Gene, 71 (1988) 491–499
Elsevier

GEN 02716

The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions

(Sequence analysis; polymerase chain reaction; diatoms; *Skeletonema costatum*)

Linda Medlin, Hille J. Elwood, Shawn Stickel and Mitchell L. Sogin

National Jewish Center for Immunology and Respiratory Medicine, Denver, CO 80206 (U.S.A.)

Received 3 June 1988
Revised and accepted 3 August 1988
Received by publisher 19 August 1988

SUMMARY

Polymerase chain reaction conditions were established for the in vitro amplification of eukaryotic small subunit ribosomal (16S-like) rRNA genes. Coding regions from algae, fungi, and protozoa were amplified from nanogram quantities of genomic DNA or recombinant plasmids containing rDNA genes. Oligodeoxynucleotides that are complementary to conserved regions at the 5′ and 3′ termini of eukaryotic 16S-like rRNAs were used to prime DNA synthesis in repetitive cycles of denaturation, reannealing, and DNA synthesis. The fidelity of synthesis for the amplification products was evaluated by comparisons with sequences of previously reported rRNA genes or with primer extension analyses of rRNAs. Fewer than one error per 2000 positions were observed in the amplified rRNA coding region sequences. The primary structure of the 16S-like rRNA from the marine diatom, *Skeletonema costatum*, was inferred from the sequence of its in vitro amplified coding region.

INTRODUCTION

Sequence comparisons of rRNAs or their genes have revolutionized our perspectives on molecular and cellular evolution (Gutell et al., 1986; Sogin et al., 1986a; Gunderson et al., 1987a; Woese, 1988). This, in turn, has led to the development of new rRNA-based technologies capable of detecting and identifying small numbers of microorganisms in clinical isolates (Gobel et al., 1987) or in complex natural populations (Olsen et al., 1986). Considerable attention has been focused upon the structural analysis of 16S-like rRNAs which contain sufficient evolutionary information to allow the measurement of both close and distant phylogenetic relationships (Sogin and Gunderson, 1987). Rapid sequencing strategies for 16S-like rRNA coding regions are based upon the interspersion of conserved and partially or non-conserved sequence regions (Elwood et al., 1985). Synthetic oligos complementary to conserved regions can be used to initiate DNA synthesis

Correspondence to: Dr. M.L. Sogin, National Jewish Center for Immunology and Respiratory Medicine, Dept. of Molecular and Cellular Biology, 1400 Jackson Street, Denver, CO 80206 (U.S.A.) Tel. (303)398-1636.

Abbreviations: Ap, ampicillin; bp, base pair(s); kb, kilobase(s) or 1000 bp; nt, nucleotide(s); *O*, *Ochromonas*; oligo, oligodeoxynucleotide; PCR, polymerase chain reaction; PCR buffer, see MATERIALS AND METHODS, section c; rRNA, ribosomal RNA; 16S-like rRNA, small subunit ribosomal RNA; RF, replicative form; *S*, *Skeletonema*; SDS, sodium dodecyl sulfate.

0378-1119/88/$03.50 © 1988 Elsevier Science Publishers B.V. (Biomedical Division)

in dideoxynucleotide chain-termination sequencing protocols (Sanger and Coulson, 1975), thus eliminating the requirement for constructing a nested set of overlapping fragments in the M13 cloning and sequencing system. For studies of rRNA primary structure, the time-limiting factor is that required to construct genomic libraries and identify recombinant clones containing rRNA coding regions.

In the case of the marine diatom, *Skeletonema costatum*, our initial efforts to clone rDNA genes were unsuccessful because of difficulties in obtaining sufficient quantities of high molecular weight nuclear DNA to construct a genomic library. *S. costatum* is an important component of coastal phytoplankton communities. In some areas it comprises 90% of the biomass and it is a valuable mariculture organism for invertebrate feeding. Because of its importance to marine life sciences, *S. costatum* has been the subject of a wide range of physiological studies but its phylogenetic relation to other eukaryotic algae has not been established.

The cloning steps for characterization of 16S-like rRNAs from species such as *S. costatum* can be circumvented by using bulk RNA as a template for reverse transcriptase-mediated sequencing reactions (Qu et al., 1983; Lane et al., 1985). Direct sequence analysis of 16S-like rRNAs is extremely rapid and the characterization of functional RNAs (as opposed to non-expressed or pseudo rRNA genes) is assured. Unfortunately, the rapidity of the direct rRNA analysis is compromised by the loss of information at approximately 1–2% of the sequence positions. Ambiguities at these positions result from band compressions or universal terminations caused by modified bases or tight secondary structures in the rRNA template. Furthermore, only a single strand is available for analysis which eliminates the redundant accuracy check inherent when two strands of a DNA sequence are determined.

