FINAL PROJECT TECHNICAL PROGRESS REPORT

MICROBIAL COMMUNITIES IN SUBSURFACE ENVIRONMENTS:
DIVERSITY, ORIGIN AND EVOLUTION

RECEIVED
AUG 23 1999
O.S.T.

US Department of Energy Contract No. DE-FG02-90ER60989
Period: (9/1/96-present)

To Rensselaer Polytechnic Institute (Sandra A. Nierzwicki-Bauer)

July 4, 1999

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

I. PROJECT DESCRIPTION

We had requested supplemental funding to complete two studies that were initiated under our DOE grant project (DE-FG02-90ER60989 to Sandra A. Nierzwicki-Bauer). The first study involved characterization of the abundance, distribution and diversity of group I introns in subsurface isolates. This research was intended to further our understanding as to the origins and evolution of bacterial species. This work builds on and extends, our preliminary results obtained from a survey of 50 *Pseudomonads* from Savannah River Site isolates. Secondly, we proposed to continue our characterization of non-culturable bacteria using 16S rRNA sequence analysis on DNA extracted from samples from the GEMHEX interval at the Hanford site at in Washington. The purpose of this work was to further our understanding of the contribution of non-culturable bacteria to the total bacterial populations in sediments where little transport was believed to have occurred.

The project results in this final report were obtained using state-of-the-art techniques in molecular biology for the exploration of the long-term survival of bacteria in sediment communities. It had been previously hypothesized that bacteria isolated from these environments may be descendants of organisms present at the time of sediment deposition. Thus, these organisms may possess novel characteristics that have allowed them to survive in a stressed environment over millions of years. Unique characteristics of these organisms may make them useful for DOE bioremediation efforts. Furthermore, a more complete understanding of the genome organization of these organisms could provide insight regarding the evolution of organisms, with general implications for microbial ecology.

During the project reporting period our research addressed the following three hypotheses:

1. Microorganisms that have survived for extensive periods of time, by reduced metabolic activity and slow growth rates, will possess group I self-splicing introns, which may be ancient features of gene structure. Whereas, microorganisms that have been actively growing in surface soils will have streamlined genomes that lack group I self-splicing introns.
2. Phylogenetically related microorganisms that survived in isolation in deep subsurface sediments and that were derived from organisms associated with sediments as they were deposited millions of years ago, will possess intron sequences that are more closely related to one another, than to sequences from the same species found in sediments deposited at different times.

3. Microbial communities in saturated paleosols and lacustrine sediments that are considered confining layers, where it is unlikely that transport occurred, will consist of unculturable organisms that are the survivors or progeny of microorganisms associated with the original geological deposits.

II. MAJOR RESEARCH ACTIVITIES (FOR FINAL GRANT PERIOD)

A. Abundance, distribution and diversity of group I introns

- Complete survey of selected subsurface bacterial isolates (from the SMCC) for the presence of putative group I self-splicing introns.
- Determine the gene that any putative group I self-splicing introns reside within.
- Verify, via sequencing and sequence analyses, the identity of any putative group I self-splicing introns found during preliminary screening.

B. Characterization of non-culturable bacteria from GEMHEX intervals

- Carry out phylogenetic analyses of 16S rRNA clones from selected GEMHEX interval samples.
- Compare identities of microbial isolates versus those from the natural microbial community, without culturing.

III. PROJECT OUTPUT: MAJOR RESEARCH ACCOMPLISHMENTS

A. Abundance, distribution and diversity of group I Self-splicing introns

1. Survey for the Presence of Group I Self-Splicing Introns in Subsurface Isolates

An objective during the grant period has been the analysis of putative group I introns identified in subsurface stains isolated from the Savannah River Site (SRS). Because group I introns are believed to be ancient features of gene structure, the presence of these motifs in subsurface microorganisms may provide evidence for the ancient origin of subsurface microbial populations. Furthermore, the discovery and genetic analysis of group I introns from these strains will provide insight into the evolution of prokaryotic genomes.

Initially 50 *Pseudomonad* strains isolated from the SRS borehole were screened for the presence of group I introns using oligonucleotide probes designed in our laboratory. On the basis of this preliminary screening, thirteen strains yielded positive hybridization results and were therefore used in subsequent analyses for further characterization of the putative group I introns. The dot blot hybridization that identified these isolates is presented below.

