

Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA

(large subunit rRNA/ribosome/protists/evolution)

ANNE BAROIN*, ROLAND PERASSO*, LIANG-HU QU†, GUY BRUGEROLLE‡, JEAN-PIERRE BACHELLERIE†, AND ANDRÉ ADOUTTE*

*Laboratoire de Biologie Cellulaire 4 (Centre National de la Recherche Scientifique, Unité Associée 1134), Batiment 444, Université Paris-Sud, 91405 Orsay Cedex, France; †Centre de Recherche de Biochimie et de Génétique Cellulaires (Centre National de la Recherche Scientifique, LP 8201), 118, Route de Narbonne, 31062 Toulouse Cedex, France; ‡Laboratoire de Zoologie et Protistologie (Centre National de la Recherche Scientifique, Unité Associée 40138), Université de Clermont, B.P. 45, 63170 Aubière, France

Communicated by Pierre Joliot, December 7, 1987 (received for review August 19, 1987)

ABSTRACT Using a rapid rRNA sequencing technique, we have determined the sequence of the 400 nucleotides located at the 5' end of the large subunit rRNA molecule from eight species of unicellular eukaryotes (protists). This region contains a pair of conservative domains well-suited for long-range phylogenetic evaluations among eukaryotes, due both to their substantial length and to their intrinsic rate of sequence variation during evolution. It also comprises a central more rapidly evolving portion, which allows for a fine tuning of distance evaluation between closely related species. Molecular distances were computed between the aligned nucleotides of all presently available protist sequences and were used to derive a tentative dendrogram. Within the limitations inherent to this approach, a number of interesting observations emerge: The various protist groups appear to have separated very early from each other. The most deeply divergent protists belong to a number of orders of flagellates (mastigotes), suggesting a very ancient origin for organelles containing a 9 + 2 microtubular arrangement. Ciliates emerged late among eukaryotes, suggesting that their peculiar genetic code was derived secondarily. Moreover, a dinoflagellate clusters with ciliates, thus making it likely that the unusual features of nuclear organization and mitosis of this group are not primitive but derived characters. Finally, within groups, taxonomic and evolutionary inferences appear to be feasible using this portion of the rRNA.

Unicellular eukaryotes (protists or protoctists; see refs. 1 and 2) have probably played an essential role in the history of life as intermediates between prokaryotes and eukaryotes on one hand and single-celled eukaryotes and multicellular plants and animals on the other hand. Many of the present day protists presumably correspond to descendants of some of the groups that proliferated during these two major evolutionary transitions. Their analysis may therefore shed light on the path followed during early cellular evolution. A preliminary requirement for the reconstruction of possible evolutionary scenarios is the availability of a general outline of the chronological emergence of the various protistan groups. A number of attempts have been made to this end by using all the information that was available (mostly ultrastructural and, to a more limited extent, biochemical), and some explicit phylogenies have been cautiously suggested (see, for example, refs. 2–6). The difficulties in deriving phylogenies of the protists have been stressed repeatedly: (i) the organisms are extremely diverse and may belong to as many as 45 phyla (1), (ii) they have left little easily identifiable fossil record (7), and (iii) in many groups biochemical data are fragmentary. Direct

sequencing of informational macromolecules would therefore appear to constitute the method of choice to derive classification schemes and phylogenetic inferences on these organisms. Data on 5S RNA have been obtained in many species during the last few years (8, 9). This molecule suffers, however, from a number of limitations when used for very distant species (discussed in ref. 9).

The development of a technique for rapidly and easily sequencing large stretches of 18S or 28S rRNA, which has been reported (10), opened the way for systematic exploitation of the remarkable properties of these molecules as phylogenetic indicators (11). In addition to having a large information content, they present the advantage of having a characteristic mosaic structure, with an interspersed of adjacent domains evolving at widely different rates (12). One can therefore use fast evolving portions for comparing species that are expected to be closely related and slowly evolving regions for distantly related species. The technique involves neither DNA cloning nor RNA end-labeling and is carried out directly on total cellular RNA (10). The general applicability of this method has recently been confirmed by Lane *et al.* (13) by using partial sequence determinations of small subunit rRNA. In this paper, we report on the use of this method for sequencing the 400 5'-terminal nucleotides of the 28S RNA of eight species of protists.[§]

Using a distance matrix method, we derive a tentative dendrogram and discuss a number of interesting biological conclusions that emerge from the sequence comparisons of all presently available large subunit rRNA sequences. The conclusions are in excellent agreement with those based on complete small ribosomal subunit sequences (14).

