EVE 161:
Microbial Phylogenomics

Class #7:
Era II: rRNA from the Environment

UC Davis, Winter 2018
Instructor: Jonathan Eisen
Teaching Assistant: Cassie Ettinger

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6955–6959, October 1985
Evolution

Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses

(reverse transcriptase/dideoxynucleotide)

DAVID J. LANE*, BERNADETTE PACE*, GARY J. OLSEN*, DAVID A. STAHL†‡, MITCHELL L. SOGIN†, AND NORMAN R. PACE*§

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1985, p. 1379–1384 Vol. 49, No. 6
0099-2240/85/061379-06$02.00/0

Characterization of a Yellowstone Hot Spring Microbial Community by 5S rRNA Sequences

DAVID A. STAHL,† DAVID J. LANE, GARY J. OLSEN, AND NORMAN R. PACE*
Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405

Received 30 November 1984/Accepted 8 March 1985

Presenters for Today?

Terminology and Questions?

Differences Between the Papers?

Paper #1

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6955–6959, October 1985
Evolution

Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses

(reverse transcriptase/dideoxynucleotide)

DAVID J. LANE*, BERNADETTE PACE*, GARY J. OLSEN*, DAVID A. STAHL†‡, MITCHELL L. SOGIN†, AND NORMAN R. PACE*§

Lane et al. Abstract

ABSTRACT Although the applicability of small subunit ribosomal RNA (16S rRNA) sequences for bacterial classification is now well accepted, the general use of these molecules has been hindered by the technical difficulty of obtaining their sequences. A protocol is described for rapidly generating large blocks of 16S rRNA sequence data without isolation of the 16S rRNA or cloning of its gene. The 16S rRNA in bulk cellular RNA preparations is selectively targeted for dideoxynucleotide-terminated sequencing by using reverse transcriptase and synthetic oligodeoxynucleotide primers complementary to universally conserved 16S rRNA sequences. Three particularly useful priming sites, which provide access to the three major 16S rRNA structural domains, routinely yield 800–1000 nucleotides of 16S rRNA sequence. The method is evaluated with respect to accuracy, sensitivity to modified nucleotides in the template RNA, and phylogenetic usefulness, by examination of several 16S rRNAs whose gene sequences are known. The relative simplicity of this approach should facilitate a rapid expansion of the 16S rRNA sequence collection available for phylogenetic analyses.

Lane et al. Background: Why Sequences?

All of the available molecular methods for evaluating phylogenetic relationships (e.g., DNA·DNA and DNA·rRNA hybridization, 5S rRNA and protein sequencing, 16S rRNA oligonucleotide cataloging, enzymological patterning, etc.) have advantages and limitations. In general, macromolecular sequences seem preferred because they permit quantitative inference of relationships (2, 3). Moreover, because they accumulate, sequences are most useful in the long term.

Lane et al. Background: Why Sequences?

All of the available molecular methods for evaluating phylogenetic relationships (e.g., DNA·DNA and DNA·rRNA hybridization, 5S rRNA and protein sequencing, 16S rRNA oligonucleotide cataloging, enzymological patterning, etc.) have advantages and limitations. In general, macromolecular sequences seem preferred because they permit quantitative inference of relationships (2, 3). Moreover, because they accumulate, sequences are most useful in the long term.

2. Zuckerkandl, E. & Pauling, L. (1965) *J. Theor. Biol.* **8**, 357–366.
3. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.

Lane et al. Background: Why rRNA

Of the macromolecules used for phylogenetic analysis, the ribosomal RNAs, particularly 16S rRNA,[1] have proven the most useful for establishing distant relationships because of their high information content, conservative nature, and universal distribution. Using RNase T1 oligonucleotide catalogs of 16S rRNA, Woese and his colleagues were able to establish a comprehensive outline of prokaryotic phylogeny (4). The principle of using rRNA sequences to characterize microorganisms has now gained wide acceptance (5), and its

Lane et al. Background: Why rRNA

Of the macromolecules used for phylogenetic analysis, the ribosomal RNAs, particularly 16S rRNA,[1] have proven the most useful for establishing distant relationships because of their high information content, conservative nature, and universal distribution. Using RNase T1 oligonucleotide catalogs of 16S rRNA, Woese and his colleagues were able to establish a comprehensive outline of prokaryotic phylogeny (4). The principle of using rRNA sequences to characterize microorganisms has now gained wide acceptance (5), and its

4. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. (1980) *Science* **209**, 457–463.
5. Murray, R. G. E. (1984) in *Bergey's Manual of Systematic Bacteriology*, eds. Kreig, N. R. & Holt, J. G. (Williams & Wilkins, Baltimore), Vol. 1, pp. 31–34.

Lane et al. Methods: How Get RNA

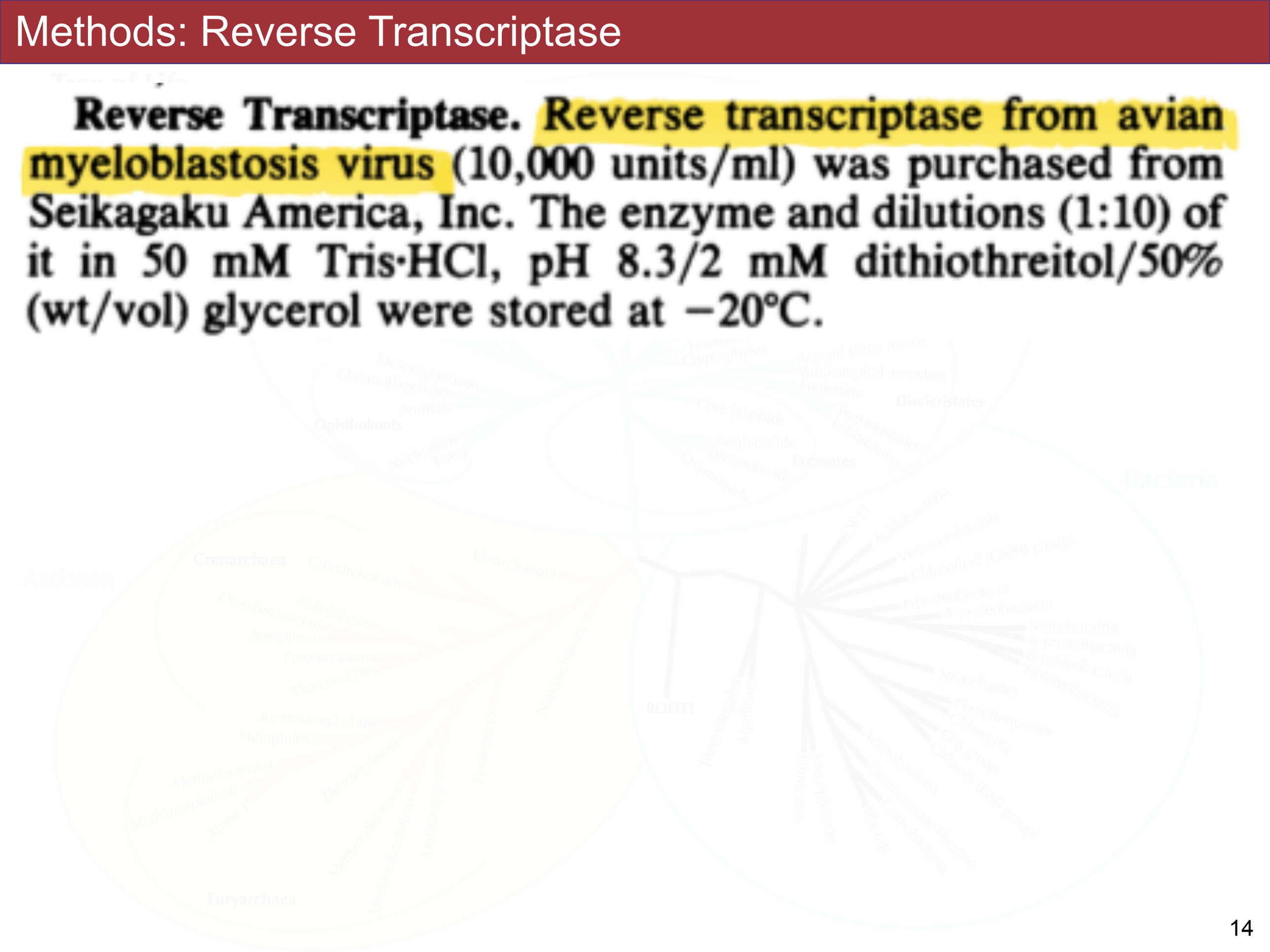
Purification of RNA Templates. Bulk, cellular RNA was purified by phenol extraction of French pressure cell lysates as detailed by Pace *et al.* (6), except that ribosomes were not pelleted before extraction. High molecular weight RNA was then prepared by precipitation with 2 M NaCl (6). Although not essential, NaCl precipitation of the RNA generally increased the amount of legible sequence data and reduced backgrounds on gels, presumably by eliminating fragmented DNA from the reactions. RNA was stored at 2 mg/ml in 10 mM Tris·HCl (pH 7.4) at −20°C.

