2008 R&D 100 AWARDS ENTRY FORM

1. Submitting Organization:

Organization	Lawrence Berkeley National Laboratory (LBNL)
Address	1 Cyclotron Road
City, State, Zip	Berkeley, CA 94720-8125
Country	USA
Submitter	Pamela Seidenman
Phone	510/486-6461
Fax	510/486-6457
E-mail	PSSeidenman@lbl.gov

AFFIRMATION: I affirm that all information submitted as a part of, or supplemental to, this entry is a fair and accurate representation of this product.

Submitter's signature: _____

2. Joint entry with:

Organization	
Address	
City, State, Zip	
Country	
Contact Name	
Phone	
Fax	
E-mail	

3. Product name:

Berkeley Lab PhyloChip

4. Briefly describe (25 words or less) what the entry is (e.g., balance, camera, nuclear assay, etc.)

The PhyloChip is a DNA chip (microarray) that quickly, comprehensively, and accurately identifies species within microbial samples from any environmental source, without any culturing required.

5. When was this product first marketed or available for order? (Must have been available in 2007.)

This product was made available to academic and research organizations on a limited basis for the purpose of testing the product. It was marketed and made commercially available in 2007.

6. Inventor or Principal Developer (List all developers from all companies):

Developer Name	Gary Andersen
Position	Staff Scientist
Organization	Lawrence Berkeley National Laboratory
Address	1 Cyclotron Road
City, State, Zip	Berkeley, CA 94720
Country	USA
Phone	(510) 495-2795
Fax	(510) 486-7152
E-mail	GLAndersen@lbl.gov

Developer Name	Todd DeSantis
Position	Software Developer
Organization	Lawrence Berkeley National Laboratory
Address	1 Cyclotron Road
City, State, Zip	Berkeley, CA 94720
Country	USA
Phone	(510) 761-6720
Fax	
E-mail	TDeSantis@lbl.gov

Key Contributor	Eoin Brodie
Position	Research Scientist
Organization	Lawrence Berkeley National Laboratory
Address	1 Cyclotron Road
City, State, Zip	Berkeley, CA 94720
Country	USA
Phone	(510) 486-6584
F*ax*	(510) 486-7152
E-mail	ELBrodie@lbl.gov

Key Contributor	Yvette Piceno
Position	Research Associate
Organization	Lawrence Berkeley National Laboratory
Address	1 Cyclotron Road

City, State, Zip	Berkeley, CA 94720
Country	USA
Phone	(510) 486-4498
Fax	(510) 486-7152
E-mail	YMPiceno@lbl.gov

7. Product price:

To be determined by licensee.

8. Do you hold any patents or patents pending on this product?

Yes—International Patent Application number, PCT/US2007/024720, filed 11/29/07, titled "Array for Detecting Microbes," LBNL Docket No. IB-2229PCT.

9. Describe your product's primary function as clearly as possible. What does it do? How does it do it? What theories, if any, are involved?

> *"[The PhyloChip] streamlines the process of sample analysis… allowing us to address previously unapproachable questions."*
>
> —Dr. Roberto Kolter, Harvard Medical School

Bacteria never live alone. Instead, they grow in complex communities, made up of hundreds of different species. Bacterial communities are found in familiar environments, such as water, soil, food, air, and within our own bodies. Knowledge about these bacterial communities—including the predominant species among them, their interaction, and their changes over time—is essential for understanding the effect of any disturbance within natural ecosystems. Deep, sudden changes in the structure of a bacterial community could represent a danger to us. These changes could take the form of an airborne biological terrorist attack, or an epidemic caused by contaminated water, or soil, or hazardous atmospheric alterations caused by climate change. But how do we know "what's there," hidden in those ecosystems?

Until now, scientists have had no fully accurate, comprehensive way of detecting the presence, quantity, and diversity of bacteria (including disease-causing *microbes* or *pathogens*) in an air, water, soil, or clinical sample. Researchers relied on bacterial cultures to identify what was present in such a sample: The problem with this method, in addition to the time (often days or weeks) that growing the culture requires, is that it leaves out all of the organisms that can't survive in the culture, which could be as much as 99 percent of the bacteria in a sample. Moreover, while recent advances in genetic detection technology

(for example, assays that detect the presence of a specific gene sequence) have been effective in identifying discrete organisms in samples, they have only been able to test for individual bacteria *that researchers anticipate* being present in the sample. As in the child's card game Go Fish, this technology could determine only whether *specific* bacteria were present, not *which* bacteria were present. We still would have little idea of "what's there."

The Berkeley Lab PhyloChip, developed by Gary Andersen, Todd DeSantis, Eoin Brodie, Yvette Piceno, and their colleagues at Berkeley Lab, provides the best answer yet to "what's there." The PhyloChip is a microarray unique in its ability to quickly and comprehensively identify multiple bacteria and archaea within microbial DNA samples. **Capable of analyzing samples from any environmental source—air, water, soil, blood, tissue—the PhyloChip is unprecedented in its ability to accurately test such samples without any culturing required, and without prior knowledge of a sample's microbial composition, all in a single test.**

Easily fitting into a person's hand (Figure 1), the PhyloChip can simultaneously (i.e., within one testing sample) detect most known microorganisms—testing for over 8,000 bacterial species, a microbial detection power previously unknown. Also, its ability to produce useful results in a matter of hours means that numerous samplings of a specific environment can be conducted virtually on a daily basis, enabling scientists to track, as never before, the progress (i.e., appearance/disappearance, increase/decrease) of a certain microorganism over a short period of time.

Figure 1. The Berkeley Lab PhyloChip profiles microbial populations at a rate and with an accuracy heretofore unknown. It simultaneously detects most known microorganisms (over 8,000 species tested in parallel) without culturing.

Berkeley Lab PhyloChip Operation

Inside the PhyloChip's disposable cartridge is a glass surface (called a microarray) divided into a grid of 356 rows × 712 columns, resulting in 253,472 separate tests, each capable of capturing a specific nucleic acid. When the DNA or RNA molecules from the sample (soil, blood, etc.) come into contact with the short pieces of DNA bound to the glass surface, each molecule adheres only to the appropriate location (Figure 2). The Berkeley Lab PhyloChip can distinguish microbes based on how well their ribosomal DNA or RNA anneals, or "sticks," to each of the many test sites (Figure 3).

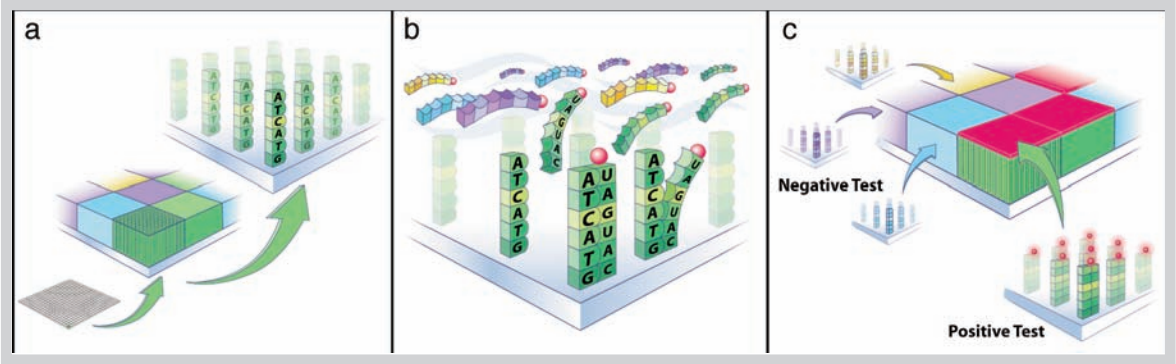


Figure 2. PhyloChip operation: (a) Multiple tests conducted on a single glass surface; (b) DNA from a sample (blood, soil, water, etc.) adheres where a match is found ("hybridization"); (c) laser scanning reveals which tests were positive (i.e., which microbes are present). In this way, PhyloChip quickly and accurately identifies microbes in complex samples. (Images provided by Affymetrix, Inc., Santa Clara, CA.)

The cartridge is inserted into a scanner that detects which tests are positive by the emission of fluorescent light only from the test sites bound by nucleic acids in the sample. The intensity of the fluorescence from each test corresponds to the quantity of organisms in the sample—allowing ecosystem comparison over time to determine which bacterial populations are changing.

All microorganisms have ribosomal genes because these genes (as protein generators) are essential for life. The small sequence differences within the ribosomal gene are what distinguish different species. The PhyloChip has divided all known sequence variations from bacterial and archaeal ribosomal genes into over 8,000 distinctive groupings, each representing a specific microbial genus or species identified from many different sampled environments—including clinical, air, soil, and water. **Given its ability to perform over 250,000 tests simultaneously, the**

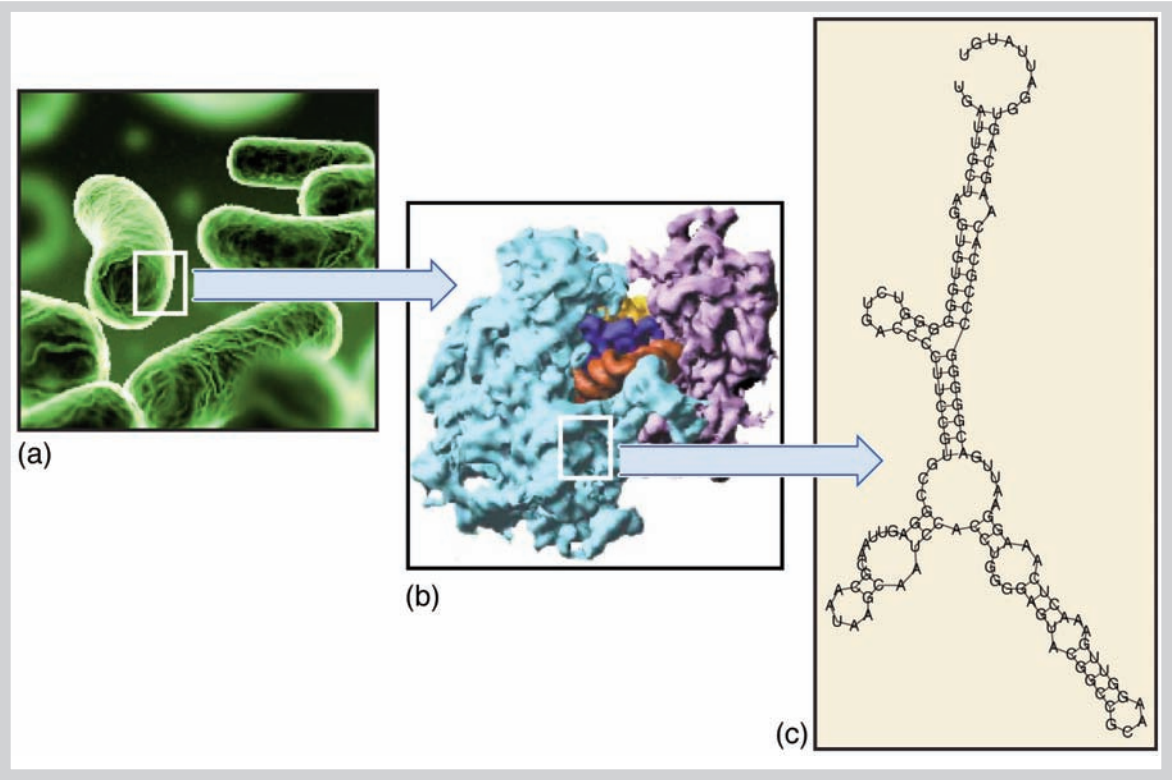


Figure 3. Bacterial types (a) contain ribosomes (b) that are composed of a specific sequence of DNA/RNA bases (c), allowing them to be differentiated. The PhyloChip detects these differences, and thus is the first tool able to classify all types of DNA/RNA sequences in complex mixtures.

PhyloChip does not require customization for different types of samples, since the total diversity of all known ribosomal genes that could be present within any media is represented within the 250,000 tests. With a comprehensive assay for all currently identified bacteria, the PhyloChip eliminates before-the-sampling guesswork as to what bacteria exist within a sample, taking advantage of (and incorporating into the PhyloChip) the vast amount of ribosomal sequence data available in public databases. As these databases grow, updated versions of the PhyloChip can easily include novel diversity, because room exists on the chip surface for over 1 million tests.

Moreover, the PhyloChip's high-density format, combining tests for each species with paired mismatch-control tests, significantly reduces the chances of misidentifying a specific microorganism. These PhyloChip capabilities allow for unprecedented accuracy in characterizing a microbial sample—all within a day.

Berkeley Lab PhyloChip Validation, Accomplishments, and Contributions

> *"The PhyloChip has given us an unparalleled view of the bacterial community…there is nothing else as comprehensive or sensitive."*
>
> —Dr. Kasthuri Venkateswaran, Jet Propulsion Laboratory

In its short existence, the Berkeley Lab PhyloChip has already achieved considerable success. It has been validated (Figure 4) through an extensive collection of air samples obtained for identifying microbial communities typically inhaled by inhabitants of U.S. metropolitan cities. In addition, preliminary PhyloChip testing of water and soil samples has shown that this technology is feasible in those environments as well. Findings made possible by the PhyloChip have led to publications in a number of scientific journals (see Appendix B) and attracted extensive press coverage (see Appendix C).

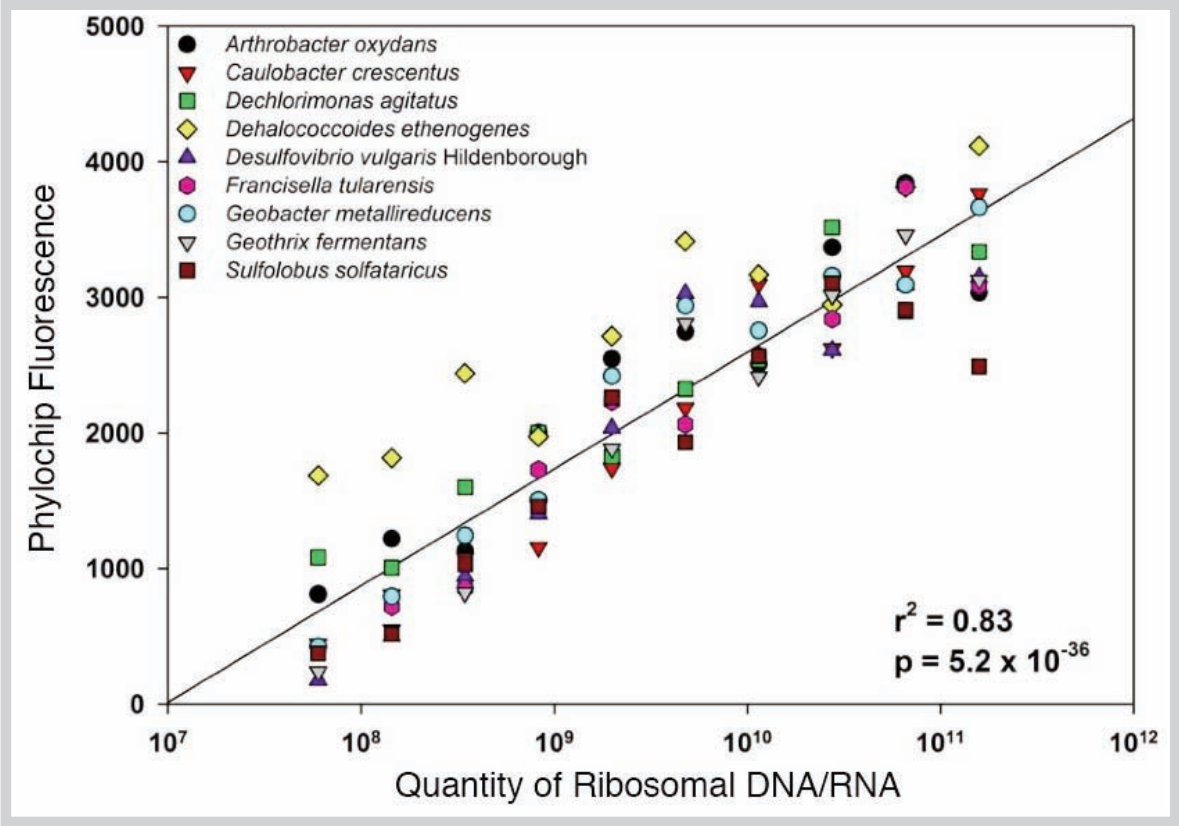


Figure 4. Validation of PhyloChip sensitivity. The figure shows how mixtures of nine bacteria (shown in the legend) in varying concentrations are accurately resolved by the PhyloChip. Fluorescence intensity correlates well with the quantity of bacteria in the sample, revealing the bacterial community structure. The PhyloChip is sensitive over a wide dynamic range—it can simultaneously detect large amounts of one bacterial species and small amounts of another. No other device has the ability to profile both the predominant and the minority bacterial populations.

In 2004, Berkeley Lab researchers conducted (for the U.S. Department of Homeland Security) a first-of-its-kind cataloguing of microbes taken from air samples above the Texas cities of San Antonio and Austin. Before this study, no one had a sense of the diversity of microbes in the air. Investigators found over 1,800 diverse bacterial types, a much richer and more varied population than anyone expected (Figure 5), rivaling that of soil microbial diversity. This research, described in a paper published in the *Proceedings of the National Academy of Sciences* and reported on in *Scientific American* in December 2006 (see Appendix B), marks the beginning of a regional bacterial census that will help the Department of Homeland Security differentiate between normal and suspicious fluctuations in airborne microbes. It also helps to establish a baseline background of airborne bacteria, which scientists can now use to track how climate change affects bacterial populations.

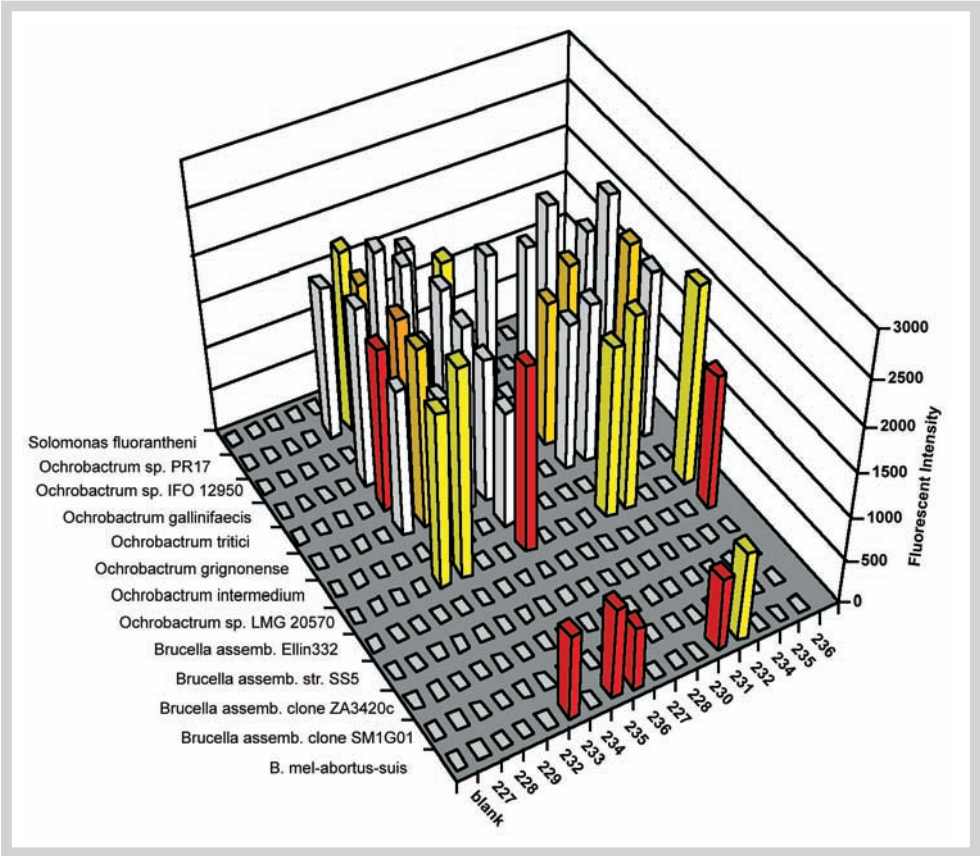


Figure 5. Closely related bacterial species can be tracked spatially and temporally. This figure shows 13 of the greater than 8,000 bacteria (left axis) profiled from 16 samples (bottom axis) from above San Antonio and Austin. Some bacteria were never found; others were detected in many samples. The height of each bar represents the abundance of each species, while the color indicates detection likelihood (in descending order from red, to orange, to yellow, to white). The PhyloChip allows rapid, inexpensive comparison among samples to reveal community differences, regardless of bacterial culturability.

In addition, the PhyloChip has already been an important part of recent critical medical studies. Respiratory infections caused by the ambient environment are a major problem in hospitals. As reported in the *Journal of Clinical Microbiology* (2007) (see Appendix B), the PhyloChip was used to analyze the microbial environment within respiratory tube airways. It was the key to discovering that a loss of bacterial diversity (resulting from antibiotic treatments) was directly associated with the development of pneumonia in ventilated patients exposed to a certain common strain of bacteria. The potential life-saving possibilities of this finding are obvious.

The PhyloChip has also shown great value in bioremediation cleanup efforts at contaminated sites. When uranium mining and processing for nuclear weapons and fuel were at their peak, in the 1950s and 1960s, uranium-containing wastes accumulated, resulting in a multitude of contaminated sites in the U.S. and worldwide. One promising approach to containing uranium migration is to catalyze the reduction of soluble U(VI) to the less-soluble U(IV). As described in a paper published in *Applied and Environmental Microbiology* (2006) (see Appendix B), the PhyloChip was central in identifying (from soil samples) those bacteria that could, through ingestion of the uranium, prevent U(IV) from converting to soluble uranium U(VI), thus forestalling the migration of this radioactive material and optimizing site remediation efforts.

Degraded water quality is also a growing environmental problem. The PhyloChip allows researchers and environmental managers to gain unprecedented knowledge about water-borne microbes to rapidly distinguish between harmful and beneficial species. The PhyloChip has empowered water-resource administrators to assert proper corrective actions within days after a pollution episode (as opposed to months). For instance, in an ongoing study, the PhyloChip was recently used to monitor California creek water suspected to be contaminated by sewage. The PhyloChip allowed researchers to pinpoint the creek locations associated with specific types of human fecal bacteria. Evidence was produced to demonstrate that untreated sewage was entering the creek. Using the results from PhyloChip sampling, public health officials will be able to specify break points in the creek to divert and sanitize the water before it reaches the ocean.

10A. List your product's competitors by manufacturer, brand name, and model number.

1. The **Berkeley Lab PhyloChip** is the first technology to both identify and quantify the full spectrum of microbes within a sample in a cost-effective

manner. Competitive technologies are: **MicroSeq**, distributed by Applied Biosystems, Inc., which determines the identity of culturable microbes one at a time by reading their ribosomal gene sequence.

2. **MicroLog**, distributed by Biolog, Inc., which determines the identity of culturable microbes one at a time by testing their ability to grow on a variety of carbon sources.
3. **Clone Sequencing Services** are provided by, for example, SeqWright, Inc. or Agencourt Biosciences, Inc. For this technology, a small portion of DNA is extracted directly from the microbes within a sample without the need for culturing. Hundreds to thousands of ribosomal genes from the DNA extraction are sequenced to identify the microbes present in the highest quantity.

10B. Supply a matrix or table showing how the key features of your product compare to existing products or technologies. Include both numerical and descriptive comparisons.

	Berkeley Lab PhyloChip	MicroSeq	MicroLog	Clone Sequencing Service	Competitive advantage of Berkeley Lab PhyloChip
Resolution	Identifies and quantifies >8,000 species.	Identifies only what is culturable (~1,000 species).	Identifies only what is culturable (~1,000 species).	Identifies and quantifies ~50% of the most abundant species within a sample.	It identifies and quantifies far more organisms within a sample than other technologies.
Sensitivity	Excellent—sensitivity over 4 orders of magnitude (i.e., detects wide variations in abundance). Detects both culturable and nonculturable bacteria.	Detects *only culturable* organisms. Produces a sequence to match against a small, proprietary database.	Detects *only culturable* organisms. Produces a carbon utilization pattern to match against a small, proprietary database.	Poor—only about 10% of the community members are detected, producing a sequence that can be matched against a public database.	It is far superior to competing technologies in its ability to distinguish microorganisms in a sample.
Reproducibility (Identification, Quantification)	Very good—coefficient of variation is ~10% (the smallest of the relevant technologies).	Reproduces only what is culturable. Unable to test for abundance.	Reproduces only what is culturable. Unable to test for abundance.	Poor reproducibility, cannot generate a reproducible distribution for scarce species within a large sample.	It gives reproducible detection across all species within a sample (culturable or nonculturable, abundant or scarce).

Output Rate/Speed	16 complete microbial communities per day.	1 partial community per month.	1 partial community per month.	1 partial community per week.	It analyzes far more samples in far less time than competitive technologies.
Ease of Operation	Requires lysis (cell disintegration), labeling, loading of array cartridge.	Requires plating (replicating), incubating, lysis, polymerase chain reaction (PCR) method (to replicate species DNA), sequencing.	Requires plating, incubating, use of microtiter plate.	Requires PCR in lab, then mailing off for analysis, receiving results, then performing in-house bioinformatics.	It provides a simple *in-house* solution for microbial detection.
Cost	$250 per microbial community.	$50 per isolated community member (typically hundreds), so about $6,000 to 20,000 per community.	$25 per isolated community member (typically hundreds), so about $3,000 to 10,000 per community).	$650 for cloning and robotic selection + $5,000 to quantitatively sample each species—ultimately ~$50,000.	It provides a comprehensive sample analysis at much lower cost than competitive technologies.

10C. Describe how your product improves upon competitive products or technologies.

The Berkeley Lab PhyloChip includes the following advantages over competing technologies:

- ***It detects more microbial species simultaneously in one test than any other method, by far.*** Unlike other technologies, it is not bound by the culturability of a specific microbe within a sample, nor does a microbe have to be present in high abundance for the PhyloChip to detect it.
- ***It is dramatically superior to competing technologies in its ability to distinguish microorganisms within a sample.*** Most other tests designed to detect many bacteria simultaneously can do so for tens to hundreds of bacteria, but the PhyloChip detects thousands of different bacterial groups.
- ***It provides a simple, in-house solution for microbial detection.*** While other methods involve a multitude of steps, and in some cases require mailing of samples back and forth to other service labs, the PhyloChip enables *in-house* analyses using a few steps (Figure 6), thus simplifying the process and reducing the chances for error or contamination.
- ***It is much faster than any competitive technologies.*** For the reasons stated in the previous sentence, and because it does not rely on culturing of bacte-

rial samples, the PhyloChip can be up to two orders of magnitude faster than other methods.

- ***It offers comprehensive sample analyses at much less cost.*** Per sampled community, it is typically an order of magnitude less expensive than competitive technologies.

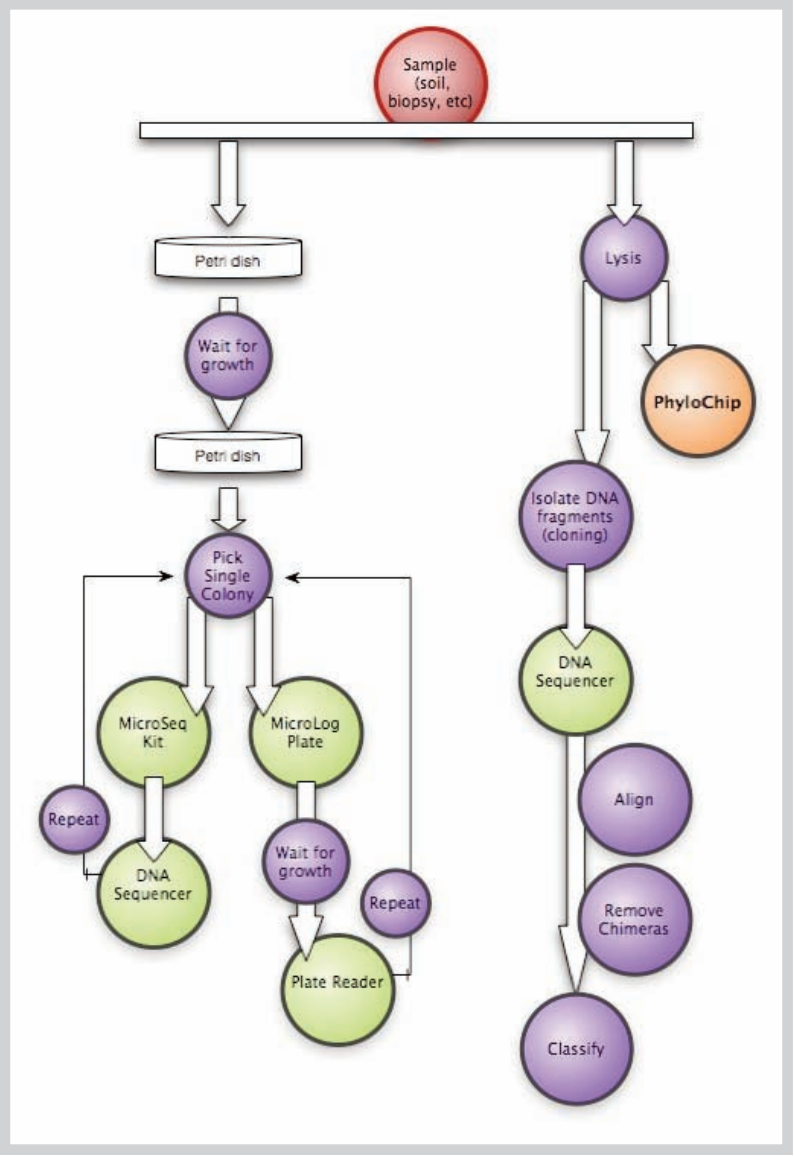


Figure 6. Overview of processing steps required for existing community profiling technologies. The PhyloChip is easier to use compared to culture and cloning methods. Its direct operational path (right path) contrasts with the redundant (MicroSeq, MicroLog) or multistep (cloning) methods (left path). The PhyloChip's direct method requires less technical training, leads to faster results, and provides fewer opportunities for error.

11A. Describe the principal applications of this product.

Environmental

The main application of the Berkeley Lab PhyloChip is to detect microbes in samples from a variety of environments, namely air, water, or soil. The recent sampling of the air over San Antonio and Austin was the first scientific program of its kind, and led to a completely new understanding of airborne microbial life, as well as what was involved in accurately assessing airborne bioterrorist threats.

Clinical/Medical

The PhyloChip has also already proven itself in hospital studies, not only in detecting dangerous bacteria within air samples but also in studies of blood and tissue samples. With its ability to detect bacteria in all such samples, it promises to be an indispensable medical-diagnostic tool for the foreseeable future.

Bioremediation

As in the uranium waste and creek sewage examples discussed in Section 9 (above), the PhyloChip is an essential tool for monitoring bacterial populations during environmental bioremediation. Unlike any previous technology, it can quickly identify the most metabolically active microorganisms in a soil sample, as well as those microorganisms most responsive to environmental changes or stresses. In this way, it can isolate what species would be most effective in urgent, ongoing bioremediation efforts, and track their abundance.

11B. List all other applications for which your product can now be used.

Biofuels

Plant bacteria are crucial in the development of those plants that will become biofuels. To properly understand the biological processes involved in a desired outcome, like energy creation, we need to understand not only the individual microorganisms responsible for any catalytic events that promote fuel creation, but also the interplay of the supporting cast of bacteria that provide metabolites and defenses for the plant. The PhyloChip is the only tool that exists for rapidly detecting the identity, diversity, and abundance of bacteria, and for observing how the bacterial community changes and interacts over time.

Rapid Diagnostics

All the applications listed here take advantage of the PhyloChip's ability to characterize a microbial sample without culturing. Given that commonly as much as 99%

of bacterial communities are not detected using culture-based techniques, this is a huge advantage, especially when researchers are studying new systems. The ability to detect a shift, for example, in a patient from a bacterial community typical of a healthy person, to one typical of someone with minor indicators of a diseased state, could allow more rapid treatment and result in shorter, less intense illness.

Carbon Sequestration

The PhyloChip could be decisive in determining the bacteria involved in the key geochemical reactions for enhanced CO_2 storage—specifically for monitoring the microbial processes that facilitate mineral trapping of CO_2, or the development of precipitates and biomass that could seal natural fractures, which might otherwise serve as conduits for CO_2 seepage.

Microbial Fuel Cell Monitoring

Microbes aid in converting carbon compounds found in wastewater directly into electricity, within chambers referred to as microbial fuel cells (MFC's). The PhyloChip assists engineers in monitoring the microbial community during the creation and maintenance of novel high-efficiency MFC's. The PhyloChip is able to identify the bacterial families that are able to grow rapidly and convert a very high percentage (nearly 90 percent) of the water-borne carbon compounds into electrical current.

12. Summary. State in layman's terms why you feel your product should receive an R&D 100 Award. Why is it important to have this product? What benefits will it provide?

Knowing that what we can't see can affect us in profound ways, we need to be able to detect what was undetectable before now, at a speed inconceivable before now. The Berkeley Lab PhyloChip provides this ability, and in the process changes the way scientists conduct certain basic, essential investigations.

The PhyloChip packs an enormous amount of analytical power into a device not much larger than a quarter. Its ability to test all manner of environmental samples for their microbial content is unprecedented. The information that it has already provided about the airborne bacterial content above American cities is a first step in distinguishing between a climate-related bacterial change and a real bioterrorist threat.

Moreover, the PhyloChip's contributions to public health, medical diagnostics, and environmental cleanup projects have already paid large dividends. It promises even more advances in the development of biofuels and carbon sequestration. In short, scientists are continually finding new ways to use the PhyloChip, and finding things that they could not have found by any other means.

ORGANIZATION DATA

13. Contact person to handle all arrangements on exhibits, banquet, and publicity.

Name	Pamela Seidenman
Position	Marketing Manager, Technology Transfer
Organization	Lawrence Berkeley National Laboratory (LBNL)
Address	1 Cyclotron Road, 90R1070
City, State, Zip	Berkeley, CA 94720-8125
Country	USA
Phone	510/486-6461
Fax	510/486-6457
E-mail	PSSeidenman@lbl.gov

2008 R&D 100 Awards
ENTRY—Berkeley Lab PhyloChip

List of Attachments

A. Letters of Support

B. Selected Publications

C. Selected Scientific and Popular Press Coverage

2008 R&D 100 Awards
ENTRY—Berkeley Lab PhyloChip

Appendix A

Letters of Support

- Dr. Roberto Kolter, Harvard Medical School
- Jizhong Zhou, Institute for Environmental Genomics, University of Oklahoma
- Kasthuri Venkateswaran, California Institute of Technology, Jet Propulsion Lab
- Dr. Wen-Tso Liu, National University of Singapore
- Dr. George Kowalchuk, Netherlands Institute of Ecology
- Patricia Holden, University of California, Santa Barbara
- Dr. Jenine P. Wiener-Kronish, University of California, San Francisco

HARVARD MEDICAL SCHOOL
Department of Microbiology
and Molecular Genetics

200 Longwood Avenue
Boston, Massachusetts 02115

Roberto Kolter, Ph.D.
Professor

Tel: (617) 432-1776
Fax: (617) 738-7664
e-mail: kolter@mbcrr.harvard.edu

Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee:

I am pleased to enthusiastically offer my full support to the 2008 R&D100 Award application for the PhyloChip developed at the Berkeley Lab. This fabulous tool is already beginning to surpass other techniques as a method to determine the composition of complex bacterial communities in both environmental and clinical samples. A high-density 16s rDNA microarray, the PhyloChip permits high-resolution analysis of microbial community composition and dynamics and has significant benefits over traditional clone-library based approaches, the current "gold-standard", particularly where communities are dominated by a few bacterial species, such as Cystic Fibrosis (CF) lung infections. More importantly, this tool streamlines the process of sample analysis from months per sample to weeks for multiple samples allowing us to address previously unapproachable questions.

Defining bacterial community composition in states of disease and health in an accurate and high-throughput manner is essential if we are to move towards predictive modeling and more rapid diagnostics. In my laboratory, we have two projects that utilize the PhyloChip. The first, spearheaded by Dr. Vanja Klepac-Ceraj, focuses on delineating the establishment and dynamics of the complex microbial community involved in chronic lung infection in children with CF. Using the PhyloChip we have detected a diverse population of bacterial species, many of which may be significant in disease progression, that had been previously overlooked due to the inadequacies other approaches. With the PhyloChip, we have the capacity to not only fully define these complex polymicrobial lung infections, but, for the first time, to monitor large numbers of individuals over time and determine the impact of antibiotic therapy on the entire bacterial population, rather than just a few known pathogens. The second project, led by Dr. Katherine Lemon, aims to determine the microbial community composition of normal flora in the nostrils and throats of healthy children. Elucidating the development and maintenance of these microbiota will provide the foundation for exploring the effects of pathogen carriage, antibiotic use, vaccination, and variation in host genetics on this resident microbial flora.

We envision virtually unlimited utility of this novel technology in a broad range of both clinical settings, where it will undoubtedly lead to a better understanding of human-microbial interactions and polymicrobial pathogenesis, and natural settings, as a means to monitor the emergence of pathogens in the environment.

Sincerely,

Roberto Kolter, Ph.D.

Department of Botany and Microbiology
Institute for Environmental Genomics
University of Oklahoma
101 David L. Boren Blvd., Norman, OK 73019
Phone: *405-325-6073;* ***Fax:*** *405-325-7552;* ***Email:*** *jzhou@ou.edu;* ***Web:*** *http://ieg/ou.edu/*

January 18, 2008

Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee:

It is my great pleasure to write this letter in support of Dr. Gary Andersen and his colleague, Mr. Todd DeSantis, for nomination for the 2008 R&D100 Award. Gary is a world-class leader in the application of microarrays to the study of microbial communities in the environment. Over the past decade, the Andersen lab has provided innovative solutions to address the challenge of identifying key microbial interactions on complex environmental communities. It is well known that a single gram of soil may contain many thousands of distinctive bacterial species. The only way, until now, to specifically identify the individual members of a population has been to serially sequence hundreds to thousands of cloned 16S rRNA gene fragments. This is costly, tedious and very time consuming. The PhyloChip, developed in the Andersen lab, is the first device that has been able to comprehensively categorize all of the individual types of bacteria in a rapid DNA hybridization reaction. This is nothing short of revolutionary for the study of how bacterial interact with each other and the environment and will have applications in all areas where bacteria exist as a multi-member community.

Prior to the Berkeley Lab PhyloChip, observations of microbial communities were limited to one or a few individual samples in a study. This was due to the prohibitive cost of completing a 16S rRNA gene sequence library and length of time to properly place individual sequence observations in the correct phylogenetic context. As the price of DNA sequencing has decreased over the years, the desire to obtain more sequence information from each sample has resulted in a steady state in cost for characterizing an individual sample. This is because billions of amplified 16S rRNA gene molecules are produced in a PCR reaction but, at best, only thousands of molecules are actually cloned into an appropriate vector and sequenced. By contrast, the entire amplified 16S gene product can be placed on the PhyloChip to categorize the individual components. The PhyloChip is particularly advantageous in identifying the rare, low abundance members of the microbial community. With orders of magnitude reduced costs for PhyloChip analysis compared with a typical 16S rRNA gene sequence library, the PhyloChip lends itself to multiple observations with replication so that, for the first time, we can get an accurate picture of how all the members of a microbial community are changing over time.

What makes this technology worth of the R&D 100 award is the impact it has had and will continue to have on the study of the interactions of microbes to each other and to their surrounding environment. The PhyloChip has already demonstrated key microbial interactions in subsurface bioremediation sites contaminated with mobile uranium, chromium contaminated groundwater, and clinical samples from patients with ventilator associated pneumonia. The PhyloChip has also, for the first time, shown the diversity and composition of bacteria in the atmosphere as well as the effect of climate change on soil-borne microbes. As great as these accomplishments are, the real value of the PhyloChip will be observed in the next couple of years as the numbers and types of studies using this technology increases. This breakthrough technology will shed light on microbial processes that could be harnessed by scientists for any number of applications including biofuels, land management, food safety, human health and climate change prediction. For these reasons, I give my highest endorsement of the innovative PhyloChip for this prestigious R&D 100 Award.

If you have any question, please do not hesitate to ask me!

Yours sincerely

Jizhong Zhou

Presidential Professor, Department of Botany and Microbiology
Director, Institute for Environmental Genomics (IEG)
University of Oklahoma
Editor for Applied and Environmental Microbiology

Jet Propulsion Laboratory
California Institute of Technology

4800 Oak Grove Drive
Pasadena, California 91109-8099

(818) 354-4321

JPL

Dr. Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125 FAX: 510-495-2263
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee:

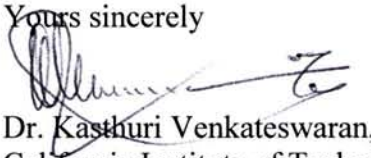
I am pleased to write a letter in support of the Berkeley Lab Phylochip for the R&D100 awards. The PhyloChip has been of tremendous use in my lab for measuring microbial diversity in clean room and spacecraft assembly environments, as well as bioaerosols in commercial aircraft cabin air. The extremely low microbial biomass in these environments makes it especially difficult to accurately measure microbial composition.

The Phylochip has given us an unparalleled view of the bacterial community when compared with more traditional typing methods. The ability to analyze multiple samples with replication has been indispensable in tracking key microbial groups. Conventional sequencing of 16S rRNA gene clone libraries is considerably more expensive and does not allow for multiple replications. There is nothing else as comprehensive or sensitive that we have found for our work in this area.

Because of its unprecedented power, the PhyloChip has the potential to uncover important bacterial species-level relationships within environmental samples. This will be of critical importance in our upcoming Mars missions. As Principal of the Biotechnology and Planetary Protection Group, we are required to protect the pristine condition of the extraterrestrial environment of several missions. While searching for novel life we take every steps in cleaning the spacecraft component to avoid contamination of other celestial bodies. The PhyloChip has been the key to the identification of any forward-contaminating organisms.

In short, I give my full and hearty recommendation for the PhyloChip for the R&D100 award. If you have any questions, please do not hesitate to contact me.

Yours sincerely



Dr. Kasthuri Venkateswaran, Ph.D.
California Institute of Technology, Jet Propulsion Laboratory
Biotechnology and Planetary Protection Group; M/S 89-2
4800 Oak Grove Dr., Pasadena, CA 91109

Faculty of Engineering
Division of Environmental Science and Engineering

NUS National University of Singapore

January 16, 2008

Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee:

My highest recommendation goes to the innovative DNA array, **the Berkeley Lab PhyloChip**. The ability of this unique device to rapidly identify and monitor the dynamics of all known bacterial populations (> 8-9000 types) in an area of less than 2x2 cm^2 over time and space has been consistently validated by researchers from the industry and academic around the world—no other platform has such an ability to comprehensively identify the interactions among multiple organisms in parallel on the earth.

As Professor at the National University of Singapore, and senior scientist at the Institute of Microelectronics at Agency of Science, Technology and Research, Singapore, my special field of interest is the use of microarrays in the study of microbial ecology. I have been using the PhyloChip in a microbial community census of bacteria, including pathogens in Singapore waters, secondary effluent discharged from local wastewater treatment plants, and lake water samples taken from Tibetan region in China.

The PhyloChip is simply better than any comparable technology: its sensitivity equivalent to "finding a needle in a hay stack" has been shown to be much greater than sequencing 1000 clones from a 16S rRNA gene library construction (US$100-200 for PhyloChip vs US$2000 for clone library sequencing). As a result, we have been able to identify many organisms (for example pathogenic *Leptospira* spp.) previously not expected to be present in a given marine water sample taken around Singapore. This information is very important in protecting public health in Singapore. Furthermore, we have also been able to identify the most metabolically active organisms using the PhyloChip by examining the native rRNA extracted from the environments and comparing the results with the rRNA genes amplified from a same environmental sample, something that we had not been able to do with any other method. We strongly feel that this technology provides the only cost-effective solution to collect data for the high number of samples employed in our project. The quick turnaround time down to 1-2 weeks for samples taken over a period of one month from 4-6 marine sites has greatly increased our-throughput for microbial community sample analysis, and thus lowered our costs.

Based on its simplicity, sensitivity and relevance to public health and science advance, I strongly support and endorse the nomination of the **Berkeley Lab PhyloChip** for the 2007 R&D 100 award. Please contact me if you would like any further information.

Prof. Wen-Tso Liu
Division of Environmental Science and Engineering
National University of Singapore
9 Engineering Drive 1, EA-03-12
Singapore 117576

NETHERLANDS INSTITUTE OF ECOLOGY

Prof. George A. Kowalchuk

Centre for Terrestrial Ecology

To: Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202
USA

date January 31st, 2008

subject R&D award recommendation letter for the PhyloChip

Dear R&D 100 Award Selection Committee:

I would like to offer my enthusiastic support for the 2008 R&D100 Award application for the Berkeley Lab PhyloChip. As the Environmental Genomics team leader of The Netherlands Institute of Ecology (NIOO-KNAW) and full Professor at Free University of Amsterdam, I have many projects examining the microbial community dynamics in a range of environments. I was introduced to the PhyloChip system in its initial testing phase, and was immediately intrigued by its potential to provide comprehensive assessments of microbial diversity in environment samples. Numerous groundbreaking applications and publications have since provided testimony to the power of the PhyloChip approach.

NIOO-KNAW
Boterhoeksestraat 48
Heteren

postal address:
P.O.box 40
6666 ZG Heteren

phone:
(+31)26 479 13 14
fax:
(+31)26 472 32 27

email:
g.kowalchuk
@nioo.knaw.nl

internet:
www.nioo.knaw.nl

The PhyloChip's performance has been so superior that we have fully integrated it into the research plans of the department, and I have sent a number of my people to the Andersen lab (at LBNL) to learn this method in detail. We are currently using the PhyloChip on a number of issues of critical importance to The Netherlands—including climate change, soil fertility, and alternative agronomic practices. We are currently writing several scientific publications, which we aim to target for high impact peer-reviewed journals, in which this method is prominently featured.

The introduction of the PhyloChip system in our laboratory has provided a quantum leap in our ability to interrogate microbial community diversity in relation to key environmental issues. Previously used

biomarker-based approaches such as 16S rRNA gene libraries and PCR-based community profiling methods (e.g. T-RFLP, DGGE) either fail to provide the necessary throughput or detail to describe microbial community diversity adequately. The PhyloChip combines the ability to identify important taxa that is possible using 16S clone library sequencing, with the ability to have multiple observations and replications of profiling methods. Thus, the rate and precision with which the PhyloChip produces results enables to conduct investigations in ways that would have been cost-prohibitive previously.

I strongly support this application for the 2008 R&D100 Award. Please contact me if you would like any more information.

Most sincerely,

Prof. George A. Kowalchuk
Netherlands Institute of Ecology
Centre for Terrestrial Ecology
P.O. Box 40
6666 ZG Heteren
the Netherlands
tel. 31 (0) 2647 91314
fax 31 (0) 2647 23227
mob. 31 06 22 546556
g.kowalchuk@nioo.knaw.nl

& Institute of Ecological Sciences
Free University of Amsterdam

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

DONALD BREN SCHOOL OF ENVIRONMENTAL SCIENCE AND MANAGEMENT SANTA BARBARA, CA 93106
http://www.bren.ucsb.edu/

February 11, 2008

Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee:

I write to enthusiastically support the nomination of the Berkeley Lab PhyloChip for consideration of the 2008 R&D 100 award. No other tool has been so expeditious and simple-to-use as the PhyloChip for detecting the thousands of types of bacteria in the streams and coastal zones that my research team monitors. Such observations are extremely crucial if we want to understand and predict how point source contamination will affect downstream water quality. The Berkeley Lab PhyloChip is unique in its ability to categorize all members of a bacterial community and to measure population shifts over time and location.

Degraded water quality is a serious environmental problem in this region as well as many other parts of the world where human populations are increasing. The PhyloChip allows researchers and environmental managers to gain immense knowledge about microbes in our natural surroundings, some of which are harmful and others beneficial. Being able to rapidly distinguish between them allows us to employ proper corrective actions within days after a pollution episode, as opposed to months.

In 2007, we used the PhyloChip for monitoring creek water suspected of sewage contamination. We were able to determine the precise creek locations associated with specific types of bacteria including those typically found in human feces. Not only were we able to demonstrate that untreated sewage was entering the creek but the resolution of the PhyloChip approach lets us target break points in the creek to divert and treat the water before it reaches the oceans.

Without the PhyloChip we would be forced to either run hundreds of separate tests or use a very low resolution fingerprinting technique to detect the diversity of waterborne bacteria. The materials and technician time would be expensive and would require separate quality assurance measures for each test. With the PhyloChip, we can test for over 8000 of types of bacteria, including the ones affecting human health and disease, with a single method.

The research and environmental protection opportunities made possible by the PhyloChip are important to any habitat since it produces rapidly applicable data that leads directly to management decisions having an immediate impact on a local community. For these reasons, I most

PHONE: (805) 893-3195 • FAX: (805) 893-7612
holden@bren.ucsb.edu

enthusiastically support this nomination. Please don't hesitate in to contact me if you have any further questions

Sincerely,

Patricia Holden, Ph.D., P.E.
Professor, Environmental Microbiology

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

Jeanine P. Wiener-Kronish, M.D.,
Professor of Anesthesia and Medicine,
Vice Chairman, Department of Anesthesia and Perioperative Care
Investigator, Cardiovascular Research Institute

513 Parnassus Avenue, Suite S-255
San Francisco, California 94143-0542
Tel: (415) 476-1935
Fax: (415) 476-8841
E-mail: wienerkj@anesthesia.ucsf.edu

Cheryl Ventimiglia
R&D 100 Lab Project Coordinator
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS 46-125
Berkeley, CA 94720-8202

Dear R&D 100 Award Selection Committee,

I am pleased to offer my enthusiastic support for the Berkeley Lab PhyloChip device in response to the 2008 R&D100 Award. This tool permits rapid identification of bacterial communities in complex environmental and clinical samples and holds great promise for providing clinicians with an accurate and comprehensive system for polymicrobial identification of bacteria in states of health and disease. The high-density PhyloChip uses a unique combination of oligonucleotide probes for the identification of virtually all known bacterial taxonomic groups and includes mismatch control probes for every perfectly matching probe on the array to minimize the effect of cross-hybridization and increase detection accuracy.

As a Professor of Anesthesia at UCSF, I am exploring the impact of microbial community composition on airway diseases, including asthma and ventilator associated pneumonia. The PhyloChip has been a unique resource for high-resolution identification of bacterial community composition and temporal dynamics associated with changing pulmonary function in a variety of patient groups. We have used this device to establish that reduced bacterial diversity in the airways of intubated patients treated with antimicrobials, was associated with a high rate of mortality. These finding have not only lead to several publication in a highly ranked journals, but are also contributing to a change in our perception of clinical infection and pathogenesis.

This novel technology has not only been of great help in my research but has generated interest among many of the physicians that I have talked to. Myself and my colleague, Dr. Susan Lynch have set up multiple collaborative projects on cystic fibrosis, interstitial pulmonary fibrosis, chronic obstructive pulmonary disease, chronic rhinosinusitis and food allergy amongst other disease states that use the PhyloChip to provide the most comprehensive polymicrobial community analysis of these disease states – particularly those of unknown etiology. This approach, which represents the cutting edge of translational research will truly revolutionarize how we perceive and treat microbial infections. The need for such a rapid comprehensive tool is great in the clinical arena and I foresee a strong role for the

PhyloChip in multiple studies and potentially as a clinical diagnostic tool. This is truly an innovative tool to harnesses the power of parallel sampling and provide critical insight into polymicrobial composition in states of health and disease, aspects fundamental to understanding and combating infectious disease.

Sincerely,

Jeanine Wiener-Kronish M.D
Professor of Anesthesia and Medicine
Vice-Chairman, Department of Anesthesia and Perioperative Care
Investigator, Cardiovascular Research Institute

2008 R&D 100 AWARDS

Berkeley Lab PhyloChip

Appendix B

Selected Scientific Publications

Brodie, Eoin L., Todd Z. DeSantis, Jordan P. Moberg Parker, Ingrid X. Zubletta, Yvette M. Piceno, and Gary L. Andersen, Urban aerosols harbor diverse and dynamic bacterial populations, *PNAS, 104* (1), 299–304, 2007.
(supporting information supplied upon request)

Flanagan, J.L., E.L. Brodie, L. Weng, S.V. Lynch O. Garcia, R. Brown, P. Hugenholtz, T.Z. DeSantis, G.L. Andersen, J.P. Wiener-Kronish, and J. Bristow, Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with *Pseudomonas aeruginosa*, *Journal of Clinical Microbiology, 45* (6), 1954–1962, 2007.

Brodie, Eoin L., Todd Z. DeSantis, Dominique C. Joyner, Seung M. Baek, Joern T. Larsen, Gary L. Andersen, Terry C. Hazen, Paul M. Richardson, Donald J. Herman, Tetsu K. Tokunaga, Jiamin Wan, and Mary K. Firestone, Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and redoxidation, *Applied and Environmental Microbiology, 72* (9), 6288–6298, 2006.

DeSantis, Todd, Eoin L. Brodie, Jordan Moberg, Ingrid X. Zubieta, Yvette M. Piceno, and Gary L. Andersen, High-density universal !6S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment, *Microbial Ecology*, DOI: 10.1007/s00248-006-9134-9, 2006.

DeSantis, T.Z., P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G.L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Applied and Environmental Microbiology, 72* (7), 5069–5072, 2006.

Lin, Li-Hung, Pei-Ling Wang, Douglas Rumble, Johanna Lippman-Pipke, Erik Boice, Lisa M. Pratt, Barbara Sherwood Lollar, Eoin Brodie, Terry C. Hazen, Gary L. Andersen, Todd Z. DeSantis, Duane P. Moser, Dave Kershaw, and T.C. Onstott, Long-term sustainability of a high-energy, low-diversity crustal biome, *Science, 314*, 479–482, 2006.

Urban aerosols harbor diverse and dynamic bacterial populations

Eoin L. Brodie, Todd Z. DeSantis, Jordan P. Moberg Parker, Ingrid X. Zubietta, Yvette M. Piceno, and Gary L. Andersen*

Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Edited by Steven E. Lindow, University of California, Berkeley, CA, and approved November 7, 2006 (received for review September 20, 2006)

Considering the importance of its potential implications for human health, agricultural productivity, and ecosystem stability, surprisingly little is known regarding the composition or dynamics of the atmosphere's microbial inhabitants. Using a custom high-density DNA microarray, we detected and monitored bacterial populations in two U.S. cities over 17 weeks. These urban aerosols contained at least 1,800 diverse bacterial types, a richness approaching that of some soil bacterial communities. We also reveal the consistent presence of bacterial families with pathogenic members including environmental relatives of select agents of bioterrorism significance. Finally, using multivariate regression techniques, we demonstrate that temporal and meteorological influences can be stronger factors than location in shaping the biological composition of the air we breathe.

16S rRNA | biosurveillance | aerobiology | microarray | climate change

Low levels of moisture and nutrients combined with high levels of UV radiation make the earth's atmosphere an extreme environment for microbial life. Little is known regarding the atmospheric microbial composition and how it varies by location or meteorological conditions. Plant canopies for example, are known to be significant sources of bacterial aerosols with upward flux of bacteria positively impacted by temperature and wind speed (1). Aerosols created at the surface of aquatic systems are known to concentrate and carry bacteria through the liquid–air interface (2, 3). The relationship between environmental conditions and bacterial aerial dispersal indicates that climate change could potentially alter the microbial composition of downwind areas, resulting in increased health risk from pathogens or allergenic components of unclassified environmental bacteria. For instance, the last decade has seen a dramatic increase in the amount of desertification and a concomitant increase in upper atmospheric particulates (4). In sub-Saharan regions of Africa, dust storms have been associated with regional outbreaks of meningococcal meningitis caused by the bacterium *Neisseria meningitidis* (5). Since the 1970s, El Nino weather events have coincided with increased flux of African Dust across the Atlantic (4) that, in turn, has been linked to coral reef disease (6) and increased exacerbations of pediatric asthma (7) in the Caribbean. Therefore, as particles from dust storms shield bacterial and fungal passengers from the inactivating effects of UV exposure, global transport of dust will have more far-reaching affects than impaired visibility.

The consequences of natural environmental variation such as meteorological shifts, combined with anthropogenic influences such as land use changes, may alter atmospheric microbial composition. To monitor the effects of climate change on aerosol microbial composition, it first is necessary to establish baselines that acknowledge the current microbial components and how they fluctuate naturally. However, the potential heterogeneity, both spatial and temporal, in species composition coupled with low microbial biomass ensures this is not a facile task.

Natural shifts in bacterial composition also have implications for atmospheric pathogen monitoring systems, such as the Department of Homeland Security effort to monitor major U.S. cites for intentional release of biowarfare agents (www.ostp.gov/html/10-20-03%20jhm%20BioSecurity%202003.pdf). Many such pathogens and other closely related bacteria with undefined pathogenicity already are endemic to the locations that are being monitored (8) and so may interfere with detection networks (9), but little is known regarding the frequency or variability of their occurrence. Most aerobiology studies to date (e.g., refs. 10–12), have used culture-based methods for determining microbial composition. Although some studies recently have applied culture-independent techniques (e.g., refs. 13 and 14), little is known of what constitutes the breadth of diversity of "typical" organisms in the atmosphere (as opposed to those capable of growth in laboratory media) and what influences their composition. To address these methodological limitations and to augment our view of aerosol microbial diversity and dynamics, we have designed a microarray (PhyloChip) for the comprehensive identification of both bacterial and archaeal organisms. We target the variation in the 16S rRNA gene, possessed by all prokaryotes, to capture the broad range of microbial diversity that may be present in the atmosphere. This tool allows bacteria and archaea to be identified and monitored in any type of sample without the need for microbial cultivation.

The two greatest obstacles to designing a 16S rRNA gene-based microarray to identify individual organisms in a complex environmental mixture are natural sequence diversity and potential cross-hybridization. Sequence diversity is an issue as we sample new and distinctive environments such as the atmosphere. There may be many undocumented organisms with 16S rRNA gene sequences that are similar, but not identical, to the sequences that were used for array design. Microarrays based on single sequence-specific hybridizations (single probes) may be ineffective in detecting such environmental sequences with one or several polymorphisms. To overcome this obstacle, we have designed a minimum of 11 different, short oligonucleotide probes for each taxonomic grouping, allowing for the failure of one or more probes. On the other hand, nonspecific cross-hybridization is an issue when an abundant 16S rRNA gene shares sufficient sequence similarity to nontargeted probes, such that a weak but detectable signal is obtained. We have found that the perfect match (PM)-mismatch (MM) probe pair approach effectively minimizes the influence of cross-hybridization. Widely used on expression arrays as a control for nonspecific binding (15), the central nucleotide is replaced with any of the

Author contributions: E.L.B. and T.Z.D. contributed equally to this work; E.L.B., T.Z.D., and G.L.A. designed research; E.L.B., T.Z.D., J.P.M.P., I.X.Z., and Y.M.P. performed research; E.L.B., T.Z.D., Y.M.P., and G.L.A. analyzed data; and E.L.B., T.Z.D., and G.L.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ129237–DQ129666, DQ236245–DQ236250, and DQ515230–DQ515231).

*To whom correspondence should be addressed. E-mail: glandersen@lbl.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0608255104/DC1.

MICROBIOLOGY

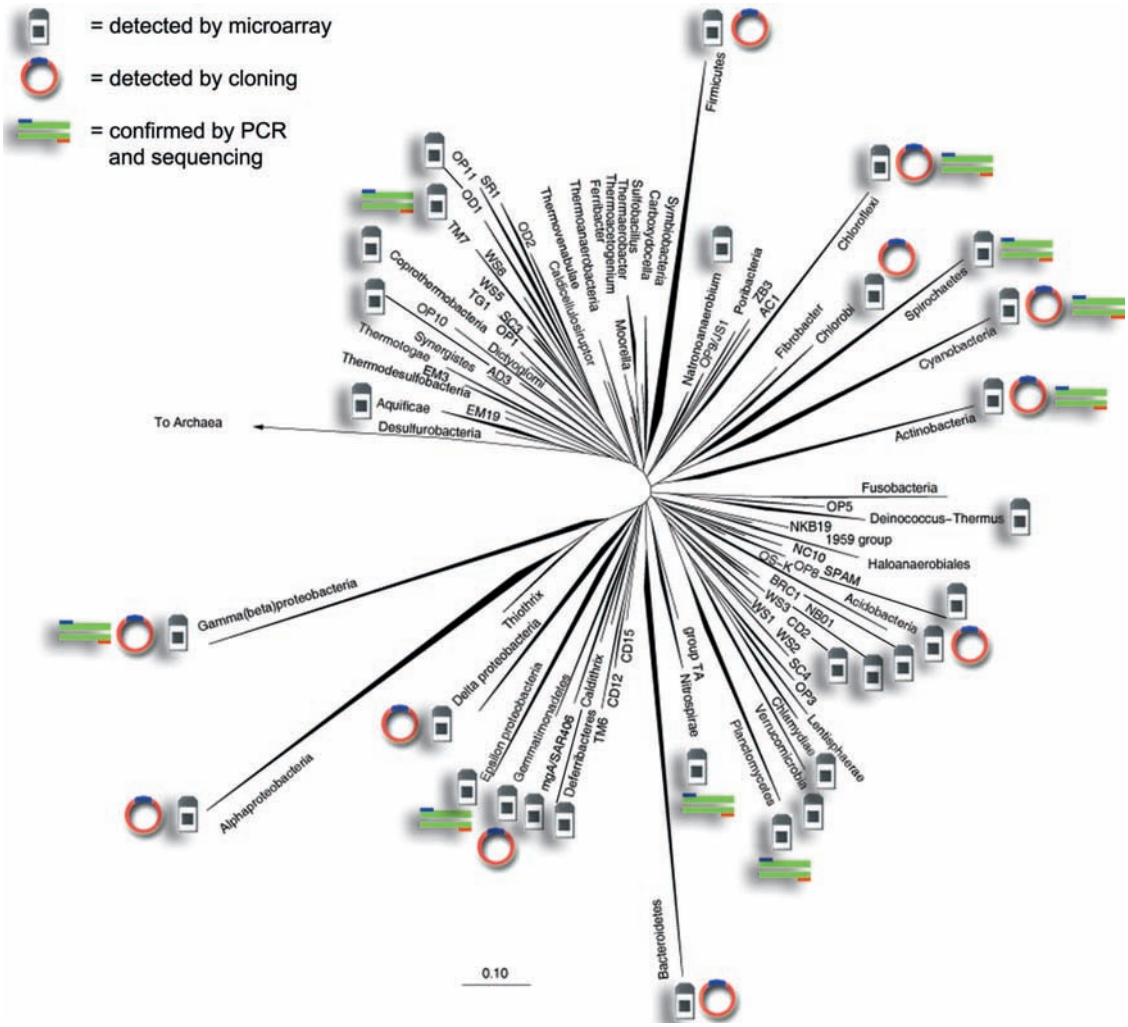


Fig. 1. Representative phylogenetic tree showing all known bacterial phyla (and individual classes in the case of proteobacteria) annotated to show 16S rRNA gene sequences detected in an urban aerosol by both microarray and cloning. Also annotated are phyla detected by microarray only that subsequently were confirmed by targeted PCR and sequencing. The Archaea are used as an outgroup. (Scale bar: 0.1 changes per nucleotide.)

three nonmatching bases so that the increased hybridization intensity signal of the PM over the paired MM indicates a sequence-specific, positive hybridization. By requiring multiple PM-MM probe pairs to have a positive interaction, we substantially increase the chance that the hybridization signal is due to a predicted target sequence.

We grouped known 16S rRNA gene sequences >600 bp into distinct taxa such that a set of at least 11 probes that were specific to the taxon could be chosen. The resulting 8,935 taxa (8,741 of which are represented on the PhyloChip), each containing ≈3% sequence divergence, represented all 121 demarcated bacterial and archaeal orders [supporting information (SI) Table 2]. For a majority of the taxa represented on the PhyloChip (5,737, 65%), probes were designed from regions of gene sequences that have been identified only within a given taxon. For 1,198 taxa (14%), no probe-level sequence could be identified that was not shared with other groups of 16S rRNA gene sequences, although the gene sequence as a whole was distinctive. For these taxonomic groupings, a set of at least 11 probes was designed to a combination of regions on the 16S rRNA gene that taken together as a whole did not exist in any other taxa. For the remaining 1,806 taxa (21%), a set of probes were selected to minimize the number of putative cross-reactive taxa. Although more than half of the probes in this group have a hybridization potential to one outside sequence, this sequence was typically from a phylogenetically similar taxon. For all three probe set groupings, the advantage of the hybridization approach is that multiple taxa can be identified simultaneously by targeting unique regions or combinations of sequence.

To assess the bacterial composition of environmental aerosols and how it changes over time and with location, we examined outdoor air collected at multiple locations in two cities, Austin and San Antonio, TX. These cities are part of the U.S. multiagency biosurveillance effort that use aerosol collectors to concentrate airborne particulate matter in search of pathogens that potentially could be indicative of a bioterrorism threat. For either city, aerosol monitors were used to draw in air and pass it through filters designed to collect submicrometer particulates for a 24-h period. The samplers were placed immediately adjacent to six Environmental Protection Agency air quality monitoring stations located throughout the urban area of each city, and the filter eluents were pooled for each day before amplification of the 16S rRNA gene products from the extracted DNA. Although PCR amplification may introduce some bias in terms of quantitative assessment of an organism's abundance due to factors such as preferential amplification (16, 17), the extremely low bacterial biomass in aerosol samples necessitate such an approach. Amplified products from 4 days within a 7-day period

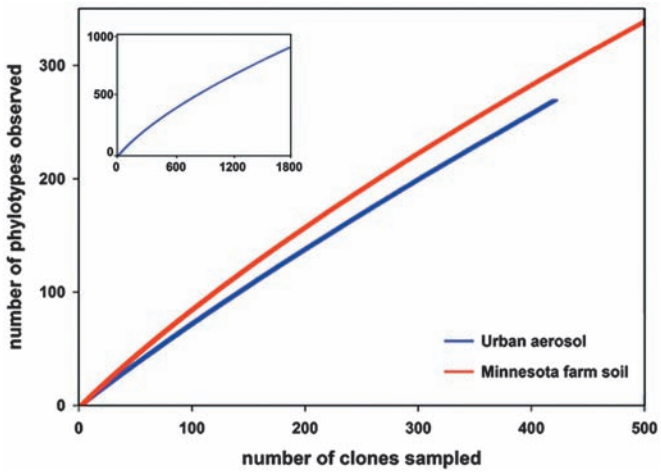


Fig. 2. Rarefaction curves comparing bacterial diversity in a Minnesota farm soil (20) and the urban aerosol in this study. (*Inset*) Complete rarefaction curve for 1,874 sequences from the Minnesota farm soil library.

were pooled into a single sample representing 1 week, and 17 consecutive weekly samples beginning May 2003 were analyzed from both cities for bacterial composition.

Results and Discussion

PhyloChip results for one sample, representing bacteria recovered from outdoor air at San Antonio from the week of July 14–20 (calendar week 29), 2003, were compared with clone library sequence results from the same pool of amplified 16S rRNA gene products (Fig. 1). A conservative comparison of the PhyloChip and cloning approaches was made at a taxonomic level below family and above species classification (see *SI Materials and Methods*), termed "subfamily" for clarity. This demonstrated that the PhyloChip correctly detected 90% of cloned subfamilies (SI Table 3) and additionally detected almost 2.5-fold more diversity at the phylum level (Fig. 1). We subsequently have confirmed many of these PhyloChip-only hits (which include known pathogenic genera), by cluster-specific PCR and sequencing (Fig. 1 and SI Table 4). The most common sequences in the air clone library (35%) were *Bacilli* most similar to the species *Bacillus bataviensis* (previously isolated from soil in a disused hay field) (18) and another *Bacillus* sp. associated with biodeterioration of mural paintings (19), suggesting dispersal through aerosolization. The diversity of the remaining clone sequences was quite high, with a majority of the clones representing distinctive 16S rRNA gene sequences (SI Fig. 4).

Because of the relative dearth of information regarding aerosol bacterial diversity, we compared the diversity detected in this aerosol sample by cloning with that found in a farm soil from a previous study (20). Soils are considered to be highly diverse microbial habitats with an estimate of up to 1 million distinct genomes per gram (21). Rarefaction analysis revealed a similar level of diversity (at the 16S rRNA gene biomarker level) in the aerosol and soil samples (Fig. 2). Predicted estimates of richness (Chao1 and ACE) indicated between 1,500 and 1,800 16S rRNA phylotypes (by using a 99% identity cutoff) in the aerosol sample (SI Fig. 5). However, because both ACE and Chao1 richness prediction curves were nonasymptotic, this is likely to be an underestimate because of insufficient clone sampling, a common problem when assessing environmental microbial diversity by using cloning approaches.

Microbial communities are characteristically dynamic, and it is expected that aerosol communities are no exception, considering the turbulent and well mixed nature of the atmosphere. Using a Latin Square type study containing mixtures of amplicons from diverse bacterial species applied to the PhyloChips in rotating concentrations (SI Table 5), we tested the ability of the PhyloChip to track 16S rRNA amplicon dynamics quantitatively. This demonstrated a strong linear relationship, spanning five orders of magnitude between PhyloChip intensities and quantities of bacterial 16S rRNA gene signatures applied to PhyloChips (SI Fig. 6). Having determined the potential of the PhyloChip for detecting changes in biomarker quantities, we analyzed intensity data for the two cities over the 17-week period of the study. We also collated a range of meteorological parameters to investigate whether local weather conditions could be correlated with the observed fluctuations in aerosol bacterial populations. Using multivariate regression tree analysis (22, 23), we examined such correlations, with tree topology and splitting parameters suggesting that sample location (in this case two geographically proximate cities) was less of a factor in explaining the variability of aerosol bacterial composition than temporal or meteorological influences (Fig. 3). The week of the year at which a sample was taken proved to be a stronger predictor of community composition with samples taken after the first three weeks in May (weeks 19–21) clustering separately from those taken before this, regardless of city sampled. Unsurprisingly, sample week was observed to correlate with weather conditions (SI Table 6). For both cities, week was positively correlated with temperature, air pressure, and visibility, whereas negatively correlated with wind speed and particulate matter. It is important to note from the composition of the PhyloChip generated tree clusters that the clone library "snapshot" taken during week 29 in San Antonio was representative of only approximately one-third of the samples collected (11 of 34 weeks clustered at this node). Therefore, caution should be used in interpreting snapshot analyses in such dynamic systems. Underlying these changes in bacterial community composition was a differential abundance of biomarkers for many spore-forming bacteria such as Actinomycetes and Firmicutes. Indeed, most of the taxa with significant correlations to weather conditions were Actinomycetes, which showed positive correlations with temperature (SI Table 7). Warmer temperatures may result in increased desiccation of soil/plant-based bacteria, leading to spore dispersal or aerosolization. Additionally, alpha-proteobacteria such as phyllosphere-inhabiting *Sphingomonas* spp (24). were correlated with sea-level pressure, week, and temperature (Fig. 3 and SI Table 7). These and other alpha-proteobacteria are typically oligotrophic and also may originate from freshwater and marine ecosystems (3) in addition to plant surfaces. The most significant correlation between any PhyloChip intensity pattern and an environmental/temporal variable was between the gammaproteobacterium *Pseudomonas oleovorans* and week ($r = 0.83$, $P = 2.1 \times 10^{-5}$). Real-time quantitative PCR of the same genomic DNA pools used for PhyloChip PCRs demonstrated that changes in PhyloChip intensity were representative of the dynamics of this organism (SI Fig. 7).

Despite the variable nature of the aerosol bacterial population, we detected some groups of organisms in every sample over the 17-week period (summarized in Table 1, with complete details in SI Table 8). Between the two cities, more types of bacteria consistently were detected in San Antonio aerosol samples (80 subfamilies) compared with Austin (43 subfamilies), although there was significant overlap in the consistent 16S rRNA signatures between the two cities. Many of these organisms (e.g., Acidobacteria and Verrucomicrobia) are major components of the soil microbiota and may be particle-associated. *Sphingomonas* species also were detected consistently, psychrotolerant strains of which have been detected in dust and air samples from the Antarctic (25). Notably, other bacteria consistently detected were spore formers such as the endospore-forming Bacilli and Clostridia and the exospore-forming Actinomycetes. Cyanobacteria such as *Plectonema* were also

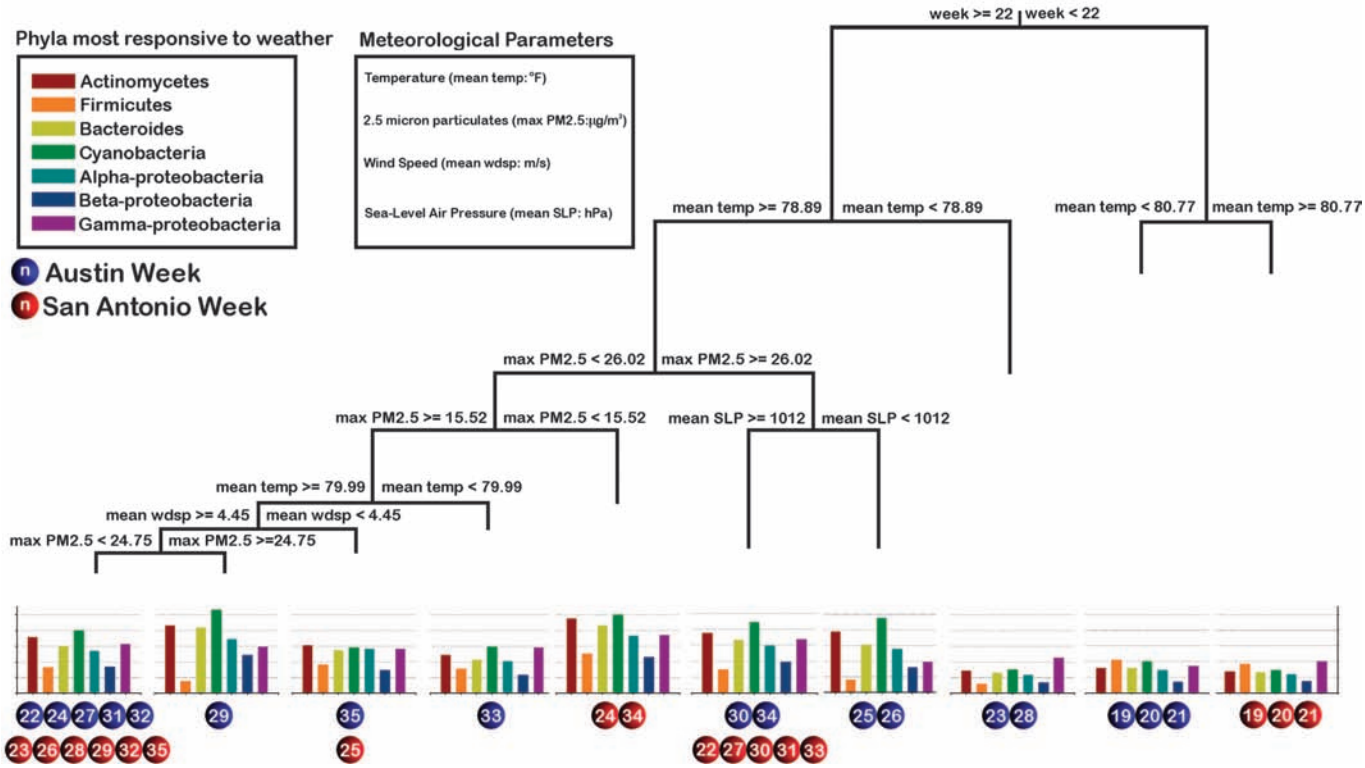


Fig. 3. Multivariate regression tree analysis of the interaction between aerosol bacterial dynamics (array intensity) and environmental parameters. The model explains 89.1% of variance in SI Data Set 1. Bars plotted under each cluster represent mean of normalized array intensities of phylogenetically related bacteria shown to be significantly correlated with environmental/temporal parameters.

frequently detected members of the aerosol community, as were plant chloroplasts (presumably from pollen). Significantly, epsilon proteobacteria were consistently detected by PhyloChip in both cities, including organisms within the families *Campylobacteraceae* and *Helicobacteraceae*, both of which contain human and animal pathogens. The exact *Campylobacteraceae* taxon detected by the PhyloChip contains the genus *Arcobacter*, whose presence we subsequently confirmed by taxon-specific PCR and sequencing (SI Table 4). This genus is known to cause bacteremia and severe gastrointestinal illnesses in humans, and together with *Helicobacter* (a causative agent of gastric ulcers), could be considered indicators of fecal contamination (26, 27), which is known to occur through aerosolization from wastewater treatment plants (11, 28).

Table 1. Bacterial groups detected in all weeks during sampling period

Phylum, class	San Antonio	Austin
Acidobacteria, Acidobacteria	Y	Y
Acidobacteria, Acidobacteria-6	Y	N
Acidobacteria, Solibacteres	Y	Y
Actinobacteria, Actinobacteria	Y	Y
Actinobacteria, BD2–10 group	N	Y
Bacteroidetes, Sphingobacteria	Y	N
Chloroflexi, Anaerolineae	Y	N
Chloroflexi, Dehalococcoidetes	Y	N
Cyanobacteria, Cyanobacteria/Chloroplasts	Y	Y
Firmicutes, Bacilli	Y	Y
Firmicutes, Catabacter	Y	N
Firmicutes, Clostridia	Y	Y
Nitrospira, Nitrospira	N	Y
OP3, Unclassified	Y	N
Proteobacteria, Alphaproteobacteria	Y	Y
Proteobacteria, Betaproteobacteria	Y	Y
Proteobacteria, Gammaproteobacteria	Y	Y
Proteobacteria, Deltaproteobacteria	Y	N
Proteobacteria, Epsilonproteobacteria	Y	Y
TM7, TM7–3	Y	N
Verrucomicrobia, Verrucomicrobiae	Y	Y

Subfamilies detected are summarized to phylum and class level. Y, detected in each of 17 weeks sampled per city; N, not detected in every sample.

The consistent detection of signatures from potentially pathogenic bacteria led us to examine taxonomic clusters containing other pathogens (and their relatives) of public health and bioterrorism significance over the 17-week period (SI Table 9). Environmental relatives of monitored pathogens have already been implicated in multiple detection events in U.S. Homeland Security monitoring systems (www.houstontx.gov/health/NewsReleases/bacteria%20detection.htm). In fact, in response to such a detection event, a recent survey of soils in Houston was carried out to determine potential reservoirs of environmental relatives (10). This study revealed a surprising diversity of *Francisella*-like organisms that may have been responsible for triggering detectors in the aerosol monitoring systems. Similarly, in the aerosol samples analyzed here, we detected taxonomic clusters containing organisms closely related to *Francisella* in one week in Austin and two weeks in San Antonio, although the causative agent of tularemia, *Francisella tularensis*, was never encountered. We also consistently detected phylogenetic near-neighbors to *Bacillus anthracis* with the taxonomic cluster containing *B. anthracis* itself (also containing common soil relatives *B. cereus*, *B. thuringiensis*, and *B. mycoides*) being detected in one week in San Antonio. Tick-borne *Rickettsia* and *Clostridium botulinum* types C (causes illness in mammals, fish, and birds)

and G (rarely illness causing) also were detected regularly, as were *Burkholderia mallei* and *Bu. pseudomallei*, which cause glanders and melioidosis respectively. Other select agents such as *Yersinia pestis* and *Brucella* spp. (*melitensis*, *suis*, and *abortus*) were never encountered. The frequent occurrence of environmental relatives of bacteria targeted by biosurveillance efforts in urban aerosols makes prediction of natural occurrences of endemic pathogens or their uncharacterized environmental relatives critical for the implementation of a robust biosurveillance network.

This study represents a comprehensive molecular analysis of airborne bacterial composition and dynamics. We have demonstrated that the atmosphere contains a diverse assemblage of microorganisms probably representing the amalgamation of numerous point sources. The composition of this habitat varies widely and may be subject to climatic regulation. A global-scale study of this uncharacterized ecosystem is necessary to determine baselines for bioaerosol transport patterns. Such data will enable an understanding of future anthropogenic impacts including pollution, bioterrorism, and climate change in altering the biological composition of the air we breathe.

Materials and Methods

Sample Collection and Pooling. Air samples were collected by using an air filtration collection system under vacuum located within six Environmental Protection Agency air quality network sites in both San Antonio and Austin. Approximately 10 liters of air per minute were collected on a Celanex polyethylene terephthalate, 1.0-μm filter (Calanese, Dallas, TX). Samples were collected daily over a 24-h period. Sample filters were washed in 10 ml buffer (0.1 M sodium phosphate/10 mM EDTA, pH 7.4/0.01% Tween-20), and the suspension was stored frozen until extracted. Samples were collected from 4 May to 29 August 2003. Sample dates were divided according to a 52-week calendar year starting January 1, 2003, with each Monday-to-Sunday cycle constituting a full week. Samples from four randomly chosen days within each sample week were extracted. Each date chosen for extraction consisted of a 0.6-ml filter wash from each of the six sampling sites for that city (San Antonio or Austin) combined into a "day pool" before extraction. In total, for each week, 24 filters were sampled.

DNA Extraction and 16S rRNA Gene Amplification. The "day pools" were centrifuged at 16,000 $\times$ *g* for 25 min, and the pellets were resuspended in 400 μl of 100 mM sodium phosphate buffer (pH 8). DNA extraction was performed as described in DeSantis *et al.* (29), but only a single bead-beating velocity and duration was used (6.5 $m \cdot s^{-1}$ for 45 s). DNA was quantified by using a PicoGreen fluorescence assay according to the manufacturer's recommended protocol (Invitrogen, Carlsbad, CA). 16S rRNA gene amplification was performed according to standard procedures as outlined in *SI Materials and Methods*.

PhyloChip Processing, Scanning, Probe Set Scoring, and Normalization. The pooled PCR product was spiked with known concentrations of synthetic 16S rRNA gene fragments and non-16S rRNA gene fragments as internal standards for normalization with quantities ranging from 5.02×10^8 and 7.29×10^{10} molecules applied to the final hybridization mix (SI Table 10). Target fragmentation, biotin labeling, PhyloChip hybridization, scanning, and staining were as described by Brodie *et al.* (30), and background subtraction, noise calculation, and detection and quantification criteria were essentially as reported in Brodie *et al.* (30), with some minor exceptions. These exceptions were as follows: For a probe pair to be considered positive, the difference in intensity between the PM and MM probes must be at least 130 times the squared noise value (N). A taxon was considered present in the sample when 92% or more of its assigned probe pairs for its corresponding probe set were positive (positive fraction $> = 0.92$). This was determined based on empirical data from clone library analyses. Hybridization intensity (referred to as intensity) was calculated in arbitrary units for each probe set as the trimmed average (maximum and minimum values removed before averaging) of the PM minus MM intensity differences across the probe pairs in a given probe set. All intensities <1 were shifted to 1 to avoid errors in subsequent logarithmic transformations. When summarizing PhyloChip results to the subfamily, the probe set producing the highest intensity was used.

Validation of PhyloChip Detection of Airborne Bacteria by Comparison with Clone Library. To compare the diversity of bacteria detected with PhyloChips to a known standard, one sample week was chosen for cloning and sequencing and replicate PhyloChip analysis. One large pool of SSU amplicons (96 reactions, 50 μl per reaction) from San Antonio week 29 was made. One milliliter of the pooled PCR product was gel-purified, and 768 clones were sequenced at the DOE Joint Genome Institute (Walnut Creek, CA) by standard methods. An aliquot of this same pooled PCR product also was hybridized to a PhyloChip (three replicate PhyloChips performed). Subfamilies containing a taxon scored as present in all three PhyloChip replicates were recorded. Individual cloned rRNA genes were sequenced from each terminus, assembled by using Phred and Phrap (31–33), and were required to pass quality tests of Phred 20 (base call error probability $<10^{-2.0}$) to be included in the comparison. Chimeric sequences were removed after Bellerophon (34) analysis, and similarity of clones to PhyloChip taxa was calculated with DNADIST (35) measurement of homology (DNAML-F84) over 1,287 conserved columns identified by using the Lane mask (36). Sequences were assigned to a taxonomic node by using a sliding scale of similarity threshold (37). These steps are described in detail in *SI Materials and Methods*, and a full comparison between clone and PhyloChip analysis is available in SI Table 3.

Validation of PhyloChip-Detected Subfamilies Not Supported by the Clone Library. Primers targeting sequences within particular taxa/subfamilies were generated by using ARB's probe design feature (38) and based on regions targeted by PhyloChip probes or were obtained from published literature (SI Table 4). Primer quality control was carried out by using Primer3 (39).

Quantitative Detection of Changes in 16S rRNA Gene Concentration in Heterogeneous Solutions. To determine whether changes in 16S rRNA gene concentration could be detected by using the PhyloChip, various quantities of distinct rRNA gene types were hybridized to the PhyloChip in rotating combinations. We chose environmental organisms, organisms involved in bioremediation, and a pathogen of biodefense relevance. 16S rRNA genes were amplified from each of the organisms shown in SI Table 5. Then each of these nine distinct 16S rRNA gene standards was tested once in each concentration category, spanning five orders of magnitude (0 molecules, 6×10^7, 1.44×10^8, 3.46×10^8, 8.30×10^8, 1.99×10^9, 4.78×10^9, 2.75×10^{10}, 6.61×10^{10}, and 1.59×10^{11}) with concentrations of individual 16S rRNA gene types rotating between PhyloChips such that each PhyloChip contained the same total of 16S rRNA gene molecules. This is similar to a Latin Square design, although with a 9×11 format matrix.

Real-Time Quantitative PCR Confirmation of PhyloChip-Observed Shifts in Taxon Abundance. A taxon (no. 9389) consisting only of two sequences of *Pseudomonas oleovorans* that correlated well with environmental variables was chosen for quantitative PCR confirmation of PhyloChip-observed quantitative shifts. Primers for this taxon were designed by using the ARB (38) probe match function to determine unique priming sites based on regions

detected by PhyloChip probes. These regions then were imputed into Primer3 (39) to choose optimal oligonucleotide primers for PCR. Primer quality was assessed further by using Beacon Designer v3.0 (Premier BioSoft, CA). Primers 9389F2 (CGACTACCTGGACTGACACT) and 9389R2 (CACCGGCAGTCTCCTTAGAG) were chosen to amplify a 436-bp fragment. Validation of primer specificity and reaction conditions are available in *SI Materials and Methods*.

Statistical Analyses. All statistical operations were performed in the R software environment (ref. 40; www.R-project.org). For each day of aerosol sampling, 15 factors including humidity, wind, temperature, precipitation, pressure, particulate matter, and week of year were recorded from the U.S. National Climatic Data Center (www.ncdc.noaa.gov) or the Texas Natural Resource Conservation Commission (www.tceq.state.tx.us). The weekly mean, minimum, maximum, and range of values were calculated for each factor from the collected data. The changes in ln(intensity) for each PhyloChip taxon considered present in the study was tested for correlation against the environmental conditions. The resulting *P* values were adjusted by using the step-up false discovery rate controlling procedure (41).

Multivariate regression tree analysis (22, 23) was carried out by using the package "mvpart" within the "R" statistical programming environment. A Bray-Curtis-based distance matrix was created by using the function "gdist." The Bray-Curtis measure of dissimilarity is generally regarded as a good measure of ecological distance when dealing with "species" abundance, or in this case, array probe-set intensity, because it allows for nonlinear responses to environmental gradients (22, 42). Large trees were calculated with splitting based on information gain and then pruned (from 13 to 10 nodes) based on 100 cross-validations to a complexity parameter of 0.025286, where cross-validation relative error had reached a plateau.

Before clone library rarefaction analysis, a distance matrix (DNAML homology) of clone sequences, was created by using an online tool at http://greengenes.lbl.gov/cgi-bin/nph-distance_matrix.cgi (43) after alignment of the sequences by using the NAST aligner (http://greengenes.lbl.gov/NAST) (44). DOTUR (45) was used to generate rarefaction curves, Chao1, and ACE richness predictions and rank-abundance curves. Nearest neighbor joining was used with 1,000 iterations for bootstrapping.

We thank Dr. Phil Hugenholtz and Dr. Paul Richardson of the Joint Genome Institute for clone library sequencing; Susannah Green Tringe for providing the soil 16S rRNA gene sequences; Sonya Murray for expert technical assistance; John Coates, Lisa Alvarez-Cohen (both of University of California, Berkeley, CA), Hoi-Ying Holman, Terry Hazen (both of Lawrence Berkeley National Laboratory), and Arthur Friedlander (U.S. Army Medical Research Institute of Infectious Diseases, Frederick, MD) for the generous gifts of bacterial cultures or DNA; and Sue Lynch, Terry Hazen, Jill Banfield, Tamas Torok, and two anonymous reviewers for helpful suggestions and comments on the manuscript. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Berkeley National Laboratory, under Contract DE-AC02-05CH11231 and was supported in part by Department of Homeland Security Grant HSSCHQ04X00037 and the Climate Change Research Division, Biological and Environmental Research, Office of Science, U.S. Department of Energy. Computational support was provided through the Virtual Institute for Microbial Stress and Survival.

1. Lindemann J, Upper CD (1985) *Appl Environ Microbiol* 50:1229–1232.
2. Baylor ER, Peters V, Baylor MB (1977) *Science* 197:763–764.
3. Aller JY, Kuznetsova MR, Jahns CJ, Kemp PF (2005) *J Aerosol Sci* 36:801–812.
4. Shinn EA, Griffin DW, Seba DB (2003) *Arch Environ Health* 58:498–504.
5. World Health Organization (2003) *Meningococcal Meningitis Fact Sheet* (World Health Org, Geneva).
6. Shinn EA, Smith GW, Prospero JM, Betzer P, Hayes ML, Garrison V, Barber RT (2000) *Geophys Res Lett* 27:3029–3032.
7. Gyan K, Henry W, Lacaille S, Laloo A, Lamsee-Ebanks C, McKay S, Antoine RM, Monteil MA (2005) *Int J Biometeorol* 49:371–376.
8. Anda P, del Pozo JS, Garcia JMD, Escudero R, Pena FJG, Velasco MCL, Sellek RE, Chillaron MRJ, Serrano LPS, Navarro JFM (2001) *Emerg Infect Dis* 7:575–582.
9. Barns SM, Grow CC, Okinaka RT, Keim P, Kuske CR (2005) *Appl Environ Microbiol* 71:5494–5500.
10. Bauer H, Fuerhacker M, Zibuschka F, Schmid H, Puxbaum H (2002) *Water Res* 36:3965–3970.
11. Griffin DW (2004) *Aerobiologia* 20:135–140.
12. Lee T, Grinshpun SA, Martuzevicius D, Adhikari A, Crawford CM, Luo J, Reponen T (2006) *Indoor Air* 16:37–47.
13. Hughes KA, McCartney HA, Lachlan-Cope TA, Pearce DA (2004) *Cell Mol Biol* 50:537–542.
14. Maron PA, Lejon DPH, Carvalho E, Bizet K, Lemanceau P, Ranjard L, Mougel C (2005) *Atmos Environ* 39:3687–3695.
15. Mei R, Hubbell E, Bekiranov S, Mittmann M, Christians FC, Shen M-M, Lu G, Fang J, Liu W-M, Ryder T, *et al.* (2003) *Proc Natl Acad Sci USA* 100:11237–11242.
16. Polz MF, Cavanaugh CM (1998) *Appl Environ Microbiol* 64:3724–3730.
17. Lueders T, Friedrich MW (2003) *Appl Environ Microbiol* 69:320–326.
18. Heyrman J, Vanparys B, Logan NA, Balcaen A, Rodriguez-Diaz M, Felske A, De Vos P (2004) *Int J Syst Evol Microbiol* 54:47–57.
19. Heyrman J, Balcaen A, Rodriguez-Diaz M, Logan NA, Swings J, De Vos P (2003) *Int J Syst Evol Microbiol* 53:459–463.
20. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter, JC, *et al.* (2005) *Science* 308:554–557.
21. Gans J, Wolinsky M, Dunbar J (2005) *Science* 309:1387–1390.
22. De'Ath G (2002) *Ecology* 83:1105–1117.
23. Larsen DR, Speckman PL (2004) *Biometrics* 60:543–549.
24. Kim H, Nishiyama W, Kunito T, Senoo K, Kawahara K, Murakami K, Oyaizu H (1998) *J Appl Microbiol* 85:731–736.
25. Busse HJ, Denner EBM, Buczolits S, Salkinoja-Salonen M, Bennasar A, Kampfer P (2003) *Int J Syst Evol Microbiol* 53:1253–1260.
26. Wesley IV, Wells SJ, Harmon KM, Green A, Schroeder-Tucker L, Glover M, Siddique I (2000) *Appl Environ Microbiol* 66:1994–2000.
27. Engberg J, On SLW, Harrington CS, Gerner-Smidt P (2000) *J Clin Microbiol* 38:286–291.
28. Hughes KA (2003) *Atmos Environ* 37:3147–3155.
29. DeSantis TZ, Stone CE, Murray SR, Moberg JP, Andersen GL (2005) *FEMS Microbiol Lett* 245:271–278.
30. Brodie EL, DeSantis TZ, Joyner DC, Baek S, Larsen JT, Andersen GL, Hazen TC, Richardson PM, Herman DJ, Tokunaga TK, *et al.* (2006) *Appl Environ Microbiol* 72:6288–6298.
31. Ewing B, Hillier L, Wendl MC, Green P (1998) *Genome Res* 8:175–185.
32. Ewing B, Green P (1998) *Genome Res* 8:186–194.
33. Gordon D, Abajian C, Green P (1998) *Genome Res* 8:195–202.
34. Huber T, Faulkner G, Hugenholtz P (2004) *Bioinformatics* 20:2317–2319.
35. Felsenstein J (1989) *Cladistics* 5:164–166.
36. Lane DJ (1991) in *Nucleic Acid Techniques in Bacterial Systematics*, eds Stackebrandt E, Goodfellow M (Wiley, New York), pp 115–175.
37. Schloss PD, Handelsman J (2004) *Microbiol Mol Biol Rev* 68:686–691.
38. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, *et al.* (2004) *Nucleic Acids Res* 32:1363–1371.
39. Rozen S, Skaletsky H (2000) in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, eds Krawetz S, Misener S (Humana, Totowa, NJ), pp 365–386.
40. R Development Core Team (2005) *R: A Language and Environment for Statistical Computing* (R Found Stat Comput, Vienna).
41. Benjamini Y, Hochberg Y (1995) *J R Stat Soc B* 57:289–300.
42. Faith DP, Minchin PR, Belbin L (1987) *Vegetatio* 69:57–68.
43. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) *Appl Environ Microbiol* 72:5069–5072.
44. DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R, Andersen GL (2006) *Nucleic Acids Res* 34:W394–W399.
45. Schloss PD, Handelsman J (2005) *Appl Environ Microbiol* 71:1501–1506.

JOURNAL OF CLINICAL MICROBIOLOGY, June 2007, p. 1954–1962 Vol. 45, No. 6
0095-1137/07/$08.00+0 doi:10.1128/JCM.02187-06

Loss of Bacterial Diversity during Antibiotic Treatment of Intubated Patients Colonized with *Pseudomonas aeruginosa*▽

J. L. Flanagan,[1]† E. L. Brodie,[2]† L. Weng,[3] S. V. Lynch,[1] O. Garcia,[1] R. Brown,[1] P. Hugenholtz,[3] T. Z. DeSantis,[2] G. L. Andersen,[2] J. P. Wiener-Kronish,[1] and J. Bristow[3]*

Department of Anesthesia and Perioperative Care, University of California, San Francisco, California 94143[1]; Earth Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720[2]; and DOE Joint Genome Institute, 2800 Mitchell Drive, Bldg. 400-404, Walnut Creek, California 94598[3]

Received 25 October 2006/Returned for modification 8 January 2007/Accepted 26 March 2007

Management of airway infections caused by *Pseudomonas aeruginosa* is a serious clinical challenge, but little is known about the microbial ecology of airway infections in intubated patients. We analyzed bacterial diversity in endotracheal aspirates obtained from intubated patients colonized by *P. aeruginosa* by using 16S rRNA clone libraries and microarrays (PhyloChip) to determine changes in bacterial community compositions during antibiotic treatment. Bacterial 16S rRNA genes were absent from aspirates obtained from patients briefly intubated for elective surgery but were detected by PCR in samples from all patients intubated for longer periods. Sequencing of 16S rRNA clone libraries demonstrated the presence of many orally, nasally, and gastrointestinally associated bacteria, including known pathogens, in the lungs of patients colonized with *P. aeruginosa*. PhyloChip analysis detected the same organisms and many additional bacterial groups present at low abundance that were not detected in clone libraries. For each patient, both culture-independent methods showed that bacterial diversity decreased following the administration of antibiotics, and communities became dominated by a pulmonary pathogen. *P. aeruginosa* became the dominant species in six of seven patients studied, despite treatment of five of these six with antibiotics to which it was sensitive in vitro. Our data demonstrate that the loss of bacterial diversity under antibiotic selection is highly associated with the development of pneumonia in ventilated patients colonized with *P. aeruginosa*. Interestingly, PhyloChip analysis demonstrated reciprocal changes in abundance between *P. aeruginosa* and the class *Bacilli*, suggesting that these groups may compete for a similar ecological niche and suggesting possible mechanisms through which the loss of microbial diversity may directly contribute to pathogen selection and persistence.

Mechanically ventilated patients develop bacterial colonization of the oropharynx and endotracheal tube within 12 h of intubation (14, 36). During mechanical ventilation, the endotracheal tube decreases normal airway defenses and allows microbe-laden oropharyngeal or gastric secretions to be aspirated around the endotracheal tube into the lower airways (14). The fate of such bacteria in the lungs of intubated patients is largely unknown. To date, investigations of intubated patients have used aerobic cultures of either endotracheal secretions or bronchoalveolar lavage (BAL) specimens to determine the presence of pathogenic bacteria. However, these techniques determine neither the dominant bacterial species nor the range of bacterial diversity within the community. Culture-independent methods provide a more comprehensive view of bacterial diversity and may be used to evaluate complex bacterial community dynamics, particularly during antibiotic therapy.

Culture-independent analyses are typically based on biomarker identification. The 16S rRNA gene is the most commonly used biomarker for bacterial community studies (44). Highly conserved regions of the 16S rRNA gene enable amplification of this gene from most bacteria with "universal" PCR primers, while variable regions within this gene permit discrimination between bacterial types (16). This approach has been applied successfully to the analysis of environmental (2, 20, 35, 37, 42) and human (18, 26, 28, 29, 38) bacterial communities and has revealed a much broader bacterial diversity than has traditional culture-based techniques (38). However, to our knowledge, no previous study has examined the extent of bacterial diversity within the lungs of intubated patients by culture-independent methods.

Here, we report the use of two 16S rRNA gene-based culture-independent methods, clone library sequencing (the current "gold standard" in microbial ecology) (1) and a novel high-density oligonucleotide microarray (PhyloChip) (3). Clone libraries typically involve sequencing of a few hundred 16S rRNA genes following PCR amplification; while this strategy provides great specificity, it may profile only the dominant organisms of a complex bacterial population, while less-abundant species that may contribute to disease pathogenesis remain undetected. For this reason, the PhyloChip approach was applied in parallel with 16S rRNA clone libraries to determine changes in the bacterial community composition of the lungs of intubated patients with hospital-acquired *Pseudomonas aeruginosa* during antimicrobial treatment.

MATERIALS AND METHODS

Patient selection and sampling protocol. Daily endotracheal aspirates (EAs) were collected from all intubated patients in the medical, surgical, neurovascular, and cardiac intensive care units of the University of California at San Francisco

* Corresponding author. Mailing address: DOE Joint Genome Institute, 2800 Mitchell Dr., Bldg. 400-404, Walnut Creek, CA 94598. Phone: (925) 296-5609. Fax: (925) 296-5272. E-mail: jbristow@lbl.gov.
† E. L. Brodie and J. L. Flanagan contributed equally to this work.
▽ Published ahead of print on 4 April 2007.

(UCSF). Daily EA samples were obtained, and all samples were screened by culture on Difco *Pseudomonas* isolation agar (Becton Dickinson). *Pseudomonas*-positive patients were approached, they gave informed consent, and daily quantitative cultures of *P. aeruginosa* were performed on their respiratory samples. Blind mini-bronchoalveolar lavage (mini-BAL) was performed with a Combi-Cath catheter when clinical infection was suspected. The right bronchoalveolar tree was irrigated with three 20-ml aliquots of nonbacteriostatic saline. The resulting lavage was used for *P. aeruginosa* culture and 16S rRNA analysis. Blind mini-BAL and EA samples were centrifuged and stored at −80°C. Clinical data were collected from patients' charts and recorded in an Access database so that severity of illness could be correlated with 16S rRNA data. Antibiotic drug history was also recorded. All protocols were approved by the Committee on Human Research of UCSF.

DNA extraction and amplification of the bacterial 16S rRNA gene. Bacterial genomic DNA was isolated from 0.5 ml of EA and BAL samples with the Promega (Carlsbad, CA) Wizard genomic DNA purification kit, according to the manufacturer's instructions for purification of both gram-negative and gram-positive bacteria. The 16S rRNA gene was amplified from extracted DNA with the universal bacterial primers Bact-27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and Bact-1492R (5′-GGTTACCTTGTTACGACTT-3′) (22). The reaction mixture (50 μl, final volume) contained 5 μl of 10× PCR buffer (Amersham, NJ), 1 μl of deoxynucleoside triphosphates (10 mM), 0.7 μl of forward primer and reverse primer (100 pmol/μl each), 0.35 μl of *Taq* polymerase (5 U/μl; Amersham), and 1 μl of template DNA. PCR was performed with the DNA Engine Tetrad thermal cycler (Bio-Rad). To maximize the number of bacterial species that could be recovered by PCR, three different annealing temperatures (48°C, 52°C, and 56°C) were used for each sample to amplify the 16S rRNA genes. The following cycling parameters were used: 3 min of initial denaturation at 95°C followed by 25 cycles of denaturation (30 s at 95°C), annealing (30 s), and elongation (120 s at 72°C), with a final extension at 72°C for 7 min. Amplified products from all samples were verified by gel electrophoresis. All PCR products were gel purified with the QIAquick gel extraction kit (QIAGEN), and, for each sample, the purified products amplified at three different annealing temperatures were pooled for cloning, sequencing, and microarray analysis.

Cloning and sequencing. To generate libraries for each sample, the respective full-length 16S RNA PCR products were cloned into pCR4-TOPO vectors (Invitrogen) according to the manufacturer's instructions. One hundred ninety-two transformants from each library were picked randomly. Double-ended sequencing reactions of the entire 16S RNA sequence were carried out with PE BigDye terminator chemistry (Perkin Elmer) and resolved with an ABI PRISM 3730 (Applied Biosystems) capillary DNA sequencer. Sequencing was performed at the DOE Joint Genome Institute.

Sequence alignment and phylogenetic analysis. Paired-end sequencing reads of 16S rRNA clones were assembled with Phrap (12, 13), and only clones in which >80% of the bases had quality scores of at least Phred 20 (indicating a probability of error of ≤1%) were included in the analysis. Chimeric sequences were then detected with an online tool at Greengenes (8) (http://greengenes.lbl.gov) by using an updated version of Bellerophon (19) and were also excluded from the analysis. For the 19 samples we report here, an average of 90% of sequenced clones met these criteria and were included in subsequent analyses.

Sequences were aligned to the Greengenes 7,682-character format with the NAST (9) web server prior to being assigned to a taxonomic node with a sliding scale of similarity thresholds (34) and the Greengenes classifying tool. Distance matrices were constructed for each library with the distance matrix tool at Greengenes, with NAST-aligned sequence data as input.

Phylotype clustering and diversity estimates. Using the distance matrices generated, numbers of 16S rRNA gene phylotypes were calculated at 99% sequence identity with the furthest neighbor clustering in the program DOTUR (33), with 1,000 iterations for bootstrapping. A representative 16S rRNA gene-based phylogenetic tree was constructed with the software package ARB (32) using data from the Greengenes database. Microbial diversity was estimated with the Shannon diversity index (H′). This index takes into account both the number of species present and the proportion of the total accounted for by each species, increasing with species number and greater unevenness of species prevalence (11, 15, 23). The number of species present and the diversity index were compared with the nonparametric Wilcoxon signed-rank test for paired samples.

PhyloChip processing, scanning, probe set scoring, and normalization. The pooled PCR product was spiked with known concentrations of synthetic 16S rRNA gene fragments and non-16S rRNA gene fragments as internal standards for normalization, with quantities ranging from 5.02×10^8 to 7.29×10^{10} molecules applied to the final hybridization mix. Target fragmentation, biotin labeling, PhyloChip hybridization, scanning, and staining were as described by Brodie et al. (3), while background subtraction, noise calculation, and detection and quantification criteria were essentially as previously reported (3), with some minor exceptions. For a probe pair to be considered positive, the difference in intensity between the perfect match (PM) and mismatch (MM) probes must be at least 130 times the squared noise value (*N*). A taxon was considered present in the sample when 90% or more of its assigned probe pairs for its corresponding probe set were positive (positive fraction, ≥0.90). Hybridization intensity (referred to as intensity) was calculated in arbitrary units for each probe set as the trimmed average (maximum and minimum values removed before averaging) of the PM minus MM intensity differences across the probe pairs in a given probe set. All intensities of <1 were shifted to 1 to avoid errors in subsequent logarithmic transformations. To account for scanning intensity variations from array to array, the intensities resulting from the internal standard probe sets were natural log transformed. Adjustment factors for each PhyloChip were calculated by fitting a linear model by the least-squares method. A PhyloChip's adjustment factor was subtracted from each probe set's $\ln_{\text{intensity}}$. Intensities for patient 1 were also normalized by total array intensity. When summarizing PhyloChip results to the subfamily, the taxon with a probe set producing the highest intensity within a subfamily was used.

Nucleotide sequence accession numbers. The nucleotide sequences generated in this study have been submitted to the GenBank database under accession numbers EF508731 to EF512008.

RESULTS

We began our study by comparing sampling techniques. To do this, the bacterial species compositions of samples obtained with blind mini-BAL were compared to those obtained with EAs from two patients with hospital-acquired *P. aeruginosa* by using 16S rRNA clone library sequencing. While there were distinct differences in community compositions between these two patients, the community compositions of the EA and BAL samples from each individual were highly similar (Fig. 1A and B). Therefore, in the remainder of this study, EAs were used due to the simplicity and cost-effectiveness of this less-invasive sample collection method.

Next, as a control for this study, EA samples were collected from three healthy individuals who had been briefly intubated for elective surgery. No 16S rRNA PCR product was detected from these patients (data not shown) under the conditions that readily yielded 16S rRNA amplicons in study patients, confirming that the normal lung is sterile and that our techniques identify organisms present only after colonization of the endotracheal tube and airway has occurred.

We next collected EAs from seven patients colonized by *Pseudomonas aeruginosa*. Patient age, sex, timing of EA sampling, periods of antibiotic administration, and antibiotic sensitivity details are presented in Table 1. All patient samples yielded a 16S rRNA PCR product, and a total of 3,278 nonchimeric 16S rRNA sequences from patient-derived 16S libraries were subjected to phylogenetic analysis. Almost all organisms detected by cloning and sequencing were from five bacterial phyla, the *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, and *Fusobacteria* (Fig. 1C). Over half (55%) of the sequences obtained were from *Pseudomonas aeruginosa*, followed by *Stenotrophomonas maltophilia* (9.7%), *Prevotella* spp. (5.8%), *Acinetobacter* spp. (5.7%), *Serratia marcescens* (5.0%), *Haemophilus* spp. (3.8%), *Neisseria* spp. (3.3%), *Mycoplasma* spp. (2.4%), and *Streptococcus* spp. (2.3%). An additional18 genera were detected but together represented less than 7% of all clones sequenced. Of these less-abundant species, many are known oral, nasal, and gastrointestinal tract inhabitants, e.g., *Porphyromonas*, *Campylobacter*, *Fusobacterium*, *Lactobacillus*, *Enterococcus*, *Rothia*, *Actinomyces*, *Abiotrophia*, *Alcaligenes*, *Corynebacterium*, *Staphylococcus*, and *Veillonella*

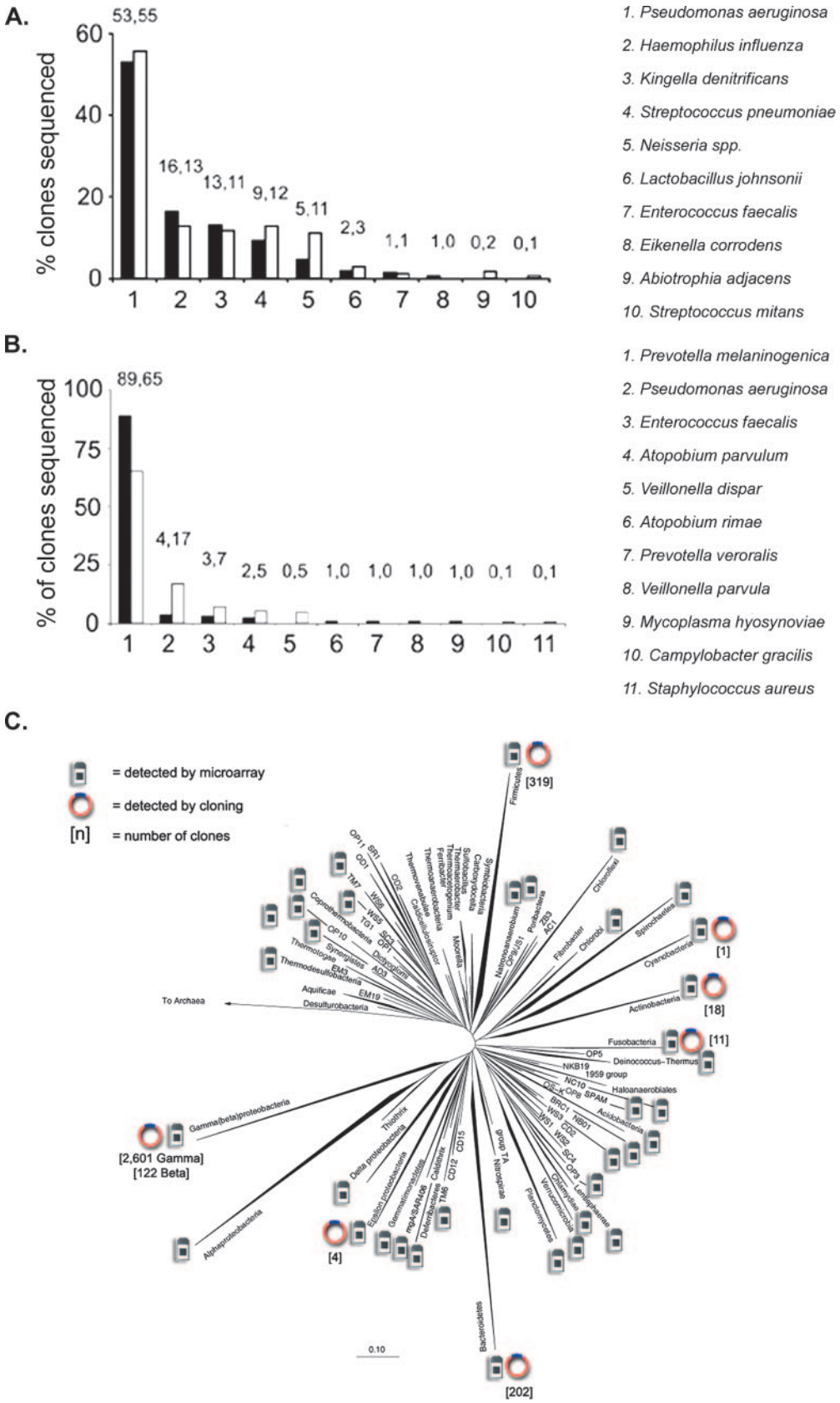


FIG. 1. (A and B) Comparisons of 16S rRNA clone libraries of EA (white bars) and BAL (black bars) patient samples. (A) A total of 173 clones were sequenced from the EA samples, and 153 clones were sequenced from the BAL samples. (B) A total of 154 clones were sequenced from the EA samples, and 141 clones were sequenced from the BAL samples. Species abundances (as percentages) of the sequenced clones are indicated above the bars. (C) Phylogenetic tree showing all recognized bacterial phyla/divisions. Phyla detected in EAs by cloning and PhyloChip microarray analyses are shown. The five main phyla detected by clone library are indicated. Many additional phyla, undetected by cloning, were detected by array analysis.

TABLE 1. Patient information and antimicrobial treatment

Patient	No. of days after enrollment (sample no.)	Sex	Patient age	Antimicrobial treatment	
				Within 24 h before study enrollment	Following sampling (sensitivity[a])
1	1 (1)	Female	57 yr	Cefazolin, piperacillin-tazobactam	Piperacillin-tazobactam (S), fluconazole, cefazolin
	7 (2)				Cefazolin (S), fluconazole (S), levofloxacin (S)
2	1 (1)	Male	79 yr	Cefazolin, ceftazidime	Antifungal, ceftazidime (S), vancomycin
	11 (2)				Ceftazidime (R), vancomycin, piperacillin-tazobactam (S), ciprofloxacin (S)
	15 (3)				Vancomycin, piperacillin-tazobactam, ciprofloxacin
3	1 (1)	Female	54 yr	None	Ciprofloxacin (S)
	5 (2)				Ciprofloxacin
4	1 (1)	Male	55 yr	None	Piperacillin-tazobactam, vancomycin
	7 (2)				Piperacillin-tazobactam
5	1 (1)	Female	85 yr	Clindamycin	Clindamycin, piperacillin-tazobactam (S)
	4 (2)				Piperacillin-tazobactam (S), vancomycin, ciprofloxacin (S)
	15 (3)				None
	18 (4)				None
	23 (5)				None
6	1 (1)	Female	45 yr	None	Meropenem (I), fluconazole, linezolid
	23 (2)				Tobramycin(S), imipenem (I), cefpirome, cefazolin, cefepime (I)
	102 (3)				Timentin, trimethoprim-sulfamethoxazole, imipenem, vancomycin, fluconazole, cefepime, cefpirome, amphotericin B, tobramycin
7	1 (1)	Female	2 mo	Ampicillin, gentamicin, trimethoprim-sulfamethoxazole	Ampicillin (R), gentamicin
	82 (2)				Gentamicin (S)

[a] S, sensitive; R, resistant; I, indeterminate.

(6, 10, 27, 39). These results support the hypothesis that oral, nasal, and gastrointestinal tract microbiota are the major reservoirs for bacteria that colonize the lower airway in intubated patients (17, 40, 41).

We next evaluated the bacterial diversity in five patients (Table 1, patients 1 to 5) for whom an initial sample was obtained within 24 h of parenteral antibiotic administration and a second sample was obtained 4 to 10 days later. Analysis of microbial diversity in the 16S clone libraries demonstrated a substantial reduction in bacterial diversity during antibiotic administration (Fig. 2). The mean number of bacterial species identified fell from 16.2 to 5.6 ($P < 0.05$), and Shannon's diversity index, a commonly used measure of microbial diversity, fell from 1.48 to 0.59 ($P < 0.05$). In each case, antibiotic therapy led to selection of a pathogenic species that dominated the community. In four of five patients, *P. aeruginosa* came to dominate the microbial community at the second time point, despite the administration of antibiotics to which it was susceptible in vitro (Table 1). In the fifth patient, the community became dominated by another pathogen, *Klebsiella pneumoniae*. Even when *P. aeruginosa* constituted the majority of the population at the initial time point, as in patient 3, who was chronically ventilated, there was initial diversity in the remainder of the community that was lost with antibiotic therapy.

To investigate how the bacterial community changed over a longer period of antibiotic treatment, we obtained additional samples at later time points for two patients (patients 2 and 5), and during prolonged antibiotic therapy in two additional patients (patients 6 and 7). All of these samples showed reduced diversity compared with the initial time point, and *P. aeruginosa* was the predominant species in six of seven samples, suggesting that once this organism is established as the dominant species, microbial diversity is slow to recover. Patient 2 did demonstrate a reduction in the relative abundance of *P. aeruginosa* following adjustment of antipseudomonal therapy 11 days after colonization, and, importantly, this was accompanied by substantial recovery of microbial diversity. This patient improved clinically, and extubation was attempted. Collectively, these data demonstrate that the loss of bacterial diversity and *Pseudomonas* dominance are highly correlated during antipseudomonal therapy.

Dominance of bacterial communities by one or a few species may result either from overgrowth of the dominant species or from loss of the nondominant species. Due to the limited number of clones that can feasibly be sampled from clone libraries, highly abundant species may mask the presence of less-abundant but clinically significant species. To determine whether the decline in diversity observed in clone libraries was

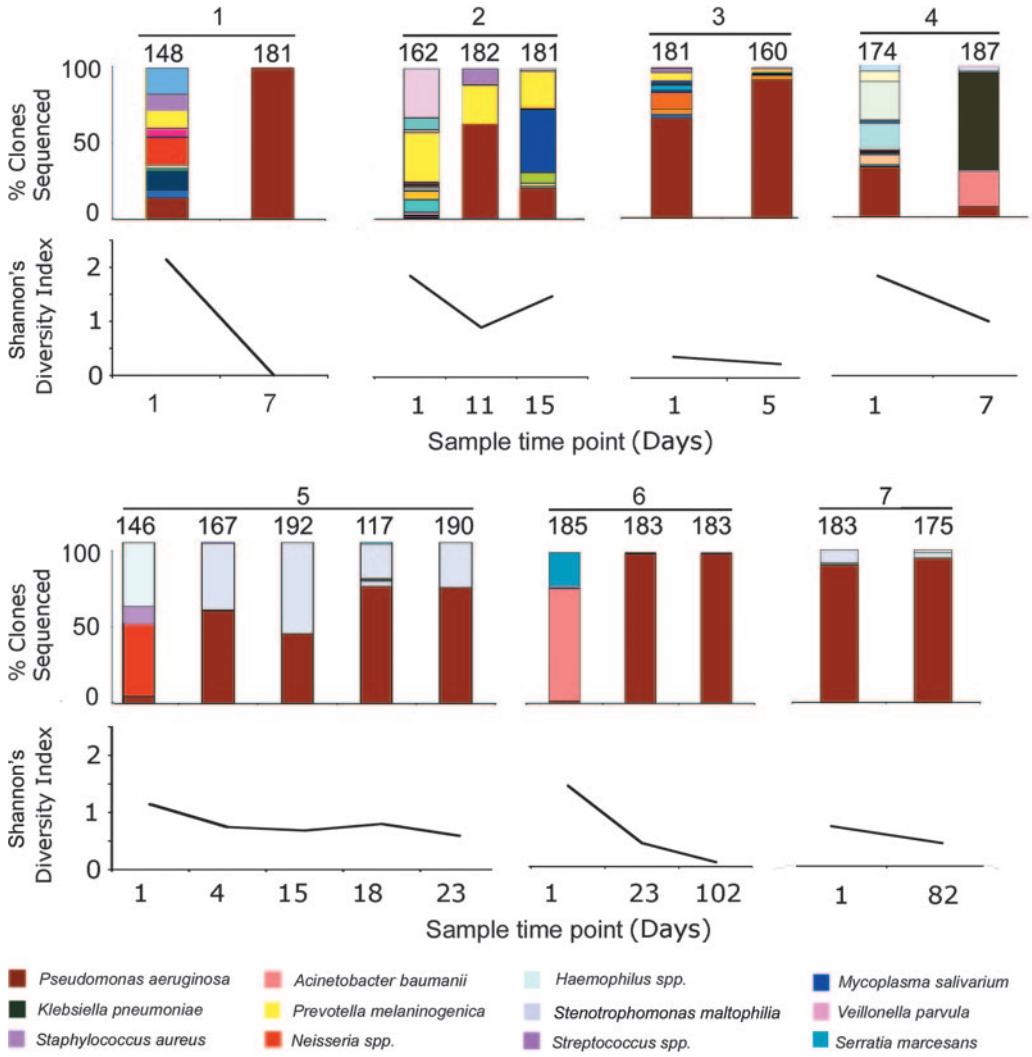


FIG. 2. Temporal changes in bacterial diversity. Each bar represents the color-coded relative abundance of bacteria in a single EA. Numbers above the horizontal bars represent individual patients, and the numbers of clones analyzed for each sample are indicated directly above each bar. For each sample, Shannon's diversity statistic, which reflects both species numbers and evenness of species distribution, is plotted below the histogram.

a true reflection of the bacterial community in these patients, bacterial diversity was also analyzed with high-density microarrays (PhyloChips), which have enhanced sensitivity for low-abundance species compared with cloning (7). For four patients, the same 16S rRNA gene amplicon pools from which clone libraries were prepared were subsequently hybridized to PhyloChips, and the bacterial communities were compared (Fig. 3). While the microarray approach detected >30 times more bacterial types than did clone library sequencing ($P < 0.0001$), the PhyloChip data clearly mirrored the loss of diversity found in the clone library data, with the mean number of observed phylotypes falling from 517 to 280 during antibiotic treatment ($P < 0.05$). These data clearly show that the loss of diversity demonstrated in clone libraries is not an artifact of sampling and confirm that antibiotic treatment leads to lower diversity.

Due to the great sensitivity of the PhyloChip, entire bacterial community responses can be monitored by this technique. Figure 4A illustrates temporal changes in the fluorescence

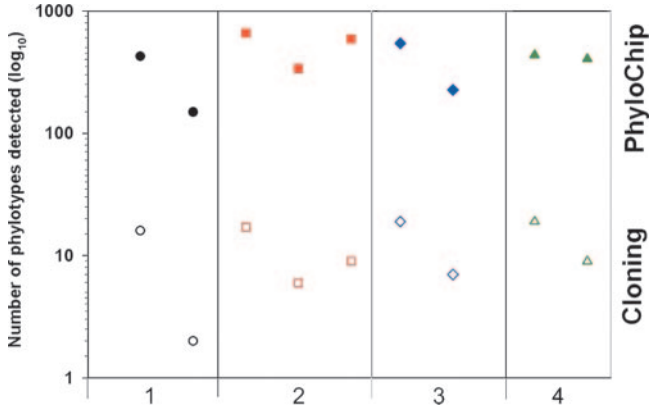


FIG. 3. Comparison of PhyloChip and clone library monitoring of bacterial diversity (phylotype numbers) over time for four patients. Closed symbols show numbers of bacterial phylotypes detected by PhyloChip analysis; open symbols show numbers of bacterial phylotypes detected by clone library on a logarithmic scale. The PhyloChip reveals many more bacterial groups, but the data exactly mirror the loss of diversity revealed by clone library sequencing.

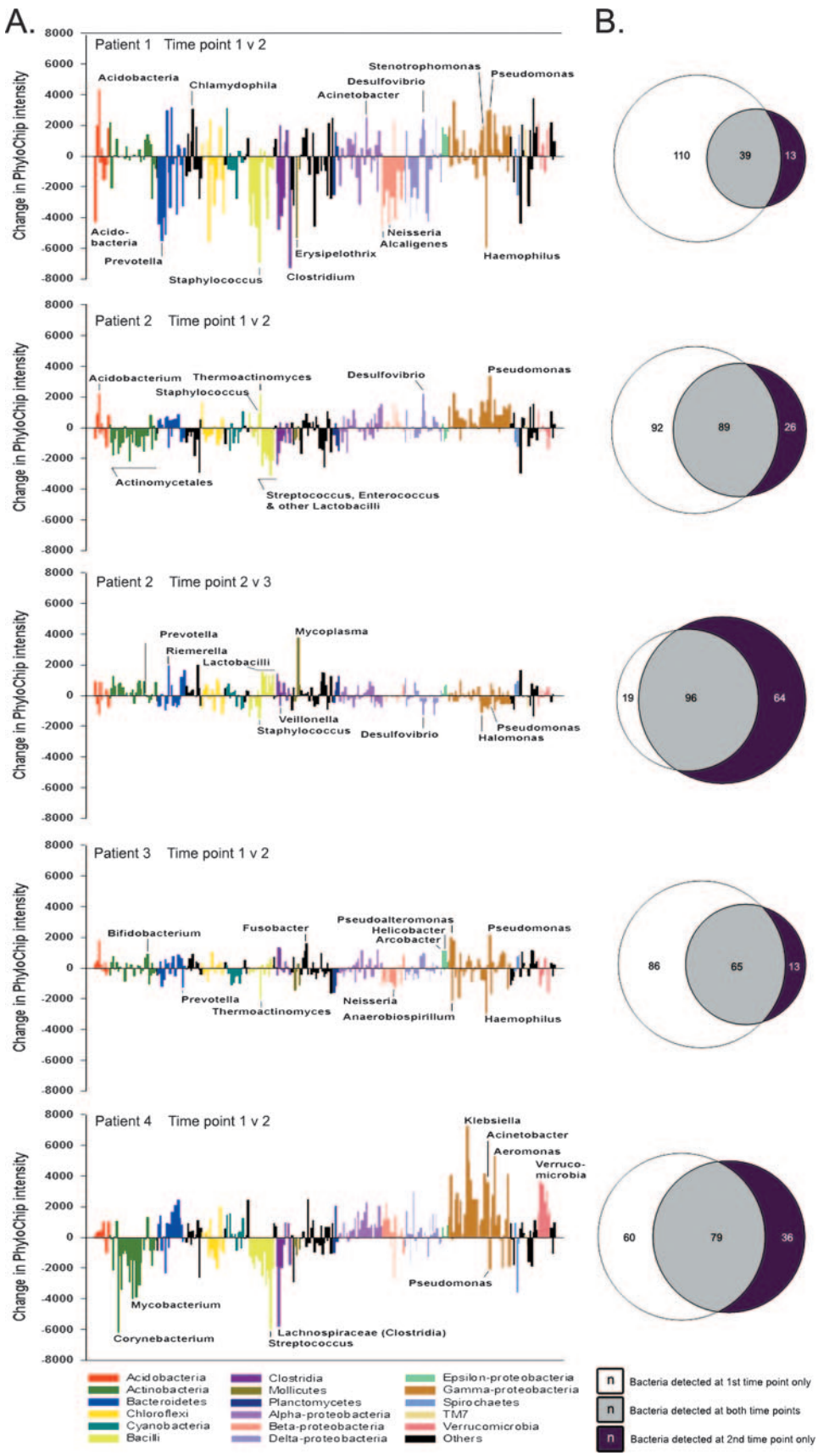


FIG. 4. PhyloChip analysis of complete bacterial communities over time in EAs. (A) Bacteria are ordered alphabetically from left to right according to taxonomic affiliation. Bars above the zero line represent bacteria that increased in abundance relative to the first EA sample being compared; bars below represent those bacteria that declined in abundance. (B) Venn diagrams demonstrate the number of bacterial subfamilies detected at each time point. Note the significant overlap in composition between successive time points.

intensity of bacteria detected by PhyloChip. Bacteria demonstrating large changes in intensity between time points are labeled, and these correspond well to the dominant bacterial species detected by clone library analysis. The dominance of a few species within a community analyzed by 16S rRNA clone library clearly reflects the limited number of clones sequenced. The PhyloChip readily demonstrated that many bacteria present at the initial sampling point were indeed still present in the subsequent sampling period (Fig. 4B). Conversely, bacteria such as *Klebsiella* that became dominant in later clone library samples were detected in the initial sample by PhyloChip but were not detected in the corresponding clone library. This underscores the potential for low-abundance species to eventually dominate bacterial communities during the course of antimicrobial administration.

The microarray analysis also allowed us to correlate changes in phylogenetic groups over time when these groups of bacteria respond in similar manners to antibiotic administration. For example the γ-*Proteobacteria* (which include *Pseudomonas aeruginosa*) generally exhibit an inverse relationship in abundance with bacteria in the phylum *Actinobacteria* and in the class *Bacilli* (which includes *Lactobacillus*, *Streptococcus*, *Staphylococcus*, and *Enterococcus*) in all patients examined (Fig. 4A). Similarly *Haemophilus* and *Pseudomonas* also demonstrated an inverse relationship. The reciprocal changes in these subgroups within the bacterial community suggest they may be competing for similar niches in the endotracheal environment or otherwise influencing each other's growth.

DISCUSSION

Management of infections caused by *P. aeruginosa* is increasingly difficult due to this bacterium's metabolic versatility, intrinsic antimicrobial resistance, and remarkable armory of virulence factors. Clinically, treatment options are becoming limited due to the rapid emergence of multidrug-resistant strains, which are now estimated to account for up to 30% of strains isolated from patients in nursing homes, hospitals, and intensive care units (30). In this study, we used 16S rRNA-based culture-independent methods to determine the effect of antibiotic therapy on bacterial community dynamics in patients colonized with *P. aeruginosa*. Compared to culture-based methods, both clone library and microarray techniques have often provided a richer picture of microbial diversity (3, 10, 44), and we found that to be true in this study. Prior to, or early in, antibiotic therapy, the airways were colonized with a remarkably wide array of oral, nasal, and gut flora that are presumably aspirated into the lung around the endotracheal tube.

Not unexpectedly, antimicrobial treatment had a pronounced effect on bacterial community composition, with bacterial diversity falling in every case. In most cases, *P. aeruginosa* began as a relatively small fraction of the observed species and came to dominate the community despite the administration of antibiotics to which it was sensitive in vitro. Alarmingly, pathogenic species became dominant in every patient during antibiotic therapy, and both the loss of diversity and *P. aeruginosa* dominance often persisted during prolonged antibiotic therapy. When *P. aeruginosa* declined with antibiotic therapy after becoming dominant (patient 2), there was a substantial recovery in microbial diversity. In contrast, when *P. aeruginosa* was replaced by a dominant pathogen (patient 4), diversity declined. Together, our data demonstrate that loss of diversity is highly correlated with pathogen selection in *P. aeruginosa*-colonized patients who are treated with antibiotics. We hypothesize that reduced microbial diversity under antibiotic selection in the airways may contribute directly to pathogen selection through the loss of microbial competition. This concept is supported by the observation that the abundance of specific organisms (notably *Actinobacteria* and *Bacilli*) always appeared to change in a manner opposite that of *P. aeruginosa.*

It has long been hypothesized that the evolution of virulence is related to the number and variety of bacterial species infecting the host (4, 31). Previously, it was predicted that increased diversity of pathogenic species would promote virulence of individual species (31). However, more recently it has been demonstrated that in polymicrobial infections, less-virulent strains are often favored, suggesting that increased diversity may in fact reduce virulence (21). This hypothesis is supported by our observation that administration of antimicrobials eliminates competition by decreasing the diversity of nontarget organisms and allowing an increase in pathogen abundance. One mechanism through which a decrease in bacterial diversity might alter virulence is quorum sensing. Quorum sensing is a means of bacterial cross-talk between individual cells incorporated into a biofilm, and it may radically affect the gene expression and virulence of pathogens, including *P. aeruginosa* (24, 25, 43). Further studies investigating the contribution of bacterial dynamics to pathogenicity will be required to fully evaluate this hypothesis.

This study highlights the importance of recognizing and understanding the complex bacterial community dynamic that exists in the airways of intubated patients and that is routinely missed by traditional culture-based analyses. Treatment strategies that rely primarily on the identification of *P. aeruginosa* in the airways may often result in suboptimal clinical outcomes. Specifically, in our study, administration of antibiotics to patients colonized with *P. aeruginosa* within a complex bacterial community led to loss of microbial diversity and selection of *P. aeruginosa* or another pathogen. In previous studies, strategies to minimize antibiotic use led to improved outcomes and the emergence of fewer multidrug-resistant strains (5). While further study is clearly needed, we believe that culture-independent surveillance of microbial diversity could be a useful adjunct in the management of intubated patients colonized by *P. aeruginosa* by helping to determine both the effectiveness and the appropriate duration of antibiotic therapy.

We used two culture-independent methods in this study, both of which showed qualitatively similar reductions in bacterial diversity during antibiotic treatment. The PhyloChip is clearly superior to clone library sequencing for assessing the complete spectrum of bacteria present in the community, including low-abundance species, while the 16S rRNA library approach provides important information on relative abundance at the species level. Both culture-independent molecular methods documented a multitude of bacteria undetected by standard identification techniques. Importantly, this was true not only for fastidious, slow-growing, and/or unculturable organisms but also for routinely cul-

tured pathogens (3, 10, 44). Unlike sequencing, the PhyloChip method does not require cloning and is therefore rapid enough that it could be used to provide information about bacterial abundance and diversity in the clinical setting.

In summary, we have documented the evolution of the complex bacterial community that exists in intubated patients, who are often subjected to a variety of antibiotics. Our data indicate that *P. aeruginosa* frequently arises to become the dominant organism in colonized patients, as competing flora are eliminated by antimicrobial therapy. This may occur even when the patient is on appropriate antipseudomonal therapy as indicated by the in vitro susceptibility profile. If bacterial community changes could be monitored in a timely and cost-effective manner, the effectiveness of antimicrobial therapy might be more readily judged. Furthermore, we hypothesize that this methodology could be utilized to gauge the effectiveness of antibiotic treatment and individualize the duration of antibiotic therapy, perhaps by treating only until microbial diversity rebounds. Ultimately, culture-independent methods such as the PhyloChip will need to be tested for the ability to help guide antibiotic therapy and improve outcomes for intubated patients.

ACKNOWLEDGMENTS

This work was supported by SCCOR grants PH50HL74005 and HL69809 to J.P.W.-K., by a UCSF Academic Senate Independent Investigator Award to S.V.L., and by NIH grant HL024075 to J.B. Part of this work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396. We declare no conflicts of interest.

We thank Susannah Tringe for helpful comments on the manuscript and Sam Pitluck and Victor Kunin for help with GenBank submissions.

REFERENCES

1. **Angenent, L. T., S. T. Kelley, A. St. Amand, N. R. Pace, and M. T. Hernandez.** 2005. Molecular identification of potential pathogens in water and air of a hospital therapy pool. Proc. Natl. Acad. Sci. USA **102:**4860–4865.
2. **Bond, P. L., S. P. Smriga, and J. F. Banfield.** 2000. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. Appl. Environ. Microbiol. **66:**3842–3849.
3. **Brodie, E. L., T. Z. Desantis, D. C. Joyner, S. M. Baek, J. T. Larsen, G. L. Andersen, T. C. Hazen, P. M. Richardson, D. J. Herman, T. K. Tokunaga, J. M. Wan, and M. K. Firestone.** 2006. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl. Environ. Microbiol. **72:**6288–6298.
4. **Brown, S. P., M. E. Hochberg, and B. T. Grenfell.** 2002. Does multiple infection select for raised virulence? Trends Microbiol. **10:**401–405.
5. **Chastre, J., M. Wolff, J. Y. Fagon, S. Chevret, F. Thomas, D. Wermert, E. Clementi, J. Gonzalez, D. Jusserand, P. Asfar, D. Perrin, F. Fieux, and S. Aubas.** 2003. Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial. JAMA **290:**2588–2598.
6. **Chhour, K. L., M. A. Nadkarni, R. Byun, F. E. Martin, N. A. Jacques, and N. Hunter.** 2005. Molecular analysis of microbial diversity in advanced caries. J. Clin. Microbiol. **43:**843–849.
7. **DeSantis, T. Z., E. L. Brodie, J. P. Moberg, I. X. Zubieta, Y. M. Piceno, and G. L. Andersen.** 2 March 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb. Ecol. doi:10.1007/s00248-006-9134-9.
8. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72:**5069–5072.
9. **DeSantis, T. Z., Jr., P. Hugenholtz, K. Keller, E. L. Brodie, N. Larsen, Y. M. Piceno, R. Phan, and G. L. Andersen.** 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. **34:**W394–W399.
10. **Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman.** 2005. Diversity of the human intestinal microbial flora. Science **308:**1635–1638.
11. **Edwards, M. L., A. K. Lilley, T. H. Timms-Wilson, I. P. Thompson, and I. Cooper.** 2001. Characterisation of the culturable heterotrophic bacterial community in a small eutrophic lake (Priest Pot). FEMS Microbiol. Ecol. **35:**295–304.
12. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.
13. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.
14. **Feldman, C., M. Kassel, J. Cantrell, S. Kaka, R. Morar, A. Goolam Mahomed, and J. I. Philips.** 1999. The presence and sequence of endotracheal tube colonization in patients undergoing mechanical ventilation. Eur. Respir. J. **13:**546–551.
15. **Gafan, G. P., V. S. Lucas, G. J. Roberts, A. Petrie, M. Wilson, and D. A. Spratt.** 2005. Statistical analyses of complex denaturing gradient gel electrophoresis profiles. J. Clin. Microbiol. **43:**3971–3978.
16. **Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature **345:**60–63.
17. **Grap, M. J., C. L. Munro, R. K. Elswick, Jr., C. N. Sessler, and K. R. Ward.** 2004. Duration of action of a single, early oral application of chlorhexidine on oral microbial flora in mechanically ventilated patients: a pilot study. Heart Lung **33:**83–91.
18. **Green, G. L., J. Brostoff, B. Hudspith, M. Michael, M. Mylonaki, N. Rayment, N. Staines, J. Sanderson, D. S. Rampton, and K. D. Bruce.** 2006. Molecular characterization of the bacteria adherent to human colorectal mucosa. J. Appl. Microbiol. **100:**460–469.
19. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20:**2317–2319.
20. **Hugenholtz, P., G. W. Tyson, and L. L. Blackall.** 2002. Design and evaluation of 16S rRNA-targeted oligonucleotide probes for fluorescence in situ hybridization. Methods Mol. Biol. **179:**29–42.
21. **Kreft, J. U., and S. Bonhoeffer.** 2005. The evolution of groups of cooperating bacteria and the growth rate versus yield trade-off. Microbiology **151:**637–641.
22. **Lane, D. J.** 1991. 16S/23S rRNA sequencing. Wiley, Chichester, United Kingdom.
23. **Ledder, R. G., P. Gilbert, S. A. Huws, L. Aarons, M. P. Ashley, P. S. Hull, and A. J. McBain.** 2007. Molecular analysis of the subgingival microbiota in health and disease. Appl. Environ. Microbiol. **73:**516–523.
24. **Levy, S. B.** 1997. Antibiotic resistance: an ecological imbalance. Ciba Found. Symp. **207:**1–14.
25. **Lewenza, S., M. B. Visser, and P. A. Sokol.** 2002. Interspecies communication between Burkholderia cepacia and Pseudomonas aeruginosa. Can. J. Microbiol. **48:**707–716.
26. **Ott, S. J., and S. Schreiber.** 2006. Reduced microbial diversity in inflammatory bowel diseases. Gut **55:**1207.
27. **Paju, S., J. M. Bernstein, E. M. Haase, and F. A. Scannapieco.** 2003. Molecular analysis of bacterial flora associated with chronically inflamed maxillary sinuses. J. Med. Microbiol. **52:**591–597.
28. **Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst.** 2001. Bacterial diversity in human subgingival plaque. J. Bacteriol. **183:**3770–3783.
29. **Paster, B. J., M. K. Russell, T. Alpagot, A. M. Lee, S. K. Boches, J. L. Galvin, and F. E. Dewhirst.** 2002. Bacterial diversity in necrotizing ulcerative periodontitis in HIV-positive subjects. Ann. Periodontol. **7:**8–16.
30. **Ray, G. T., R. Baxter, and G. N. DeLorenze.** 2005. Hospital-level rates of fluoroquinolone use and the risk of hospital-acquired infection with ciprofloxacin-nonsusceptible Pseudomonas aeruginosa. Clin. Infect. Dis. **41:**441–449.
31. **Read, A. F., and L. H. Taylor.** 2001. The ecology of genetically diverse infections. Science **292:**1099–1102.
32. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
33. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501–1506.
34. **Schloss, P. D., and J. Handelsman.** 2004. Status of the microbial census. Microbiol. Mol. Biol Rev. **68:**686–691.
35. **Schmidt, T. M., E. F. DeLong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371–4378.
36. **Sottile, F. D., T. J. Marrie, D. S. Prough, C. D. Hobgood, D. J. Gower, L. X. Webb, J. W. Costerton, and A. G. Gristina.** 1986. Nosocomial pulmonary infection: possible etiologic significance of bacterial adhesion to endotracheal tubes. Crit. Care Med. **14:**265–270.

37. **Stahl, D. A., D. J. Lane, G. J. Olsen, and N. R. Pace.** 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. Appl. Environ. Microbiol. **49:**1379–1384.
38. **Suau, A., R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dore.** 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. **65:**4799–4807.
39. **Tanner, A. C., P. M. Milgrom, R. Kent, Jr., S. A. Mokeem, R. C. Page, C. A. Riedy, P. Weinstein, and J. Bruss.** 2002. The microbiota of young children from tooth and tongue samples. J. Dent. Res. **81:**53–57.
40. **Torres, A., M. El-Ebiary, N. Soler, C. Monton, N. Fabregas, and C. Hernandez.** 1996. Stomach as a source of colonization of the respiratory tract during mechanical ventilation: association with ventilator-associated pneumonia. Eur. Respir. J. **9:**1729–1735.
41. **Torres, A., M. el-Ebiary, N. Soler, C. Monton, J. Gonzalez, and J. Puig de la Bellacasa.** 1995. The role of the gastric reservoir in ventilator-associated pneumonia. Clin. Intensive Care **6:**174–180.
42. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345:**63–65.
43. **Waters, C. M., and B. L. Bassler.** 2005. Quorum sensing: cell-to-cell communication in bacteria. Annu. Rev. Cell Dev. Biol. **21:**319–346.
44. **Zoetendal, E. G., E. E. Vaughan, and W. M. de Vos.** 2006. A microbial world within us. Mol. Microbiol. **59:**1639–1650.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Sept. 2006, p. 6288–6298 Vol. 72, No. 9
0099-2240/06/$08.00+0 doi:10.1128/AEM.00246-06

Application of a High-Density Oligonucleotide Microarray Approach To Study Bacterial Population Dynamics during Uranium Reduction and Reoxidation†

Eoin L. Brodie,[1,2]* Todd Z. DeSantis,[2] Dominique C. Joyner,[2] Seung M. Baek,[1] Joern T. Larsen,[2] Gary L. Andersen,[2] Terry C. Hazen,[2] Paul M. Richardson,[3] Donald J. Herman,[1] Tetsu K. Tokunaga,[2] Jiamin M. Wan,[2] and Mary K. Firestone[1]

Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720[1]; Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720[2]; and Department of Energy, Joint Genome Institute, Walnut Creek, California 94598[3]

Received 1 February 2006/Accepted 28 June 2006

Reduction of soluble uranium U(VI) to less-soluble uranium U(IV) is a promising approach to minimize migration from contaminated aquifers. It is generally assumed that, under constant reducing conditions, U(IV) is stable and immobile; however, in a previous study, we documented reoxidation of U(IV) under continuous reducing conditions (Wan et al., Environ. Sci. Technol. 2005, 39:6162–6169). To determine if changes in microbial community composition were a factor in U(IV) reoxidation, we employed a high-density phylogenetic DNA microarray (16S microarray) containing 500,000 probes to monitor changes in bacterial populations during this remediation process. Comparison of the 16S microarray with clone libraries demonstrated successful detection and classification of most clone groups. Analysis of the most dynamic groups of 16S rRNA gene amplicons detected by the 16S microarray identified five clusters of bacterial subfamilies responding in a similar manner. This approach demonstrated that amplicons of known metal-reducing bacteria such as *Geothrix fermentans* (confirmed by quantitative PCR) and those within the *Geobacteraceae* were abundant during U(VI) reduction and did not decline during the U(IV) reoxidation phase. Significantly, it appears that the observed reoxidation of uranium under reducing conditions occurred despite elevated microbial activity and the consistent presence of metal-reducing bacteria. High-density phylogenetic microarrays constitute a powerful tool, enabling the detection and monitoring of a substantial portion of the microbial population in a routine, accurate, and reproducible manner.

Uranium contamination is a persistent legacy of the cold war era. When uranium mining and processing for nuclear weapons and fuel were at their peak, uranium-containing wastes accumulated, resulting in a multitude of contaminated sites worldwide. In the United States specifically, there are more than 120 uranium contaminated sites, containing approximately 6.4 trillion liters of waste (33). The dominant uranium isotope in this waste, ^{238}U, has a half-life of approximately 4.5 billion years; additionally, uranium is a heavy metal and as such is toxic to cellular function. Uranium remediation strategies in recent years have focused on containment, to minimize migration of uranium in groundwater and prevent infiltration into surrounding water courses and potable water supplies. A promising approach to minimizing uranium migration is to catalyze the reduction of soluble U(VI) to the less-soluble U(IV) (31, 39). This process can be accelerated by the action of indigenous microorganisms fueled through the addition of exogenous carbon. Organic carbon addition stimulates biomass and microbial activity in these typically nutrient-poor environments and has a profound impact on microbial community composition (18, 35).

Although a wide range of bacteria are capable of uranium reduction (27), the reduction of U(VI) to U(IV) typically coincides with an increase in populations of metal-reducing bacteria such as members of the *Geobacteraceae* and others within the *Deltaproteobacteria* (18, 22, 28, 54). A recent pilot scale field application has suggested that uranium reduction is transient, with remobilization apparently associated with shifts to microbial populations possibly less capable of U(VI) reduction (2). While it has been shown that reoxidation of U(IV) can occur in the presence of molecular O_2, NO_3^-, and denitrification intermediates (NO_2^- and N_2O) (15), it is generally assumed that reduced U is stable under constant low redox conditions.

In an effort to monitor the long-term stability of bioreduced U(IV), we carried out flowthrough column incubations for more than 500 days using soil from Area 2 of the uranium-contaminated NABIR Field Research Center (FRC) at Oak Ridge, Tenn. (60). U(VI) reduction and immobilization appeared to be successful for the first 100 days; however, following this period of net U(VI) reduction, significant reoxidation of U(IV) and remobilization of U(VI) occurred. Surprisingly, during the period of net U(IV) reduction, columns were under constant reducing (methanogenic) conditions. To determine if the remobilization of U(VI) was associated with alterations in microbial populations, we used a high-density oligonucleotide microarray-based approach (10) which permits simultaneous monitoring of the population dynamics of almost 9,000 distin-

* Corresponding author. Mailing address: Ecology Department, Earth Sciences Division, Lawrence Berkeley National Lab, 1 Cyclotron Rd., MS 70A-3317, Berkeley, CA 94720. Phone: (510) 486-6584. Fax: (510) 486-7152. E-mail: elbrodie@lbl.gov.

† Supplemental material for this article may be found at http://aem.asm.org/.

guishable prokaryotic taxa/units (operational taxonomic units [OTUs]). To validate this approach, we also analyzed identical samples using a more routine clone library approach in addition to confirmatory tests using quantitative PCR. Microarrays targeting functional genes have been employed for analysis of biodegradative capabilities in contaminated sites (43), while 16S rRNA gene microarrays have been used successfully to differentiate bacteria in specific groups, such as *Enterococcus* (26), *Cyanobacteria*, (4), nitrifying bacteria (24), fish pathogens (61), and human colon microflora (37). Here we report the first application of high-density array technology in profiling the complex microbial communities of soils or sediments.

MATERIALS AND METHODS

Column setup and analyses. The experiment was conducted on soils heavily contaminated from past U waste disposal (U concentration of 206 mg kg^{-1}), obtained from the NABIR Field Research Center (FRC) at Oak Ridge National Laboratory (Tennessee) (http://www.esd.ornl.gov/nabirfrc/). Wet soils were passed through a 5.6-mm sieve, homogenized, packed into columns (200-mm length, 32-mm inner diameter [ID]) to a bulk density of 1.35 Mg m^{-3} and permeated with 10.7 mM Na-lactate solution (32 mM organic carbon [OC], pH 7.2) to stimulate U bioreduction by the native microbial community. The solution was supplied at a pore fluid velocity of 13 mm day^{-1} for 215 days, and thereafter decreased to 10 mm day^{-1}. Uranium reoxidation and remobilization was noted after 100 days, 115 days prior to the reduction in fluid pore velocity. Throughout the experiment, OC concentrations in the column effluents did not decline below 1 mM, demonstrating a nonlimiting supply of OC despite decreased pore fluid velocity. Further details are presented by Wan et al. (60).

Microbial activity. Dehydrogenase activity was determined by 2-(*p*-iodophenyl)-3-(*p*-nitrophenyl)-5-phenyl tetrazolium chloride (INT) reduction (57) under anaerobic conditions at room temperature.

DNA extraction from soil. DNA was extracted from 500 mg (wet weight) of soil using a BIO101 soil DNA extraction kit (QBiogene, Irvine, Calif.) according to the manufacturer's protocol. For samples taken prior to lactate stimulation ("Area 2" soil) and during the net U(VI) reduction phase ("Red."), 3 subsamples were taken from homogenized material and each extracted independently. For samples taken during the net U(IV) oxidation phase ("Ox."), 3 regions of the column were sampled, bottom, middle, and top, to test for variability in microbial community composition. Extracted DNA was quantified by absorption at 260 nm, and quantities are expressed in terms of μg DNA g soil (dry weight)$^{-1}$.

PCR amplification of 16S rRNA genes. 16S rRNA genes were amplified from each DNA extract in triplicate, and amplicons were pooled for each extract. PCRs were performed in a final volume of 100 μl and contained 1× Takara ExTaq PCR buffer, 2 mM $MgCl_2$, 300 μM primers (27F and 1492R) (62), 200 μM deoxynucleoside triphosphates, 50 μg bovine serum albumin, 0.5 μl of DNA extract, and 2.5 U ExTaq DNA polymerase (Takara Mirus Bio Inc., Madison, Wis.), and sterile milliQ H_2O up to 100 μl. For terminal restriction fragment length polymorphism (T-RFLP) analysis, the forward primer 27F was labeled with 6-carboxyfluorescein for detection by capillary electrophoresis. Cycling was performed using an iCycler (Bio-Rad, Hercules, Calif.) with initial denaturation at 95°C (5 min), followed by 35 cycles of 95°C (30 s), 53°C (30 s), and 72°C (1 min) using maximum temperature ramp rates, and a final extension at 72°C (7 min). Amplicons were run on 1% agarose gels. Amplicons were purified using an UltraClean PCR clean up kit (MoBio, Inc., Carlsbad, Calif.) according to the manufacturer's protocol, eluted in a final volume of 50 μl, and quantified by gel electrophoresis.

Microbial community screening by T-RFLP. Terminally labeled 16S rRNA gene amplicons were generated, purified, and quantified as described above. Approximately 400 ng of amplicon was digested overnight at 37°C in a 50-μl reaction mixture containing 2 U of restriction endonuclease AluI (NEB, Beverly, Mass.) in 1× NEB buffer 2. Digested amplicons were precipitated and desalted prior to capillary electrophoresis using glycogen as a coprecipitant and standard salt-alcohol procedures. Pellets were dried under a vacuum in the dark and resuspended in Hi-Di formamide (10 μl; Applied Biosystems, Foster City, Calif.). Aliquots (1 μl) were mixed with a GeneScan 500-Rox size standard (0.25 μl; Applied Biosystems) and 10 μl Hi-Di formamide. Immediately prior to electrophoresis, samples were denatured at 95°C for 5 min and cooled rapidly on ice. Electrophoresis was carried out using an ABI 3100 automated capillary sequencer (Applied Biosystems) run in GeneScan mode, with an electrokinetic injection time of 60 s. Terminal restriction fragments (TRFs) were sized using GeneScan v3.7 software (Applied Biosystems). All profiles consistently gave cumulative peak heights of over 10,000 fluorescence units, and only peaks of 50 fluorescence units or greater were considered further. TRFs were considered identical if they differed by <0.5 bp. Individual TRF heights were normalized as a percentage of total peak height. Principal component analysis (PCA) ordination was performed using the software package PC-Ord v.4.01 (MjM Software, Gleneden Beach, Oreg.).

Oligonucleotide probe selection and 16S PhyloChip design. The microarray probe design approach previously described for differentiating *Staphylococcaceae* (10) was applied to all known prokaryotic sequences of substantial length. Briefly, 16S rRNA gene sequences (*Escherichia coli* base pair positions 47 to 1473) were obtained from over 30,000 16S rRNA gene sequences that were at least 600 nucleotides in length in the 15 March 2002 release of the 16S rRNA gene database "Greengenes" (greengenes.lbl.gov). This region was selected because it is bounded on both ends by universally conserved segments that can be used as PCR priming sites to amplify bacterial or archaeal genomic material using only 2 to 4 primers (11). