Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8644–8648, November 1986
Evolution

Dinoflagellate 17S rRNA sequence inferred from the gene sequence: Evolutionary implications

(ribosomal gene/rRNA secondary structure/*Prorocentrum micans*/phylogenetic tree/primitive eukaryote)

MICHEL HERZOG* AND LUC MAROTEAUX†

Laboratoire Arago, Paris VI University, Centre National de la Recherche Scientifique UA 117, 66650 Banyuls sur Mer, France

Communicated by Lynn Margulis, July 21, 1986

ABSTRACT **We present the complete sequence of the nuclear-encoded small-ribosomal-subunit RNA inferred from the cloned gene sequence of the dinoflagellate *Prorocentrum micans*. The dinoflagellate 17S rRNA sequence of 1798 nucleotides is contained in a family of 200 tandemly repeated genes per haploid genome. A tentative model of the secondary structure of *P. micans* 17S rRNA is presented. This sequence is compared with the small-ribosomal-subunit rRNA of *Xenopus laevis* (Animalia), *Saccharomyces cerevisiae* (Fungi), *Zea mays* (Planta), *Dictyostelium discoideum* (Protoctista), and *Halobacterium volcanii* (Monera). Although the secondary structure of the dinoflagellate 17S rRNA presents most of the eukaryotic characteristics, it contains sufficient archaeobacterial-like structural features to reinforce the view that dinoflagellates branch off very early from the eukaryotic lineage.**

In the five-kingdom system for classification of organisms (1) the kingdom Monera contains all the prokaryotes (including the archaeobacteria). The remaining four kingdoms (Protoctista, Fungi, Animalia, and Planta) are all eukaryotic. The dinoflagellates (dinomastigota) are thought by some to be the oldest extant eukaryotes. Their fossil cysts have been dated to the Precambrian period (2). The dinoflagellates are a group of diverse eukaryotic protists, generally classified with the eukaryotic algae (3) and possessing a number of peculiar traits of nuclear organization (4) suggestive of the prokaryotic state. Among this unusual primitive pattern of nuclear organization, the most striking feature is the now well-documented absence of histones and nucleosomal structures. Dinoflagellate chromatin fibrils, which appear as smooth filaments in electron microscopy, are associated with a low amount of one or two major basic proteins whose amino acid composition is different from known prokaryotic basic proteins or eukaryotic histones (5–7). Nuclease digestion of purified nuclei lends to a smear with no recognizable discrete DNA fragments (8, 9). Permanently condensed chromosomes are stabilized by divalent cations and RNA (10, 11). In eukaryotes histones and especially the histone H1, which is thought to be primarily responsible for organizing higher-order structure of nucleosomes, are involved in gene repression (12). Thus, the problem of gene regulation in such organisms is still unsolved, with a prokaryotic-like chromatin organization and some of the following eukaryotic nuclear traits: (*i*) The portion of repeated sequences (around 60%) is typical for eukaryotes (13). (*ii*) The rRNA is synthesized as a precursor of 38S, and a 5.8S rRNA has been reported to be released from the 27S precursor to give rise to the final 24S RNA (14). (*iii*) The 5S rRNA sequence shows greater homology with that of higher eukaryotes than that of eubacteria (15), but additional molecular data (16) indicate that it contains some conserved residues specific for lower eukaryotes (unpublished data) and that the general structure of the 5S rRNA from sulfolobale and thermoplasmale archaeobacteria is also closer to that of eukaryotes than that of eubacteria (17). (*iv*) Poly(ADP-ribose) polymerase activity is detectable (18). (*v*) The small nuclear RNAs U1–U5 have been isolated (19). The U5 dinoflagellate sequence shows strong homology (63%) with that of vertebrates (20), but appears to be more homologous to that of *Tetrahymena* (70%) (ref. 21, and unpublished data). Thus these molecular data, in addition to the interpretation of Zillig *et al.* (17) on the archaeobacterial origin of the eukaryotic cytoplasm, renew the interest of whether the absence of nucleosomes in dinoflagellates is a secondary loss (15).

We decided to use rRNA as an evolutionary marker of dinoflagellate evolution. rRNA occurs in every living organism, and its functions have remained constant throughout evolution. The large- and the small-ribosomal-subunit rRNA retain the core structure common to prokaryotic rRNA, although their size increases substantially (22). Since these sequences are very highly conserved during evolution, they contain information on phylogenetic relationships during the early evolution of the eukaryotic kingdoms (23).

We have isolated clones from a genomic library of *Prorocentrum micans* containing the rRNA gene. We present here the complete sequence of the *P. micans* 17S rRNA inferred from the gene sequence. We propose a potential secondary structure for this molecule and compare it with other published small ribosomal subunit rRNA sequences.

MATERIALS AND METHODS

Strain. *P. micans* Erhenberg was provided by the Cambridge algae collection strain LB 113614.

Growth of Organisms and Isolation of Nucleic Acids. The organisms were grown under standard conditions, and their nuclei were isolated from cultures in logarithmic phase as described (9). They were sequentially digested by RNase A and proteinase K, and after phenol extraction, the DNA was further purified by two steps of CsCl gradient centrifugation. Total RNA was extracted from cells homogenized in the presence of 4 M guanidinium isothiocyanate, by pelleting through a 5.7 M CsCl cushion as described (24).

Cloning and DNA Sequencing. A genomic DNA library of *P. micans* has been constructed by cloning of 10- to 20-kilobase nuclear DNA fragments, digested by *Mbo* I, into purified arms of phage λEMBL4 (25) that had been digested with *Bam*HI. Plaques containing *P. micans* DNA were detected by hybridization with a flax ribosomal probe (pBG35) (26) and purified through three cycles of plating. Other cloning methods were used as described in ref. 27.

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.

Abbreviation: nt, nucleotide.

*Present address: Department of Biochemistry, 539 Biological Sciences West, University of Arizona, Tucson, AZ 85721.
†To whom reprint requests should be addressed.

DNA sequences were determined by the dideoxy chain-termination method (28) after subcloning into M13 vectors.

Sequence Alignments and Tree Construction. The *P. micans* sequence was aligned according to Nelles *et al.* (29), based on obvious sequence homologies and refined by use of secondary structure features. The calculation of sequence homologies and the inference of the tree most consistent with the data were performed as described (30).

RESULTS AND DISCUSSION

Primary Sequence of the 17S rRNA of *P. micans*. The rRNA genes from *P. micans* were isolated from a genomic λEMBL4 library. Ten positive clones were initially detected, and three representative clones were chosen for detailed studies (Fig. 1). Genomic hybridizations (data not shown) using either the flax probe or the end-labeled *P. micans* rRNA allowed the construction of the restriction map (Fig. 1*A*) with the position of the different rRNAs. R-loop experiments (data not shown) confirm the organization found by restriction mapping: the 17S, 5.8S, and 24S rRNA genes are encoded in this order, in a 6-kilobase tandemly repeated unit, probably chromosomal (14), separated by a nontranscribed spacer of variable length (4–8 kilobases). The number of ribosomal gene copies has been estimated by Southern blot hybridization as being at least 200 copies per haploid genome. Thus, ribosomal genes of dinoflagellate protists are organized in a way typical of other eukaryotes.

The λPr10 insert was subcloned into an M13 vector (31), and we sequenced the 17S rRNA gene. The analysis strategy is presented in Fig. 1*B*. The rRNA sequence inferred from the DNA sequence is presented in Fig. 2. The exact localization of the 5′ and 3′ ends has been determined by comparison with other sequences and by S1-nuclease mapping (data not shown), confirming that this clone probably does not encode a pseudogene. The 17S rRNA sequence is 1798 nucleotides (nt) long, which is longer than that of *Tetrahymena thermophila* (1753 nt) and close to that of *Saccharomyces cerevisiae* (1789 nt), but shorter than the sequence of *Dictyostelium discoideum* (1873 nt) and of *Xenopus laevis* (1825 nt) (Table 1). The G+C content of *P. micans* 17S rRNA (46%) also lies between that of *T. thermophila*, of *D. discoideum* (43%), and of vertebrates, plants (>50%), and close to that of yeast (45%). This value is very similar to that found earlier for another dinoflagellate, *Crypthecodinium cohnii* (14). Thus, assuming an increase in length and in the G+C content of small ribosomal subunit rRNA during evolution as for large ribosomal subunit rRNA (35), the most primitive would be the ciliates. *D. discoideum* is difficult to position with the lowest G+C content and the longest sequence. *P. micans* would be close to the fungi. However, comparisons of several small ribosomal subunit rRNA sequences reveal that "universal" sequences are interspersed among semiconserved and nonconserved sequences. The variable regions display much higher rates of genetic drift and different G+C content than the rest of the molecule (36).