MATERIALS AND METHODS

Sources of Protist Cultures. Fresh cultures of *Blepharisma japonicum*, *Stentor coeruleus*, and *Amoeba proteus* (grown on bacterized rice medium, *Chlorogonium*, and *Chilomonas*, respectively) were generously supplied by G. Fryd-Versavel (Laboratoire de Zoologie 2, Université Paris-Sud, Orsay) and were harvested after disappearance of the food organism. *Crithidia fasciculata* (ATCC no. 117) was grown on brain/heart extract (Difco) supplemented with rabbit blood. *Trichomonas vaginalis* (human isolate) was grown on trypticase/yeast extract/maltose (TYM) medium. *Paramecium caudatum* (unreferenced wild-type strain) was grown in bacterized wheat grass powder medium. Total RNA from *Paramecium primaurelia* (strain 156) and *Tetrahymena pyriformis* was kindly provided by François Caron (Centre de

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[§]The sequences reported in this paper are being deposited in the EMBL/GenBank data base (Intelligenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03648).

Génétique Moléculaire, Centre National de la Recherche Scientifique, Gif-sur-Yvette) and J.-J. Cury (Biologie Animale, Université de Lille 1), respectively.

RNA Isolation and Sequencing. The starting material usually consisted of 0.5–1.0 ml of washed packed cells that were either fresh (all the ciliates) or had been frozen in liquid nitrogen (*C. fasciculata* and *T. vaginalis*). Total RNA was extracted from cells homogenized either in guanidium thiocyanate medium at 60°C (4 M guanidium thiocyanate/50 mM Tris Cl, pH 7.6/4 mM EDTA/2% *N*-lauryl-sarcosinate/1% 2-mercaptoethanol) or in NaDodSO₄ medium at room temperature (5% NaDodSO₄/50 mM Tris Cl, pH 7.4/0.15 M NaCl/5 mM EDTA). All subsequent steps were carried out as described elsewhere (10, 15). Briefly, RNA sequencing was carried out by reverse transcriptase elongation of ³²P 5' end-labeled synthetic DNA primers in the presence of chain terminating dideoxynucleotides. Three oligonucleotides, complementary to the evolutionarily conserved 28S rRNA segments 84–106, 278–302, and 382–404 [numbers referring to mouse sequence coordinates (16)], respectively, were used systematically as primers.

Sequence Comparisons. Alignment of sequences, computation of the observed as well as the corrected number of nucleotide differences [using Kimura's Knuc correction (16)], and derivation of the resulting distance matrix were carried out exactly as described (15, 17). Dendrograms were constructed from the distance data both manually and by using the Fitch program of Felsenstein's Phylip 3.0 package. In both cases, one aims at finding a tree whose branch lengths reflect, as closely as possible, the distance values that were provided in the initial distance matrix. The procedure is therefore basically that used by Fitch and Margoliash (18)

with some modifications described in the Phylips documentation.

RESULTS

Strategy. The presence of invariant (or almost invariant) tracts in rRNA allows for the selection of "universal" DNA primers, which can hybridize to rRNA of any species and thus be used for sequencing large portions of the molecule by reverse transcriptase elongation in the presence of dideoxynucleotides (10, 13, 15). We have recently shown (15) that a region of the eukaryotic large subunit rRNA, which extends over about 400 nucleotides from the 5' end, contains two conservative domains. These domains appear ideally suited for evaluating distances among very distant eukaryotes since substitutions do not reach saturation levels within them (unlike the situation with 5S RNA) but remain, nevertheless, sufficiently numerous to allow meaningful calculations. A few "hot spots" (denoted by short vertical lines in Fig. 1) have been detected within these two domains by comparison of all the presently available sequences. These noninformative sites have been excluded from the homology measurements. As for the D1 domain (see Fig. 1 for delineation), which corresponds to one of the rapidly evolving areas of the 28S RNA molecule (12), changes become saturating for even medium-range comparisons among eukaryotes (15). This domain is therefore excluded from the present analysis except in the case of more closely related species, where it becomes very useful.