Methods: How Make Primers

Oligodeoxynucleotide Primers. Oligodeoxynucleotide primers were synthesized manually by using the appropriate blocked and protected nucleoside diisopropylphosphoramidites and established coupling protocols (7). Deblocked products were purified by polyacrylamide gel electrophoresis, eluted from gels in 1 mM EDTA/50 mM NH_4OAc, and adsorbed onto C_8 Bond Elut columns (Analytichem International, Harbor City, CA). The columns were washed with the loading buffer and then washed with 5 mM NH_4OAc, and finally the primers were eluted with acetonitrile/H_2O, 1:1 (vol/vol). After drying *in vacuo*, primers were dissolved in H_2O to 0.1 mg/ml. (Requests for these primers should be addressed to Susan Andrews, Institute for Molecular and Cellular Biology, Indiana University, Bloomington, IN 47405.)

Methods: Reverse Transcriptase

Reverse Transcriptase. Reverse transcriptase from avian myeloblastosis virus (10,000 units/ml) was purchased from Seikagaku America, Inc. The enzyme and dilutions (1:10) of it in 50 mM Tris·HCl, pH 8.3/2 mM dithiothreitol/50% (wt/vol) glycerol were stored at −20°C.



Methods: Nucleotides

Nucleotides. Deoxyadenosine 5′-[α-thio]triphosphate labeled with ^{35}S in the α-thio position (dATP[α-^{35}S]) was purchased from New England Nuclear. The 2′-deoxy- (d-) and 2′,3′-dideoxy- (dd-) nucleoside triphosphates were from P-L Biochemicals. Stock solutions of unlabeled nucleotides [10 mM in 10 mM Tris·HCl (pH 8.3) as determined spectrophotometrically] were stored frozen at −70°C. Nucleotide mixtures for the reverse transcription reactions contained 10 mM Tris·HCl (pH 8.3), 250 μM dCTP, 250 μM dGTP, 250 μM dTTP, 125 μM [α-thio]dATP, and either one dideoxynucleotide (30 μM ddCTP, 19 μM ddGTP, 30 μM ddTTP, or 1.25 μM ddATP) or no dideoxynucleotide. Aliquots sufficient for 10 reactions were stored at −20°C.

Methods: Sequencing

Reverse Transcription Reactions. The sequencing protocol described here is a base-specific, dideoxynucleotide-terminated chain elongation method (8, 9), modified for the use of reverse transcriptase and RNA templates (10). Since sequencing reactions were only 5.0 μl, common components were mixed and then aliquots were taken to reduce pipetting errors. Untreated 0.5-ml Microfuge tubes were used throughout. DNA primers were hybridized to RNA templates in 7.5-μl reaction mixtures containing 1.5 μl of 5× hybridization buffer (500 mM KCl/250 mM Tris·HCl, pH 8.5), 3.5 μl of high molecular weight RNA (2.0 mg/ml), 1.5 μl of the desired primer (0.1 mg/ml), and 1.0 μl of H_2O. This annealing mixture was heated at 90°C for 1 min and then allowed to cool slowly over 10 min to 25°C. Of the hybridized template–primer mixture, 6.5 μl was then transferred to a tube containing 30 μCi (1 Ci = 37 GBq) of dried dATP[α-^{35}S], 6.5 μl of 5× reverse transcription buffer (250 mM Tris·HCl, pH 8.3/250 mM KCl/50 mM dithiothreitol/50 mM $MgCl_2$), and 6.5 μl of reverse transcriptase (1,000 units/ml). Three microliters of this final mixture was then added to each of six tubes that contained, respectively, 2.0 μl of ddCTP-, ddATP-, ddTTP-, ddGTP-, ddCTP-, or no-dideoxynucleotide-containing nucleotide mixture (above). The reactions were incubated for 5 min at room temperature and then 30 min at 37°C. One microliter of chase mix [1.0 mM each of dATP, dCTP, dTTP, and dGTP in 10 mM Tris·HCl (pH 8.3) containing 1,000 units/ml of reverse transcriptase] was then added to each reaction, and incubation at 37°C was continued for another 15 min. The reactions were chilled on ice and stopped by adding 6.0 μl of stop mix (86% formamide/10 mM EDTA/0.08% xylene cyanol/0.08% bromophenol blue). Reaction mixtures were heated for 2 min at 90°C immediately prior to loading onto sequencing gels. Two gels were usually run on each set of sequencing reactions: one was a 0.3-mm-thick, 40-cm-long, 8% polyacrylamide buffer gradient gel (9) (bromophenol blue run to 40 cm), and the other was an equivalent, 8% polyacrylamide nongradient gel (xylene cyanol run off the bottom to a calculated "distance" of 70 cm).

RESULTS AND DISCUSSION

In assessing the relationships of organisms to one another by the comparison of their 16S rRNA sequences, it is not important that the complete sequences of the molecules be determined. What is important is that the number of nucleotides compared is statistically meaningful and that com-

Priming

truly homologous sequence positions are considered.

The 16S rRNAs vary in their nucleotide sequences, but they contain regions that are conserved perfectly, or nearly so, among all organisms so far inspected. Certain of these conserved sequences, adjacent to less-conserved regions that are useful for phylogenetic evaluations, provide broadly applicable initiation sites for primer elongation sequencing techniques. Oligodeoxynucleotides, 15–20 residues in length, that are complementary to certain of the conserved sequences were synthesized and tested as primers for dideoxynucleotide-terminated sequencing reactions with reverse transcriptase and 16S rRNA templates.

Fig. 1 shows the location of three of the most useful priming sites in representative eubacterial, archaebacterial, and eukaryotic small-subunit rRNAs and indicates the extent of phylogenetically useful sequence routinely obtained from

Priming

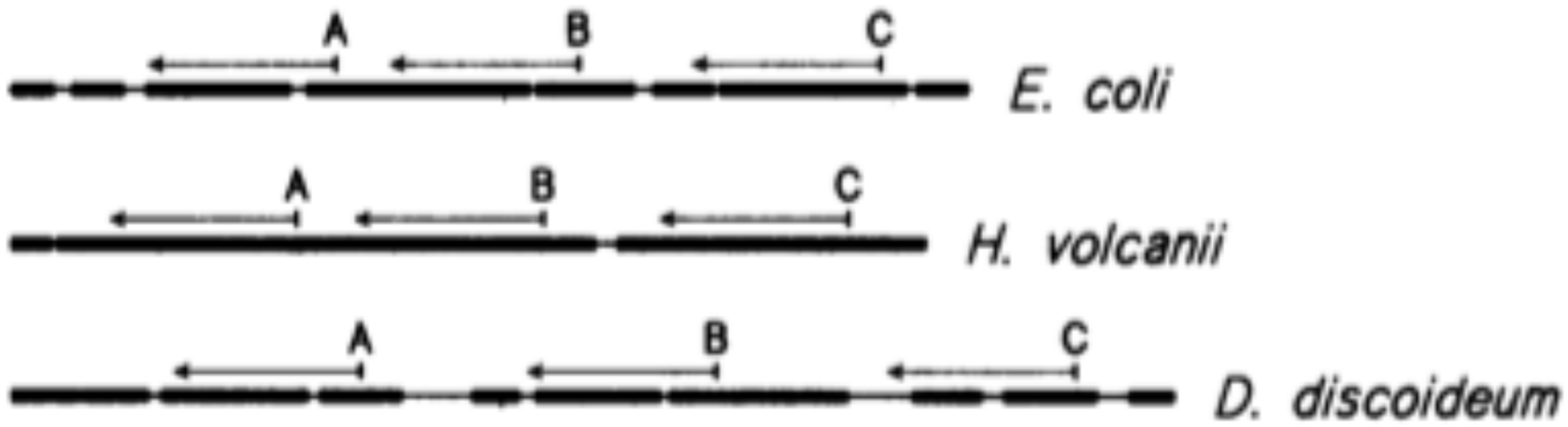


FIG. 1. Hybridizing sites of "universal" small-subunit rRNA primers. The locations of three particularly useful priming sites (A, B, and C) are shown on linear representations of the small-subunit rRNAs from *Escherichia coli* (a eubacterium) (11), *Halobacterium volcanii* (an archaebacterium) (12), and *Dictyostelium discoideum* (a eukaryote) (13). The primer sequences and their hybridizing locations in the *E. coli* 16S rRNA are G-W-A-T-T-A-C-C-G-C-G-G-C-K-G-C-T-G, positions 519–536 (A); C-C-G-T-C-A-A-T-T-C-M-T-T-T-R-A-G-T-T-T, 907–926 (B); and A-C-G-G-G-C-G-G-T-G-T-G-T-R-C, 1392–1406 (C). In these sequences K = G or T, M = A or C, R = A or G, and W = A or T. The solid boxes along the sequence lines are regions that display sufficient intrakingdom structural conservation to be generally useful in the inference of phylogenies. The arrow pointing to the left from each priming site indicates the approximate extent of the sequence data (300 nucleotides) accessible from each primer.

