Selective cloning of 16S rRNA molecules to describe naturally occurring microorganisms
by Roland Weller

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Microbiology
Montana State University

Abstract:
Microorganisms, and among those especially the bacteria, seldom have morphological characteristics which would allow a quick and simple identification. The most commonly used technique to assess the species composition of natural microbial communities relies on culturing the organisms, and subsequent biochemical or physiological characterization. Unfortunately, microorganisms have very diverse and very rigid growth requirements. Unless their environmental conditions are exactly reproduced, the organisms will not grow in culture. Thus a large fraction of the naturally occurring microorganisms might not have been isolated by culture-dependent approaches and the diversity of microorganisms might be largely unknown. I have developed a novel technique allowing identification of microorganisms from natural habitats directly, circumventing the inherently biasing nature of culture-dependent techniques. The new approach makes use of a unique molecule which is found in every living cell. This molecule, the 16S ribosomal RNA (rRNA), can be used for the identification and classification of the organisms. The 16S rRNA sequences retrieved from a natural community are compared to 16S rRNA sequence collections. Sequences which do not match sequences from previously isolated organisms indicate the presence of organisms which have escaped isolation by culture-dependent methods. Further analysis of these 16S rRNA sequences from uncultured community members allows the elucidation of their phytogeny, i.e. their relationship to other known microorganisms. The 16S rRNA sequences were selectively retrieved by cloning of cDNA, synthesized from RNA extracted from the community. Using a well-studied hot spring community as a model system, it was demonstrated that the formerly isolated species comprise only a fraction of the community. The presence of several unknown community members was revealed. Phylogenetic analysis placed the retrieved 16S rRNA sequences into the known eubacterial phyla. Ten analyzed sequences represented seven unique sequence types. Of these three sequences originated from cyanobacteria, one from a green nonsulfur bacterium, and the remaining sequences are possibly from two spirochetes and a proteobacterium. This work confirms the suspicion of many researchers in the field that our knowledge of the naturally occurring microflora is rather limited. The definite proof of the presence of uncultured organisms, together with information derived from the use of oligonucleotide probes directed against 16S rRNA for in situ hybridization, could eventually lead to the isolation of novel organisms.

SELECTIVE CLONING OF 16S rRNA MOLECULES TO DESCRIBE NATURALLY OCCURRING MICROORGANISMS

by

Roland Weller

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

June 1990

APPROVAL

of a thesis submitted by

Roland Weller

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for the submission to the College of Graduate Studies.

11 May 1990
Date

Chairperson, Graduate Committee

Approved for the Major Department

11 May 1990
Date

Norman D. Reed
Head, Major Department

Approved for the College of Graduate Studies

June 5, 1990
Date

Henry L Parsons
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute copies of the dissertation in and from microfilm and the right to reproduce and distribute by abstract in any format."

Signature R. Weller

Date 6/4/90

ACKNOWLEDGMENTS

I wish to acknowledge all the people who have made possible this long and sometimes strenuous but most of all fun endeavour in the world of science. Most of all I express my thanks to my wife Jennifer, who endured endless discussions about scientific problems, helped to develop the concept of the hexamer priming approach, and assisted in the isolation of ribosomes. Next I would like to thank my dear parents, who have supported me in every aspect of life and during many years of financial hardship. I am grateful to my advisor Dr. David M. Ward who was very patient with a not so traditional graduate student. I thank Mary Bateson, Dr. Clifford Bond, and Linda Mann for excellent advice throughout the work. I would further like to thank the Pacific Northwest Laboratories for a Department of Energy Research Fellowship awarded for the last year of this study.

This work was supported by grants BSR-8506602, BSR-8818358, and BSR-8907611 from the National Science Foundation.