Analysis of the Data. The sequences of the eight protist species we have determined are shown in Fig. 1 aligned with three previously reported protist sequences [i.e., *Dictyostelium discoideum* (19), *Physarum polycephalum* (20) and

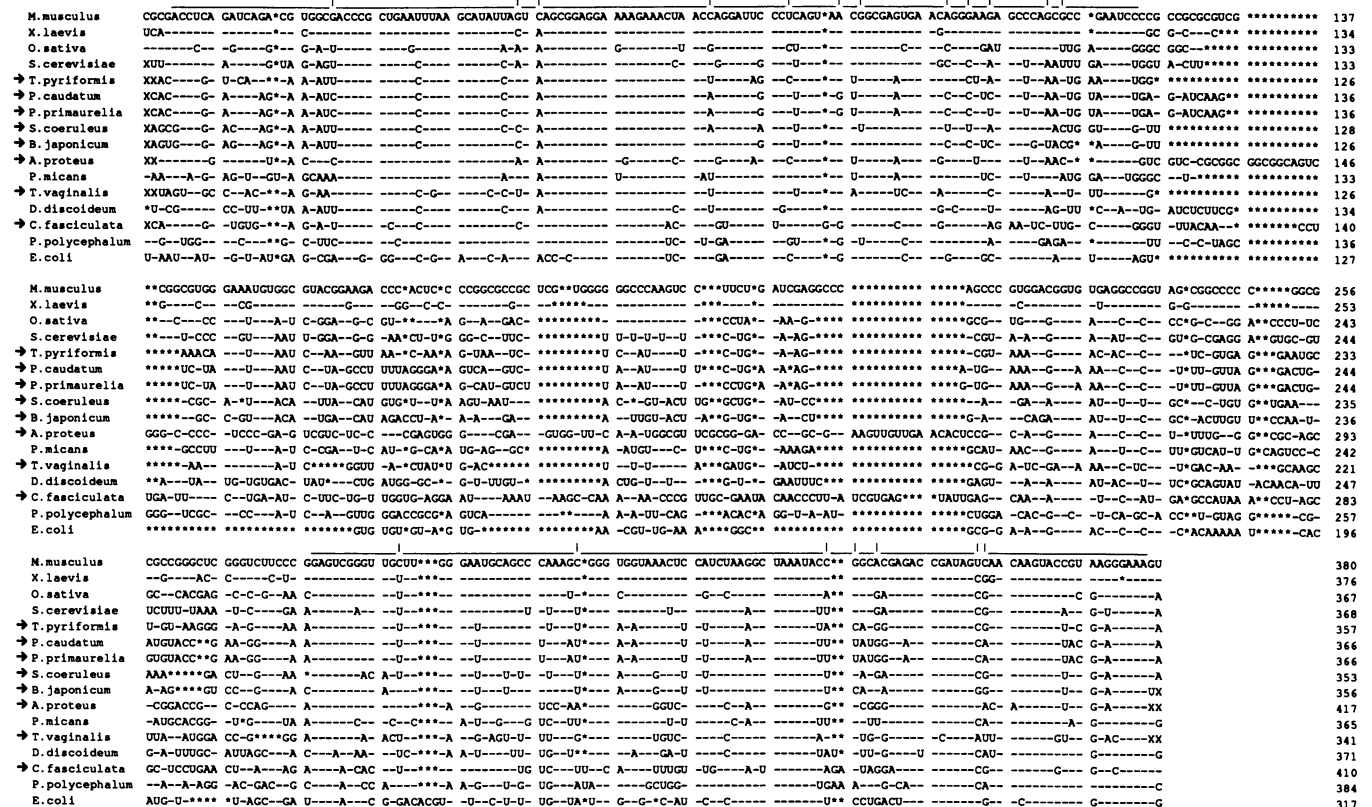


FIG. 1. Sequence alignments from the 5'-terminal region of large subunit rRNA. The protist species that we have sequenced in this study are indicated by arrows; the other sequences serve as references. Only the nucleotides that differ from those of the mouse are shown (identities are denoted by hyphens and deletions are denoted by stars). The 5'-terminal nucleotide positions that could not be identified are denoted by an "X." The two conservative domains used for evaluating phylogenetic distances are overlined (the vertical bars denote more variable nucleotide positions, which have been excluded from the measurements). The D1 domain corresponds to the portion that has not been overlined.

