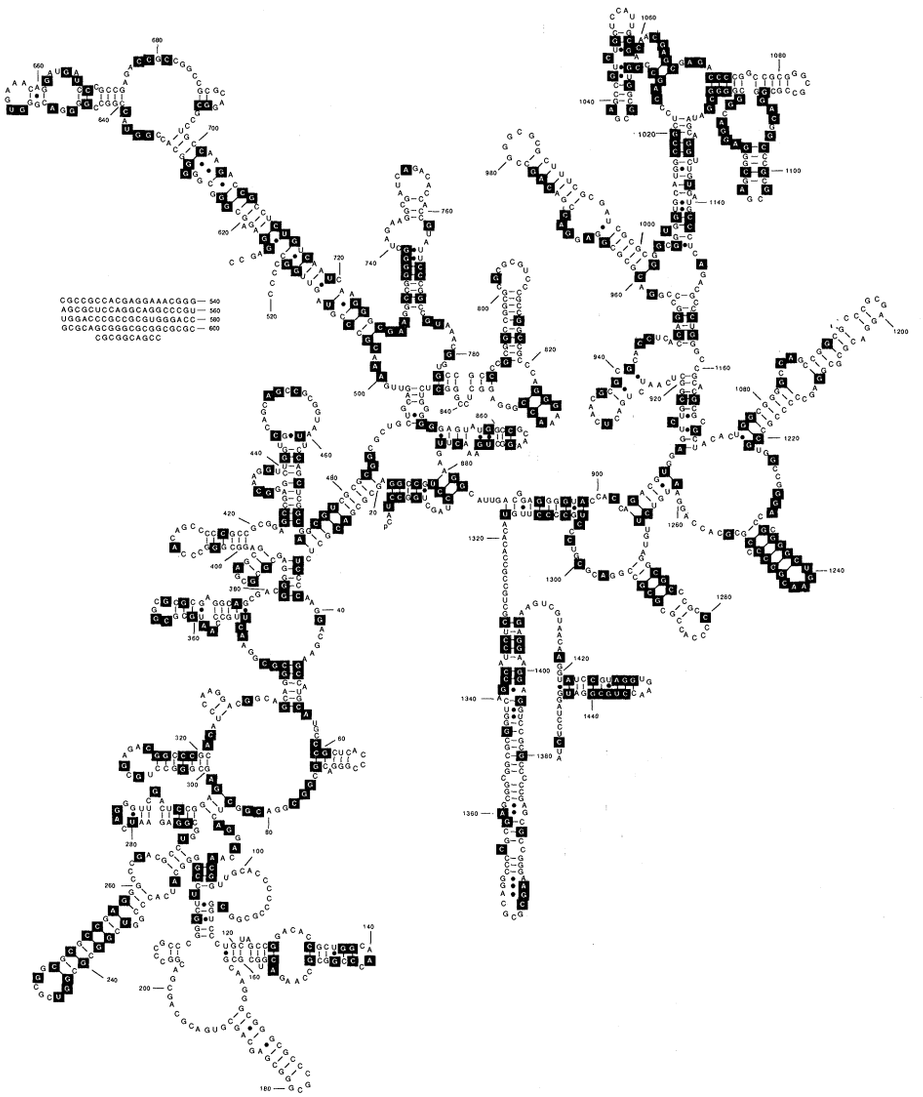

Fig. 2. Secondary structure for the *G. lamblia* 16*S*-like rRNA. Positions in the *G. lamblia* 16*S*-like rRNA that are identical to modified consensus positions in either the archaebacteria or eubacteria (Table 1, column 3) are indicated by "reversed" letters. The helical regions are based on their phylogenetic conservation as described by Woese (*3*).

clear that protists are not a cohesive phylogenetic assemblage and that the diversity within certain protistan groups approaches or exceeds that seen in other "higher" eukaryotic kingdoms. For example, we have previously shown that ciliated protozoans are as genetically diverse as the Plantae or Animalia (*20*). This finding raises questions about what kingdom level designations should represent. Major divisions in traditional classification systems are generally defined by the largest measurable differences in structural organization and life-style. However, there is no reason why deep genealogical separations must be marked by major structural differences. Indeed, for phylogenetic schemes based on comparative analyses of phenotypic characters, the boundaries of large taxonomic groups sometimes disagree with evolutionary distances inferred from macromolecular sequence comparisons. In the case of *G. lamblia*, a very deep branch in the rRNA-based tree leads to a small group of organisms more isolated from other eukaryotes than groups frequently recognized as separate kingdoms (animals, plants, and fungi). Are the diplomonads, then, to be given kingdom status? The answer will vary according to whether genealogical or morphological differences are used to recognize major groups of organisms. We are led to the view that eukaryotes should be regarded as a single kingdom made up of a progression of diverging lineages.