TABLE OF CONTENTS

TABLE OF CONTENTS continued

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Microorganisms, and among those especially the bacteria, seldom have morphological characteristics which would allow a quick and simple identification. The most commonly used technique to assess the species composition of natural microbial communities relies on culturing the organisms, and subsequent biochemical or physiological characterization. Unfortunately, microorganisms have very diverse and very rigid growth requirements. Unless their environmental conditions are exactly reproduced, the organisms will not grow in culture. Thus a large fraction of the naturally occurring microorganisms might not have been isolated by culture-dependent approaches and the diversity of microorganisms might be largely unknown. I have developed a novel technique allowing identification of microorganisms from natural habitats directly, circumventing the inherently biasing nature of culture-dependent techniques. The new approach makes use of a unique molecule which is found in every living cell. This molecule, the 16S ribosomal RNA (rRNA), can be used for the identification and classification of the organisms. The 16S rRNA sequences retrieved from a natural community are compared to 16S rRNA sequence collections. Sequences which do not match sequences from previously isolated organisms indicate the presence of organisms which have escaped isolation by culture-dependent methods. Further analysis of these 16S rRNA sequences from uncultured community members allows the elucidation of their phylogeny, i.e. their relationship to other known microorganisms. The 16S rRNA sequences were selectively retrieved by cloning of cDNA, synthesized from RNA extracted from the community. Using a well-studied hot spring community as a model system, it was demonstrated that the formerly isolated species comprise only a fraction of the community. The presence of several unknown community members was revealed. Phylogenetic analysis placed the retrieved 16S rRNA sequences into the known eubacterial phyla. Ten analyzed sequences represented seven unique sequence types. Of these three sequences originated from cyanobacteria, one from a green nonsulfur bacterium, and the remaining sequences are possibly from two spirochetes and a proteobacterium. This work confirms the suspicion of many researchers in the field that our knowledge of the naturally occurring microflora is rather limited. The definite proof of the presence of uncultured organisms, together with information derived from the use of oligonucleotide probes directed against 16S rRNA for in situ hybridization, could eventually lead to the isolation of novel organisms.

INTRODUCTION

Description of Natural Microbial Communities

The field of microbial ecology suffers from the lack of methods for the description of a microbial community in terms of its species composition as well as the numerical importance of the community members. Unlike plants or animals in macrobiotic ecosystems, microorganisms, with few exceptions, do not show distinctive morphological characteristics which would allow identification of the species. Microscopic observations can be helpful in some instances but in general are only used to confirm or strengthen a conclusion about the presence or dominance of a microorganism. Robert Koch (1843-1910) was the first to realize the importance of a pure culture of a microbe in order to characterize the organism biochemically. Ever since, cultivation and analysis of the biochemical potential of microorganisms has been the preferred method for their identification and classification. Sergei Winogradsky (1858-1953) refined this principle for ecological purposes by including selective reagents into the culture medium to enrich particular physiological types of microorganisms. This strategy of the "selective enrichment culture" is still the most widely used approach to obtain information about the composition of microbial communities.

In 1987, Brock (10) attempted a critical evaluation of the field:

> "During the past twenty years, the real field of ecology, macroecology, that is, has grown up and matured. The field of microbiology has changed beyond all recognition. But microbial ecology?"

He continues:

> "...,many studies that pretend to be ecological are still using antiquated, discredited, or meaningless methods. Many studies are unfocused, or do not deal with important questions."

In his detailed look at the current state of microbial ecology, his major emphasis centers around the speculation that culture-dependent methods cannot retrieve all or even the most important community members. Determination of the full species diversity of a microbial community by the enrichment of

physiological groups remains impossible. The reasons for this are manyfold. In order to reproduce an organisms' ecological niche, the investigator must be able to preconceive all the nutritional requirements of the organism. Not only the physiological requirements of the cells are important but also their physiological state. A viable but nonculturable stage in the life cycle of bacteria has commonly been observed (59). On the other hand, dormant cells or cells not active in a habitat, into which they might have been introduced by accident, can be grown on a suitable medium leading to the erroneous conclusion that the organism is an important, active community member.

Many other microbiologists have pointed out that our view of the microbial world is heavily biased by the culture-dependent methodology employed in the study of microbial habitats (1, 58, 73, 80, 96). To be able to understand the function of a complex ecosystem and the interactions among community members, the structure of the ecosystem must be defined. Thus one of the most important questions in microbial ecology still unanswered is: "Who lives and prospers in a given microbial community?". The conclusion of a committee of microbiologists and microbial ecologists (89) was:

> "...ecologically relevant characterization of the members of complex bacterial populations requires the identification of [new chemotaxonomic markers] in a burgeoning field of biochemical/molecular/genetic research."