The speed of rRNA primer extension analyses and the accuracy afforded by double-strand sequencing determinations might be coupled if PCR methods (Mullis and Faloona, 1987; Saiki et al., 1988) were used to amplify rRNA genes for sequence analyses. We have explored the use of PCR protocols to amplify 16S-like rRNA coding regions from previously characterized cloned rRNA genes and from ng quantities of genomic DNA isolated from the marine diatom, *S. costatum* and the ascomycete, *Kluveromyces lactis*. Oligos complementary to conserved 5′ and 3′ proximal regions in eukaryotic 16S-like rRNA genes were used to initiate DNA synthesis in the amplification step. The resulting product could be sequenced directly or subcloned into the single-stranded DNA phage M13. As judged by identity to previously characterized rRNA coding regions or to partial rRNA sequences determined by the direct rRNA methods, the amplified coding regions appear to represent faithful transcripts. The sequence of the 16S-like rRNA amplified coding region from *S. costatum* was used to establish its relationship to other eukaryotes.

MATERIALS AND METHODS

(a) Algal cultures and isolation of RNA/DNA

Clones of *S. costatum* (Pacifica, isolated from the Straits of Georgia, British Columbia, by Rosemary Walters, University of British Columbia Culture No. 18/C) were grown in artificial sea water with 50 μg Ap/liter at 24°C under continuous light and stirred manually on a daily basis. Cells were harvested by filtration, washed in extraction buffer (100 mM Tris, pH 8.5, 100 mM NaCl, 50 mM EDTA), and disrupted by vortex treatment in the presence of 2% SDS and glass beads. Bulk nucleic acids were extracted with phenol–chloroform–isoamyl alcohol (50 : 48 : 2) and concentrated by ethanol precipitation. The DNA was purified by CsCl equilibrium gradient centrifugation (Maniatis et al., 1982). Banding of DNA preparations in CsCl was necessary for consistent results in the amplification procedures.

(b) Primers for amplifying 16S-like rRNA coding regions

Comparison of 45 eukaryotic 16S-like rRNA sequences reveals that there are conserved sequence elements proximal to the 5′ and 3′ termini which can be used as start points in PCR experiments. Sequences between the conserved elements can be exponentially amplified by repetitive cycles of denaturing duplex DNA, annealing primers complementary to the conserved sequence elements, and then primer extension using DNA polymerase. The

products of the primer extension as well as the original duplex DNA can serve as templates in successive amplification cycles. Amplification primers for the 16S-like rRNA genes were designed with polylinkers at their 5′ termini to facilitate the cloning of the PCR amplification products into the single-stranded phage, M13. The synthetic oligos for amplifying eukaryotic 16S-like rRNA coding regions are presented in Fig. 1. The sequence of primer 'A' is complementary to 21 nt in the coding strand at the 5′ terminus of eukaryotic 16S-like rRNAs and contains restriction sites for *Eco*RI and *Sal*I. Primer 'B' is complementary to 24 nt in the non-coding stand at the 3′ terminus of eukaryotic 16S-like rRNAs and includes restriction sites for *Sma*I, *Bam*HI, *Hin*dIII, and *Pst*I.

(c) DNA amplification

A modification of the PCR methods previously reported (Mullis and Faloona, 1987; Saiki et al., 1988) was used to amplify the 16S-like rRNA coding regions from genomic DNA or from recombinant plasmids known to harbor rDNA coding regions. Incubations were carried out in 1.5-ml eppendorf tubes. Reaction mixtures containing 10 ng of DNA in 100 μl of 1 × PCR buffer (10 mM Tris · HCl, pH 8.4, 50 mM KCl, 2 mM $MgCl_2$, 200 μM in dATP, dTTP, dCTP, and dGTP, and 0.1 mg/ml gelatin) were made 1 μM in each of the amplification primers. The samples were layered with 200 μl of mineral oil to minimize evaporation. After heating the reactions at 95°C for 5 min to denature the DNA, 0.1 units of the heat-stable *Taq* DNA polymerase (New England Biolabs) were added. The samples were transferred between three incubations: (1) primer annealing to the DNA template at room temperature for 4.25 min; (2) primer extension at 65–68°C for 7.5 min; and (3) denaturation at 90–92% for 2.5 min. The sequential incubations were repeated between 40 and 90 cycles using a Shandon Elliott Duplex Processor retrofitted with a 15-s interval/1 h timer to move the samples between temperature baths.

(d) M13 cloning and sequence analysis

The PCR products were extracted with phenol and concentrated by precipitation with ethanol (Maniatis et al., 1983). The samples were suspended in 20 μl of REact 3 buffer (Bethesda Research Laboratories) and digested with appropriate restriction enzymes. The digested samples were re-extracted with phenol, concentrated by ethanol precipitation, and then ligated into the RF of M13mp18 or M13mp19 (Messing, 1983). Single-stranded templates for directing DNA synthesis in sequencing protocols were prepared from recombinant M13mp18 and M13mp19 phages containing amplified 16S-like rRNA coding regions. Synthetic

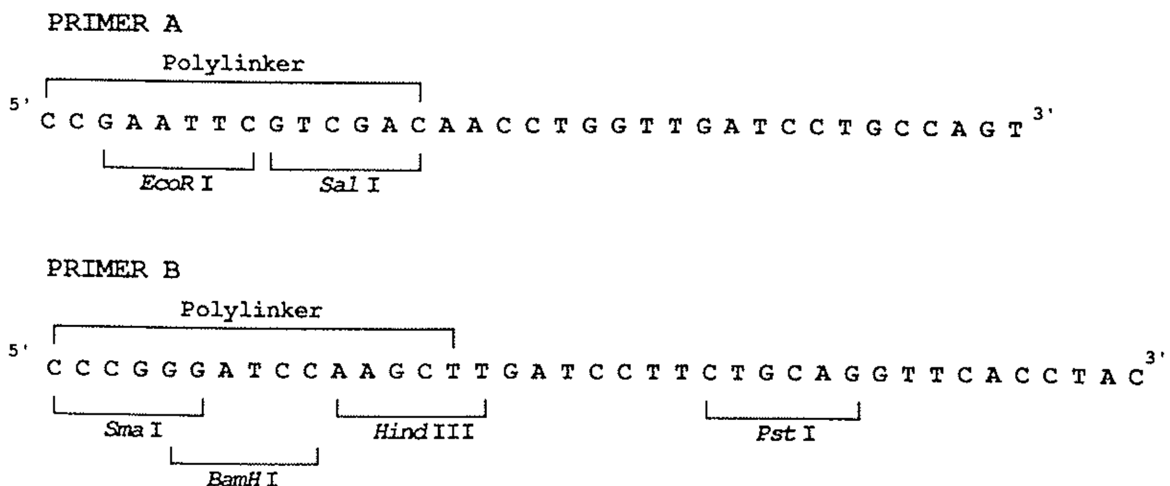

Fig. 1. Synthetic oligos used for in vitro amplification of eukaryotic 16S-like rRNA coding regions. Primer 'A' contains a polylinker plus a conserved 21-nt sequence which is complementary to the coding strand at the 5′ terminus of eukaryotic 16S-like rRNAs. Primer 'B' contains a polylinker plus a conserved sequence that is complementary to 24 nt in the noncoding strand at the 3′ terminus of eukaryotic 16S-like rRNAs.

oligo primers that are well conserved in eukaryotic 16S-like rRNA genes (Elwood et al., 1985) were used to initiate DNA synthesis in the dideoxynucleotide chain-termination sequencing protocols (Sanger and Coulson, 1975). Reverse transcriptase-mediated primer extension sequencing protocols were used to directly characterize the *S. costatum* 16S-like rRNA (Lane et al., 1985). Primers complementary to conserved regions at *Dictyostelium discoideum* nt positions 377–393, 892–906, and 1504–1519 (Sogin and Gunderson, 1987) were used to initiate the reverse transcription reactions.

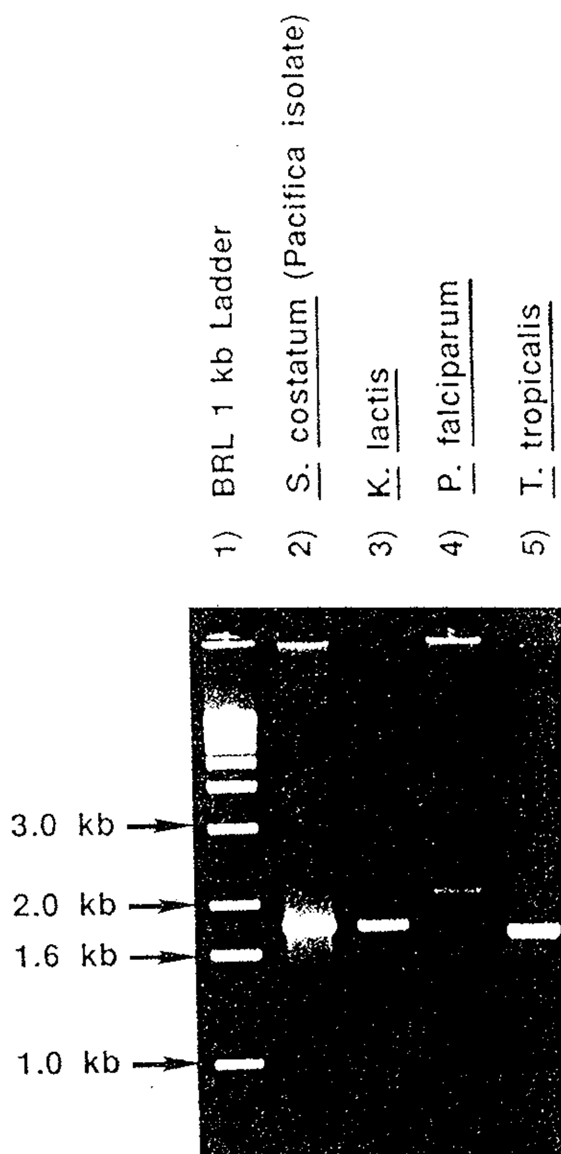

Fig. 2. Agarose gel electrophoresis of amplified small subunit rRNA coding regions. Polymerase chain reactions directed by genomic DNA or cloned rRNA templates were primed with oligos complementary to conserved sequences in eukaryotic 16S-like rRNAs as described in MATERIALS AND METHODS, part c. Aliquots (1–5 μl) from each of the reactions were mixed with 10 μl LT buffer (10 mM Tris, pH 7.2, 10 mM NaCl, and 0.5 mM EDTA, pH 7.2) plus 5 μl stop solution (25% Ficol, 0.05% bromphenol blue, 40 mM Tris, pH 8.3, and 20 mM EDTA, pH 8.0). The samples were electrophoresed for 2 h at 7 V/cm in 2% agarose gels built in E buffer (40 mM Tris/CH_3COOH, pH 7.0, 1 mM EDTA). The bands were located by staining with 2% ethidium bromide. Lane 1 is a 1-kb BRL (Bethesda Research Laboratories) ladder. Products from 90 amplification cycles using 10 ng of genomic DNA from *S. costatum* are shown in lane 2, and 40 amplification cycles using 10 ng of genomic DNA from *Kluveromyces lactis* are shown in lane 3. Lanes 4 and 5 show the product of 40 amplification cycles using 10 ng of cloned rDNA genes from *Plasmodium falciparum* (McCutchan et al., 1988) and *Tetrahymena tropicalis* (Sogin et al., 1986), respectively.

RESULTS AND DISCUSSION

(a) Polymerase chain reaction of 16S-like rRNA coding regions

DNA samples from in vitro amplification of 16S-like rRNA coding regions were electrophoretically fractionated in 2% agarose gels. The mobilities of DNA fragments that were amplified using 10 ng of genomic DNA from *S. costatum* (lane 2) and *K. lactis* (lane 3) are shown in Fig. 2. The prominent reaction products are represented by single bands with chain lengths of 1.8 kb. Direct sequence analysis (see RESULTS AND DISCUSSION, section **b**) demonstrated that both are 1798 nt in length. Lanes 4 and 5 of Fig. 2 display PCR products whose synthesis was directed by recombinant DNA templates containing rDNA genes from *Tetrahymena tropicalis* (Sogin et al., 1986b) or *Plasmodium falciparum* (McCutchan et al., 1988). We have previously reported chain lengths of 1753 and 2091 nt for the 16S-like rRNAs of *T. tropicalis* and *P. falciparum*, which is consistent with the electrophoretic mobilities of the PCR products. The yields from the in vitro amplification reactions varied with cloned rRNA genes from different organisms. For example, despite nearly perfect complementarity between the *P. falciparum* coding region and the amplification primers, the PCR yields were substantially less than

with any other tested clone. The reason for this variability was not systematically examined. In most cases, with as little as 1 ng of genomic DNA, the yield for 40 PCR cycles was in the order of 5 μg of total nucleotides. No attempts were made to amplify rDNA genes from less than 1 ng of genomic DNA. Increasing the number of cycles beyond 40 does not significantly improve the yield and can affect the quality of the product. The *S. costatum* PCR products appear more heterogeneous (as judged by higher backgrounds) than the *K. lactis* PCR products. The *S. costatum* reaction cycle was repeated 90 times which compares to only 40 cycles for *K. lactis* and *P. falciparum*. A lower number of cycles significantly reduces the background level in the *S. costatum* amplification reactions (data not shown).

PCR amplification for rDNA structure analyses compares favorably with the primer extension methods employed in direct analyses of rRNAs. Microgram quantities of 16S-like rRNA coding regions can be produced from as little as 1 ng of genomic DNA even in the presence of high concentrations of mitochondrial and/or chloroplast DNA. Complementarity of the primers to regions that are conserved in all eukaryotic 16S-like coding regions assures their specific amplification. (Sequences that are conserved in prokaryotic rRNA genes may be useful for specifically amplifying mitochondrial or chloroplast 16S-like rRNA coding regions against a background of nuclear DNA). With the PCR methods, the number of cells required for characterizing rRNA coding regions is remarkably low. For example, reverse transcriptase-mediated primer extension analyses require approximately 3–4 μg of RNA for each sequencing reaction. This corresponds to a minimum of 10^6 cells for a typical eukaryotic microorganism such as *D. discoideum* (Ashworth and Watts, 1970). The PCR amplification of a 16S-like rRNA coding region from a typical eukaryotic microorganism with a genome complexity of 1×10^8 bp may require fewer than 100 cells to produce 0.5 μg of product. Thus the analysis of rDNA genes from organisms that are difficult or impossible to propagate in the laboratory, including obligate parasites and symbionts, is now possible.

In addition to facilitating structural studies of rRNAs, the potential impact of this technology upon the analysis of natural microbial populations is considerable. Measurements of molecular diversity, e.g., using rRNA or its coding region, can be used to estimate the number of different species in a given population (Olsen et al., 1986). Unfortunately, the reverse transcriptase-mediated primer extension technique is not directly applicable to sequencing mixed populations of RNAs (without first fractionating the RNAs, which is difficult, at best, for 16S-like rRNAs). Similarly, molecular diversity studies based upon isolation of rRNA genes are subject to potential bias from variant cloning efficiencies. (In our survey of protists we have found that some rDNA genes are much easier to clone than others. Estimates of numbers of rDNA genes for a given species requires re-probing of the original, unfractionated DNA population with probes/clones specific to individual population members.) These problems are readily solved by the amplification methods. The rDNA genes from different organisms can be simultaneously amplified from complex DNA populations, and the products, which vary in size (as demonstrated by distinct electrophoretic mobilities for amplification products from *P. falciparum*, *T. tropicalis*, and *S. costatum*) can be resolved by gel electrophoresis. The number of products displaying variant electrophoretic mobilities can be taken as an estimate of the organismal diversity in the natural population.

(b) Sequence analysis of the polymerase chain reaction products

The capability of rapidly analyzing rRNA genes in complex DNA populations must be weighed against the potential for artifacts introduced during PCR amplification. The fidelity of synthesis was evaluated by comparing the primary structures of the amplification products with previously reported 16S-like rRNA gene sequences or with partial rRNA sequences determined by reverse transcriptase mediated primer extension methods. For each of the amplification products, the coding and non-coding strands of the rRNA genes were characterized. The PCR products for *T. tropicalis* and *S. costatum* were digested with *Bam*HI + *Eco*RI and *Bam*HI + *Sal*I, respectively. The *P. falciparum* amplification product was digested with *Xba*I + *Hin*dIII (restriction sites for these enzymes are separated by 384 nt in the *P. falciparum* rRNA gene). Polylinker sites were used to clone the amplified DNA segments into double-

stranded replicative forms of the M13mp18 and M13mp19 DNA phage. Single-stranded templates were prepared from representative M13 recombinants (one M13 template for the coding and another containing the non-coding strand) and sequenced by the dideoxynucleotide methods (see MATERIALS AND METHODS, section **d**). Sequences were determined for the entire *T. tropicalis* amplification product and for a 384-nt region in the amplified 16S-like rRNA coding region of *P. falciparum*. No differences were detected when the sequence of the *T. tropicalis* amplification product was compared to that of the previously reported coding region (Sogin et al., 1986b). The M13 clone containing the coding strand of *P. falciparum* gene was identical to the published sequence (McCutchan et al., 1988); however, the M13 clone containing the non-coding strand differed from the published sequence at a single position. The disparity in the non-coding sequence does not represent a sequencing error (six well-resolved, overlapping sequence determinations in this region of the non-coding strand were identical with each other). The difference between the published sequence and our representative M13 clone containing the non-coding strand must reflect an error which occurred during the amplification procedure.