Priming

universal.

Reverse transcriptase has been used previously to copy rRNA templates (14–17). Initial attempts to produce cDNA reverse transcripts from 16S rRNA templates using appended polyadenylate tails to create priming sites met with only partial success (14); reverse transcription was found to terminate abruptly about 30 nucleotides from the 3′-end, corresponding to the location of a widely conserved N^6-dimethyladenosine dinucleotide, m$^6_2$A-m$^6_2$A. However, it was subsequently demonstrated with *E. coli* 16S rRNA as template that if reverse transcription were initiated 5′ proximal to that modified dinucleotide, then the enzyme would proceed to the 5′ end of the template (15). Other nucleoside modifications that terminate reverse transcription from some 16S rRNAs have been identified (16), but these fall outside the sequences routinely accessible from the universal rRNA primers shown in Fig. 1.

Errors

the enzyme.

Since most reverse transcriptase molecules read through the anomalous positions, the template nucleotide often is revealed as the most prominent band (Fig. 2). Positions where the residue assignment is not clear are scored as "N," for sequence alignment and comparison purposes. Ambiguities in the derived nucleotide sequence merely reduce the extent of sequence available for comparison with other sequences. For phylogenetic purposes, even minor inaccuracies in nucleotide identification incur little penalty. Random errors impart the appearance of a fast "evolutionary clock." On the other hand, assignment of nucleotide identity on the basis of homology with a sequence from another organism (which might be assumed related) results in systematic errors, perhaps forcing incorrect affilations.

Errors

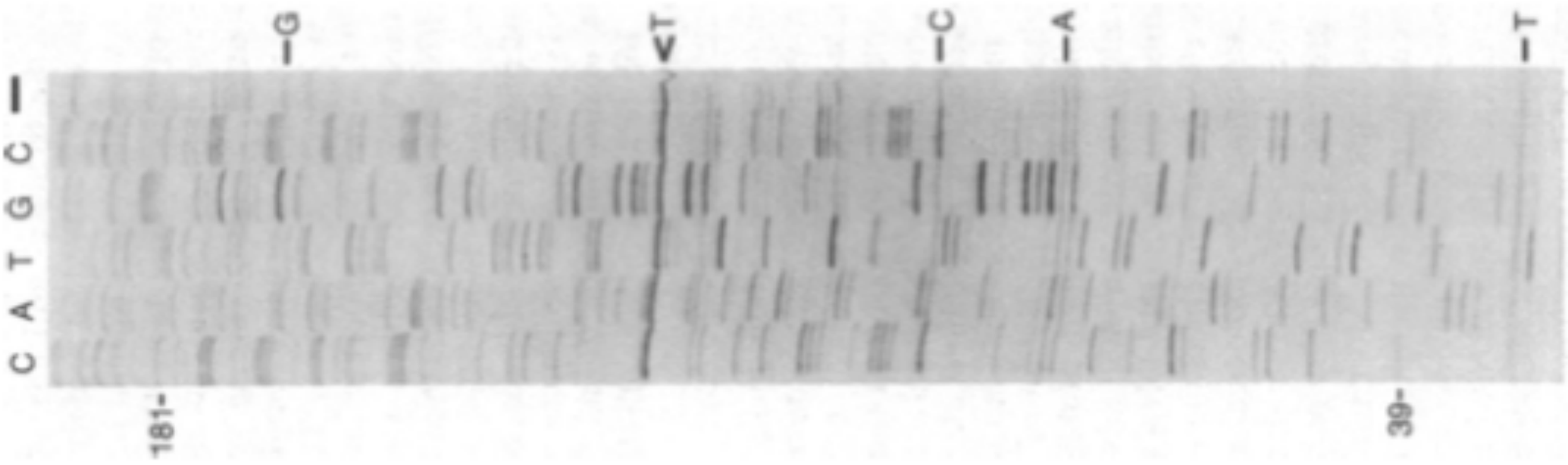


FIG. 2. Reverse transcriptase sequencing "anomalies." An autoradiogram of a typical 8% acrylamide/8 M urea buffer-gradient sequencing gel analysis (9) of reverse transcriptase reactions using bulk, high molecular weight *H. volcanii* RNA as template and the 1392–1406 primer is shown. Distances from the 3′ terminus of the primer are indicated on the left. Terminations not mediated by dideoxynucleotide incorporations in which the correct nucleotide is evident (−) and those where it is not (<) are indicated on the right. Sequencing reactions containing dideoxycytosine (C), dideoxyadenosine (A), dideoxythymidine (T), dideoxyguanosine (G), or no dideoxynucleotides (−) are indicated at the top. The two cytidine reactions facilitated alignment of bands across the gel.

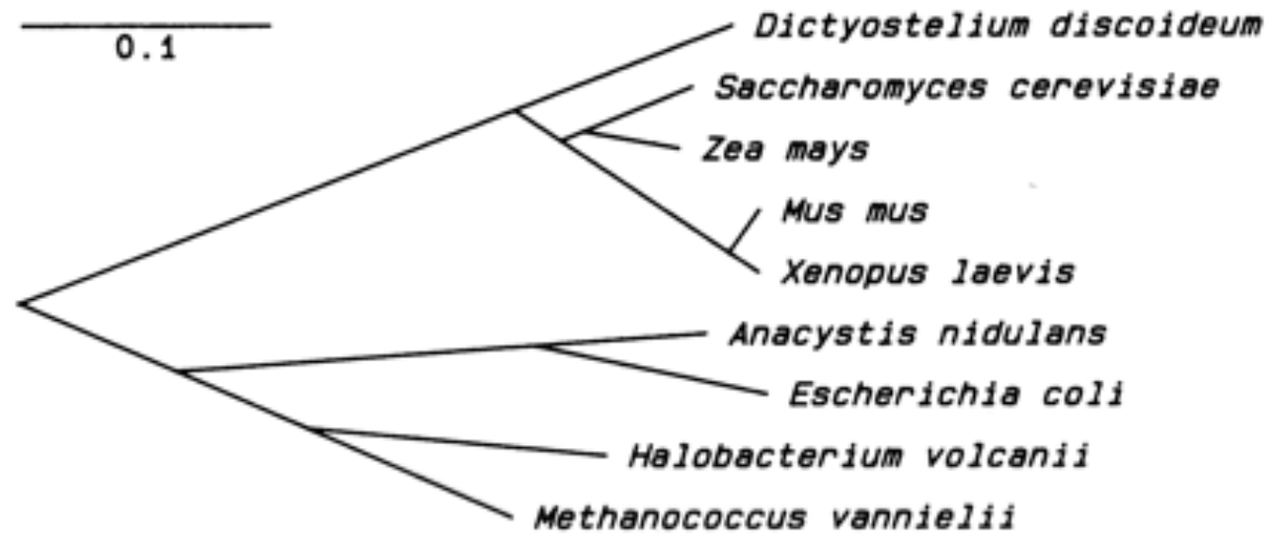
Alignment

After the collection of partial sequence data from a novel organism, they must be aligned with other available sequences. Variabilities in the primary and secondary structures of small subunit rRNAs are now sufficiently well understood that alignment of a novel sequence on the known skeleton is straightforward, keying on regions of conserved sequence and secondary structure (23). By alignment of novel sequences on the existing rRNA framework, many sequence errors due to band compressions or rearrangements on sequencing gels may also be detected. For establishing phylogenetic relationships, it is imperative to employ only sequences that are unambiguously homologous (24). Regions of ambiguous homology must be eliminated from analyses.