Culture-independent Assessment of Natural Microbial Communities

Knowing about the shortcomings of the culture-dependent approaches, researchers have developed several alternative techniques using biomarkers to assess the community structure of microbial ecosystems. A biomarker is a molecule or a set of molecules which is characteristic for a specific organism or a group of organisms. Lipids for example have been used to infer the presence of microbial groups in selected habitats (55, 81, 83). A lipid spectrum from a natural population can indicate the presence of a community member provided it has been demonstrated that these particular compounds are exclusively synthesized by the organism in question. Since we suspect that only a small percentage of the extant microorganisms is in culture it is currently hard to prove that lipid biomarkers come from one particular source organism. Thus the presence of lipid biomarkers must

be interpreted very carefully and considered group-specific at most. Cell surface antigens constitute another class of biomarkers. Antibodies against these markers can be species-specific (14) and have been used successfully in ecological studies (39, 74). The disadvantage of this approach is the necessity to have the organism in pure culture in order to raise antibodies for a species-specific identification. Again we see that only organisms can be detected whose presence is preconceived by the investigator.

A biomarker that is species-specific, present in all organisms, and can be identified without prior culturing of the organisms would be desirable for an unbiased look at microbial ecosystems. A class of molecules which may come close to meeting these requirements resides in the protein-synthesizing machinery of every living cell. Ribosomal RNAs (rRNA) are an essential and functional component of the ribosome. Because of functional constraints some regions of the rRNA molecule have changed extremely slowly, while other regions have evolved much more rapidly (29). Due to the fast mutation in the variable regions most procaryotic species investigated have a unique rRNA sequence which can be used to identify the organism. The more conserved regions are very useful in phylogenetic analysis (22, 97) and for the alignment of the sequences which have to be compared in a position-by-position fashion (47). Another advantage of rRNA biomarkers is the fact that homologous molecules are easily identified by their sizes. There are small ribosomal subunit [SSU] rRNAs or 16S-like rRNAs, and large subunit [LSU] rRNAs or 23S-like and 5S-like rRNAs. The rRNA genes (and the products thereof) also seem to be free of the artefacts of lateral gene transfer (68). All the above holds true for all the rRNAs including the 5S, the 16S, and the 23S rRNA (in procaryotes).

The general approach for using rRNA sequences in ecological studies has been developed by Pace and colleagues (Figure 1.) (49, 54). A database of rRNA sequences from organisms in pure culture is established by sequencing isolated rRNA with the enzyme reverse transcriptase (36). rRNA sequences from the community members are retrieved and separated by cloning nucleic acids which code for the sequence. The sets of sequences can then be compared to reveal organisms which have formerly been isolated from the community and to demonstrate the presence of organisms not yet in

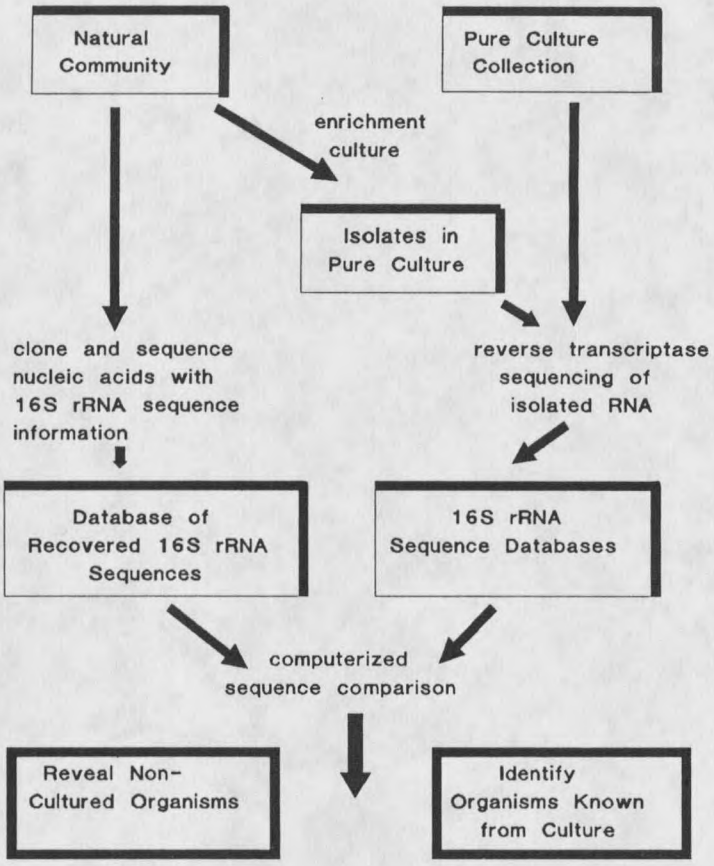

Figure 1. Overview of the 16S rRNA biomarker approach. rRNA sequences obtained from the natural community by cloning (left side) are compared to a collection of 16S rRNA sequences (database) from organisms in pure culture. Known organisms are identified through sequence identity.