Model of Secondary Structure of *P. micans* 17S rRNA. Models have been derived for small-ribosomal-subunit rRNA mainly on the basis of phylogenetic-sequence comparisons, where evidence for the presence of a double helix provided by base pairing in one organism is replaced by alternative base pairing in another (compensating base changes). The different models for eubacteria (22, 37, 38) show a large degree of agreement. Models have also been proposed for eukaryotic small-ribosomal-subunit rRNA (22, 29, 39, 40) that show extensive structural homology with the bacterial models. None of these models is completely analogous to the prokaryotic one, and the experimental evidence presently available in support of the eukaryote structure is very limited. The model shown in Fig. 2 for *P. micans* 17S rRNA comes closest to the model proposed for *D. discoideum* (40). Several of the following structural features, which distinguish the dinoflagellate, have been identified and are shown in Fig. 3.

(*i*) Helix 8. This helix, close to a variable region, has been shown to be more stable in archaeobacteria than in most eubacteria and eukaryotes (41). It is clear in Fig. 3*A* that the organization of the *P. micans* helix is closer to that of archaeobacteria. There are the same bulges and unpaired nucleotides, except one more G·G unpairing. The ciliate organization (32, 36) of this region is identical to that of *P. micans* except for one compensatory change, consistent with the existence of this helix that is not present in all models. All other eukaryotic sequences have shorter stems as in *Dictyostelium*, and/or bigger bulges as in *Saccharomyces*.

(*ii*) Helix E9. The helix E9-1 is not supported by any compensatory change. The helix 9 of Nelles *et al.* (29) cannot exist in the dinoflagellate sequence.

(*iii*) Helix 11. The central region of this helix is also more stable in archaeobacteria than in eubacteria (41). For the dinoflagellate sequence, the first internal bulge is two nucleotides longer than that of *Halobacterium* (42), and the second

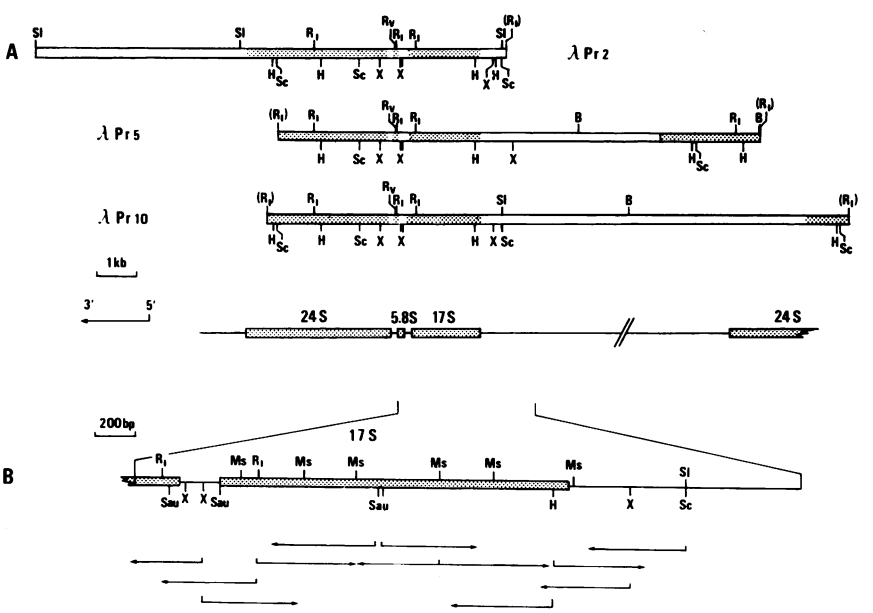

FIG. 1. (*A*) Restriction map of *P. micans* ribosomal clones. The three first lines represent the maps of three positive clones, λPr2, λPr5, and λPr10. Restriction enzyme cleavage sites are represented as follows: *Eco*RI by Rl, *Eco*RV by Rv, *Sal* I by Sl, *Bam*HI by B, *Hin*dIII by H, *Sac* I by Sc, and *Xho* I by X. The two distal *Eco*RI sites are from the cloning site of the phage λEMBL4. The bottom line represents the general organization of the *P. micans* ribosomal gene cluster, as shown by hybridization mapping and R-loop. Dashed boxes correspond to the different rRNA coding regions. (*B*) Sequencing strategy. The expanded region illustrates the dideoxy-sequencing strategy used for the 17S rRNA gene region. The arrows indicate the length of the sequence read, with the arrowhead corresponding to the 5′ end of the insertion into the M13 vector. *Msp* I cleavage sites are noted by Ms, and *Sau*3A by Sau.