There are two approaches to establishing the phylogenetic relationships of an aligned sequence to others in the data collection. In the first approach, phylogenetic "trees" are generated using the available 16S rRNA sequences. Nucleic acid and protein sequences have been extensively used in phylogenetic tree construction, although use of the method with 16S rRNA is limited at this time by the small number of sequences available; only about 25 complete 16S rRNA sequences have been reported. As the reference collection of complete and partial 16S rRNA sequences expands, however, quantitative phylogenetic trees may provide the best means for relating organisms.

means for relating organisms.

It has been shown previously that phylogenetic trees constructed by using certain limited regions of the 16S rRNA had topologies identical to that obtained by using complete sequences (13, 25). Fig. 3 illustrates that this is also true for the sequence blocks accessed by reverse transcription from the universal rRNA primers. The upper tree was constructed



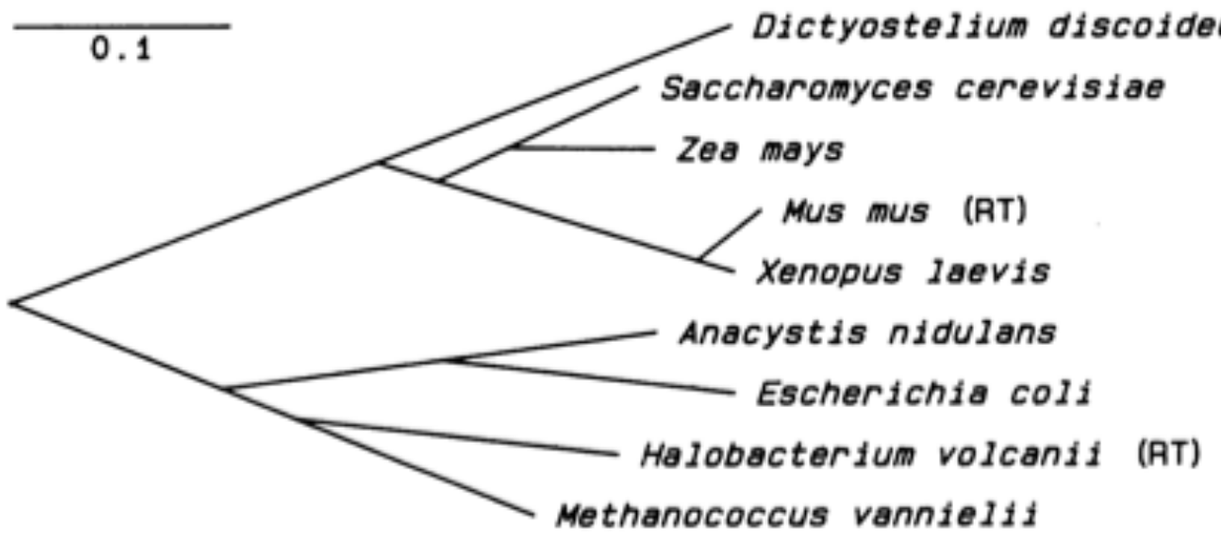


FIG. 3. Comparison of phylogenetic relationships inferred by using complete and partial 16S rRNA sequences. The illustrated networks ("trees") were deduced by the matrix method (24), omitting regions of ambiguous sequence alignment from homology calculations. Alignment gaps were assigned half the weight assigned to a nucleotide in that position. While determining the optimal tree, the mean square difference between the estimated evolutionary distance separating each pair of sequences and the corresponding tree distance was weighted by the statistical uncertainty of the distance estimate (26). An iterative program that follows the path of "steepest descent" in the optimization parameter (25, 27) was used to determine the tree topology which best fit, by the weighted least-mean-square difference criterion, the homology data. The scale bar represents an evolutionary distance of 0.1 K_{nuc} (K_{nuc} = average number of nucleotide changes per sequence position). The networks are based on complete 16S rRNA sequences (11–13, 18, 28–32) using about 950 nucleotide positions after elimination of regions of ambiguous alignment (*Upper*) or are based on blocks of sequence determined by reverse transcription from the 1392–1406 and 907–926 primers of *H. volcanii*, *D. discoideum*, and mouse rRNA and homologous blocks from the other sequences (*Lower*). Only about 350 nucleotide positions are compared. However, it should be noted that because of the lesser amount of interkingdom sequence conservation, fewer unambiguously alignable blocks of sequence are available for these comparisons than are available (see Fig. 1) for intrakingdom comparisons.

impressive correspondence for interkingdom comparisons.

The second approach to relating organisms by using sequence information gathered as described here uses the collection of RNase T1 oligonucleotide catalogs established by Woese and his colleagues over the past decade (4). RNase

to the appropriate sequence information.

The general sequencing procedures employed here are in wide use. The novel aspect is the use of evidently universally applicable, oligodeoxynucleotide primers to access 16S rRNA sequences for phylogenetic characterizations. Because of its rapidity (from cell pellet to 800–1,000 nucleotides of sequence in *ca*. 3 days) and technical simplicity, the approach is amenable to the screening of large numbers of organisms. Therefore, phylogenetic analysis by 16S rRNA sequences could become a standard tool of laboratories concerned with characterizing organisms of uncertain affiliation.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1985, p. 1379–1384
0099-2240/85/061379-06$02.00/0

Vol. 49, No. 6

Characterization of a Yellowstone Hot Spring Microbial Community by 5S rRNA Sequences

DAVID A. STAHL,† DAVID J. LANE, GARY J. OLSEN, AND NORMAN R. PACE*
Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405

Received 30 November 1984/Accepted 8 March 1985

Stahl et al. Abstract

The microorganisms inhabiting a 91°C hot spring in Yellowstone National Park were characterized by sequencing 5S rRNAs isolated from the mixed, natural microflora without cultivation. By comparisons of these sequences with reference sequences, the phylogenetic relationships of the hot spring organisms to better characterized ones were established. Quantitation of the total 5S-sized rRNAs revealed a complex microbial community of three dominant members, a predominant archaebacterium affiliated with the sulfur-metabolizing (dependent) branch of the archaebacteria, and two eubacteria distantly related to *Thermus* spp. The archaebacterial and the eubacterial 5S rRNAs each constituted about half the examined population.

Stahl et al. Introduction

The description of pure cultures is a foundation of experimental microbiology. It seems possible, however, that much of the biological diversity of the earth has not been, or cannot be, brought into pure culture. We are developing technical strategies for analyzing fundamental aspects of naturally occurring microbial populations without the need for their cultivation. The methods use recent advances in nucleic acid sequencing and recombinant DNA technology to determine nucleotide sequences of 5S or 16S rRNA genes from microbial communities containing multiple organisms. By comparing these sequences with known ones, we can define the phylogenetic status of any organism residing in the communities. Because only the naturally available biomass is required for these methodologies, the analyses project a relatively unbiased picture of an in situ microbial community.

Stahl et al. Introduction

The description of pure cultures is a foundation of experimental microbiology. It seems possible, however, that much of the biological diversity of the earth has not been, or cannot be, brought into pure culture. We are developing technical strategies for analyzing fundamental aspects of naturally occurring microbial populations without the need for their cultivation. The methods use recent advances in nucleic acid sequencing and recombinant DNA technology to determine nucleotide sequences of 5S or 16S rRNA genes from microbial communities containing multiple organisms. By comparing these sequences with known ones, we can define the phylogenetic status of any organism residing in the communities. Because only the naturally available biomass is required for these methodologies, the analyses project a relatively unbiased picture of an in situ microbial community.

Stahl et al. Introduction

The determination of phylogenetic relationships by quantitative comparison of macromolecular sequences is well established (24). Among biopolymers, the rRNAs are possibly uniquely suited as gauges of phylogenetic relatedness because of their ubiquitous distribution, functional homology, high conservation of primary structure, and apparent freedom from artifacts of lateral transfer (8). The direct isolation and sequencing of the 5S rRNAs from environments of reasonably plentiful biomass and limited complexity offers a relatively convenient determination of the dominant community members. We have used this approach to characterize bacteria symbiotically associated with certain marine invertebrates in sulfide-rich habitats (20), and we now apply it to a Yellowstone hot spring microbial community.

The presence of microbial life in near-boiling waters has long been recognized (2, 18); the hydrothermal systems of Yellowstone National Park have provided a rich area for the study of thermophily. One of the more intensively studied thermal habitats in the park is Octopus Spring (3). Here, most effort has been applied to the cooler effluent waters. Although microbial accumulation in the 91°C source of Octopus Spring is evident, there are no reports of the cultivation of characteristic microorganisms. Indeed, the inhabitants have resisted attempts at cultivation (3), a common theme with natural microbial populations. As detailed in this report, sequence analysis of the naturally available 5S rRNAs establishes the presence of a complex community with three dominant members: two representatives of eubacteria distantly related to *Thermus* spp. and one representative of the archaebacteria.