culture. The sequence information of the non-cultured organisms can be used in phylogenetic analysis or to construct specific oligonucleotide probes for autecological studies (16, 25, 71). In addition such sequences will aid in the development of a universal phylogeny since phylogenetic analysis has so far been limited to sequences from organisms in laboratory culture. The obvious bias against organisms which live in a symbiotic or parasitic relationship, and are therefore more difficult to culture, can be overcome by the inclusion of novel sequences retrieved from un-cultured species (53). At a time when sequencing of long RNA molecules was still a problem the short 5S rRNA molecule found application as a biomarker in ecological investigations (54, 70, 72). The emphasis has now been shifted to the

16S rRNA, since longer sequences allow statistically more valid comparisons (54). The 23S rRNA would increase the power of resolution even further but due to its length full sequence determination remains difficult and the number of known and compiled sequences (i.e. database) is currently very limited.

Two studies of relatively simple communities demonstrate the power of the use of the 16S rRNA biomarker. In one investigation of sulfur-oxidizing endosymbionts of marine invertebrates the 16S rRNA was directly sequenced with reverse transcriptase from bacteria purified from the invertebrate gut tissue (18). The approach was successful since in each case one endosymbiont was so dominant that rRNA sequencing resulted in a nonambiguous sequencing gel pattern. The phylogenetic analysis of the partial sequences resulted in the placement of the endosymbionts in two clusters within the gamma-subdivision of the phylum proteobacteria (formerly the gamma-group of the "purple bacteria" [69]). In the second study the 16S rRNA genes of endosymbiotic bacteria were cloned after restriction endonuclease digestion of DNA isolated from whole pea aphids and excision of the DNA bands which hybridized with a eubacterial 16S rRNA probe (77). Two different 16S rRNA gene-bearing recombinants were recovered from two procaryotic intracellular symbionts which were again shown to belong to the gamma-subdivision of the proteobacteria.

The original approach to the analysis of complex communities was to retrieve 16S rRNA genes by shotgun cloning DNA obtained from the environment (54). Even though this proposal was put forth in 1986, no data have yet been published of shotgun-cloned 16S rRNA sequences. Only one paper reports the isolation of DNA suitable for the construction of recombinant libraries from a more complex environmental sample (23); the recovery of rRNA gene-bearing recombinants is not mentioned. One of the major limitations of shotgun-cloning is the need to screen thousands of recombinants to find a representative number of 16S rRNA genes, as these comprise only a small percentage of the total genes in such a library (40). My preliminary work may demonstrate this point (Table 1). Shotgun cloned DNA libraries were prepared with DNA isolated either from a natural microbial mat community (Nymph Creek Cyanidium mat) or from a pure culture of the

archaebacterium Thermoplasma acidophilum. Extensive hybridization screening finally lead to the recovery of a single 16S rRNA gene bearing recombinant clone from the T. acidophilum library. A 16S rRNA gene bearing recombinant from the natural community could not be found after screening of approximately 100,000 recombinants. Shotgun cloning is further complicated by the fact that the genes of abundant community members will outnumber the genes of other community members. A more selective cloning of only the 16S rRNA sequences would definitely increase the range of applications of this very powerful biomarker method, especially allowing the analysis of complex microbial communities with a large number of unique community members.

Table 1. Screening of shotgun cloned DNA libraries

DNA library	Approximate number of recombinants screened	16S rDNA-bearing recombinants recovered
T. acidophilum	30,000	1
Nymph Creek Cyanidium mat	100,000	0

(Weller and Ward, unpublished results).

The development of the polymerase chain reaction (PCR) (61) has allowed the million-fold amplification of specific genes. Amplification of the gene of interest before cloning can substantially reduce the screening effort required to locate the right clones. PCR has been used to recover 16S rRNA genes from organisms where the recovery of large amounts of DNA for cloning is difficult or impossible (43). Amplification of 16S rRNA genes from a marine picoplankton community has also been achieved (24). The fact that procaryotic organisms have different and usually unknown numbers of rRNA coding regions in conjunction with the non-linear amplification process, may make a quantitative analysis of the community structure with the PCR-based approach impossible.

I have decided to develop another selective cloning strategy, based on the synthesis of complementary DNA (cDNA) from 16S rRNA for the following reasons:

1) Rather than amplifying the rRNA gene in vitro, the method makes use of the natural multiple transcription of the gene in the cell during the synthesis of ribosomes.

2) Each rRNA template can be used only one time for the synthesis of cDNA, thus preserving a record of the abundance of the rRNA sequence type. This abundance of 16S rRNA in a community is a function of the numerical importance of the organisms and their growth rate, and thus might give an indication of the protein synthetic capacity of the species.

An outline of the cloning strategy is shown in Figure 2. The method is based on cDNA synthesis from rRNA (35). Specific priming of cDNA synthesis requires only the isolation of total RNA from a cell lysate. The selectivity of the method is achieved by the use of an oligonucleotide complementary to a highly conserved region (36) within the 16S rRNA molecule as a primer. Only molecules which possess this region will serve as templates for the synthesis of cDNA. This specific priming is the basis for the selective retrieval of the 16S rRNA biomarker sequence. Under optimal conditions this results in the synthesis of a cDNA strand (16S rcDNA or rcDNA) which preserves the sequence of more than 90% of each 16S rRNA molecule. The second strand of the cDNA is made against the first strand by the use of Escherichia coli DNA polymerase I. Next a homopolymer tail consisting of deoxycytidine is synthesized onto the double-stranded 16S rcDNA. The rcDNA is annealed to a cloning vector with a complementary oligo-(dG) tail and transformed into E. coli. The use of cloning vectors with antibiotic resistance genes allows only recombinant cells to form colonies on a medium with the appropriate antibiotic. The formation of a PstI restriction endonuclease recognition site upon insertion of the oligo-(dC) tailed rcDNA makes a size analysis of the cloned rcDNAs possible.

Even though the approach described above does work very well several theoretical and practical considerations lead me to develop a second protocol for the synthesis of 16S rcDNA.

1) Posttranscriptional modifications of nucleotides of 16S rRNA molecules close to the priming region, or a high degree of secondary structure, might prevent efficient readthrough of the reverse transcriptase resulting in short rather than long rcDNA sequences.

2) It has not been established whether small subunit rRNAs from all natural occurring organisms do contain the conserved region required for specific priming.

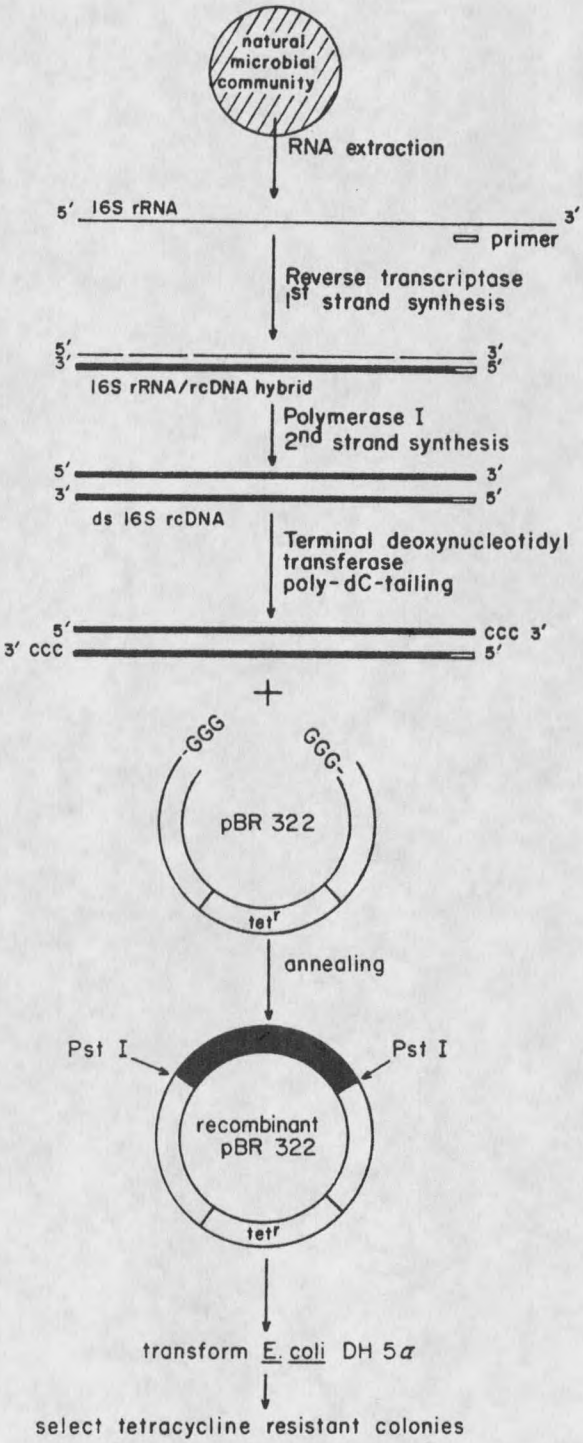

Figure 2. Outline of the 16S rcDNA cloning approach for the selective recovery of 16S rRNA sequences. ds, Double-stranded (from reference 90).