REFERENCES AND NOTES

1. J. O. Corliss, *Prog. Protistol.* **1**, 11 (1986).
2. E. Zuckerkandl and L. Pauling, *J. Theor. Biol* **8**, 357 (1965).
3. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
4. D. E. Feely, S. L. Erlandsen, D. G. Chase, in Giardia *and Giardiasis*, S. L. Erlandsen and E. A. Meyer, Eds. (Plenum, New York, 1984), pp. 3–31.
5. The *G. lamblia* (Portland 1) 16*S*-like rRNA sequence was determined from the rDNA clone pGK7 (M. G. Ortega-Pierres, R. A. Alonso, C. Montanez, D. A. Peattie, in preparation) and has been submitted to GenBank (GenBank accession number M19500). The rDNA clone pGK7 contains a 6-kb Kpn I-Kpn I fragment that includes the entire rRNA transcription unit. Restriction fragments containing the 16*S*-like rRNA gene were cloned into the single-strand phages M13/mp18 or M13/mp19. The dideoxynucleotide chain termination techniques initiated with consensus oligonucleotide primers for 16*S*-like rRNA genes (*6*) were used to determine the sequence of both strands. Band compressions were resolved by using 7-deaza–deoxyguanosine triphosphate (7-deaza–dGTP) instead of dGTP in the sequencing reactions, incubating the chain elongation reactions at 50°C rather than at 37°C, incorporating 40% formamide in the 8% acrylamide, 8*M* urea sequencing gels, or a combination of all of the above.
6. H. J. Elwood *et al.*, *Mol. Biol. Evol.* **2**, 399 (1985).
7. Eukaryotic 16*S*-like rRNAs from *Acanthamoeba castellanii, Achyla bisexualis, Artemia salina, Chalmydomonas reinhardtii, Crithidia fasciculata, Dictyostelium discoideum, Euglena gracilis, Euplotes aediculatus, Glycine max, Homo sapiens, Mus musculus, Neurospora crassa, Ochromonas danica, Oryctolagus caniculus, Oryza sativa, Oxytricha nova, Paramecium tetraurelia, Plasmodium berghei, Prorocentrum micans, Ratus ratus, Saccharomyces cerevisiae, Stylonychia pustulata, Tetrahymena hegewischi, Tetrahymena thermophila, Trypanosoma brucei, Vairimorpha necatrix, Xenopus laevis,* and *Zea mays*, as compiled in Dams *et al.* (*8*); *Leishmania donovani* [D. Looker *et al.*, *Nucleic Acids Res.* **16**, 7198 (1988)]; *Nanochlorum eucaryotum* [M. Sargent, R. Zahn, B. Walters, R. Gupta, B. Kaine, *Nucleic Acids Res.* **16**, 4156 (1988)]; *Plasmodium falciparum* [T. F. McCutchan *et al.*, *Mol. Biochem. Parasitol.* **28**, 63 (1988)]; *Pneumocystic carinii* [J. C. Edman *et al.*, *Nature* **334**, 519 (1988)]; *Skeletonema pacifica* [L. Medlin *et al.*, *Gene* **71**, 491 (1988)]; *Zamia pumila* [C. J. Nairn and R. J. Ferl, *J. Mol. Evol.* **27**, 133 (1988)]; and sequences from *Blastocladiella emersonii, Entamoeba histolitica, Mucor racemous, Kluveromyces lactis, Podospora anserina, Schizophyllum commune*, and *Schizosaccharomyces pombe* (*9*).
8. E. Dams *et al.*, *Nucleic Acids Res.* **16**, r87 (1988).
9. M. L. Sogin, unpublished data.
10. Archaebacterial 16*S*-like rRNAs from *Halobacterium cutirubrum, Halococcus morrhuae, Halobacterium volcanii, Methanoccocus vannielii*, and *Sulfolobus solfataricus*, as compiled in Dams *et al.* (*8*), and *Pyrodictum occultum* (C. R. Woese, unpublished data).
11. Eubacterial 16*S*-like rRNAs from *Agrobacterium tumefaciens, Anacystis nidulans, Bacillus subtilis, Escherichia coli, Mycoplasma capricolum, Proteus vulgaris*, and *Pseudomonas testosteroni*, as compiled in Dams *et al.* (*8*).
12. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
13. D. L. Swofford computer program, "Phylogenetic analysis using parsimony" (PAUP-version 2.4) (Illinois Natural History Survey, Champaign, IL, 1985).
14. J. Felsenstein, *Syst. Zool.* **27**, 401 (1978).
15. J. A. Lake, *Mol. Biol. Evol.* **4**, 167 (1987).
16. J. Shine and L. Dalgarno, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342 (1974).
17. J. C. Boothroyd, A. Wang, D. A. Campbell, C. C. Wang, *Nucleic Acids Res.* **15**, 4065 (1987).
18. D. A. Peatrie, unpublished data.
19. The *G. lamblia/Plasmodium berghei* and *G. lamblia/Trypanosoma brucei* structural similarity values are 0.675 and 0.677, respectively. This compares to valus of 0.700 and 0.711 for *Sulfolobus solfataricus/Escherichia coli* and *Halobacterium volcanii/E. coli*, respectively.
20. M. L. Sogin and H. J. Elwood, *J. Mol. Evol.* **23**, 53 (1986).

21. Supported by grants from NIH (M.L.S.), the John D. and Catherine T. MacArthur Foundation (D.A.P.), and CONACYT, Mexico, to M. G. Ortega for the support of R.A.A.

9 June 1988; accepted 23 September 1988