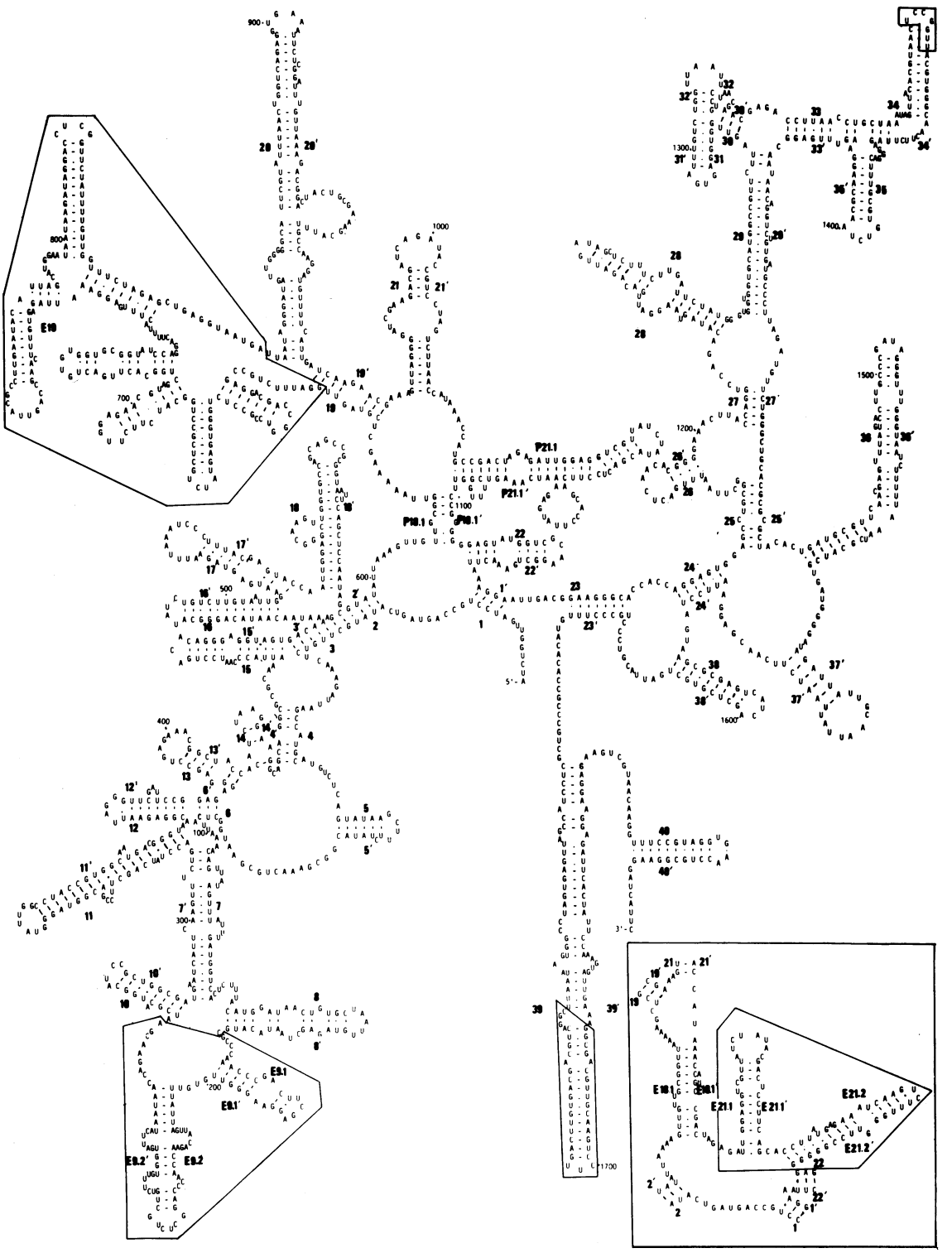

FIG. 2. Sequence and potential secondary structure of the *P. micans* 17S rRNA inferred from the gene sequence. The secondary structure map of the 17S rRNA is based, with minor modifications, on the proposed universal secondary structure map of Woese *et al.* (22). We have used the helix numbering system proposed by Nelles *et al.* (29). Nucleotide positions are numbered in intervals of 100 nucleotides, and 5′ and 3′ ends are indicated. Boxed areas represent the variable regions of the small subunit ribosomal RNAs, which were not taken into account for calculation of Table 2. (*Insert*) Possible alternative conformation of the central area, between helices 1, 2, 19, 21, and 22 is shown.

one is identical in length. In all other eukaryotes, helix 11 is 6 nt longer. The terminal loop of *P. micans* has 5 out of 7 nt in common with *Halobacterium volcanii* (42) (Fig. 3*B*).