3) In mitochondria, which could be considered highly specialized endosymbionts, noncontinuous SSU rRNAs have been discovered (28). It might very well be the case that other symbiotic organisms have evolved a similar SSU rRNA structure. This is an important consideration for the broad application of the cDNA approach to a wide variety of natural microbial communities.

In the second protocol random hexanucleotides are used to initiate cDNA synthesis at random along all molecules present in the reaction (60). The retrieved rcDNA molecules are on average smaller than the longest rcDNAs primed with the specific primer. This disadvantage however might well be outweighed by the unbiased recovery of SSU rRNA sequences. To restrict cDNA synthesis to 16S-like rRNAs the RNA has to be extracted from small ribosomal subunits (Figure 3).

The Octopus Spring Cyanobacterial Mat Community

To develop and test these molecular approaches to evaluating the composition of a microbial community it was decided to investigate a thoroughly-studied hot spring cyanobacterial mat community, found in Octopus Spring. Geothermal habitats are very stable throughout the seasons which might result in only minor fluctuations in the species composition of the microbial ecosystem. Alkaline geothermal springs are much more stable than acidic ones (8) and thus lend themselves better to long term investigations. The relatively high temperatures exclude higher life forms and limit the number of species adapted to the environment (9). No grazing eucaryotes are observed in the Octopus Spring mat and no evidence points at the presence of eucaryotic microbes.

Octopus Spring, in the earlier work called pool A, is located in the White Creek area in Yellowstone National Park (YNP), Wyoming (19). At the source the geothermally heated, slightly alkaline (pH 8.1) water emerges with a temperature of 92°C. The water cools rapidly in the shoulder areas of the pool and at temperatures between 42 to about 72°C a dark green cyanobacterial mat develops. The species composition as it has been determined by traditional methods (enrichment cultures and microscopy) is given in Table 2. The mat is laminated with an uppermost green layer (0.2-1.0 mm) where mostly phototrophic mat-building organisms are thought to thrive (19). Synechococcus lividus, a cyanobacterium, has been identified in the mat by microscopic observation

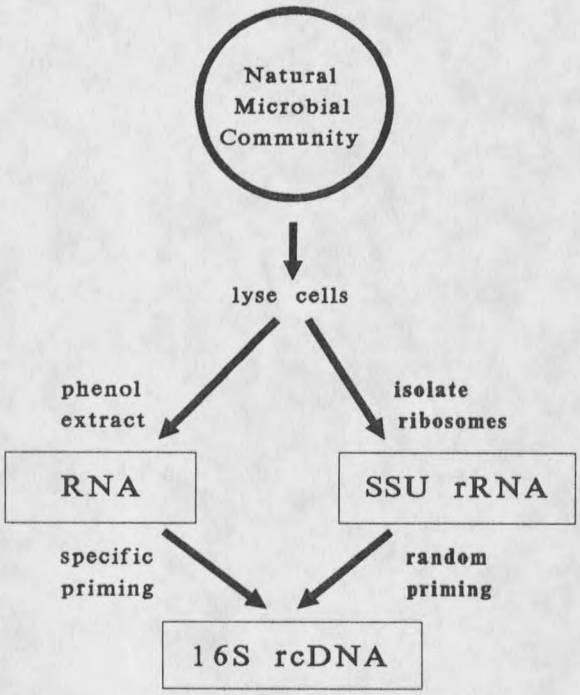

Figure 3. Comparison of random priming and specific priming approaches for selective cDNA synthesis from 16S rRNA. While total cellular RNA can be used in the presence of a specific primer which brings about the selectivity, 16S-like rRNA has to be obtained from the small ribosomal subunit to assure exclusive synthesis of 16S rcDNA when random priming is employed.

of its fairly "unique" morphology and characteristic autofluorescence of the cells' chlorophyll a. The cell density has been estimated, again by microscopic means, to be 10^{10} cells per milliliter (5). A green nonsulfur bacterium, Chloroflexus aurantiacus, stabilizes the mat due to its filamentous morphology. Observations made with a Chlorofexus-specific immunological probe (74) indicate that the abundance of C. aurantiacus is also quite high. The same study indicates that Chloroflexus is not the sole filamentous organisms in the photic zone of the mat, since filaments of different widths are observed and not all of these react with the antibody. One strain of C. aurantiacus (Y-400) has been isolated from the study site (M. T. Madigan, personal communication). In the deeper layers (1-4 mm) of the cyanobacterial mat mostly moribund S. lividus (as judged by absence of autofluorescence in the cells), C. aurantiacus filaments, and the aerobic planctomycete bacterium Isosphaera pallida (formerly misclassified as a cyanobacterium, Isocystis spp.) are reported. The latter can be recognized

Table 2. Microorganisms identified in the Octopus Spring cyanobacterial mat.

Organism	Physiological Type (Phylogenetic Type)	Abundance
Synechococcus lividus	Photosynthetic Cyanobacterium (Cyanobacterial Phylum)	High[a] (ca. $10^{10}ml^{-1}$)
Chloroflexus aurantiacus	Photosynthetic Bacterium (Green Nonsulfur Phylum)	High[a]
Thermus aquaticus	Aerobic Heterotrophic Eubacterium (Thermus/Deinococcus Phylum)	Unkown[b]
Isosphaera pallida	Aerobic Heterotrophic Eubacterium (Planctomycete Phylum)	High[a]
Thermomicrobium roseum	Aerobic Heterotrophic Eubacterium (Green Nonsulfur Phylum)	Moderate[c]
Thermobacteroides acetoethylicus	Anaerobic Fermentative Eubacterium (Gram Positive Phylum)	High[d] (ca. $10^{7}ml^{-1}$)
Thermoanaerobium brockii	Anaerobic Fermentative Eubacterium (Gram Positive Phylum)	Low[e]
Thermoanaerobacter ethanolicus	Anaerobic Fermentative Eubacterium (Gram Positive Phylum)	Unknown[b]
Clostridium thermohydrosulfuricum	Anaerobic Fermentative Eubacterium (Gram Positive Phylum)	Low[e] ($<10^{3}ml^{-1}$)
Clostridium thermosulfurogenes	Anaerobic Fermentative Eubacterium (Gram Positive Phylum)	Low[e] ($<10^{3}ml^{-1}$)
Thermodesulfobacterium commune	Sulfate-reducing Eubacterium (Novel Phylogeny)	Unknown[b]
Methanobacterium thermoautotrophicum	Methanogenic Bacterium (Archaebacterium)	High[d] (ca. $10^{7}ml^{-1}$)

[a] Direct microscopic observation
[b] Enriched from undiluted sample
[c] Suggested by lipid analysis (83)
[d] Enriched from highly diluted sample
[e] Enriched from low dilution samples

(modified from [84]).

microscopically due to its "unique" morphology (19). The only other aerobic organism, Thermus aquaticus, has been cultured from Octopus Spring from an undiluted sample (11). Since no attempt has been made to grow T. aquaticus from a diluted sample, which would allow an estimate of its abundance (by dilution to extinction), its numerical importance is unclear.

More is known about the anaerobic decomposition of the mat (82) and the organisms carrying out these processes. Thermobacteroides acetoethylicus was isolated from a high dilution sample (6, 101), indicating that this obligately anaerobic fermentative bacterium may be an important decomposer. Other fermentative bacteria, like Thermoanaerobium brockii, Thermoanaerobacter ethanolicus, Clostridium thermohydrosulfuricum, and C. thermosulfurogenes might be less important, since they have only been isolated by low-dilution enrichment (63, 94, 95, 101, 103). The fermentation products of the anaerobic chemoorganotrophic bacteria may be used by either acetogenic bacteria, sulfate-reducing bacteria, or methanogenic bacteria (88). Sulfate reduction seems to play a role in the decomposition of the mat (19), even though the concentration of sulfate is low (84), and methanogenesis should dominate sufur reduction (82). The only sulfate reducing bacterium known to inhabit the mat is Thermodesulfobacterium commune (102), which has been enriched from undiluted samples. A methanogenic archaebacterium, Methanobacterium thermoautotrophicum, has been isolated from the Octopus Spring mat (62) and methanogens have been detected microscopically due to the green autofluorescence of the coenzymes characteristic for this group of organisms. Since M. thermoautotrophicum has been isolated from a high-dilution sample the abundance is estimated to be on the order of 10^7 cells per milliliter (62, 101).

