J Mol Evol (1991) 32:53–63

Journal of Molecular Evolution

A Molecular Phylogeny of Dinoflagellate Protists (Pyrrhophyta) Inferred from the Sequence of 24S rRNA Divergent Domains D1 and D8

Guy Lenaers,[1] Christopher Scholin,[2] Yvonne Bhaud,[1] Danielle Saint-Hilaire,[1] and Michel Herzog[1]

[1] Département de Biologie Cellulaire et Moléculaire, Laboratoire Arago, Université de Paris VI, CNRS UA 117, F-66650 Banyuls-sur-Mer, France
[2] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Summary. The sequence of two divergent domains (D1 and D8) from dinoflagellate 24S large subunit rRNA was determined by primer extension using total RNA as template. Nucleotide sequence alignments over 401 bases have been analyzed in order to investigate phylogenetic relationships within this highly divergent and taxonomically controversial group of protists of the division Pyrrhophyta. Data are provided confirming that dinoflagellates represent a monophyletic group. For 11 out of the 13 investigated laboratory grown species, an additional domain (D2) could not be completely sequenced by reverse transcription because of a hidden break located near its 3'-terminus. Two sets of sequence alignments were used to infer dinoflagellate phylogeny. The first [199 nucleotides (nt)] included conservative sequences flanking the D1 and D8 divergent domains. It was used to reconstruct a broad evolutionary tree for the dinoflagellates, which was rooted using *Tetrahymena thermophila* as the outgroup. To confirm the tree topology, and mainly the branchings leading to closely related species, a second alignment (401 nt) was considered, which included the D1 and D8 variable sequences in addition to the more conserved flanking regions. Species that showed sequence similarities with other species lower than 60% on average (Knuc values higher than 0.550) were removed from this analysis. A coherent and convincing evolutionary pattern was obtained for the dinoflagellates, also confirmed by the position of the hidden break within the D2 domain, which appears to be group specific. The reconstructed phylogeny indicates that the early emergence of *Oxyrrhis marina* preceded that of most Peridiniales, a large order of thecate species, whereas the unarmored Gymnodiniales appeared more recently, along with members of the Prorocentrales characterized by two thecal plates. In addition, the emergence of heterotrophic species preceded that of photosynthetic species. These results provide new perspectives on proposed evolutionary trees for the dinoflagellates based on morphology, biology, and fossil records.

Key words: Dinoflagellates — Molecular phylogeny — Large subunit RNA — Divergent domains — Direct RNA sequencing

Introduction

Ribosomal RNAs provide molecular markers that are particularly informative in the study of phylogeny because their function and structure have been conserved to a large extent through the evolutionary history of organisms. They constitute, for subsequent reference, a growing data base consisting of partial rRNA sequences determined by primer extension methods using reverse transcriptase (Qu et al. 1983; Lane et al. 1985), as well as of cloned rDNA sequences. Sequences of the largest rRNAs (16–18S and 23–28S) are thus far the most powerful molecular markers to evaluate either long- or short-range phylogenies. Pro- and eukaryotic large subunit rRNA

(23–28S LsuRNA), for instance, display a common, largely conserved structural core, which in eukaryotes is interspersed with 12 divergent, more rapidly evolving domains (D1–D12) (Hassouna et al. 1984; Michot et al. 1984). The conservative core [>2000 nucleotides (nt)] constrained by heavy selective pressure, is suitable for phylogenetic evaluation among distant taxa and has been used recently to infer evolutionary relationships among archaebacteria (Leffers et al. 1987; Gouy and Li 1989a) and eukaryote kingdoms (Cedergren et al. 1988; Lenaers et al. 1989; Gouy and Li 1989b). Partial sequences limited to conservative domains near the 5′ end of the LsuRNA have been used to infer phylogenetic relationships among helminths (Qu et al. 1986; Gill et al. 1988), protists (Baroin et al. 1988), and algae (Perasso et al. 1989).

On the other hand, divergent domains display a high rate of sequence variation and in some cases (D2, D8, and D12) have dramatically increased in length during evolution (Michot and Bachellerie 1987). Therefore, they do not provide useful information for the comparison of distant organisms. However, some of these domains (mainly D1, D3, D8, and, to some extent, D2) have the potential to be useful for phylogenetic and taxonomic analyses of closely related species. So far, a limited number of analyses have been conducted in order to test the information of divergent domains for such short-range phylogenies. Qu et al. (1988) have investigated the 5′ terminal domain of eukaryotic LsuRNAs. They demonstrated that the D1 divergent domain can be successfully used to infer phylogenetic relationships among related taxa, provided the clustering of groups of closely related species has been established by comparison of the two flanking conservative tracks. In their analysis of relationships between the protists, Baroin et al. (1988) showed that the clustering of five ciliate species was identical when comparing the sequences of either the D1 divergent domain or flanking conservative regions. In a previous comparison of primary and secondary structure of divergent domains (Lenaers et al. 1988) we defined a 277-nt-long sequence alignable among protists, representing all or part of the fast evolving D3, D8, D9, and D10 domains. This analysis allowed us to derive close relationships between yeast, ciliates, and dinoflagellates, which were later confirmed by conservative core sequence comparisons (Lenaers et al. 1989).

Here, we further investigate two variable domains (D1 and D8) and their flanking conservative regions and evaluate their usefulness for the inference of phylogenetic relationships among the dinoflagellates.

Dinoflagellate protists (Pyrrhophyta) represent a highly diversified phylum, consisting of two classes (Dodge 1984), Dinophyceae (about 130 genera and 4000 species) including mostly free-living sea-water species (photosynthetic or heterotrophic), and intracellular parasitic Syndiniophyceae (about 40 species). Members of the main class, Dinophyceae, are mainly characterized by a complex cellulose cell covering (thecae) and a peculiar nucleus having (almost) permanently condensed chromosomes and a chromatin devoid of histones and nucleosomes (Herzog and Soyer 1981; Rizzo 1987; Vernet et al. 1990).

The great ultrastructural and biochemical diversity among dinoflagellate species has resulted in conflicting classifications and phylogenetic models. Tentative classifications of extant and fossil dinoflagellates [the earliest unequivocal records are from the Late Silurian, 400 million years (Myr) ago] are primarily based on morphology of motile cells and cysts, number, shape, and tabulation of the thecal plates, motility, ultrastructure, and limited biochemical data (Loeblich 1976, 1984; Taylor 1980, 1987; Dodge, 1984; Sournia 1986). Evolutionary patterns proposed for Dinophyceae are mainly based upon the five recognized organizational types (prorocentroid, dinophysoid, gymnodinioid, gonyaulacoid, and peridinioid), which have been taken to represent lineages derived from ancestral desmokonts (apical flagella; Taylor 1980).

Three conflicting models (reviewed in Bujak and Williams 1981; Goodman 1987) have been proposed to describe phylogenetic relationships between main orders within the division, with emphasis on thecal plate numbers and arrangement: (1) a plate increase model mainly inferred from biological considerations (Loeblich 1976; Taylor 1980) suggesting that thecate species (Prorocentrales, Peridiniales) represent a more primitive condition, whereas unarmored forms (Gymnodiniales) are a derived, or more advanced, state; (2) a plate reduction model supported by Mesozoic to Tertiary fossil records (limited to cyst-forming species: Dörhörfer and Davis 1980; Eaton 1980), later adopted by Loeblich (1984), suggesting that a large number of thin plates preceded smaller numbers so that the Gymnodiniales are considered the most primitive forms and the Prorocentrales the more advanced; and (3) a plate fragmentation model (Bujak and Williams 1981) combining biological and paleontological patterns, according to which Prorocentrales (anteriorly flagellated, cellulosic thecae primarily composed of two valves) represent the most ancient lineage. Differentiation of the thecae into numerous plates and a change in the flagella insertion gave rise to the Gonyaulacales and most Peridiniales, and ultimately in one line to the Gymnodiniales through reduction of the thecae to the unarmored condition.

In order to provide independent criteria for eval-

Table 1. Reference list of the investigated dinoflagellate species

Species	Strain	Abbreviation	Order[a]	Culture collection source	Medium[b]
Prorocentrum micans	1136/1	*P.m.*	Prorocentrales	Cambridge, UK	Erdschreiber
Cachonina niei	ME46	*C.n.*	Peridiniales incertae sedis	Helgoland, FRG	F/2
Heterocapsa pygmea	ME71	*H.p.*	Peridiniales incertae sedis	Helgoland, FRG	F/2
Amphidinium carterae	1102/1	*A.c.*	Gymnodiniales	Cambridge UK	F/2
Gymnodinium sp.	—	*G.*sp.	Gymnodiniales	Banyuls-sur-Mer, France	F/2
Woloszynskia coronata	1117/2	*W.c.*	Gymnodiniales	CCAP, Scotland	B.B.M.
Alexandrium[c] *catenella*	Bgt1	*Al.c.*	Peridiniales	Woods Hole, USA	F/2
Alexandrium[c] *tamarense*	Pgt183	*Al.t.*	Peridiniales	Woods Hole, USA	F/2
Gonyaulax polyedra	Gp70	*G.p.*	Peridiniales	Woods Hole, USA	F/2
Pyrocystis lunula	ME36	*P.l.*	Pyrocystales	Helgoland, FRG	Provasoli
Noctiluca scintillans	—	*N.s.*	Noctilucales	Banyuls-sur-Mer, France	Local bloom
Crypthecodinium cohnii	WHD	*C.c.*	Peridiniales	Woods Hole, USA	M.L.H.
Oxyrrhis marina	—	*O.m.*	Oxyrrhinales	Villefranche-sur-Mer, France	F/2

[a] Order based on Sournia (1986)

[b] Medium references: Erdschreiber (Starr 1964); F/2 (Guillard and Ryther 1963); B.B.M. (Bischoff and Bold 1963); Provasoli (Provasoli 1963); M.L.H. (Tuttle and Loeblich 1975)

[c] Formerly *Protogonyaulax*

uating these models, we have determined and compared, for 13 dinoflagellates, the nucleotide sequence of two rapidly evolving domains (D1 and D8) from 24S RNA. The secondary structure of the domains was determined in order to improve sequence alignments. Combination of the two domains significantly increased the number of informative nucleotides for phylogenetic analysis of dinoflagellates, using distance matrix methods.

Material and Methods

Sources of Strains, Cultures, and RNA Purification. Taxonomic position of the various dinoflagellate species (Sournia 1986) and references for the growth media are indicated in Table 1.

Cells from 2–4-l cultures were harvested by centrifugation and the pellet rinsed three times with fresh medium. The pellet was resuspended in 5 vol of 7 M guanidinium isocyanate made up in 5 mM sodium citrate (pH 7.0), 10 mM 2-mercaptoethanol, and 0.5% Sarkosyl (Maniatis et al. 1982). Disruption of cells was achieved by passing the suspension through a French press at 4000 psi. Insoluble residue was eliminated by centrifugation at 12,000 × *g* for 30 min. Total RNA was purified essentially as described by Maniatis et al. (1982), by centrifugation through a 5.7 M CsCl cushion and subsequent phenol/chloroform extraction. Alternatively, when polysaccharides and polyphenols interfered with RNA isolation, a selective precipitation method was preferred. Broken cells were phenol extracted, nucleic acids were ethanol precipitated, and RNA was separated by precipitation in the presence of 2 M $LiCl_2$ (Ausubel et al. 1987).

RNA Sequencing. Sequences of the divergent domains (D1, D2, and D8) were determined by reverse transcriptase-mediated primer extension reaction on total rRNA using a modification of the original methods of Qu et al. (1983) and Lane et al. (1985), which we describe elsewhere (Mouches et al. 1990). Three ^{32}P-end-labeled primers complementary to the 24S rRNA sequence of the dinoflagellate *Prorocentrum micans* (Lenaers et al. 1989) corresponding to the coordinates 369–388 (D1), 714–733 (D2), and 2151–2175 (D8), were used to prime the reactions.

Sequence Comparison and Phylogeny Inference. Sequences were aligned on the basis of maximum nucleotide similarity. Best alignment was achieved using secondary structure criteria. The phylogenetic distance has been estimated by means of nucleotide difference (Kimura 1968) and Knuc values (Kimura 1980). Deletion of one nucleotide was counted as a transversion (Qu et al. 1988). Trees were inferred by the neighbor-joining method (Saitou and Nei 1987; modified by Studier and Keppler 1988) computerized for an IBM-PC, and by the Fitch–Margoliash program of Felsenstein's PHYLIP package (Fitch and Margoliash 1967; Fitch 1981).

Results and Discussion

During the course of evolution, divergent domains of the large rRNA molecules generally have not been constrained by heavy selective pressure. However, structural comparisons have shown that some of these divergent domains appear conservative as far as related taxa are concerned (Michot and Bachellerie 1987; Lenaers et al. 1988). They represent, therefore, potential markers to evaluate phylogenetic relationships among closely related species. We have determined the sequence of three divergent domains (D1, D8, and part of D2) of LsuRNA from 13 species of free-living dinoflagellates representing six main orders (Oxyrrhinales, Prorocentrales, Peridiniales, Pyrocystales, Noctilucales, and Gymnodiniales) in order to provide additional criteria for evaluating the phylogenetic relationships among these organisms.

For 11 out of the 13 species, we were unable to directly sequence the D2 domain due to the presence of a hidden break located near its 3′ end, previously documented for *P. micans* (Lenaers et al. 1989). This break, denoted by arrowheads in the D2 alignment of Fig. 3, appeared at three defined locations

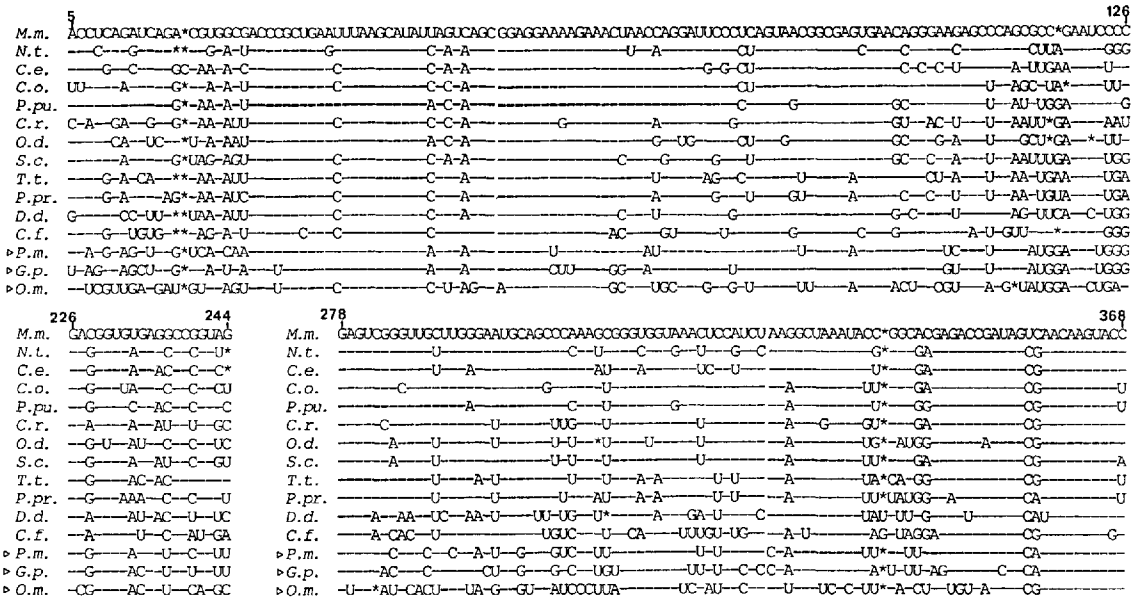


Fig. 1. Sequence alignment of the conservative domains (235 nt) from the 5′ end of the nuclear-encoded LsuRNA. The alignment was taken from Perasso et al. (1989) and completed with the corresponding sequences from the ciliate *Tetrahymena thermophila* (T.t.) (Engberg et al. 1990) and the dinoflagellates *Prorocentrum micans* (P.m.), *Gonyaulax polyedra* (G.p.), and *Oxyrrhis marina* (O.m.) (this work). Numbers refer to the listed sequence of *Mus musculus* (M.m.). Identical nucleotides are indicated by dashes, and deletions or alignment gaps are represented by stars. Other abbreviations are for: *Nicotiana tabacum* (N.t.), *Chlorogonium elongatum* (C.e.), *Cryptomonas ovata* (C.o.), *Porphyridium purpureum* (P.pu.), *Saccharomyces cerevisiae* (S.c.), *Crycosphaera roscoffensis* (C.r.), *Ochromonas danica* (O.d.), *Parамecium primaurelia* (P.pr.), *Dictyostelium discoideum* (D.d.), and *Crithidia fasciculata* (C.f.).

depending on species (see below). Only two species (*Crypthecodinium cohnii* and *Gonyaulax polyedra*) did not contain the hidden break. Various organisms display a cut LsuRNA molecule, which in some cases is associated with a processing event leading to the elimination of several nucleotides. It has been mapped, for instance, to the D7a region in *Drosophila melanogaster* (de Lanversin and Jacq 1983, 1989) and *Tetrahymena thermophila* (Engberg et al. 1990), or to the D7b and D7c in trypanosomids (Campbell et al. 1987; Spencer et al. 1987). In all these cases, like in dinoflagellates, hidden breaks are located in unpaired sequences of divergent domains.

Monophyletic Origin of Dinoflagellates

In a previous work (Lenaers et al. 1989) based on available sequences of the LsuRNA conservative core (1900 nt), we examined the phylogenetic relationships among protists (including one dinoflagellate) and higher eukaryotes. The dinoflagellate *P. micans* clustered with the ciliate *T. thermophila* and yeast *Saccharomyces cerevisiae* when the alignment was analyzed by distance methods. The tree topology inferred with parsimony slightly differed, with *S. cerevisiae* emerging from the main branch leading to higher eukaryotes rather than clustering with *T. thermophila* and *P. micans*. Close relationships among the three species had also been suggested by comparison of partial sequences from the 5′ end of the molecule (Baroin et al. 1988) and of several divergent domains alignable among protists (Lenaers et al. 1988).

The aim of the present work was to analyze relative positions among dinoflagellates. It was important to first confirm that dinoflagellates behave as a monophyletic group, by comparing partial sequences from several species with the corresponding regions from a variety of eukaryotes. We have used the sequence alignment of the 5′ end of LsuRNA recently published by Perasso et al. (1989) in their analysis of the origin of algae. We have selected 10 species representative of metaphytes, algae, fungi, and protists and added three distant dinoflagellates (see below) that were sequenced up to the 5′ end: *Oxyrrhis marina* (Oxyrrhinales), *G. polyedra* (Peridiniales), and *P. micans* (Prorocentrales). We also added the ciliate *T. thermophila* for which the complete LsuRNA sequence has recently been published (Engberg et al. 1990). The alignment shown in Fig. 1 only comprises the conservative domains (235 nt) that have been used by Perasso et al. (1989) to reconstruct a phylogeny of the algae. A distance and Knuc values matrix (Fig. 2A) was derived from this alignment and analyzed by the neighbor-joining method of Saitou and Nei (1987). The reconstructed phylogenetic tree (Fig. 2B), which has a topology similar to that initially proposed and discussed by Perasso et al. (1989), clearly shows that the three dinoflagellates, representing distant orders as demonstrated below, share a direct common ancestor. However, the present analysis limited to partial se-

A

	N.t.	Ca.	Co.	P.pu	C.r.	Od.	Sc.	T.t.	P.pr	Dd.	C.f.	P.m.	G.p.	O.m.
N.t.	0	37	41	40	57	51	43	56	57	65	69	57	66	95
Ca.	.184	0	37	35	48	37	39	46	44	65	72	62	76	90
Co.	.202	.177	0	31	44	45	34	45	44	64	71	57	69	96
P.pu	.203	.172	.146	0	40	45	34	43	44	61	64	53	66	92
C.r.	.301	.236	.214	.190	0	51	35	51	52	59	76	56	67	89
Od.	.271	.292	.226	.233	.259	0	42	54	56	64	80	56	73	93
Sc.	.217	.188	.157	.164	.162	.208	0	42	42	60	66	52	69	85
T.t.	.300	.231	.218	.213	.256	.280	.214	0	28	57	70	56	74	91
P.pr	.310	.220	.214	.222	.267	.296	.217	.135	0	61	73	52	73	89
Dd.	.364	.352	.350	.328	.306	.348	.315	.294	.322	0	70	66	77	95
C.f.	.394	.410	.395	.354	.426	.475	.360	.398	.414	.391	0	79	84	100
P.m.	.303	.333	293	.276	.280	.292	.261	.286	261	.352	.459	0	38	83
G.p.	.358	.427	384	.357	.354	.406	.369	.410	.402	.436	.492	.185	0	94
O.m.	.613	.554	606	.585	.531	.586	.503	.492	.538	.591	.663	.489	.578	0

B

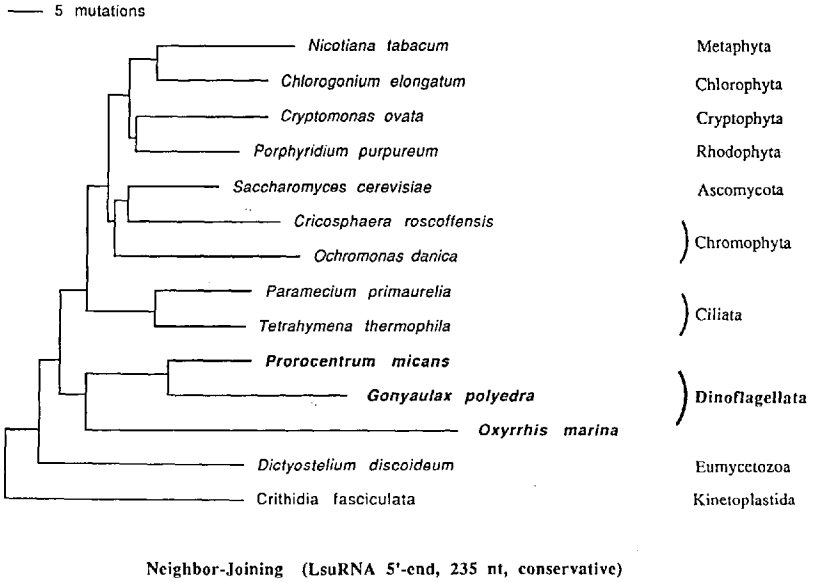


Fig. 2. Phylogenetic relationships among representatives of metaphytes, algae, fungi, and protists indicating monophyletic origin of the dinoflagellates. **A** Evolutionary distance matrix deduced from the 235-nt alignment of Fig. 1. Nucleotide distances between each pair of organisms are above the diagonal; Knuc values are below. Abbreviations are as in Fig. 1. **B** Phylogenetic tree based on the distance matrix, using the neighbor-joining method developed by Saitou and Nei (1987). The tree topology is similar to that originally derived by Perasso et al. (1989) and was rooted accordingly.

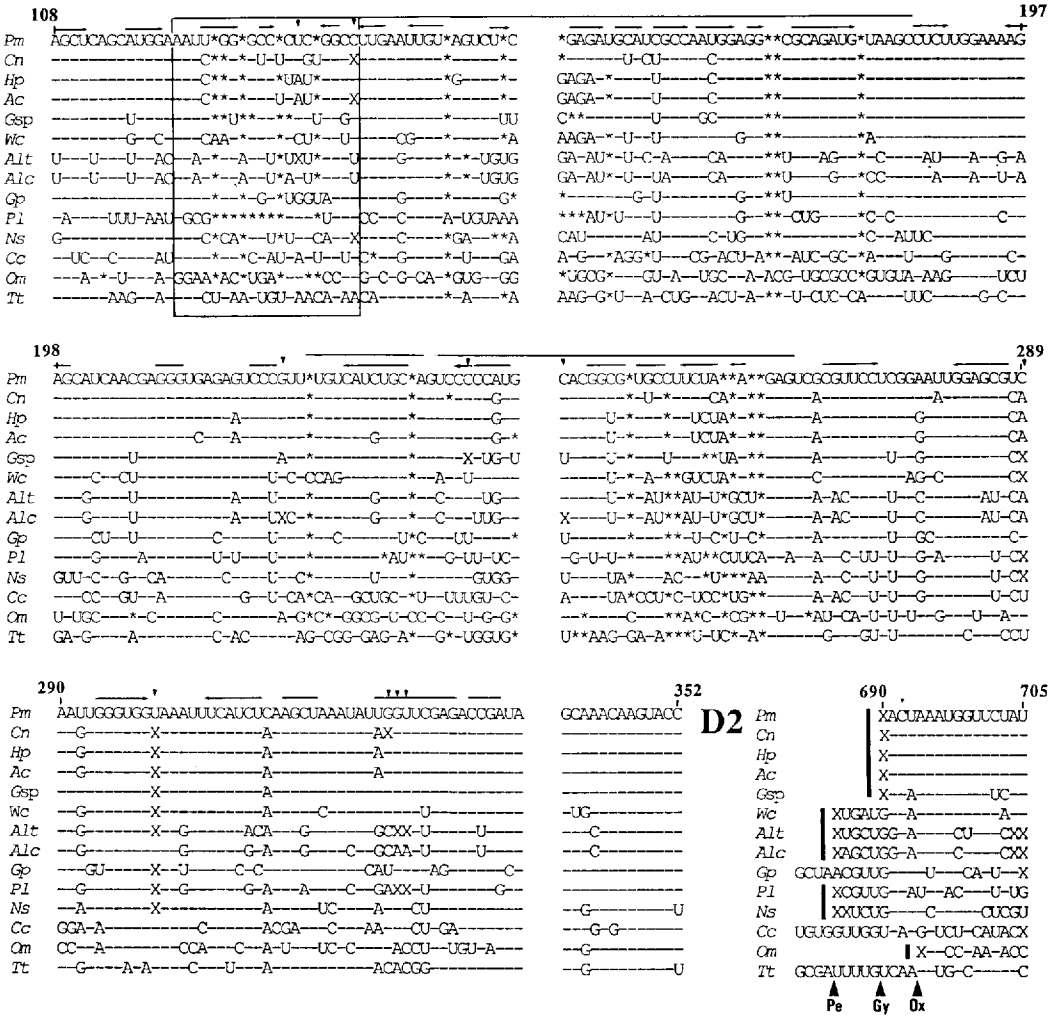
quences does not confirm the clustering of ciliates and dinoflagellates evidenced by the analysis of the whole conservative core (Lenaers et al. 1989), although the two groups remain related. We believe that the clustering of the two groups is more significant statistically than the present topology, because it has been inferred from larger sequences.

This preliminary analysis thus confirms that dinoflagellates represent a monophyletic group which emerged close to the ciliates. In the following analysis of dinoflagellate phylogeny we have investigated the sequence of several divergent domains (D1, D2, and D8) in order to increase the number of informative sites, and used *T. thermophila* as an outgroup sequence.

Primary and Secondary Structure of Dinoflagellate, D1 and D8 Divergent Domains

Nucleotide sequences of D1 and D8 of the various species were aligned along with conservative flanking regions (Fig. 3). Rapidly evolving sequences (overlined in Fig. 3) represent about 200 nt, whereas conservative flanking regions investigated here amount to an additional 200 nt. Both D1 and D8 domains display comparable nucleotide length among dinoflagellates and ciliates, thus allowing proper alignments based on maximum sequence homology and secondary structure criteria (helices are denoted by arrows). It is clear from the alignment that two dinoflagellates (*O. marina* and *C. cohnii*) display high levels of nucleotide differences. *Oxyrrhis marina,* for instance, displays the most divergent D1 and D8 sequences when compared to *P. micans,* with only 54% and 46% nucleotide similarity, respectively.

The secondary structure of the two domains was inferred for the various species in order to improve sequence alignment. The synoptic representation of secondary structure conservation and sequence variation shown in Fig. 4 clearly indicates that D1 and D8 domains have retained a highly comparable secondary structure among dinoflagellates. The rapidly evolving portion of stem D1a (boxed area in Figs. 3 and 4A) displays saturating nucleotide changes and deletions among dinoflagellates and therefore was not considered for phylogenetic inference. Figure 4 also presents the nucleotide variations among investigated species. In divergent domain D1 (as defined in Hassouna et al. 1984; and Fig. 4A), 17 nt positions are conserved (denoted by capital A, C, G, or U) among all the 13 species, and an additional 6 (a, c, g, u) if the most divergent sequence from *O. marina* is not taken into account.


D1
108
197
198
289
290
352
D2
690
705
Pe
Gy
Ox
D8
1980
2071
2072
2143
Pm
Cn
Hp
Ac
Gsp
Wc
Alt
Alc
Gp
Pl
Ns
Cc
Om
Tt

Likewise, in the D8 domain (Fig. 4B), 12 positions are invariant among all the species, plus an additional 9 nt if the *O. marina* sequence is not considered. In the D1 and the D8 variable domains, 36 and 27 positions, respectively, have a conserved purine or pyrimidine state (R, Y), plus an additional 14 and 6 positions (r, y) if the *O. marina* sequence is omitted. Finally, highly variable positions (denoted by dots) represent as much as 51% and 58% of D1 and D8 sequences, respectively.

At this level of analysis, the data suggest that these rapidly evolving domains may be too divergent to derive an accurate phylogeny of the highly diversified dinoflagellates. However, when the two most divergent sequences are removed (*O. marina* and *C. cohnii*), variable positions then represent 46% and 47% of D1 and D8 sequences, respectively, providing potential information for evaluating distances among closely related species.

Dinoflagellate Phylogeny

Detailed phylogeny of the dinoflagellates was inferred by a two-step analysis of the combined D1 and D8 sequences. A first set of aligned sequences was used to infer a broad phylogeny including all 13 investigated species and *T. thermophila*, the outgroup sequence. The portion of the molecule that was compared (199 nt in combined length, not overlined in Fig. 3) included conservative domains known to evolve at a relatively constant rate (Leffers et al. 1987; Gutell and Fox 1988), flanking the more rapidly evolving regions. A distance and Knuc values matrix (Fig. 5A) was derived from this alignment and analyzed by the neighbor-joining method, leading to the tree shown in Fig. 5B. Similar tree topology was obtained when the Knuc values were analyzed using the Fitch and Margoliash program of Felsenstein's PHYLIP package. According to this

←

Fig. 3. Sequence alignment of the D1 and D8 divergent domains and positioning of the hidden break in D2. The sequence of *Prorocentrum micans* is listed. The corresponding sequence from the ciliate *Tetrahymena thermophila* (Engberg et al. 1990) was used as an outgroup sequence. Identical nucleotides are indicated by dashes, and deletions or alignment gaps are represented by stars. The nucleotide positions that could not be identified are denoted by X. The upper arrows indicate the sequences involved in double helical structures. The boxed area near the 5′ end of D1 corresponding to the distal part of stem D1a where nucleotide changes reach saturating levels, as well as ambiguous positions (denoted by upper arrowheads) have been removed from subsequent analysis. The conservative regions (199 nt in combined length) have not been overlined, whereas the rapidly evolving sequences (202 nt) have been overlined. Position of the hidden break in the D2 domain, which appears to be group specific, is indicated by arrowheads; Oxyrrhinales-like (Ox), Gymnodiniales-like (Gy), and Peridiniales-like (Pe). Species abbreviations are indicated in Table 1.

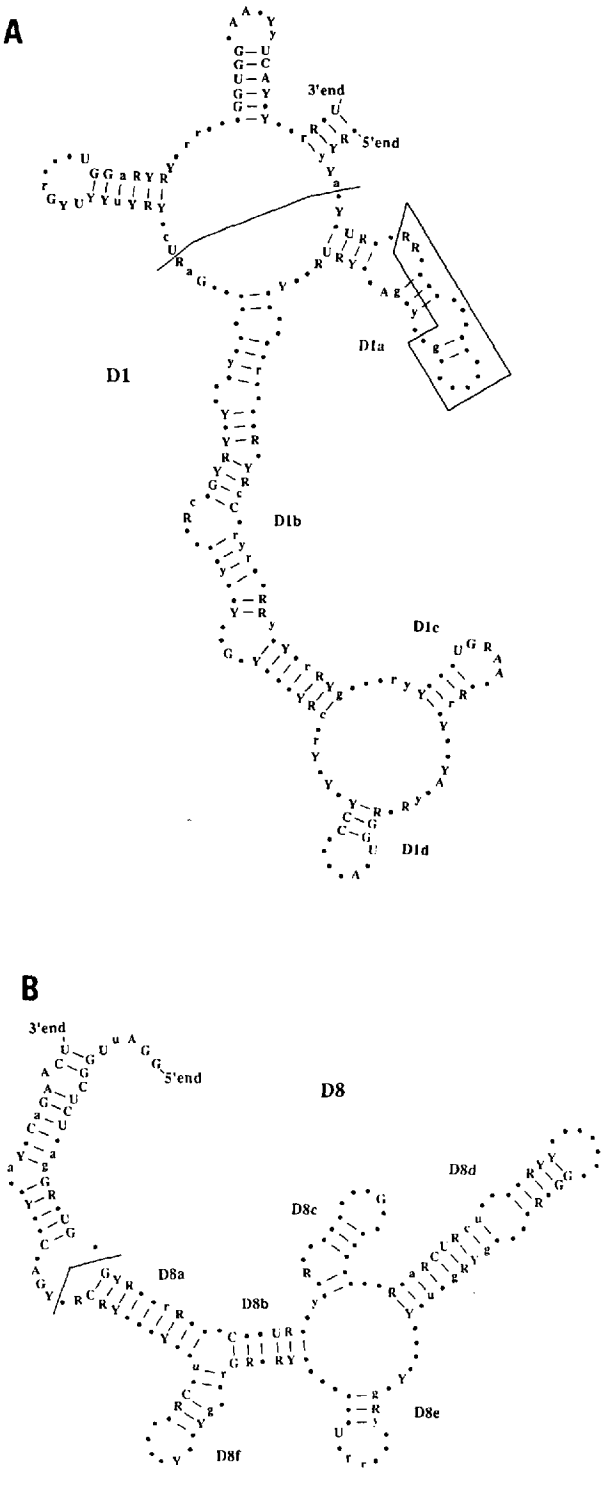


Fig. 4. A synoptic representation of sequence variations and secondary structure conservation of 24S rRNA divergent domains (**A**) D1 and (**B**) D8 from the various dinoflagellates listed in Table 1. Highly variable nucleotide positions are represented by dots, whereas the invariant nucleotides are indicated by their uppercase initial (A, C, G, U). Positions in which conservative purines (R) or pyrimidines (Y) are evident have also been marked. Corresponding lowercase letters indicate conservative positions when the most divergent species, *Oxyrrhis marina*, is removed from the analysis. The boxed area corresponding to the distal part of stem D1a is highly divergent and partially deleted in *Pyrocystis lunula* (cf. Fig. 3).

A

	P.m.	C.n.	H.p.	A.c.	G.sp.	W.c.	Al.t.	Al.c.	G.p.	P.l.	N.s.	C.c.	O.m.	T.t.
P.m.	0	7	9	10	10	25	39	39	25	34	42	52	75	52
C.n.	.036	0	4	5	9	22	35	35	25	28	38	46	76	51
H.p.	.047	.020	0	1	7	20	34	34	25	29	38	48	79	51
A.c.	.052	.026	.005	0	6	24	35	35	26	30	39	49	80	52
G.sp.	.053	.047	.036	.042	0	23	34	34	24	29	40	48	76	53
W.c.	.141	.123	.129	.135	.129	0	41	38	36	38	43	50	77	59
Al.t.	.237	.210	.204	.210	.202	.251	0	10	45	34	53	59	81	63
Al.c.	.236	.208	.202	.208	.200	.227	.053	0	45	31	53	56	81	65
G.p.	.141	.141	.141	.147	.134	.214	.281	.279	0	44	47	55	81	66
P.l.	.200	.161	.167	.174	.165	.228	.203	.181	.274	0	51	56	79	58
N.s.	.260	.232	.232	.239	.244	.269	.350	.347	.297	.330	0	56	76	59
C.c.	.334	.303	.304	.311	.300	.315	.393	.363	.356	.367	0	0	80	76
O.m.	.543	.556	.591	.602	.554	.563	.606	.606	.613	.582	.550	.604	0	84
T.t.	.339	.331	.333	.341	.345	.400	.437	.459	.475	.391	.395	.578	.640	0

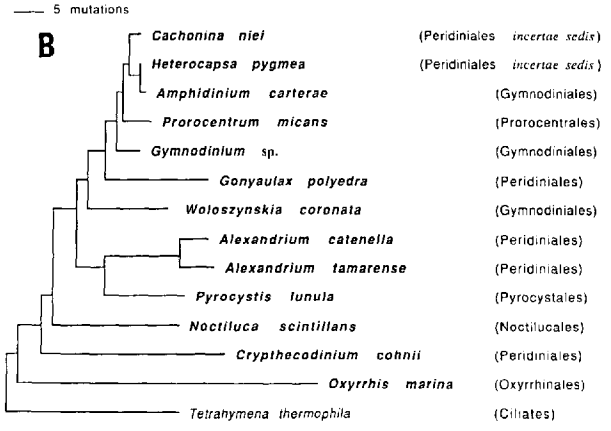


Fig. 5. Dinoflagellate phylogeny based on conservative sequence comparison. **A** Evolutionary distance matrix deduced from the 199-nt alignment corresponding to the conservative regions not overlined in Fig. 3. Nucleotide distances between each pair of organisms are above the diagonal; Knuc values are below. Abbreviations are as in Table 1. **B** Phylogenetic tree based on the distance matrix, using the neighbor-joining method developed by Saitou and Nei (1987). The tree was rooted using *Tetrahymena thermophila* as the outgroup sequence. Similar tree topology was obtained using Knuc values.

A

	P.m.	C.n.	H.p.	A.c.	G.sp.	W.c.	Al.t.	Al.c.	G.p.	P.l.	N.s.	C.c.	O.m.	T.t.
P.m.	0	28	35	37	40	61	96	97	62	97	113	162	193	180
C.n.	.078	0	18	20	36	55	91	92	67	89	106	155	200	184
H.p.	.099	.050	0	4	33	48	87	86	71	85	102	154	198	181
A.c.	.105	.056	.010	0	34	50	87	86	71	87	105	155	198	180
G.sp.	.114	.102	.093	.096	0	66	91	90	71	92	106	154	200	188
W.c.	.183	.164	.141	.148	.199	0	101	97	85	99	118	153	201	191
Al.t.	.318	.296	.283	.282	.294	.342	0	18	111	112	141	175	207	197
Al.c.	.321	.299	.277	.277	.289	.322	.050	0	112	105	137	172	209	202
G.p.	.187	.204	.219	.219	.216	.270	.390	.392	0	113	127	162	202	195
P.l.	.314	.285	.269	.276	.297	.322	.389	.354	.386	0	138	170	207	198
N.s.	.384	.351	.334	.346	.349	.404	.518	.495	.443	.496	0	148	215	181
C.c.	.654	.607	.606	.609	.591	.588	.745	.715	.642	.686	.571	0	212	217
O.m.	.832	.892	.875	.880	.882	.900	.963	.983	.913	.947	>1	>1	0	215
T.t.	.742	.778	.756	.752	.793	.838	.874	.920	.858	.876	.767	>1	>1	0

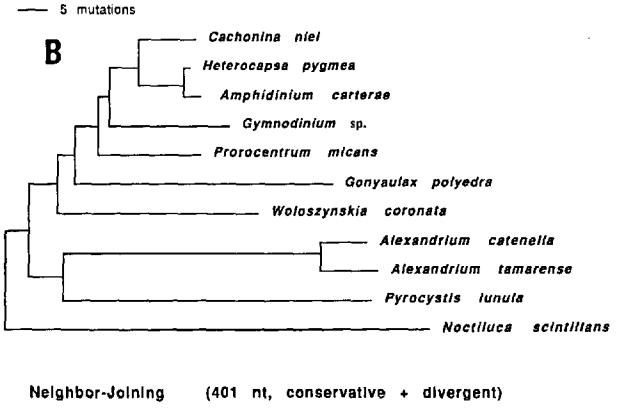


Fig. 6. Dinoflagellate partial phylogeny based on both conservative and more rapidly evolving sequences from the D1 and D8 regions. **A** Evolutionary distance matrix deduced from the 401-nt alignment of Fig. 3 including both conservative and divergent regions of D1 and D8. Highly divergent distal part of stem D1a and ambiguous positions denoted in Fig. 3 were not considered. Nucleotide distances between each pair of organisms are above the diagonal; Knuc values are below. The eleven closely related species have been used for subsequent phylogenetic analysis. The other species, which display Knuc values higher than 0.5 (boxed), have been omitted. Abbreviations as in Table 1. **B** Phylogenetic tree deduced from sequences of the 11 more closely related species, inferred with the neighbor-joining method. The tree was rooted with *Noctiluca scintillans*, according to the overall tree presented in Fig. 5B. Similar tree topology was obtained using Knuc values.

broad dinoflagellate phylogeny rooted with *T. thermophila, O. marina* emerged first, followed by *C. cohnii* and *Noctiluca scintillans,* all three being nonphotosynthetic species. The main branch then divided into two photosynthetic clusters, one including mainly Peridiniales, the second, more taxonomically heterogeneous, mainly comprising members of the orders Gymnodiniales and Prorocentrales as well as two species assigned to as Peridiniales incertae sedis (Sournia 1986).

To confirm the tree topology and achieve better resolution of the branching pattern among closely related species, an extended sequence alignment (414 nt) was considered, which included rapidly evolving sequences from D1 and D8 in addition to the more conservative stretches used initially. This analysis utilized the complete D1 and D8 sequence alignment shown in Fig. 3, with the exception of ambiguous nucleotides and a 20-nt stretch within D1, which was ignored because mutations and alignment gaps reached saturating levels (boxed area in Fig. 3 and Fig. 4A). Nucleotide distances and Knuc values calculated for each pair of organisms (Fig. 6A) confirm that for the most divergent species (*O. marina, C. cohnii,* and the outgroup *T. thermophila*) nucleotide changes reached saturating levels. For these species, calculated Knuc values were close to or higher than 0.6, indicative of random sequences not suitable for phylogenetic interpretations (Eckenrode et al. 1985).

When the analysis was limited to the 11 more closely related dinoflagellate species (Knuc values lower than 0.55) as depicted in the matrix of Fig. 6A, the topology of the tree inferred using the neighbor-joining method (Fig. 6B) or the Felsenstein's PHYLIP package, appeared identical to that based on conservative sequence comparisons (cf. Fig. 5B) and was rooted accordingly. In addition to the confirmation of the tree topology, the use of more rapidly evolving domains has significantly improved the resolution of most of the branchings, that is, the distance between nodes.

The position of the hidden break in D2 (cf. Fig. 3) also provided indirect confirmation of the tree topology. Its nonrandom location for the various

dinoflagellates appeared to be group specific. In *O. marina* which first emerged, the break is located at nucleotide coordinate 694. On the other hand, later emerging Peridiniales and *N. scintillans* are characterized by a D2 break at position 684. However, two members of the Peridiniales (*C. cohnii* and *G. polyedra*) contain an intact 24S rRNA not interrupted by any break. Finally, all members of the polyphyletic cluster comprising Gymnodiniales, Prorocentrales, and Peridiniales incertae sedis display a hidden break at position 689, consistent with their grouping near the top of the phylogenetic tree.

Phylogenetic and Taxonomic Implications

Special comments are needed with respect to the branching of *Woloszynskia coronata* and the two species assigned to as Peridiniales incertae sedis. The gymnodiniale *W. coronata,* the only freshwater species investigated here, shows several morphological features (thin hexagonal thecal plates, for instance) rather characteristic of the Peridiniales. This led Sournia (1986) to suggest that the family Woloszinskiaceae is intermediate between the Peridiniales and Gymnodiniales and thus could be assigned to either of the two orders. This statement is supported by the molecular phylogeny described here where the emergence of *W. coronata* is intermediate between the Peridiniales and Gymnodiniales. However, the position of the D2 hidden break is consistent with that of the Peridiniales. Reexamination of the taxonomic position of this family may therefore be justified.

The second comment concerns the branching of *Heterocapsa pygmea* and *Cachonina niei* among the Gymnodiniales. Again, the taxonomic position of these two species remains controversial (Morrill and Loeblich 1981; Sournia 1984). According to Sournia (1986), they cannot be assigned to any of the 13 families of the order Peridiniales and meanwhile should be noted as incertae sedis. From the phylogenetic trees (Figs. 5B and 6B), these two species appear more closely related to Gymnodiniales than to Peridiniales. Again, molecular data may help refine the taxonomy of these species.

Finally, important phylogenetic implications arise from the early emergence of *O. marina,* a species, which on account of several ultrastructural features, had already been suggested as perhaps representative of the ancestral dinoflagellate (Loeblich 1984). The primitive status of *O. marina* (which displays a deformable cell covering without apparent tabulation), the later emergence of most Peridiniales (covered with multiple thecal plates), and the recent appearance of the Prorocentrales (two single plates), Gymnodiniales (mostly unarmored), and Peridiniales incertae sedis (covered with many plates) suggest that the number of plates may not represent a valuable marker for dinoflagellate evolution. Former models mainly based on either increase, decrease, or fragmentation of the thecal plates (reviewed in Bujack and Williams 1981; and Goodman 1987) should be reexamined in light of these recent molecular sequence comparisons.

Conclusion

As demonstrated here for the highly diversified and taxonomically controversial class of dinoflagellates, two divergent domains (D1 and D8) of the LsuRNA can be used in combination to infer reliable phylogenetic relationships among closely related species. With the direct RNA sequencing method extending selected primers, rapidly evolving sequences can be easily determined along with the more conservative flanking stretches. The latter is most useful in delineating overall phylogenetic patterns. Subsequent combination of conservative and divergent regions allows precise evaluation of the distance between closely related taxa.

This first attempt to analyze dinoflagellate phylogeny using rRNA sequence data yielded a coherent and convincing evolutionary pattern. Good evidence now exists that selected domains of the LsuRNA molecule provide quantitative taxonomic markers, which we hope will be used to describe precise relationships among dinoflagellates and other diversified divisions, in combination with reliable morphological, biological, and paleontological character states.

Acknowledgments. We are grateful to Dr. M.-O. Soyer-Gobillard for her constant support, and to Dr. D.M. Anderson for support and critical discussions. This work was supported by the French Centre National de la Recherche Scientifique (ATP evolution) and by the US National Science Foundation (OCE-8911226), which provided for C.S.. G.L. had a scholarship from the French Ministère de l'Education et de la Recherche Scientifique. This is contribution number 7520 from the Woods Hole Oceanographic Institution.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Siedman JG, Smith JA, Struhl K (1987) Current protocols in molecular biology, vol. 1. Wiley Interscience, New York

Baroin A, Perasso R, Qu LH, Brugerolle G, Bachellerie JP, Adoutte A (1988) Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc Natl Acad Sci USA 85:3474–3478

Bischoff H, Bold H (1963) Phycological studies IV. Some soil algae from Enchanted Rock and related algal species. University of Texas Publications, no. 6318

Bujak J, Williams G (1981) The evolution of dinoflagellates. Can J Bot 59:2077–2087

Campbell DA, Kubo K, Clark CG, Boothroyd JC (1987) Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J Mol Biol 196:113–124

Cedergren R, Gray MW, Abel Y, Sankoff D (1988) The evolutionary relationships among known life forms. J Mol Evol 28:98–112

de Lanversin G, Jacq B (1983) Séquence de la région de coupure centrale du précurseur de l'ARN ribosomique 26S de drosophile. CR Acad Sci Ser III 296:1041–1044

de Lanversin G, Jacq B (1989) Sequence and secondary structure of the central domain of *Drosophila* 26S rRNA: a universal model for the central domain of the large rRNA containing the region in which the central break may happen. J Mol Evol 28:403–417

Dodge JD (1984) Dinoflagellate taxonomy. In: Spector DL (ed) Dinoflagellates. Academic Press, New York, pp. 17–42

Dörhöfer G, Davies E (1980) Evolution of archeopyle and tabulation in rhaetogonyaulacinean dinoflagellate cysts. R Ont Mus Life Sci Misc Publ, pp 1–91

Eaton G (1980) Nomenclature and homology in peridinialean dinoflagellate patterns. Palaeontology 23:667–688

Eckenrode VK, Arnold J, Meagher RB (1985) Comparison of the nucleotide sequence of soybean 18S rRNA with the sequences of other small-subunit rRNAs. J. Mol Evol 21:259–269

Engberg J, Nielsen H, Lenaers G, Murayama O, Fujitani H, Higashinakagawa T (1990) Comparison of primary and secondary 26S ribosomal RNA structures in two *Tetrahymena* species: evidence for a strong evolutionary and structural constraint in expansion segments. J Mol Evol 30:514–521

Fitch WM (1981) A non-sequential method for constructing trees and hierarchical classifications. J Mol Evol 18:30–37

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Gill LL, Hardman N, Chappell L, Qu LH, Nicoloso M, Bachellerie JP (1988) Phylogeny on *Onchocerca volvulus* and related species deduced from rRNA sequence comparison. Mol Biochem Parasitol 28:69–76

Goodman DK (1987) Dinoflagellate cysts in ancient and modern sediments. In: Taylor JFR (ed) The biology of dinoflagellates. Bot Monog 21:649–722

Gouy M, Li WH (1989a) Phylogenetic analysis based on ribosomal RNA sequences supports the archaebacterial tree rather than the eocyte tree. Nature 339:145–149

Gouy M, Li WH (1989b) Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. Mol Biol Evol 6:109–122

Guillard R, Ryther J (1963) Studies on marine planktonic diatoms. I. *Cyclotella nana* Husted and *Detonula confervacea* Cleve. Gran. Can J Microbiol 8:229–239

Gutell RR, Fox EF (1988) A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res 16(suppl):r175–r269

Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12:3563–3583

Herzog M, Soyer MO (1981) Distinctive features of dinoflagellate chromatin. Absence of nucleosomes in a primitive species *Prorocentrum micans*. Eur J Cell Biol 23:295–302

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions. J Mol Evol 16:111–120

Lane D, Pace B, Olsen G, Sthal D, Sogin M, Pace N (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82:6955–6959

Leffers H, Kjems J, Ostergaard L, Larsen N, Garret R (1987) Evolutionary relationships amongst archaebacteria. A comparative study of 23S ribosomal RNAs of a sulphur-dependant extreme thermophile, an extreme halophile and a thermophilic methanogen. J Mol Biol 195:43–61

Lenaers G, Nielsen H, Engberg J, Herzog M (1988) The secondary structure of large subunit rRNA divergent domains, a marker for protist evolution. BioSystems 21:215–222

Lanaers G, Maroteaux L, Michot B, Herzog M (1989) Dinoflagellates in evolution. A molecular phylogenetic analysis of large-subunit ribosomal RNA. J Mol Evol 29(1):40–51

Loeblich AR (1976) Dinoflagellate evolution: speculation and evidence. J Protozool 23:13–28

Loeblich AR (1984) Dinoflagellate evolution. In: Spector DL (ed) Dinoflagellates. Academic, New York, pp 481–522

Maniatis T, Fritsch E, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p 545

Michot B, Bachellerie JP (1987) Comparisons of large subunit rRNAs reveal some eukaryote specific elements of secondary structure. Biochemie 69:11–23

Michot B, Hassouna N, Bachellerie JP (1984) Secondary structure of mouse 28S rRNA and general models for the folding of the large RNA in eukaryotes. Nucleic Acids Res 12:4259–4279

Morrill LC, Loeblich AR (1981) A survey of body scales in dinoflagellates and a revision of *Cachonina* and *Heterocapsa*. J Phycol 16(suppl):29

Mouches C, Pauplin Y, Agarwal M, Lemieux L, Herzog M, Abadon M, Beyssat-Arnaouty V, Hyrien O, de Saint Vincent BR, Georghiou GP, Pasteur N (1990) Characterization of amplification core and esterase B1 gene responsible for insecticide resistance in *Culex*. Proc Natl Acad Sci USA 87:2574–2578

Perasso R, Baroin A, Qu LH, Bachellerie JP, Adoutte A (1989) Origin of the algae. Nature 339:142–144

Provasoli L (1963) Growing marine seaweeds. Proc Int Seaweed Symp 4:9–17

Qu LH, Michot B, Bachellerie JP (1983) Improved methods for structure probing in large RNAs: a rapid heterologous sequencing approach is coupled to the direct mapping of nuclease accessible sites. Implication to the 5′ terminal domains of eukaryotic 28S rRNA. Nucleic Acids Res 17:5903–5920

Qu LH, Hardman N, Gill LL, Chappell L, Nicoloso M, Bachellerie JP (1986) Phylogeny of helminths determined by rRNA sequence comparison. Mol Biochem Parasitol 20:93–99

Qu LH, Nicoloso M, Bachellerie JP (1988) Phylogenetic calibration of the 5′ terminal domain of large rRNA achieved by determining twenty eucaryotic sequences. J Mol Evol 28:113–124

Rizzo PJ (1987) Biochemistry of the dinoflagellate nucleus. In: Taylor FJR (ed) The biology of dinoflagellates. Bot Monogr 21:143–173

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sournia A (1984) Classification et nomenclature de divers dinoflagellés marins (classe des Dinophyceae). Phycologia 23: 345–355

Sournia A (1986) Dinophyceae. In: Sournia A (ed) Atlas du phytoplancton marin, vol 1. Edition du CNRS, Paris, 219 p

Spencer DF, Collings JC, Schnare MN, Gray MW (1987) Multiple spacer sequence in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*. EMBO J 6:1063–1071

Starr R (1964) The culture collection of algae at Indiana University. Am J Bot 51:1013–1044

Studier JA, Keppler KJ (1988) A note on the neighbor-joining algorithm of Saitou and Nei. Mol Biol Evol 6:729–731

Taylor F (1980) On dinoflagellate evolution. Biosystems 13: 65–108

Taylor F (1987) Taxonomy and classification. In: Taylor F (ed) The biology of dinoflagellates. Bot Monogr 21:723–731

Tuttle R, Loeblich A (1975) An optimal growth medium for the dinoflagellate *Cryptecodinium cohnii*. Phycologia 14(1): 1–8

Vernet G, Sala-Rovira M, Maeder M, Jacques F, Herzog M (1990) Electrophoretic and DNA-binding properties of the major basic nuclear proteins from the histone-less eukaryote *Crypthecodinium cohnii* (Pyrrhophyta). Biochim Biophys Acta 1048:281–289

Received February 26, 1990/Revised June 22, 1990