20	27	38	52	53	54	106	132	137	138
141	190	198	205	222	251	252	253	259	267
272	280	287	301	307	310	326	329	333	441
442	447	473	474	521	528	578	581	590	608
624	625	628	631	638	641	651	666	698	715
2.0	0.2		2.0	1.0	0.5		10.0	5.0	2.5

µg DNA

A. tumefaciens

Ps. putida

E. coli

Figure 1. Screening of subsurface isolates for the presence of group I self-splicing introns. DNA extracted from subsurface isolates was applied to nylon membranes, in duplicate, using a dot blot apparatus following standard methods. Each dot contained 4 ug of DNA , with the exception of the control DNA that was applied as indicated. Each blot was hybridized with a P³² end-labeled oligonucleotide probe "Q": 5'-atg gca atc ctg agc caa gc-3', which targets the conserved Q-region for group I self-splicing introns.

2. Sequencing Group I Introns from Subsurface *Pseudomonad* Strains:

Initially, it was necessary to isolate and determine the nucleotide sequences of the putative group I introns of the subsurface bacteria that were identified as described above. This was essential to conclusively demonstrate the presence of these introns. Additionally, this information was needed so that comparisons could then be made to group I self-splicing introns previously reported for cyanobacteria, alpha-proteobacteria and eukaryotes. In order to accomplish this goal we took the approach of utilizing PCR

primers that were targeted against either the tRNA^{fMet}, tRNA^{Leu}(UAA), or tRNA^{Arg}(CCU) genes. These primer sets were chosen because they have previously been shown to contain a group I intron in at least one bacterial species. Thus, this seemed the most logical approach that would allow us to elucidate whether group I self-splicing introns are also present in these genes in subsurface bacteria. Additionally, if present in one of these genes, the PCR amplified product would contain the entire group I self-splicing intron sequence. In total twelve different primer sets were used in PCR amplification reactions: one for the tRNA^{fMet} gene, nine for the tRNA^{Leu}(UAA) gene and two for the tRNA^{Arg}(CCU) gene. The sequences for these primer sets are listed below in Table 1.

Table 1. Primers Used to Amplify Intron Sequences

Target Sequence	Primer	Primer Sequence
tRNA ^{fMet}	Forw.: SM1	5'-GCG GGG TAG AGC AGC CTG GTA CCT CGT CGG G-3'
	Rev.: SM2	5'-GCG GGA TCC GGA TTT GAA CCA CCG ACC TTC GGG-3'
tRNA ^{Leu} (UAA) ¹	Forw.: M5	5'-TGG (C/T)GA AAT (C/T)GG TAG ACG C(A/T)(A/G) CGG AC-3'
	Rev.: M3	5'-TGG GGA TAG AGG GAC TTG AAC CCT CAC GAT TTT TA-3'
	Forw.: B5	5'-TGG (T/C)GG AAT (A/T/C)GG (T/C)AG AC(A/G) C(A/T)(A/C/G/T) (A/C/G/T)GG A(C/T)T-3'
	Rev.: B3	5'-TGG T(G/C)(G/C) (G/C)(G/C)(A/G) (A/T/G)G(A/G) C(G/C)G GAC TTG AAC C(G/C)G (T/C)AC-3'
	Forw.: TL25	5'-GGG G(G/A)T (G/A)T ₆ G(C/T)G (A/G)AA T(C/T)G GTA GAC GC-3'
	Rev: TL23	5'-TGG GG(G/A) (C/T)(G/A)G (G/A)GG GAC TTG AAC CC(C/A/T) CAC G-3'
tRNA ^{Arg} (CCU)	F: Arg-5'	5'-GTC C(G/A/T)C GAT GCT CA(G/A) (C/T)A(A/G) GAT A-3'
	F: Arg-5'ext	5'-CC(G/A/T) CGA TGC TCA (G/A)(C/T)A (A/G)GA TAG AGC-3'
	Rev.: Arg-3'	5'-TGG TGT CC(G/A/C) C(G/T)(G/A) (G/C)(G/A/T)G GA(A/T) TCG AAC C-3'

¹ Forward and reverse tRNA^{Leu} primers were used in nine different pair combinations: 1 – M5-M3; 2 – B5-B3; 3 – TL25-TL23; 4 – M5-B3; 5 – B5-M3; 6 – B5-TL23; 7 – TL25-B3; 8 – M5-TL23; 9 – TL25-M3.

The aforementioned primer sets were used with each of the isolates initially identified as containing a putative group I intron. PCR amplification products that were in the size range of 280-400 bp (the appropriate size for many of the known group I self-splicing

introns) were further characterized. An example of the PCR amplification results of putative introns in the subsurface bacteria, using one of the tRNA^{Leu} gene primer sets, is provided below (Fig. 2).

Figure 2. PCR amplification products of subsurface isolates tRNA^{Leu} genes. Migration in a 2.5% agarose gel of the amplification products of PCR with the tRNA^{Leu} primer pair B5-B3 (see Table 1) and 15 subsurface isolates is shown. Lane 1, B138; lane 2, B208; lane 3, B251; lane 4, B252; lane 5, B255; lane 6, B259; lane 7, B307; lane 8, B310; lane 9, B329; lane 10, B333; lane 11, B361; lane 12, B608; lane 13, B628; lane 14, B631; lane 15, B715; lane 16, *Anabaena* sp. PCC7120 (positive control); lane 17, *A. tumefaciens* (Leu intron negative); lane M, 1-kb DNA marker (Gibco/BRL) (sizes are indicated on the sides of the gel). The arrow (→) indicates the size of PCR amplification products that contain only the tRNA, lacking an intron. The arrow (←) indicates the location of the PCR amplification product for the tRNA^{Leu} of *Anabaena* 7120, that contains an intron, and an amplification product of a similar size in B715, that could contain an intron.

Based on the above results, the amplification product for subsurface isolate B715 was cloned and sequenced. The sequence data demonstrated that this tRNA^{Leu} gene did contain a sequence that is a bona fide group I self-splicing intron having all of the required conserved elements. Additionally, the secondary structure of this intron folds properly (see below, Fig. 3).

Figure 3. The secondary structure of the tRNA^{Leu} (UAA) gene group I self-splicing intron from subsurface isolate B715 based on the format described by Cech et al. (1994). To follow the sequence, start at the 5' end and follow the arrows to the 3' end. P1 – P9b are regions of base pairing that result in the conserved helices of group I introns (Michel and Westhof, 1990). P, Q, R and S refer to conserved primary sequence elements of group I self-splicing introns.

This finding represents the first group I self-splicing intron that has been found in a tRNA^{Leu} gene in a group of organisms other than cyanobacteria. To date, other PCR amplified products using subsurface microbial DNA and other tRNA^{Leu} primer sets (that were in the size range of 280-400bp), which could have contained group I intron sequences, have turned out instead to be the result of multiple tRNA copies, that are expressed in tandem with a varied intergenic spacer.

Although DOE funding for this work was not renewed, we are currently continuing these studies with new funding obtained from NASA.

B. Characterization of non-culturable bacteria from GEMHEX intervals

1. PCR amplification, cloning and sequencing of 16S rRNA genes from total community DNA:

Total microbial community DNA was extracted, in triplicate from microorganisms in lacustrine sediment sample G16A (YB02-28; 184.9M) from the GEMHEX project. The extraction procedure that was used is one that has been demonstrated to produce DNA that is amenable to PCR amplification. Three replicate PCR reactions were performed using universal oligonucleotide primers (previously developed) to amplify the 16S rRNA genes (approx. 1500 bp) from community DNA samples from each of the three extractions. Thus, this resulted in 9 (3x3) separate pools of PCR amplified products from a single lacustrine GEMHEX sample interval. Amplification products were size fractionated by gel electrophoresis to verify that products of the appropriate size were obtained. Following this verification, amplification products were subcloned into pCR 2.1-TOPO using Invitrogen's TOPO TA cloning kit.

Approximately 100-150 clones were obtained from each subcloning event. Plasmids were then isolated and inserts of the appropriate size were identified following restriction digestion using *EcoRI* and agarose gel electrophoresis. RFLP analysis of full-length amplification products was performed using the restriction enzymes *Hae* III and *Hha* I were used for this purpose. Approximately 25 different RFLP groups were observed, with a large percentage of the subclones falling into 5-6 RFLP types. Representatives from each of the RFLP groups, which were presumed to contain unique 16S rRNA genes, were sequenced.

The same procedure as described above was then carried out using lacustrine sediment samples G8 (YB02-20; 178.9M) and G12 (YB02-24; 182.1M), also from the GEMHEX interval. In total, approximately 60 16S rRNA gene products have been sequenced in their entirety (from the three lacustrine sediment samples).

2. Phylogenetic Analyses:

Sequence data was initially analyzed using the Basic Local Alignment Search Tool (BLAST). This analysis permitted comparisons to other 16S rRNA sequences present in the Ribosomal Database Project (RDP) small rRNA subunit database, as well as, the SMCC isolate sequence database. The results for these similarity analyses provided below.