(*iv*) Helices 16 and 17. These helices are surrounded by helices 2 and 3, which may be involved in intramolecular switching in the rRNA with helices 28–30 (43). The secondary structure of this region is not clearly defined in eukaryotes, but it is probably organized in a different way than in prokaryotes (usually helix 17 is not stable) (22). In the case of *P. micans*, as shown in Fig. 2, helix 17 is formed with six canonical base pairings, two G·U, and two G·A (which may exist in rRNA) (22). As discussed (22), the eukaryotic counterpart does not occupy a strictly homologous portion of the sequence and may be idiosyncratic rather than homologous. In the case of dinoflagellates, Fig. 3*C* shows that helix 17 seems to be more homologous to that of archaeobacteria.

Table 1. Nucleotide compositions and lengths of various small-ribosomal-subunit rRNAs

Species	Nucleotides, no. A	U	G	C	Length, nt	G+C content, %
R.r.	421	408	543	497	1869	56
X.l.	432	411	516	466	1825	54
Z.m.	447	439	500	419	1805	51
S.c.	475	509	458	347	1789	45
P.m.	474	496	469	359	1798	46
T.t.	518	485	433	317	1753	43
D.d.	531	546	448	348	1873	43

The data for *T. thermophila* (*T.t.*) are from ref. 32, for *D. discoideum* (*D.d.*) from ref. 33, and for *Rattus ratus* (*R.r.*), *X. laevis* (*X.l.*), *Zea mays* (*Z.m.*), and *S. cerevisiae* (*S.c.*) from ref. 34.

All other eukaryotes are of the yeast type. Helix 16 of *P. micans* is very similar to that of *H. volcanii* (42): both loops present a UAU sequence.

(*v*) Helix E19. This eukaryotic region is highly variable, while the E19 secondary structure of *P. micans* is hypothetical, but it does show some similarities with that proposed by Atmadja *et al.* (44), for *X. laevis*.

(*vi*) Central region. This region can be folded into two alternative structures, P18-1 and P18-2 for prokaryotes (Fig. 2) and E18-1, E21-1, and E21-2 for eukaryotes (Fig. 2 *Insert*). It may confer a fundamental difference in the folding pattern to prokaryotic and eukaryotic small-ribosomal-subunit rRNA (for discussion see ref. 29). In the case of *P. micans*, the two topologies are possible (as for ciliates, data not shown). But the prokaryotic-like bulged helix of *Dictyostelium* (33) and *Artemia* (29) cannot be constructed. This may indicate a more distant position for these two species from prokaryotes.

(*vii*) Helix 28. This helix is quite similar in archaeobacteria and eukaryotes, but not in eubacteria (22). The sequence of the loop is GAU in archaeobacteria (41, 42, 45) and in *P. micans* (Fig. 3*D*). But all other protoctists have different sequences (GAG for yeast and ciliates, AAA for slime mold), which means that the function of this part of the molecule is more conserved between dinoflagellates and archaeobacteria.

(*viii*) Helix 31. This helix is conserved in both prokaryotes and eukaryotes. The loop sequence is GUGA in prokaryotes (22, 41, 42, 45), dinoflagellates, yeast (Fig. 3*E*), and ciliates (data not shown), but in slime mold and higher eukaryotes it is GCGA. This difference is presumably due to a mutation that appeared after the divergence of dinoflagellates, yeasts, and ciliates from the eukaryotic lineage.

(*ix*) Helix 39. This helix is 2 nt longer for *P. micans* than *Dictyostelium*, but shorter than that of any other eukaryote and longer than that of any prokaryote. The 3′ end of the 17S rRNA from *P. micans* resembles that of other eukaryotes, as it does not present the prokaryotic Shine–Dalgarno sequence, but contains the region homologous to the 5′ end of mRNA (46). This region, in *Escherichia coli*, has been shown to be involved in a possible alternative conformation with helix 25 (47), which is also possible in *P. micans*.

Table 2. Percent homologies and evolutionary distances

	X.l.	S.c.	Z.m.	D.d.	P.m.	H.v.
X.l.	0/100%	218	207	318	242	601
S.c.	84.5%	0/100%	176	291	180	579
Z.m.	85.2%	87.4%	0/100%	274	156	592
D.d.	77.3%	79.2%	80.4%	0/100%	277	600
P.m.	82.7%	87.2%	88.8%	80.2%	0/100%	561
H.v.	57%	58.7%	57.7%	57.2%	60%	0/100%

Percentage homologies are below the diagonal and evolutionary distances are above the diagonal between various small-ribosomal-subunit rRNAs. Evolutionary distances are calculated by means of nucleotide differences. All percentages were calculated by taking the total number of similarities between two sequences (using the alignment of ref. 29) and dividing by 1410, the total number of positions that we used. Gaps and insertions were counted as a single mismatch. We have deleted the variable regions (boxed in Fig. 2) We introduced some minor modifications to the alignment of ref. 29, as shown in Fig. 3. Abbreviations are as in Table 1 and *H.v.*, *H. volcanii*.

Sequence Comparison Among Small-Ribosomal-Subunit rRNA Sequences. We added the *P. micans* sequence shown in Fig. 2 to those aligned by Nelles *et al.* (29). This sequence comparison included one archaeobacterium, *H. volcanii* (Monera) (42); one Acrasiomycota, *D. discoideum* (Protoctista) (33); one Ascomycota, *S. cerevisiae* (Fungi) (48); one Angiospermophyta, *Z. mays* (Plantae) (49); one Chordata, *X. laevis* (Animalia) (44); and *P. micans* (this work). This alignment (not shown) reveals extensive homologies between the dinoflagellate sequence and other eukaryotic sequences, and some strong homologies with the archaeobacterial sequence. It is now well known that some regions of small-ribosomal-subunit rRNA are highly variable and are valuable only for resolving close phylogenetic relationships (33). Thus, we have calculated a different matrix for these six sequences (Table 2), excluding their variable regions (boxed in Fig. 2), corresponding to 1410 nt; gaps or insertions were counted as a single mismatch (whatever the length). The few modifications that we introduced to the alignment of ref. 29 are those presented in Fig. 3. Maximum homology is obtained between *P. micans* and *Z. mays* (88%) and the minimum between *X. laevis* and *D. discoideum* (77%). With the archaeobacterial sequence, all five eukaryotes share between 57% (*Xenopus*) and 60% (*Prorocentrum*) homology. It may be

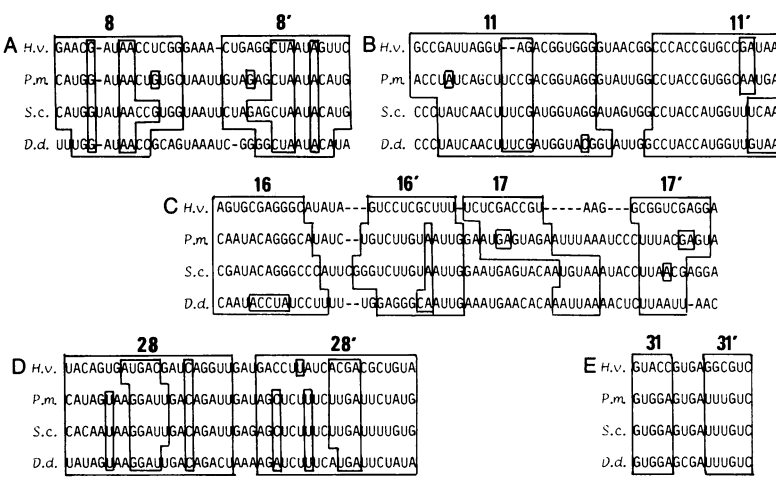

FIG. 3. Alignment of parts of the small-ribosomal-subunit rRNA sequences that are distinctive for dinoflagellates. The following sequences are aligned and indicated by the initials of genus and species name (from top to bottom): *H. volcanii* (*H.v.*), *P. micans* (*P.m.*), *S. cerevisiae* (*S.c.*), and *D. discoideum* (*D.d.*). Boxes labeled 8–8′, etc., correspond to the helices, as in Fig. 2. The interior loops and bulges are defined by nested boxes.

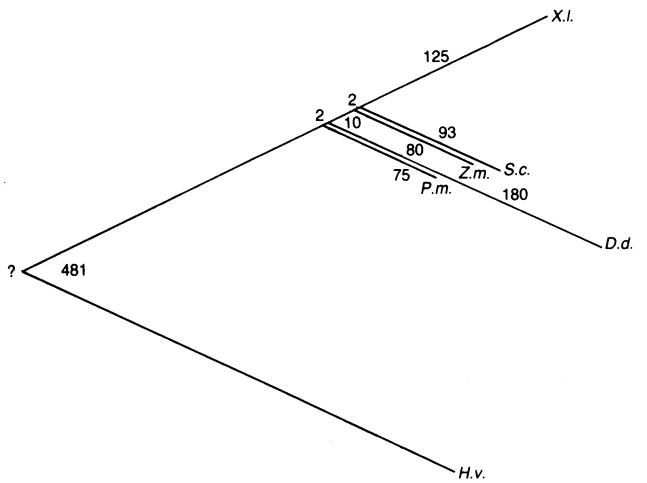

FIG. 4. Phylogenetic tree. Optimum phylogenetic tree was calculated from the difference matrix (Table 2) as described (30). The branch lengths are the number of mutational differences calculated. Abbreviations of the species name are as in Table 1.

striking that the maximum homology is shared by *H. volcanii* and *P. micans*. This would correspond either to evolutionary convergence or to a significant conservation of structural similarities common to archaeobacteria and to dinoflagellates. As discussed by Eckenrode *et al.* (34), the percentage homology between two random sequences is around 50%, which is close to that between *E. coli* and eukaryotes. So, the 60% homology between *H. volcanii* and *P. micans* is probably significant. This homology may be related to the structural homology found in helices 8, 17, 28, and 31, between the dinoflagellate and the archaeobacterial sequence. On the other hand, the similarity between *Dictyostelium* and the other eukaryotes is the lowest. This has been interpreted by McCaroll *et al.* (33) as the earliest branching in the eukaryotic lineage, rather than convergent evolution or rapid evolutionary drift. These authors have proposed a eukaryotic tree where *D. discoideum* branches off first. Using the same procedure as that which we have used for the 5.8S rRNA sequence (30), we obtained a very similar tree calculated from the data of Table 2, where dinoflagellates branch first, followed by *D. discoideum* (Fig. 4). This tree is very homologous to that of Hasegawa *et al.* (23) and presents the same ambiguity concerning the respective departure of plants and fungi; this is also the case for the slime mold and the dinoflagellates. This early branching agrees well with the data obtained from the sequence comparison of Fig. 3, where some structural features are more similar to the archaeobacteria than any other eukaryote. However, the published (50) *Euglena gracilis* sequence appears more divergent than other eukaryotes, but structural comparison with archaeobacteria has not been presented.

The data obtained from the analysis of the dinoflagellate 17S rRNA sequence seem to confirm morphological data about their primitive eukaryotic organization and suggest an early branching from the eukaryotic lineage. In addition, by isolation of the rRNA gene, we may use the promoter as a tool to check further functional evidence (on gene regulation) of the primitive position of dinoflagellates.

We thank Dr. M. Delseny for gift of pBG 35 plasmid. We are indebted to Drs. A. Picard and R. Cooke for critical reading of the manuscript and to Dr. M. O. Soyer for constant support. The technical assistance of D. St. Hilaire and M. Albert is gratefully acknowledged. This work was supported by grants from Centre National de la Recherche Scientifique (AIP 95311).

1. Margulis, L. & Schwarz, K. V. (1982) *Five Kingdoms* (Freeman, San Francisco).
2. Bujak, J. P. & Williams, G. L. (1981) *Can. J. Bot.* **59**, 2077–2087.
3. Spector, D. L. (1984) *Dinoflagellates* (Academic, New York).
4. Herzog, M., Boletzky, S. & Soyer, M. O. (1984) *Origins Life* **13**, 205–215.
5. Rizzo, P. J. (1984) *BioSystems* **16**, 211–216.
6. Jing-Yang, L. (1984) *BioSystems* **16**, 217–222.
7. Rizzo, P. J. (1981) *BioSystems* **14**, 433–443.
8. Bodansky, S., Mintz, L. B. & Holmes, D. S. (1979) *Biochem. Biophys. Res. Commun.* **88**, 1329–1336.
9. Herzog, M. & Soyer, M. O. (1981) *Eur. J. Cell Biol.* **23**, 295–302.
10. Herzog, M. & Soyer, M. O. (1983) *Eur. J. Cell Biol.* **30**, 33–41.
11. Soyer-Gobillard, M. O. & Herzog, M. (1985) *Eur. J. Cell Biol.* **36**, 334–342.
12. Weintraub, H. (1985) *Cell* **42**, 705–711.
13. Allen, J. R., Roberts, T. M., Loeblich, A. R., III, & Klotz, L. C. (1975) *Cell* **6**, 161–169.
14. Rae, P. M. M. (1970) *J. Cell Biol.* **46**, 106–113.
15. Hinnebusch, A. G., Klotz, L. C., Blanken, R. L. & Loeblich, A. R., III (1981) *J. Mol. Evol.* **17**, 334–347.
16. Erdmann, V. A., Walters, J., Huysmans, E. & De Wachter, R. (1985) *Nucleic Acids Res.* **13**, Suppl., r105–r153.
17. Zillig, W., Schnabel, R. & Stetter, K. O. (1985) *Curr. Top. Microbiol. Immunol.* **114**, 1–18.
18. Werner, E., Sohst, S., Grapp, F., Simon, D., Wagner, H. & Kroger, H. (1984) *Eur. J. Biochem.* **139**, 81–86.
19. Reddy, R., Spector, D., Liu, M. H. & Busch, H. (1983) *J. Biol. Chem.* **258**, 13965–13969.
20. Liu, M. H., Reddy, R., Henning, D., Spector, D. & Busch, H. (1984) *Nucleic Acids Res.* **12**, 1529–1542.
21. Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego Dias, L. & Pousada, C. (1983) *Nucleic Acids Res.* **11**, 8359–8367.
22. Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. (1983) *Microbiol. Rev.* **47**, 621–669.
23. Hasegawa, M., Iida, Y., Yani, T., Takaiwa, F. & Iwabuchi, M. (1985) *J. Mol. Evol.* **22**, 32–38.
24. Chirgwin, J. H., Przybla, A. E., McDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
25. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
26. Goldsbrough, P. B. & Cullis, C. A. (1981) *Nucleic Acids Res.* **9**, 1301–1309.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
29. Nelles, L., Fang, B. L., Volkaert, G., Vanderberghe, A. & De Wachter, R. (1984) *Nucleic Acids Res.* **12**, 8749–8768.
30. Maroteaux, L., Herzog, M. & Soyer-Gobillard, M. O. (1985) *BioSystems* **18**, 307–319.
31. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
32. Spangler, E. A. & Blackburn, E. H. (1985) *J. Biol. Chem.* **260**, 6334–6340.
33. McCaroll, R., Olsen, G. J., Stahl, Y. D., Woese, C. R. & Sogin, M. L. (1983) *Biochemistry* **225**, 5858–5868.
34. Eckenrode, V. K., Arnold, J. & Meagher, R. B. (1985) *J. Mol. Evol.* **21**, 259–269.
35. Gerbi, S. A., Gourse, R. L. & Clarck, C. G. (1982) *Cell Nucl.* **10**, 351–393.
36. Elwood, H. J., Olsen, G. J. & Sogin, M. L. (1985) *Mol. Biol. Evol.* **2**, 399–410.
37. Brimacombe, R. & Stiege, W. (1985) *Biochem. J.* **229**, 1–17.
38. Stiegler, P., Carbon, P., Zucke, M., Ebel, J. P. & Ehresmann, C. (1981) *Nucleic Acids Res.* **9**, 2153–2172.
39. Chan, Y. L., Gutell, R. R. & Noller, H. F. (1984) *J. Biol. Chem.* **259**, 224–230.
40. Olsen, G. J., McCaroll, R. & Sogin, M. L. (1983) *Nucleic Acids Res.* **11**, 8037–8049.
41. Leffers, H. & Garett, R. A. (1984) *EMBO J.* **3**, 1613–1619.
42. Gupta, R., Lauter, J. M. & Woese, C. R. (1983) *Science* **221**, 656–659.
43. Atmadja, J., Brimacombe, H., Blocker, H. & Frank, R. (1985) *Nucleic Acids Res.* **13**, 6919–6936.
44. Atmadja, J. R., Brimacombe, R. & Maden, H. (1984) *Nucleic Acids Res.* **12**, 2649–2667.
45. Lechner, K., Wich, G. & Bock, A. (1985) *Syst. Appl. Microbiol.* **6**, 157–163.
46. Hagenbuchle, O., Santer, M., Steitz, J. A. & Mans, J. (1978) *Cell* **13**, 551–563.
47. Thompson, J. F. & Hearst, J. E. (1983) *Cell* **33**, 19–24.
48. Hogan, J. J., Gutell, R. R. & Noller, H. F. (1984) *Biochemistry* **23**, 3322–3330.
49. Messing, J., Carlson, J., Hagen, G., Rubenstein, I. & Oleson, A. (1984) *DNA* **3**, 31–40.
50. Sogin, M. L., Elwood, H. J. & Gunderson, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383–1387.