The sequence (determined on both the coding and non-coding strand) of the PCR amplification product for *S. costatum* is presented in Fig. 3. Three sites (nt positions 127, 677, and 800) display sequence variation between the coding and non-coding strands. In all three cases the ambiguities are between cytosine and thymidine. Redundant analyses in these regions suggest that the disagreements are not because of sequencing errors. The discrepancies could reflect errors in the amplification process or micro-heterogeneity in the rRNA genes of *S. costatum*. The nt positions 127 and 677 reside in regions which may tolerate variation within the same species because these domains are poorly conserved (because of minimal functional constraints) in eukaryotic 16S-like rRNAs. In contrast, nt position 800 is located in a region that is conserved in most eukaryotic 16S-like rRNAs. Except for position 800 which was determined to be a cytosine in the RNA analysis, differences could not be demonstrated between the sequence of the *S. costatum* PCR product and a partial sequence of its rRNA (approx. 750 nt) determined by reverse transcriptase-mediated primer extension methods. The nt positions 127 and 677, which differed in comparisons of sequences for the coding and noncoding strands, also appeared as ambiguous positions in the direct rRNA sequence. We interpret this as representing micro-heterogeneity in the coding regions for the *S. costatum* 16S-like rRNAs rather than errors introduced during PCR

```
1    aaccugguug auccugccag uAGUCAUACG CUCGUCUCAA AGAUUAAGCC AUGCAUGUGU AAGUAUAAGA UACUUUAUCG AAACUGCGGA CGGCUCAUUA
101  UAUCAGUUAU AGUUUAUUUG AUAAUCYCUC ACUACUUGGA UAACCGUAGU AAUUCUAGAG CUAAUAUAUG CAUCAAAGCG GAACUCUCGG GAACCGCCGU
201  GUUUAUUAGU AUUAAACCUU CACUCUUCGG AGUUGAUUUG GUGAGUCAUA AUAACCUUUC GAAUUGCAUG CACAUGUUGC GGCAAUGGAU CAUUCAAGUU
301  UCUGCCCUAU CAGCUUGGGA UGGUAGUGUA UUGGACUACC AUGGCUUUAA CGGGUAACGA AUUGUUAGGG CAAGAUUCCG GAGAGGGAGC CUGAGAGACG
401  GCUACCACAU CCAAGGAAGG CAGCAGGCGC GUAAAUUACC CAAUCCUGAC ACAGGGAGGU AGUGACAAUA AAUAACAAUG CCGGGCCUUU ACAGGUCUGG
501  CAAUUGGAAU GAGAACAAUU UAAAUCCCUU AUCGAGUAUC AAUUGGAGGG CAAGUCUGGU GCCAGCAGCC GCGGUAAUUC CAGCUCCAAU AGCGUAUAUU
601  AAAGUUGUUG CAGUUAAAAA GCUCGUAGUU GGAUUUCUGG CAGGAGUGAC CGGCCACACA CACUGUGCGU GAGUUGYGUC AUUCUGGCCA UCCUUGGUGA
701  GAUCCUGUUU GGCAUUAAGU UGUCGGGCAG GGGAUAACCA UCGUUUACUG UGAAAAAAUU AGAGUGUUUA AAGCAGGCUU AUGCCGUUGA AUAUAUUAGY
801  AUGGAAUAAU AAGAUAGGAC UUUGAGUCUA UUUUGUUGGU UUGCGAGUCA AAGUAAUGAU UAAUAGGGAC AGUUGGGGGU AUUCGUAUUU CAUUGUCAGA
901  GGUGAAAUUC UUGGAUUUCU GAAAGACGAA CUACUGCGAA AGCAUUUACC AAGGAUGUUU UCAUUAAUCA AGAACGAAAG UUAGGGGAUC GAAGAUGAUU
1001 AGAUACCAUC GUAGUCUUAA CCAUAAACUA UGCCGACUCA GGAUUGGCGG UUGUUUUUUG ACUCCGUCAG UACUGUAUGA GAAAUCAAAG UCUUUGGGUU
1101 CCGGGGGGAG UAUGGUCGCA AGGCUGAAAC UUAAAGAAAU UGACGGAAGG GCACCACCAG GAGUGGAACC UGCGGCUUAA UUUGACUCAA CACGGGAAAA
1201 CUUACCAGGU CCAGACAUAG UGAGGAUUGA CAGAUUGAGA GUUCUUUCUU GAUUCUAUGG GUGGUGGUGC AUGGCCGUUC UUAGUUGGUG GAGUGAUUUG
1301 UCUGGUUAAU UCCGUUAACG AACGAGACCG CCGCCUGCUA AAUAGACCCG CGAAUAGCUU UUUAUUGGCG AGGUCUUCUU AGAGGGACGU UCAUUCUACA
1401 AGAUGAAGGA AGAUGGCGGC AAUAACAGGU CUGUGAUGCC CUUAGAUGUC CUGGGCCGCA CGCGCGUUAC ACUGAUGCAC UCAACGAGCA UAUAACCUUG
1501 GCCGAGAGGC CUGGGUAAUC UUGUUAACUU GCAUCGUGAU AGGGAUAGAU UAUUGCAAUU AUUAAUCUUG AACGAGGAAU UCCUAGUAAU CGCAGUUCAU
1601 CAAACUGCAA UGAUUACGUC CCUGCCCUUU GUACACACCG CCCGUCGCAC CUACCGAUUG AAUGGUCCGG UGAGGAGUCG AGAUUGUGGA UUAGCUCCUU
1701 UAUUGGGGCU UAUCCGCGAG AACCUCUCCA AACCUUAUCA UUUAGAGGAA GGUGAAGUCG UAACAAGGUU UCCguaggug aaccugcaga aggaucaa
```

Fig. 3. Sequence of a 16S-like rRNA region amplified from *S. costatum* genomic DNA. The lower case letters indicate positions that correspond to the amplification primers and therefore cannot be interpreted as being present in the *S. costatum* 16S-like rRNA coding region. Ambiguities at nt positions 127, 677, and 800 are indicated in the figure by the letter 'Y'.

amplification. Similar heterogeneity has been demonstrated in the rRNA coding regions of other organisms including *Escherichia coli* (Sogin et al., 1972), *P. falciparum* (McCutchan et al., 1988) and *Plasmodium berghei*, where structurally distinct 16S-like rRNAs appear to be differentially expressed during various stages of its life cycle (Gunderson et al., 1987b).