Table 1. Pairwise comparisons of the RNA sequences

	<i>M.m.</i>	<i>X.l.</i>	<i>O.s.</i>	<i>S.c.</i>	<i>T.p.</i>	<i>P.c.</i>	<i>P.p.</i>	<i>S.c.</i>	<i>B.j.</i>	<i>A.p.</i>	<i>P.m.</i>	<i>T.v.</i>	<i>D.d.</i>	<i>C.f.</i>	<i>P.p.</i>	<i>E.c.</i>
<i>M. musculus</i>		4	25	32	38	41	40	36	33	36	44	51	48	58	44	65
<i>X. laevis</i>	0.019		28	33	41	44	43	39	36	35	47	52	48	59	46	68
<i>O. sativa</i>	0.129	0.146		38	42	42	42	39	37	51	45	58	55	63	59	64
<i>S. cerevisiae</i>	0.171	0.178	0.208		32	32	31	34	34	42	41	53	48	60	58	70
<i>T. pyriformis</i>	0.206	0.225	0.233	0.169		17	16	30	30	43	44	53	49	63	66	78
<i>P. caudatum</i>	0.231	0.251	0.235	0.170	0.084		1	28	24	46	37	55	51	68	67	73
<i>P. primaurelia</i>	0.225	0.244	0.236	0.164	0.079	0.004		27	25	45	38	54	50	67	66	73
<i>S. coeruleus</i>	0.194	0.212	0.213	0.182	0.158	0.147	0.142		12	52	49	46	48	59	59	64
<i>B. japonicum</i>	0.175	0.193	0.199	0.181	0.157	0.124	0.130	0.058		49	47	52	53	62	61	65
<i>A. proteus</i>	0.195	0.188	0.297	0.234	0.240	0.259	0.253	0.301	0.279		55	51	54	65	57	74
<i>P. micans</i>	0.245	0.265	0.253	0.224	0.243	0.199	0.205	0.279	0.264	0.319		61	57	68	72	77
<i>T. vaginalis</i>	0.290	0.298	0.341	0.306	0.304	0.319	0.312	0.256	0.297	0.290	0.363		52	66	64	76
<i>D. discoideum</i>	0.272	0.272	0.324	0.271	0.278	0.292	0.286	0.270	0.305	0.317	0.332	0.297		63	59	72
<i>C. fasciculata</i>	0.341	0.348	0.380	0.356	0.381	0.425	0.417	0.349	0.372	0.397	0.419	0.403	0.378		64	76
<i>P. polycephalum</i>	0.244	0.257	0.352	0.344	0.410	0.421	0.413	0.351	0.368	0.336	0.454	0.386	0.350	0.388		72
<i>E. coli</i>	0.394	0.418	0.386	0.435	0.507	0.464	0.465	0.386	0.394	0.469	0.496	0.487	0.452	0.488	0.455	

Measurements have been carried out over the conservative regions delineated in Fig. 1. The uncorrected number of nucleotide differences between all the species analyzed is given in the right hand upper triangle, and the corresponding Knuc values in the left hand lower one.

Prorocentrum micans (kindly provided by G. Lenaers and M. Herzog)]. We have also included in the comparison representatives of other eukaryotic kingdoms to serve as references: a fungus, *Saccharomyces carlsbergensis* [whose sequence is identical to that of *Saccharomyces cerevisiae* (21)]; a plant, *Oryza sativa* (rice) (22); two metazoans, the amphibian *Xenopus laevis* (23) and the mammal *Mus musculus* (12); and one prokaryote, *Escherichia coli* (24). Nucleotide differences were computed and corrected and then used to derive the phylogenetic tree, as indicated in *Materials and Methods*.

All the new species, in spite of their very broad systematic distribution, can be easily aligned with previously sequenced eukaryotes over the two domains flanking the D1 domain, thus confirming the overall conservation of these domains and their usefulness for comparative purposes. Within these two domains, the most frequent type of difference occurring between eukaryotic sequences is, by far, nucleotide substitution; only a few point insertions (or deletions) are observed, and larger scale differences are absent. Any ambiguity in sequence alignment that could have resulted from such point changes has been solved by analyzing the sequence data in terms of the secondary structure of these domains of the rRNA molecule, which appears to be even more conservative than the primary structure (17, 25, 26). All these events have

been given equal weight in the computation of the distance values.