Table 2. Preliminary Identification of Subsurface Microorganisms from GEMHEX Intervals

Sediment #	Clone ID	Preliminary ID	Similarity Rank	Phylogenetic Group	# of Clones
YB02-20	VES 2	<i>Aureobacter liquefaciens</i>	0.756-0.79	Hi G+C	5/60
YB02-20	VES 4	<i>Xanthomonas campestris</i>	0.852-0.861	Gamma	4/60
YB02-20	VES 7	<i>Salinococcus roseus</i>	0.727-0.731	Bacillus (Lo G+C)	2/60
YB02-20	VES 12	<i>Lactobacillus saki</i>	0.949-0.958	Bacillus (Lo G+C)	3/60
YB02-24	VES 15	<i>Aureobacter liquefaciens</i>	0.754-0.78	Hi G+C	3
YB02-24	VES 36	<i>Sphingobacter heparinum</i>	0.905-0.968	Cytophaga	5
YB02-24	VES 44	<i>Stenotrophomonas africae</i>	0.91-0.924	Gamma	4
YB02-24	VES 28	<i>Xanthomonas campestris</i>	0.82-0.856	Gamma	3
YB02-24	VES 51	Str. BO272 (SMCC)	0.931	Beta	1
YB02-24	VES 29	<i>Mycoplana segnis</i>	0.947	Alpha	2
YB02-24	VES 66	<i>Flexibacter sancti</i>	0.608	Cytophaga	1
YB02-24	VES 68	<i>Flexibacter filiformis</i>	0.606-0.609	Cytophaga	2
YB02-24	VES 70	<i>Methylobacter radiotolerans</i>	0.977	Gamma	1
YB02-24	VES 73	<i>Burkholderia</i> sp.	0.994	Beta	2
YB02-24	VES 75	<i>Lactobacillus saki</i>	0.951-0.959	Bacillus (Lo G+C)	3
YB02-24	VES 79	<i>Agrobacterium tumefaciens</i>	0.988	Alpha	1
YB02-28	VES 84	<i>Aureobacter liquefaciens</i>		Hi G+C	5
YB02-28	VES 91	<i>Sphingobacter heparinum</i>	0.906	Cytophaga	2
YB02-28	VES 88	<i>Stenotrophomonas africae</i>	0.924-0.936	Gamma	6
YB02-28	VES 93	<i>Xanthomonas campestris</i>	0.871-0.911	Gamma	4
YB02-28	VES 81	<i>Agrobacterium tumefaciens</i>	0.996	Alpha	1

It is interesting to note that two of the organisms identified, *Aureobacter liquefaciens* and *Xanthomonas campestris*, were present within all three of the lacustrine interval sediment samples examined. Additionally, three of the organisms identified, *Spingobacter heparinum*, *Stenotrophomonas africae* and *Agrobacterium tumefaciens*, were present in both the 24 and 28 interval samples. Also, based on the number of clones that represented the different microorganisms in total, *Aureobacter liquefaciens* comprised 13/60 clones that were sequenced. *Xanthomonas campestris* comprised 11/60 clones. Finally the greatest overall diversity in types of organisms found was in the 24 interval sample, which contained 12/28 unique organisms based on sequence similarity. Based on the preliminary classification of the organisms, we have determined the location or environment type that they are characteristically found. This information is presented below (Table 3).

Table 3. Description of the Environment/Location Where Organisms Have Been Found

Preliminary ID	Location Typically Found
<i>Aureobacter liquefaciens</i>	Soil and water
<i>Xanthomonas campestris</i>	Plant pathogen
<i>Salinococcus roseus</i>	Similar to <i>Staphylococcus</i> & <i>Bacillus</i> (various)
<i>Lactobacillus saki</i>	Found in saki (related to bacillus)
<i>Sphingobacter heparinum</i>	Marine and freshwater sediments
<i>Stenotrophomonas africana</i>	Human pathogen
Str. BO272 (SMCC)	Deep subsurface sediments (Savannah River Site)
<i>Mycoplana segnis</i>	Water, soil
<i>Flexibacter sancti</i>	Marine, freshwater sediments
<i>Flexibacter filiformis</i>	Marine, freshwater sediments
<i>Methylobacter radiotolerans</i>	Various
<i>Burkholderia</i> sp.	Soil
<i>Agrobacterium tumefaciens</i>	Agricultural soils

While the majority of the organisms that were identified could be expected given the location and /or environments that they have previously been found there are also a number of less expected results. Most noticeable is the presence of *L. saki* and the human pathogen *S. africana*. Additional information is currently being gathered to better understand the possible significance of these organisms in the subsurface environment.

Following classification and similarity analysis by BLAST, phylogenetic analyses were performed. Sequences were manually aligned based on known secondary structure considerations using the Genetic Data Environment (GDE) multiple sequence editor. Aligned sequences are being further phylogenetically analyzed using the software, TREECON.