There are two strategies for guarding against potential errors in the amplification process. The first is to sequence the products of the amplification reaction directly using procedures for analyzing double-stranded DNA (Wrischnik et al., 1987). Positions which are incorrectly copied during early rounds of amplification will appear to be ambiguous in the DNA sequencing gels while errors in late rounds will not contribute to the sequence interpretation. The products can be sequenced directly if the amplification primers are dialyzed away or removed by chromatography prior to initiating DNA synthesis in the DNA sequencing reactions (data not shown). As an alternative, single-stranded templates can be prepared from mixtures of 30 or 40 recombinant M13 clones. This provides a sampling of variation in the amplification products. As above, errors which occur early in the amplification procedure will be reflected by ambiguous positions in the M13 sequence analysis. We have successfully used this strategy in the sequence analysis of *K. lactis* (M.L.S., unpublished data). Regardless of which method is employed, the error rate appears to be sufficiently low to not confuse interpretations of phylogenetic relationships based upon sequence comparisons of PCR amplification products.

(c) Phylogenetic status of *Skeletonema costatum*

The *S. costatum* 16S-like rRNA sequence was aligned with a collection of 13 other eukaryotic 16S-like rRNA sequences as previously described (Elwood et al., 1985). Pairwise comparisons of the sequences (only those position which could be

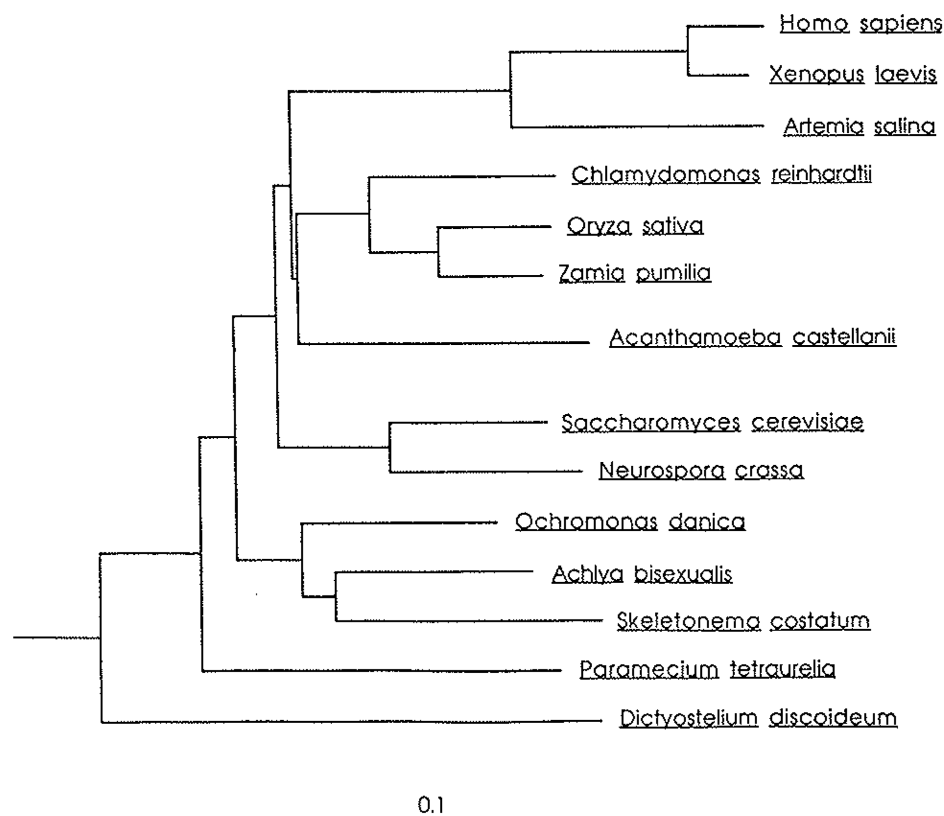

Fig. 4. Eukaryotic phylogeny inferred from 16S-like rRNA sequence similarities. A phylogenetic tree was inferred using structural distance data as previously described (Elwood et al., 1985). The evolutionary distances between nodes of the tree are represented by the horizontal component of their separation. The distance which corresponds to ten changes per 100 nt positions is indicated.

unambiguously aligned were incorporated into the analysis) were used to compute structural similarities and these were converted into the phylogenetic tree shown in Fig. 4 by the distance matrix method (Fitch and Margoliash, 1976). In this analysis, the diatom *S. costatum* is a member of an assemblage which includes the oomycete *Achlya bisexualis* and the chrysophyte, *Ochromonas danica*. The relationship between chrysophytes and oomycetes is consistent with other analyses based upon ultrastructure similarities of their flagellar apparatus (Gunderson et al., 1987a). The inclusion of diatoms in the group agrees with Cavalier-Smith's 1986 proposal for the kingdom Chromista which embraces brown algae and all protozoans having either tubular ciliary mastigonemes or chloroplast endoplasmic reticulum.

(d) Conclusions

High quality sequences on both DNA strands can be generated from amplified rDNA coding regions. The number of cells required for the analysis is much less than that necessary for primer extension analysis of rRNAs or for constructing genomic libraries. The specific amplification of nuclear rRNA genes is possible using complex DNA populations containing rRNA genes from mitochondria and/or chloroplast genomes. Using the PCR methods, the 16S-like rRNA coding region from *S. costatum* was amplified and found to be most similar to those of *A. bisexualis* and *O. danica*.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Health (GM32964) to M.L.S., and NERC Special Topics in Marine Genetics GST/02/223 to L.M.

REFERENCES

Ashworth, J.M. and Watts, D.J.: Metabolism of the cellular slime mould *Dictyostelium discoideum*. Biochem. J. 119 (1970) 175–182.

Cavalier-Smith, T.: The kingdom Chromista: origin and systematics. Prog. Phycol. Res. 4 (1986) 309–347.

Elwood, H.J., Olsen G.J. and Sogin M.L.: The small subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. J. Mol. Biol. Evol. 2 (1985) 399–410.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R. and Raff, R.: Molecular phylogeny of the animal kingdom. Science 239 (1988) 748–753.

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

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T., Wolfe, R.S., Balch, W., Tanner, R., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R.: The phylogeny of prokaryotes. Science 209 (1980) 457–463.

Gobel, U.B., Geiser, A. and Stanbridge, E.J.: Oligonucleotide probes complementary to variable regions of ribosomal RNA discriminate between *Mycoplasma* species. J. Gen. Microbiol. 133 (1987) 1969–1974.

Gunderson, J.H., Elwood, H.J., Ingold, A., Kindle, K. and Sogin, M.L.: Phylogenetic relationships between chlorophytes, chrysophytes and oomycetes. Proc. Natl. Acad. Sci. USA, 84 (1987a) 5823–5827.

Gunderson, J.H., Sogin, M.L., Wollett, G., Hollingdale, M., de la Cruz, V.F. and McCutchan, T.F.: Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. Science 238 (1987b) 933–937.

Gutel, R., Weiser, B., Woese, C.R. and Noller, H.F.: Comparative anatomy of 16S-like ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol. 32 (1986) 155–216.

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

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

McCutchan, T.F., de la Cruz, V.F., Lal, A.A., Gunderson; J.H., Elwood, H.J. and Sogin, M.L.: The primary sequences of the small subunit ribosomal RNA genes from *Plasmodium falciparum*. Mol. Biochem. Parasitol. 28 (1988) 63–68.

Messing, J.: New M13 vectors for cloning. Methods Enzymol. 101 (1983) 20–78.

Mullis, K.B. and Faloona, F.A.: Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155 (1987) 335–350.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R. and Stahl, D.A.: Microbial ecology and evolution: a ribosomal approach. Annu. Rev. Microbiol. 40 (1986) 337–365.

Qu, L.H., Michot, B. and Bachellerie, J-P.: Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Applications to the 5′ terminal domain of eukaryotic 28S rRNA. Nucleic Acids Res. 11 (1983) 5903–5920.

Saiki, R., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A.: Primer-directed

enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 (1988) 487–491.
Sanger, F. and Coulson, A.R.: A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94 (1975) 441–448.
Sogin, M.L. and Gunderson, J.H.: Structural diversity of eukaryotic small subunit ribosomal RNAs: evolutionary implications. Endocytobiology III. Ann. N.Y. Acad. Sci. 503 (1987) 125–139.
Sogin, M.L., Pechman, K.J., Zablen, L., Lewis, B.J. and Woese, C.R.: Observations on the post-transcriptionally modified nucleotides in the 16S ribosomal ribonucleic acid. J. Bacteriol. 112 (1972) 13–22.
Sogin, M.L., Elwood, H.J. and Gunderson, J.H.: Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc. Natl. Acad. Sci. USA 83 (1986a) 1383–1387.
Sogin, M.L., Ingold, A., Karlok, M., Nielsen, H. and Engberg, J.: Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major *Tetrahymena* groups. EMBO J. 5 (1986b) 3625–3630.
Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F.: Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. 47 (1983) 621–669.
Wrischnik, L.A., Higuchi, R.G., Stoneking, M., Erlich, H.A., Arnheim, N. and Wilson, A.C.: Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids. Res. 2 (1987) 529–541.

Communicated by G.N. Gussin.