Examination of the distance matrices (Table 1) allows the easy identification of a number of clusters: The metazoans, the plant, and the yeast cluster together with relatively small distance values [this clustering is confirmed when analyzing a more complex set of species, as described elsewhere (15)]. The five ciliates cluster with each other and are close to the previous group. The amoeba and the dinoflagellate are also close to these two large clusters; and, finally, the last four protists (*Trichomonas*, *Dictyostelium*, *Crithidia*, and *Physarum*) are both quite distant from each other and from all the other species. These relationships are graphically displayed on the dendrogram of Fig. 2, which was derived using the Fitch program of the Phylip package (not allowing negative branch lengths and using the G option of this program, which reconsiders the position of every species and group of species). Depending on whether crude or corrected distance values were used, the program examined 410 or 750 trees, but the topologies obtained were quite similar. The representation should, nevertheless, be considered only as one possible way to summarize the data by using a specific method of tree construction. A number of alternative procedures for treating the data are available (27), including a cladistic approach to rRNA sequences (28), which, however, is difficult to apply

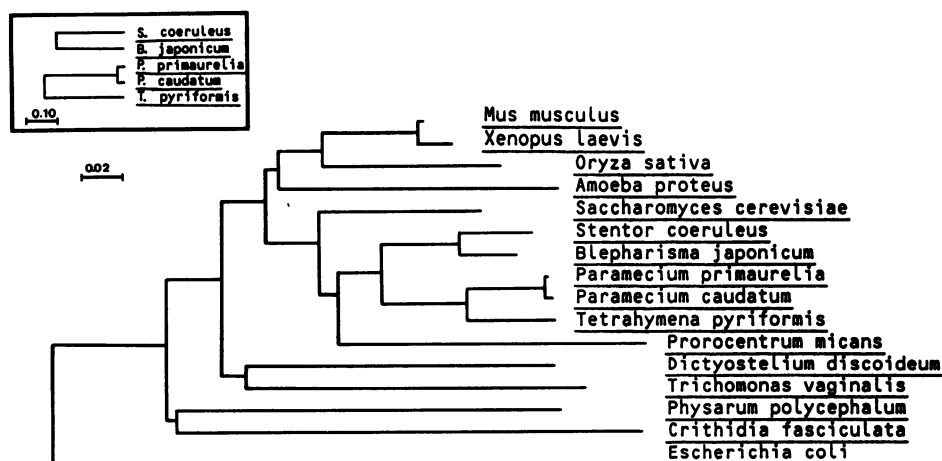


FIG. 2. Phylogenetic relationships between the protists. The tree was constructed using Knuc values from Table 1. In this representation, the distance between two species is proportional to the sum of the projections on the x (Knuc) axis of the branch lengths, while distances on the y axis are arbitrary. Note that any two (or more) branches are free to rotate over the axis leading to them. (Inset) Representation based on Knuc values derived exclusively from the D1 domain of ciliates.

on the stretch that we have sequenced because of an insufficient number of suitable sites. We note however that the dendrogram of eukaryotic species proposed by Wolters and Erdmann is similar in its topology to those derived by distance matrix methods (except for the position of *Euglena* and *Trypanosoma*).

Flagellated protozoa (mastigotes) appear to correspond to the earliest branching among the eukaryotes analyzed in this paper. A similar conclusion was recently reached by Sogin *et al.* (14) by using complete small subunit rRNA sequences of *Euglena gracilis* and *Trypanosoma brucei*. Quite recently, Vossbrinck *et al.* (29), using similar methods, have shown that microsporidia may in fact branch much earlier. Both in our dendrogram and in those based on small subunit sequences (14), the emergence of a number of major eukaryotic groups appears to be quite ancient and to extend over a long period. Ciliates, one dinoflagellate, and one amoeba are not among the earliest branching eukaryotes; they form a cluster that branched before the metazoa–metaphyte–fungi radiation.

The length of the branches within the ciliate cluster is sometimes considerable, suggesting early subdivisions within the group. The two subgroups identified in this paper fit quite well with traditional ciliate systematics (30): one corresponds to the polyhymenophora (the heterotrichs *Stentor* and *Blepharisma*, which belong to the same class, order, and suborder) and the other corresponds to the oligohymenophora (the hymenostome *Tetrahymena* and the peniculine *Paramecium*, which belong to two distinct subclasses). Examination of the D1 domain sequences provides considerable support for the individualization of these two ciliate subgroups and also for the evaluation of relative intergroup distances: within each of the ciliate subgroups, the D1 sequences can be reliably aligned over most parts of the domain and the rates of substitutions can be estimated. Although the Knuc values are about 7 times higher than those for the flanking conservative domains, the relative positioning of the different species (Fig. 2 *Inset*) is very similar to that obtained by using the conserved stretches. For such close relationships, the D1 domain becomes more informative than its flanking conservative regions.

DISCUSSION

Validity of the Methods Used. The quality of the sequence data derived by the technique used in this paper has been discussed (13, 15). The techniques of dendrogram construction have been the subject of lively debates among biologists. We are aware of a number of possible limitations and pitfalls inherent to our methodologies, both in the way nucleotide substitutions have been counted and in the use of distance matrix data. The clusters of species identified in the present study appear to be largely independent of the clustering procedure adopted (A. Hénaut and M.-O. Delorme, personal communication). The major problem probably resides in the conversion of a distance matrix into a phylogeny. The procedure we have used does not imply identity of substitution rates along all the branches of the tree. However, the extent to which inequalities can be detected depends on the presence of suitable external reference points and on the overall density of the tree. These conditions are all difficult to meet when one is dealing with species that are both very old and very distant from each other (possibly lying at the basis of the eukaryotic tree). Thus, some uncertainty about the topological order of the earliest offshoots of the tree as well as on that of closely branching species is inevitable at this stage. A number of arguments, most of them empirical, nevertheless, lend credit to the type of broad analysis carried out in this paper:

(i) McCarroll *et al.* (31) and Elwood *et al.* (32) have advocated the use of bacteria as an outgroup to check for possible inequalities in rates of substitutions. They noted that the distances between prokaryotes and protists were not larger than those between prokaryotes and metazoa, which argues in favor of the idea that the early branching of the protists was not due to an accelerated rate of substitution. The same is true for our data using large subunit sequences. The concern that this argument might be weak because the distances between bacteria and eukaryotes could have reached a saturated level of substitution is relaxed by the recent discovery of a eukaryote (the microsporidium *Vairimorpha*) that can be used as a very old outgroup, even for protists. Inclusion of the microsporidian data in a global dendrogram of the eukaryotes does not modify the previously established topology (29).

(ii) The internal consistency of the tree shown here is also evidenced by the fact that when a newly sequenced species belonging to a well characterized group is added it always links with previous representatives of the group. Thus, if accelerations occur, they similarly affect all the members of a group.

(iii) An extensive survey of metazoa and metaphytes recently achieved by sequence determinations of the same portion of the 28S rRNA molecule for 20 additional species (15, 17) does not reveal major inequalities in substitution rates among the species studied, and the upper part of the dendrogram is in excellent agreement with phylogenetic inferences based on comparative anatomy and paleontology. The molecule, therefore, is at least appropriate for deriving phylogenetically significant information on species spanning approximately one billion years [paralleling the remarkable properties of cytochrome *c* demonstrated 20 years ago by Fitch and Margoliash (18)].

(iv) Similar large scale phylogenies constructed by using 18S rRNA sequences both by Sogin *et al.* (14) and by Lempereur (33) yield topologies that are similar to those displayed here on the basis of 28S sequences. The agreement with the dendrograms recently published by Sogin *et al.* extends to details in spite of the fact that the sequenced species are not actually the same. Thus not only is the order of emergence of the various groups identical, but details of branch lengths are similar (depth of branching within the ciliates for example). In contrast, a number of topological discrepancies with phylogenies based on distance data derived from 5S RNA sequences are observed. However, severe limitations to the usefulness of this molecule are apparent when very large intereukaryote distances are involved (9, 15) since substitutions may reach saturation levels, which requires a more complex and specific treatment of the sequence data (see ref. 28, for example).

On the whole, then, it appears that the dendrogram displayed can be considered as a reasonable working tool subject to improvement as additional sequences and alternative treatment of the data are incorporated.

Biological Inferences. (i) As expected, the various protist groups surveyed in this study appear to be both very ancient and deeply divergent from each other. The species analyzed here do not cover the whole range of the protist kingdom, yet the molecular diversity uncovered is already considerable. This fits with the limited amount of molecular data that was already available (see ref. 5). Instead of considering them as a group of related organisms on the basis of their unicellular organization, the community of biologists should be aware that there may be as much (if not more) distance between some flagellates (mastigotes) and ciliates, for instance, as there is between ciliates and metazoa.

(ii) The clear subdivisions observed within the limited number of ciliates analyzed, which at the moment fit into traditional classifications, opens the possibility of a complete

renewal of systematics within such single, large, and sufficiently ancient phyla. The region that has been sequenced in the present work appears to be quite suitable for such a purpose. Finer analyses can be carried out at the order, suborder, or even family level by using other divergent domains of the molecule, which may display rates of substitution that are even higher than that of the D1 domain analyzed here. This is exemplified by the extensive reconstruction of tetrahymenine ciliate phylogeny recently achieved by Nanney and coworkers (personal communication) by using the D2 domain, one of the two fastest evolving regions of the whole 28S rRNA molecule (12, 25).

(iii) A number of flagellate (mastigote) groups emerge very early from the eukaryotic tree both in our dendrogram and in those of Sogin and coworkers (14). By assuming that the basal body (kinetosome) and its associated 9 + 2 microtubular arrangement has arisen only once in evolution (see ref. 34), this would imply that cilia and flagella (undulipodia) are of very ancient origin and this has several implications on the origin and evolution of the flagellar apparatus (kinetid), an organelle that has been used abundantly as a phylogenetic indicator (35). We note that *Physarum*, which also appears to emerge quite early, displays a flagellated stage during its cell cycle. The late emergence of nonflagellated amoeba as well as the absence of flagella in entire groups, such as fungi and higher plants, may therefore correspond to loss of the flagellar apparatus (assuming monophyly of the nuclear compartment of all the eukaryotes). All these organisms would also have lost basal bodies and centrioles and would have devised alternative microtubule-organizing centers to be used at the spindle poles during mitosis. We also note that *Trichomonas*, a representative of the parabasalia (parasitic or symbiotic flagellates, which are devoid of mitochondria and possess four flagella and a "parabasal" apparatus), emerges quite early, possibly within the same period range as that of the earliest flagellates. This not only yields some information for positioning this group, which had been difficult to insert into traditional phylogenies, but also provides elements for the discussion of the possible occurrence of a long period of eukaryotic life preceding mitochondrial symbiosis (36).

(iv) In contrast, two groups, ciliates and dinoflagellates, that have often been considered as quite primitive (see ref. 37 for an analysis of 18S rRNA from a dinoflagellate) emerge late and close to each other. This is an important point deserving special comments.

Ciliates, as a group, display a remarkable exception to the universal genetic code in that two of the stop codons code for glutamine (reviewed in ref. 38). This has sometimes been interpreted as a reflection of the antiquity of the group (i.e., as a possible relic of a primitive code). It is possible, however, to account for this coding situation by postulating the appearance and progressive improvement of glutamine nonsense suppressors in the line leading to ciliates, as first suggested by Hanyu *et al.* (39). Our results, which place several ciliates relatively close to the metazoan-metaphyte radiation, provide strong support for the idea of a late emergence of the peculiar genetic code of ciliates. Similarly, the position of the dinoflagellate on our dendrogram (based on a sequence kindly provided by Lenaers and Herzog) would suggest that the very peculiar nuclear characteristics of these organisms correspond to derived and not to primitive characters.

Note Added in Proof. The DNA parsimony program of the Phylip package yielded a topology identical to that shown in Fig. 2. The complete sequence of *Crithidia* large subunit rRNA appeared after submission of this paper (40).

We are grateful to François Michel for having induced the collaboration between our groups; Ghislaine Fryd-Versavel and Jean Génarmont for protist cultures; François Caron and Jean-Jacques

Curgy for the gift of *Paramecium* and *Tetrahymena* RNA; Monique Nicoloso for help in RNA isolation; Guy Lenaers and Michel Herzog for communicating the *Proterocentrum* sequence; and Alain Hénaut, Marie-Odile Delorme, Bernard Michot, and Philippe Ramond for information and help about the computing problems. We thank Joe Felsenstein (University of Washington, Seattle, WA) for his generous gift of the complete Phylip package. This work was supported by the Centre National de la Recherche Scientifique, the Universities of Paris 11, Clermont-Ferrand, and Toulouse, and by specific grants from the ATP and ASP "Evolution" of the Centre National de la Recherche Scientifique.