MATERIALS AND METHODS

Sample collection. To obtain sufficient microbial biomass for analysis, we devised a simple collection device consisting of cotton or glass fiber batting sandwiched between nylon screening. This collector was immersed in the main pool of Octopus Spring for 1 week and then harvested by scissoring into conveniently sized segments and freezing on dry ice.

Stahl et al. Methods

5S rRNA isolation and sequences. Total low-molecular-weight RNA was isolated directly from the fiber batting. After several freeze-thaw cycles in the presence of sodium dodecyl sulfate, total nucleic acids were extracted by using hot phenol and sodium dodecyl sulfate. In our experience, this method offers good yields of the low-molecular-weight species (5S and 4S). The nucleic acids partition in the aqueous phase and are recovered as ethanol precipitates. The 3′ termini of total nucleic acids were labeled by the RNA ligase-catalyzed appendages of [5′-^{32}P]pCp, and the total, labeled, low-molecular-weight RNA population (5S and 4S) was fractionated on either one- or two-dimensional gels (20). Bands corresponding in approximate size to 5S rRNA were excised and eluted for sequence analysis by both enzymatic and chemical protocols (6, 16).

Stahl et al. Methods

Determination of relative 5S rRNA abundance. From a broad region of a preparative acrylamide gel, the RNA roughly corresponding in size to 5S rRNA was eluted, digested to completion with RNase T_1, and 5′-^{32}P end labeled with [γ-^{32}P]ATP and polynucleotide kinase. After labeling, an excess of 3′-UMP and an additional 0.7 U of polynucleotide kinase were added, and the incubation was continued to scavenge remaining [γ-^{32}P]ATP. The end-labeled oligonucleotides were resolved by two-dimensional electrophoresis according to the method of Sanger and Brownlee (17) and were located by autoradiography. All spots were excised, and their radioactive contents were determined by scintillation counting. Spots destined for sequence determination were eluted from the paper with either 1 M NH_4HCO_3 or 1 M formic acid adjusted to pH 4.3 with pyridine. The eluted materials were repeatedly lyophilized and hydrated to remove residual salt. Sequencing of the eluted oligonucleotides was by partial enzymatic digestions as outlined in the sequencing description and resolution on acrylamide gels or by polyethyleneimine thin-layer chromatography with the above pyridine-formate solvent.

Stahl et al. Methods

Determination of phylogenetic relationships. The 5S rRNA nucleotide sequences were aligned according to a common secondary structure and by universal positions in the primary structure, as summarized by Erdmann et al. (7). Regions of base pairing, as defined by the 5S rRNA consensus secondary structure, were labeled according to Stahl et al. (21) (see Fig. 2). The corresponding phylogeny was deduced by the matrix method essentially as described by Hori and Osawa (12). Regions of terminal length variation were omitted from homology calculations, base-paired positions were given half the weight of unpaired positions, and alignment gaps were given half the weight assigned to a nucleotide in the same position. In determining the optimal tree, we weighted the difference between the estimated evolutionary distance separating each pair of sequences and the corresponding tree distance by the statistical uncertainty of the distance estimate (12).

Stahl et al. Results

RESULTS

Our initial interest in the Octopus Spring microbial community was prompted by observations of visible microbial biomass accumulation in this hot spring (3). This accumulation, previously described as pink tufts or streamers, is attached to the substratum in the source effluent. Objects immersed in the source overflow are rapidly and visibly colonized. There is one reported analysis of pink tuft accumulations for lipid composition (1) but no reported cultivation of characteristic microorganisms from these materials.

Stahl et al. Results

We were unable to extract appreciable nucleic acid from pink tufts collected from Octopus Spring, apparently because the bulk of this material is dead or moribund. Probably only a surface lamella of the fibrous pink tuft accumulations

is composed of viable mass. This is consistent with observations by Brock that these accumulations incorporate little or no radioisotopically labeled substrates (3). We therefore used a simple collection device consisting of glass or cotton fiber batting sandwiched between nylon mesh. The basic notion here was to provide an extensive surface area (the fibers) for the development of a microbial film analogous to that seen on contact slides. After immersion for 1 week in the source waters, these devices yielded ample biomass for analysis. Total nucleic acids were extracted directly from mesh segments as described above.

Stahl et al. Results

No fractionation or treatment of nucleic acids was carried out before isotopic labeling with RNA ligase and [5′-^{32}P]pCp. Because RNA ligase requires a 3′ hydroxyl acceptor, its use (in conjunction with [5′-^{32}P]pCp) favors the labeling of native rRNAs. Most degradation products, such as those that result from random hydrolysis or nonspecific nuclease action, possess 3′-phosphorylated termini and so are not labeled. An example of a preparative fractionation of 5S rRNA from total labeled nucleic acid is shown in Fig. 1. The bands indicated were excised, eluted, and sequenced as detailed above. The sequences of these isolated molecules are shown in Fig. 2. Two of the sequences are eubacterial in nature, and the third characterized is archaebacterial.

Stahl et al. Results

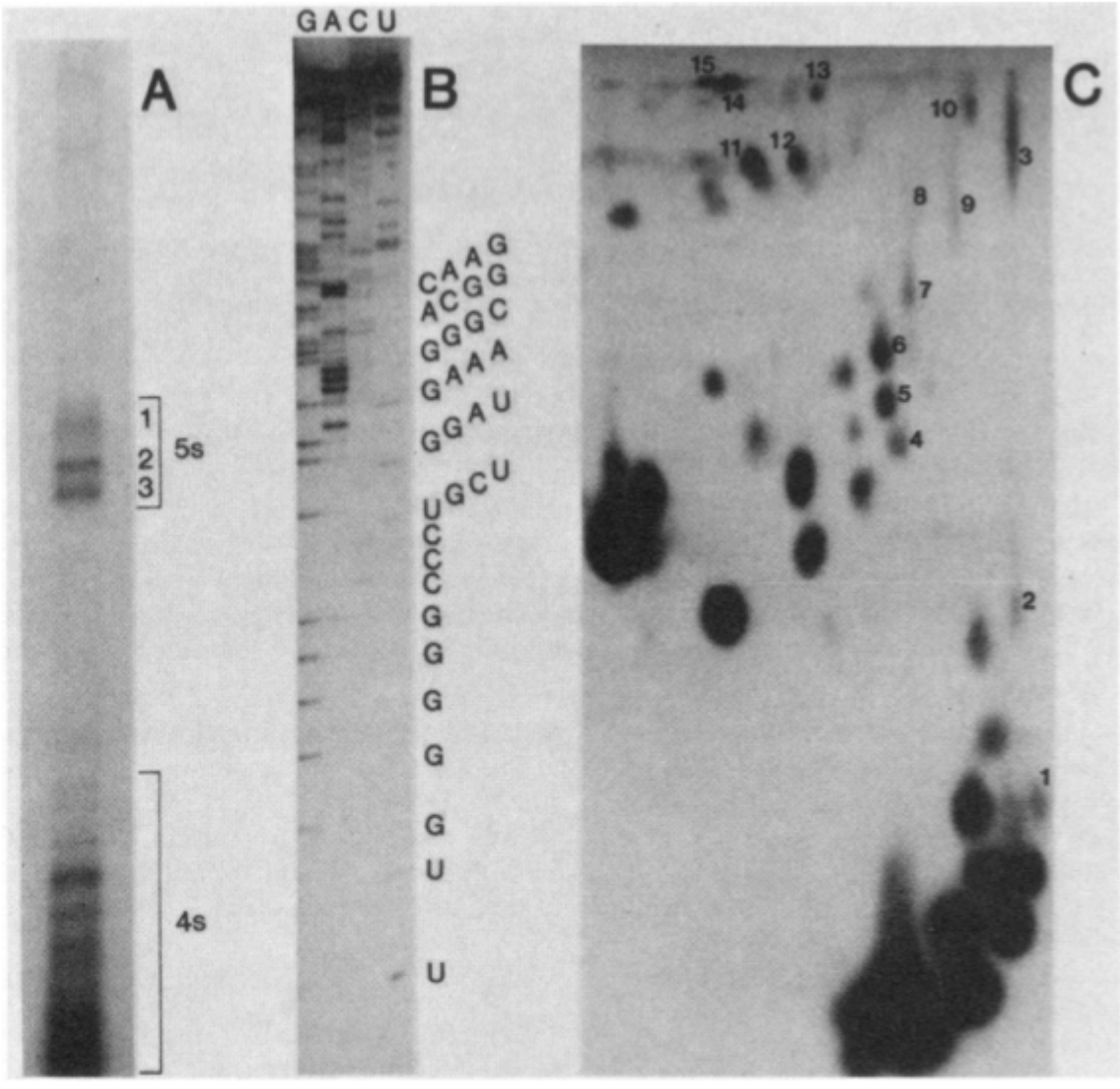


FIG. 1. Sequence and abundance analysis of the Octopus Spring 5S rRNAs. (A) Autoradiograph of total 3′-end-labeled low-molecular-weight RNAs separated on a high-resolution 10% polyacrylamide gel. (B) A representative sequencing gel autoradiograph of an Octopus Spring 5S rRNA subjected to base-specific partial chemical cleavages. (C) Autoradiograph of a two-dimensional fractionation (fingerprint) of RNase T_1 generated and 5′-end-labeled oligonucleotides derived from total 5S-sized rRNAs.