The presence of several other organisms in the mat has been suggested. Doemel et al. (19) report the presence of a filamentous cyanobacterium similar to Pseudoanabaena. The photosynthetic bacterium "Heliothrix oregonesis" is another filamentous organism which could be an Octopus Spring community member (84). This organism does not grow axenically and has only been obtained as a co-culture with I. pallida. 5S rRNA sequence analysis suggested a relationship between "H. oregonensis" and C. aurantiacus (56). Lipid biomarkers, specifically mono- and bicyclic biphytanyl ethers, indicate

the possible presence of a sulfur-dependent archaebacterium (81). A nonphotosynthetic green nonsulfur bacterium (defined phylogenetically [50]), Thermomicrobium roseum, synthesizes very characteristic long-chain diols, which have been detected in the mat (83). None of these bacteria have yet been cultured from this habitat.

There are other reasons to study the Octopus Spring cyanobacterial mat. Thermophilic environments are of interest since they might harbor novel organisms which can be exploited in biotechnological processes (46, 90). Novel organisms are also of profound interest for the evolutionary disciplines (9, 53, 97). Geochemists and paleobiologists consider the laminated mats to be modern equivalents of the ancient, fossilized stromatolites (85).

In summary, the Octopus Spring cyanobacterial mat combines several features which make the microbial habitat an attractive model system. The community is very stable, and microbiologically well characterized, although several community members including the mat-building cyanobacterium have only been suggested on the ground of microscopic observations. Information about organisms isolated from the mat, or observed in the mat allowed our laboratory to construct a sequence database (3, 4, 86) against which we can compare sequences directly retrieved from the habitat. Once more reliable information about the composition of the Octopus Spring mat community has been gained with culture-independent approaches, our knowledge about physiological processes should allow a better interpretation and integration of all results from this community.

Objectives

1) The major goal was to develop culture-independent methods for the description of the species composition of microbial communities based on specific retrieval of 16S rRNA sequences.

2) The methods were used to test the hypothesis that traditional culture-dependent approaches have not completely described the microflora of a well-studied microbial community found in Octopus Spring, Yellowstone National Park. 16S rRNA sequences from pure cultured organisms, enriched from the cyanobacterial mat, were compared to 16S rRNA sequences directly retrieved from the community.

3) The phylogeny of the community members, whose sequences were recovered, was established.

4) The methods were also applied to the selective cloning of 16S rRNA sequences from an organism, thought to be a possible Octopus Spring cyanobacterial mat community member. This organism, "Heliothrix oregonensis", can currently only be grown in a coculture with the planctomycete bacterium Isosphaera pallida. This makes reverse transcriptase sequencing of its 16S rRNA, for inclusion in the sequence database, impossible.

MATERIALS AND METHODS

General Protocols

Neutral and alkaline gel analysis of the reaction products of cDNA synthesis, analysis of RNA on formaldehyde-agarose gels, and restriction enzyme digests were performed according to standard protocols (40). Several strategies which were followed to guard against RNase activity in the work with RNA are described by Blumberg (7).

Isolation and Purification of RNA

Collection of Samples and Lysis of Cells

Samples from Octopus Spring were taken from the shoulder area where the temperature is about 55°C. For the rcDNA libraries OS-I and OS-II the top 1 cm was collected. To collect mostly phototrophic community members only the top 2 mm were used for the library OS-III. Since all physiological processes occur in the top 5 mm (84), only the top 5 mm were collected for the libraries OS-V L and OS-VI L. For enzymatic lysis of the cells the mat sample was homogenized in the field in a Wheaton tissue grinder in lysis buffer (10 mM Tris [pH 7.6], 0.5 M NaCl, 1% sodium dodecyl sulfate [SDS], 30 mM ethylenediaminetetraacetic acid [EDTA]). Proteinase K (Boehringer Mannheim Biochemicals, Indianapolis, IN [BMB]) was added to a concentration of 60 µg/ml; the homogenate was incubated at 50°C for 20 minutes and then mixed with phenol-chloroform-isoamylalcohol (25:24:1) to stop all enzymatic activities. This mixture was placed on dry ice, transported to the laboratory, and stored at -70°C until needed.

Alternatively the sample was immediately frozen in liquid nitrogen, placed on dry ice, transported to the laboratory, and stored as above. For mechanical cell disruption the sample was thawed out, ground with mortar and pestle in either lysis buffer (for extraction of total cellular RNA) or ribosome buffer (for isolation of ribosomes, see below), and subjected to 2 cycles through the