1. Corliss, J. O. (1984) *BioSystems* **17**, 87–126.
2. Margulis, L. & Schwartz, K. V. (1982) *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth* (Freeman, San Francisco), p. 338.
3. Taylor, F. J. R. (1978) *BioSystems* **10**, 67–89.
4. Cavalier-Smith, T. (1983) in *Endocytobiology II: Intracellular Space as Oligogenetic Ecosystem*, eds. Schenk, H. E. A. & Schwemmler, W. (De Gruyter, Berlin), pp. 1027–1034.
5. Ragan, M. A. & Chapman, D. J. (1978) *A Biochemical Phylogeny of the Protists* (Academic, New York), p. 317.
6. Sleight, M. A. (1979) in *The Origin of Major Invertebrate Groups*, ed. House, M. R. (Academic, New York), Systematics Association Special Vol. 12, pp. 23–43.
7. Vidal, G. (1984) *Geol. Soc. Am. Mem.* **161**, 265–277.
8. Hori, H. & Osawa, S. (1986) *BioSystems* **19**, 163–172.
9. Walker, W. F. (1985) *BioSystems* **18**, 269–278.
10. Qu, L. H., Michot, B. & Bachelier, J. P. (1983) *Nucleic Acids Res.* **11**, 5903–5920.
11. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
12. Hassouna, N., Michot, B. & Bachelier, J. P. (1984) *Nucleic Acids Res.* **12**, 3563–3583.
13. Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. & Pace, N. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6955–6959.
14. Sogin, M. L., Elwood, H. J. & Gunderson, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383–1387.
15. Qu, L. H., Nicoloso, N. & Bachelier, J. P. (1988) *J. Mol. Evol.*, in press.
16. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
17. Qu, L. H. (1986) Dissertation 1273 (Université Paul Sabatier, Toulouse, France), p. 230.
18. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.
19. Ozaki, T., Hoshikawa, Y., Iida, Y. & Iwabuchi, M. (1984) *Nucleic Acids Res.* **12**, 4171–4184.
20. Otsuka, T., Nomiya, H., Yoshida, H., Kubita, T., Kuhara, S. & Sakaka, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3163–3167.
21. Veldman, G. M., Klootwijk, J., de Regt, V. C. H. F., Planta, R. J., Branlant, C., Krol, A. & Ebel, J. P. (1981) *Nucleic Acids Res.* **9**, 6935–6952.
22. Takaiwa, F., Oono, K., Iida, Y. & Sugiura, M. (1985) *Gene* **37**, 255–259.
23. Ware, V. C., Tague, B. W., Clark, C. G., Gourse, R. L., Brand, R. C. & Gerbi, S. (1983) *Nucleic Acids Res.* **11**, 7795–7817.
24. Brosius, J., Dull, T. J. & Noller, H. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 201–204.
25. Michot, B. & Bachelier, J. P. (1987) *Biochimie* **69**, 11–23.
26. Michot, B., Hassouna, N. & Bachelier, J. P. (1984) *Nucleic Acids Res.* **12**, 4259–4279.
27. Felsenstein, J. (1981) *Quart. Rev. Biol.* **57**, 379–404.
28. Wolters, J. & Erdmann, V. A. (1986) *J. Mol. Evol.* **24**, 152–166.
29. Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. (1987) *Nature (London)* **326**, 411–414.
30. de Puytorac, P., Grain, J. & Mignot, J. P. (1987) *Précis de Protistologie* (Société Nouvelle des Editions Boubée, Paris), p. 581.
31. McCarroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R. & Sogin, M. L. (1983) *Biochemistry* **22**, 5858–5868.
32. Elwood, H. J., Olsen, G. J. & Sogin, M. L. (1985) *Mol. Biol. Evol.* **2**, 399–410.
33. Lempereur, L. (1986) Dissertation 26 (Université Paul Sabatier, Toulouse, France), p. 150.
34. Szathmary, E. (1987) *BioSystems* **20**, 115–131.
35. Melkonian, M. (1982) *Taxon* **31**, 255–265.
36. Cavalier-Smith, T. (1987) *Nature (London)* **326**, 332–333.
37. Herzog, M. & Maroteaux, L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8644–8648.
38. Caron, F. (1986) *Microbiol. Sci.* **3**, 36–40.
39. Hanyu, N., Kuchino, Y., Nichimura, S. & Beier, H. (1986) *EMBO J.* **5**, 1307–1311.
40. Spencer, D. F., Collings, J. C., Schnare, M. N. & Gray, M. W. (1987) *EMBO J.* **6**, 1063–1071.