Stahl et al. Results

```
                          I           II            III                   III'
S. acidocaldarius     GCCC-ACCCGGUCACAGU-GAGCGGGCAACACCCGGACUCAUUUCGAACCCGGAAGUUAAGCCG
Octopus Spring 1     GCCCCUAACCCGGCAAUAGGCGCCGGUGCUACGCCCGGUCUC-UUCAGAACCCGGAAGCUAAGGCC

                     II'        V          IV                   IV'      V'       I'
S. acidocaldarius    CUC-ACGUUAGUGGG-GCCGUGGAUACCGUGAGGAUCCGCAGCCCCACUAAGCUGGGAUGGGUUUU
Octopus Spring 1     GGCGCCGCGGACGGGAGUACUGGGGUCCGCGAGGCCCCGGGAAACCGCCGUGCUGGGAGGGGGCUU

                         I            II          III                   III'
Th. aquaticus        AAUCCCCCGUGCCCUUAGCGGCGUGGAACCACCCGU-UCCCAUUCCGAACACGGAAGUGAAACGCG
Th. thermophilus     aAUCCCCCGUGCCCAUAGCGGCGUGGAACCACCCGU-UCCCAUUCCGAACACGGAAGUGAAACGCG
Octopus Spring 2        UCCUCGGACCAUAGCGGAGGGAAACACCCGGUUUCCAUUCCGAACCCGGCAGUUAAGCCCU
Octopus Spring 3        CUCGGACCACCAUACCGGGGGGAAACACCCGG-UCCCAUUCCGAACCCGGAAGUUAAGCCCC

                     II'     V        IV                 IV'          V'       I'
Th. aquaticus        CCAGCGCCGAUGGUACUGGGACC-GCAG-GGUCCUGGGAGAGUAGGUCGGUGCGGGGGAU
Th. thermophilus     CCAGCGCCGAUGGUACUGGGCGG-GCGA-CCGCCUGGGAGAGUAGGUCGGUGCGGGGGAu
Octopus Spring 2     CCAGCGCCGAUGAUACUGUGCCGGGCGAACGGCACGGAAAGUAGGUCGUCCCGGGGUU
Octopus Spring 3     CCAGGGCCGAUGAUAGCCUCGCC-CCGA--GCGAGGUGAAAGUAGGUCGUGGUCCGGGCAC
```

FIG. 2. Alignment of the 5S rRNA sequences isolated from Octopus Spring with those of *Sulfolobus acidocaldarius* and *Thermus* spp. Regions of base pairing as defined by the 5S consensus secondary structure (9, 23) are indicated by horizontal bars and are labeled as previously described (21).

Stahl et al. Results

The phylogenetic relationships of the Octopus Spring organisms to one another and to better characterized microorganisms are best seen as a phylogenetic tree. The one presented in Fig. 3 spans the purple photosynthetic line of eubacterial descent (10). The inset summarizes most of the lines of eubacterial descent so far defined by partial 16S rRNA sequence characterization and is offered as reference to more familiar organisms. Two of the sequenced 5S rRNAs are of eubacterial descent (Fig. 3), and they most closely

http://organismsfortheseThermusspp.is

Thermus spp., albeit as fairly distant relatives. The 5S rRNA analysis suggests that the represented *Thermus* spp. and the eubacterial Octopus Spring organisms radiated from a common ancestor near the origin of the purple photosynthetic line of descent and should possibly be included in it. No 16S rRNA sequence information is yet available for *Thermus* spp., and because phylogenetic analysis based on 5S rRNA does not satisfactorily bridge longer phylogenetic distances, the suggested placement should be considered tentative until the data base of reference sequences in this region of the phylogenetic map is fleshed out. The phylogenetic affinities of the Octopus Spring organisms for these *Thermus* spp. is, however, clear.

Stahl et al. Results

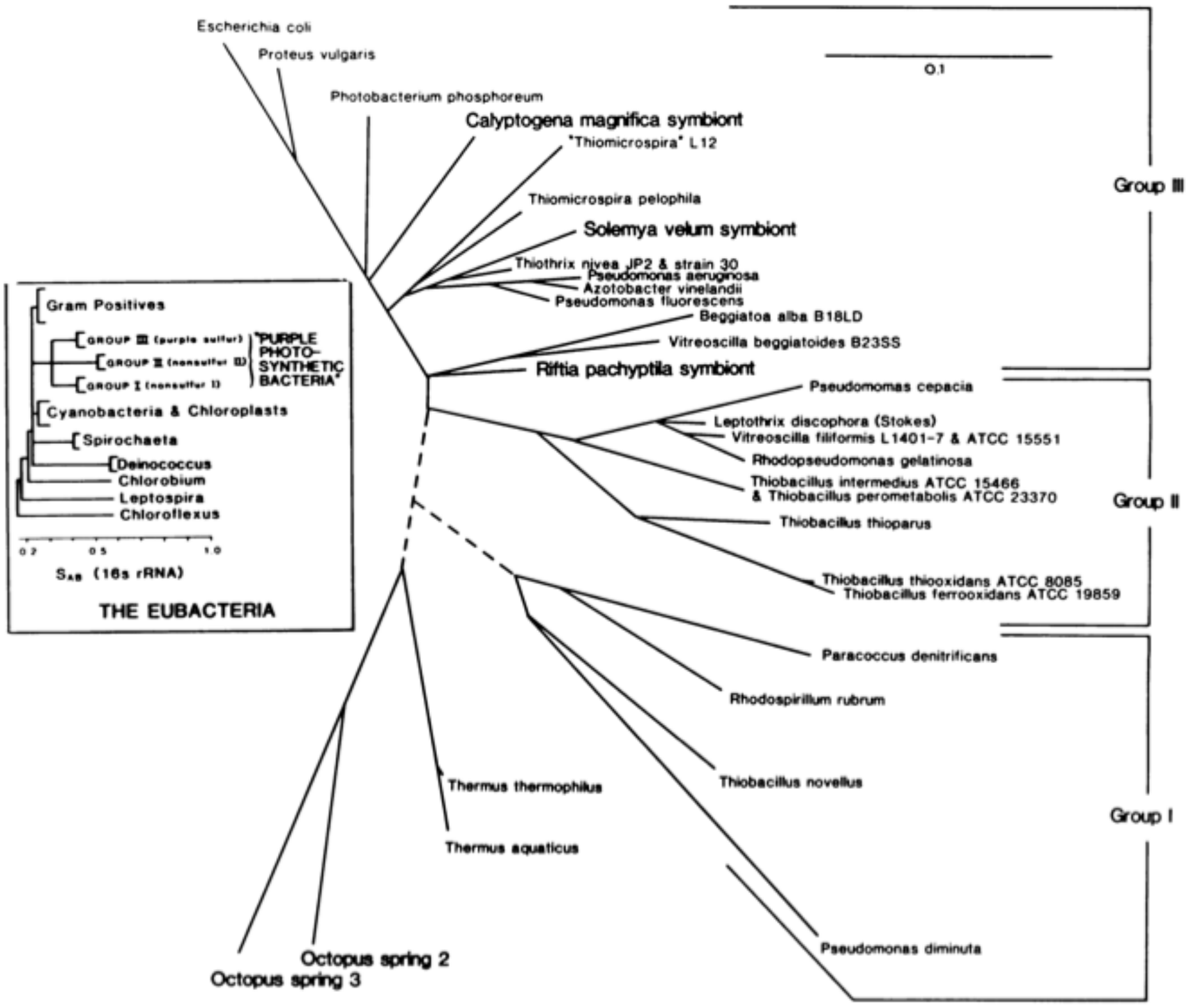


FIG. 3. Phylogeny of representative purple photosynthetic bacteria and *Thermus* spp. 5S rRNA sequences (20; D. A. Stahl, D. J. Lane, D. Heller, and N. R. Pace, unpublished data). The line of descent including the *Thermus* spp. and the Octopus Spring eubacterial 5S rRNAs (in boldface) is not necessarily included in the purple photosynthetic assemblage (see text). The scale bar represents an evolutionary distance of 0.1 nucleotide change per sequence position. The root of the tree, based on relative distances within the tree, is suggested to lie within the dotted segments. The inset (19) illustrates most of the major phylogenetic groupings (phyla) so far defined within the eubacteria. For a discussion of the symbionts (also in boldface), see reference 20.

Stahl et al. Results

The remaining 5S rRNA isolated from the source waters is archaebacterial; its phylogenetic position is depicted in Fig. 4. Although there currently are no 5S rRNAs in our sequence collection which can be considered close relatives of this organism, it clearly belongs with an assemblage recently defined as the sulfur-metabolizing branch of the archaebacteria (K. O. Stetter and W. Zillig, *in* C. R. Woese and R. S. Wolfe, ed., *The Bacteria*, Vol. 8, in press). This group includes the *Thermoproteales* and organisms similar to *Sulfolobus* spp. (*Sulfolobales*).

Stahl et al. Results

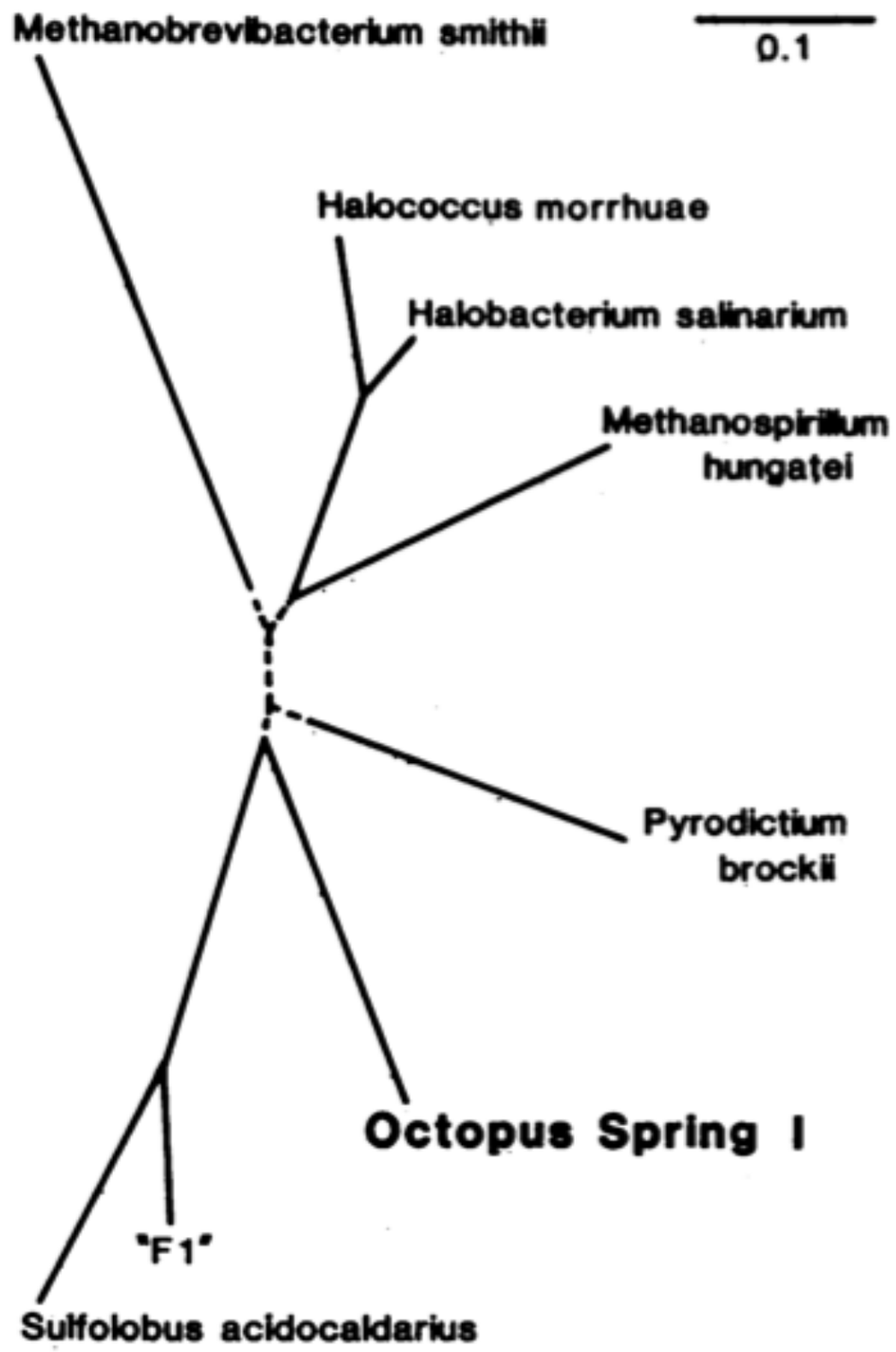


FIG. 4. Phylogeny of the Octopus Spring archaebacterium and other selected archaebacterial 5S rRNA sequences. The format is the same as that of Fig. 3.

Stahl et al. Results

The relative abundances of these major Octopus Spring 5S rRNAs and hence some appreciation of the relative abundance of the resident microorganisms were established by quantitation of oligonucleotides unique to each representative 5S rRNA. Different 5S rRNAs label at their termini with different efficiencies, so the relative incorporation of radioactivity into the various members of a population of intact 5S rRNAs does not necessarily correlate well with the relative abundances of the donor organisms. To ameliorate this bias, the total collection of 5S-sized material was first digested to

Stahl et al. Results

completion with RNase T_1, which cleaves specifically after guanosine residues. The oligonucleotides derived from the mixed population of 5S rRNAs were then end labeled at their 5′ termini with [γ-^{32}P]ATP and polynucleotide kinase and fractionated by two-dimensional, high-voltage paper electrophoresis according to the method of Sanger and Brownlee (17). The derived oligonucleotides labeled with uniformly good efficiencies, thus eliminating the uncertainty of differential incorporation into the intact molecules. All spots on the autoradiogram were quantitated by radioactive content, and a selected collection of these oligonucleotides were sequenced. Table 1 displays the sequences and relative abundances of some of the 5S-derived oligonucleotides. By the criterion of specific oligonucleotide abundance, the Octopus Spring archaebacterium possesses about half the extractable 5S rRNA, the remainder being divided nearly equally between the two thermus-like rRNAs. Importantly,

Stahl et al. Results

there are no dominant oligonucleotides unaccounted for in the nucleotide sequences of the three isolated 5S rRNAs.

Stahl et al. Results

TABLE 1. Sequence listing and quantitation of selected RNase T_1-generated oligonucleotides[a]

	Oligonucleotide	Unique 5S rRNA			cpm	Relative abundance
		I	II	III		
1.	CCCCG	1		1	13,329	0.75
2.	AAACCG	1			36,775[b]	2.1
	AACCCG	1	1	1		
3.	AAACACCCG		1	1	26,748	1.5
4.	UCCCG		1		15,320[b]	0.79
	CCUCG			1		
5.	CUACG	1	1		22,956	1.8
6.	CUAAG	1	1		26,716	1.5
7.	CAAUAG	1	1		19,227	1.1
8.	ACCAUAG					
9.	CCCUCCAG				15,551	0.9
10.	CCCCUAACCCG	1	1		9,827	0.6
11.	UUAAG			1	26,668	1.5
12.	AUACUG				14,062	0.8
13.	UCCCAUUCCG			1	6,202	0.37
14.	UCUCUUCAG	1	1		13,322	0.75
15.	UUUCCAUUCCG				6,248	0.35

[a] See Fig. 1C. The quantitation has been normalized in the relative abundance column to reflect the relative proportion of individual oligonucleotides and by inference the parent 5S rRNAs.

[b] Figure applies to both corresponding nucleotides.

DISCUSSION

The use of biochemical markers to define microbial community structure has been largely restricted to cell envelope and membrane components (lipids, carbohydrates, etc.) (15). These markers often offer some categorization of populations, but they are limited to available biochemical descriptions. Novel constituents, representing unknown microbes, cannot be interpreted. We can say with good confidence that the rRNAs are ubiquitous, and a phylogenetic framework based on their evolutionary drift in structure is established. Thus, the analysis of rRNA nucleotide sequences offers a powerful and generally applicable description of any life form. The precision with which an organism of unknown phylogenetic affiliation can be ordered relative

DISCUSSION

The use of biochemical markers to define microbial community structure has been largely restricted to cell envelope and membrane components (lipids, carbohydrates, etc.) (15). These markers often offer some categorization of populations, but they are limited to available biochemical descriptions. Novel constituents, representing unknown microbes, cannot be interpreted. We can say with good confidence that the rRNAs are ubiquitous, and a phylogenetic framework based on their evolutionary drift in structure is established. Thus, the analysis of rRNA nucleotide sequences offers a powerful and generally applicable description of any life form. The precision with which an organism of unknown phylogenetic affiliation can be ordered relative

Stahl et al. Discussion

to characterized isolates is limited only by the availability of the corresponding rRNA sequences of the isolates. There are now about 250 5S rRNA sequences available, spanning all three primary kingdoms (eubacteria, archaebacteria, and eucaryotes). The 5S rRNAs extracted from environmental samples, therefore, can be related at some level to those of better characterized microorganisms.

Stahl et al. Discussion

One intent of studying microorganisms in their natural environs is to discern their contributions to geochemical transformations. In principle, phylogenetic placement can be interpreted in terms of physiology. However, our ability to describe physiology based on a phylogenetic assignment is at this time limited both practically and conceptually. The precision with which we can describe the members of natural populations is limited practically by the availability of well-described, phylogenetically close microorganisms in our reference sequence collection. At present, our collection of rRNA sequences is quite limited, and we do not yet understand the phylogenetic significance of important physiological attributes. For example, although sulfur-based chemolithotrophy has frequently been used to define traditional taxonomic hierarchies, heterotrophs and sulfur-based autotrophs may be rather close phylogenetic relatives (20). However, as rRNA sequences are added to our reference collection, we will be better able to discern phylogenetically significant physiological attributes and also to describe more precisely those organisms encountered in natural settings.

Stahl et al. Discussion

One intent of studying microorganisms in their natural environs is to discern their contributions to geochemical transformations. In principle, phylogenetic placement can be interpreted in terms of physiology. However, our ability to describe physiology based on a phylogenetic assignment is at this time limited both practically and conceptually. The precision with which we can describe the members of natural populations is limited practically by the availability of well-described, phylogenetically close microorganisms in our reference sequence collection. At present, our collection of rRNA sequences is quite limited, and we do not yet understand the phylogenetic significance of important physiological attributes. For example, although sulfur-based chemolithotrophy has frequently been used to define traditional taxonomic hierarchies, heterotrophs and sulfur-based autotrophs may be rather close phylogenetic relatives (20). However, as rRNA sequences are added to our reference collection, we will be better able to discern phylogenetically significant physiological attributes and also to describe more precisely those organisms encountered in natural settings.

Stahl et al. Discussion

We selected Octopus Spring for this first analysis of a mixed microbial habitat by rRNA sequences because it offers a vigorous community of apparently limited complexity. Also, although Brock and colleagues (3, 4) characterized to some extent the in situ activity of the community, characteristic microorganisms had not been obtained in culture. The temperature of the source waters is a nearly constant 91°C and matches the temperature optimum for incorporation of radioisotopically labeled substrates (leucine, lactate, aspartate, phenylalanine, thymidine) (3, 4). The resident microbial population therefore appears to be optimally adapted for growth at the ambient temperature. The

slightly alkaline waters are low (ca. 5 μM) in sulfide, and addition of sulfide inhibits the uptake of labeled substances (4), so it seems unlikely that sulfide serves the metabolic needs of the community. These earlier studies also demonstrated that bubbling with CH_4, N_2, CO_2, or air does not stimulate uptake of labeled substrates. Consistent with our observation of an archaebacterium in the Octopus Spring population are transmission electron micrographs that reveal microorganisms lacking the typical eubacterial cell wall structure (3).

Stahl et al. Discussion

The two characterized eubacterial 5S rRNAs obtained from Octopus Spring most closely resemble the 5S rRNAs of the two representatives of the genus *Thermus* (*T. aquaticus* and *T. thermophilus*) in our reference collection. The members of this genus described thus far have a maximum temperature for growth of about 85°C (13, 14). The rudiments of the *Thermus* phylogeny emerging from 5S sequence comparisons suggest it to be of relatively ancient origin. The *Thermus* line of descent so far defined is at least as ancient as any of the three primary divisions within the purple photosynthetic bacteria. Over such broad phylogenetic distances, credible predictions of physiology cannot be made at this time. Members of the genus *Thermus* are formally described as filamentous, nonsporulating, nonmotile, obligately aerobic, oligotrophic heterotrophs (5). Thermus-like filaments have been observed in flowing thermal environments with organic concentrations as low as 2 ppm (2 μg/liter) and are conspicuous on contact slides immersed in Octopus Spring source waters. Contact slides immersed with the fiber mattings used here for biomass accumulation contained roughly half thermus-like filaments, the remainder being nondescript rods resembling the above-mentioned, probable archaebacterium. This observation is consistent with the quantitation derived from the 5S rRNA oligonucleotide content. Although we have no measured value for the organic contents of Octopus Spring, dense microbial mats (primarily *Chloroflexus* and *Synechoccus* species [3]) thriving in the cooler shallows surrounding the source waters probably supply adequate organics to sustain oligotrophic growth. The *Thermus* relatives encountered in the Octopus Spring source may, therefore, adhere to the physiological profile of this genus.

Stahl et al. Discussion

genus.

The Octopus Spring archaebacterium which accounts for about 50% of the isolatable 5S rRNA is not closely related to any organism in our sequence collection; the list of available archaebacterial 5S rRNA sequences is scant. It is, however, clearly affiliated with the branch of the archaebacteria designated the sulfur-metabolizing (dependent) branch, a physiologically diverse assemblage. Representatives of the genus *Sulfolobus* grow heterotrophically or as sulfur-oxidizing autotrophs. Other members of this branch grow by sulfur-dependent respiration of hydrogen or organic compounds, either heterotrophically or autotrophically. The discovery reported here of a representative of this assemblage in an alkaline source is noteworthy. To our knowledge, there so far has been no successful isolation of a member of the sulfur-metabolizing archaebacteria from an alkaline environment, and until recently, this line of descent was termed the thermoacidophilic branch of the archaebacteria. With the recognition that many representatives of this collection continue to grow or grow optimally near neutrality, the alternative designation proposed by Stetter and Zillig (in press), based on sulfur utilization, might seem a more comprehensive description. However, we point out that most if not all methanogenic archaebacteria so far characterized have the capacity to substitute sulfur as an electron acceptor for the oxidation of hydrogen (22). Thus, the designation of sulfur utilization may not address a fundamental biochemical difference between these deep branchings of archaebacteria. Instead, the broad distribution of sulfur-metabolizing capabilities may reflect the ancient origin of sulfur-based energetics. Octopus Spring is a low-sulfide (ca. 5 μM) environment, suggesting that this archaebacterium, like the *Thermus* relatives, is growing heterotrophically.

Stahl et al. Discussion

like the *Thermus* relatives, is growing heterotrophically.

The Octopus Spring source pool microbial community is relatively simple in major constituent numbers and thus is amenable to the direct isolation and fractionation of the 5S rRNAs. The practical limitation on this methodology by community complexity has not yet been established, although the fractionation of 10 or so unique species of 5S rRNA is easily within the range of analysis. However, an alternative approach which does not seem to have limitations of population complexity involves the direct cloning of the rRNA genes from naturally available DNA. Individual population members are represented by unique clones, so that the isolation of unique rDNAs (population members) requires only the sorting of these clones. In concert, these techniques should be generally applicable to questions regarding microbial ecology.

Stahl et al. Discussion

The Octopus Spring source pool microbial community is relatively simple in major constituent numbers and thus is amenable to the direct isolation and fractionation of the 5S rRNAs. The practical limitation on this methodology by community complexity has not yet been established, although the fractionation of 10 or so unique species of 5S rRNA is easily within the range of analysis. However, an alternative approach which does not seem to have limitations of population complexity involves the direct cloning of the rRNA genes from naturally available DNA. Individual population members are represented by unique clones, so that the isolation of unique rDNAs (population members) requires only the sorting of these clones. In concert, these techniques should be generally applicable to questions regarding microbial ecology.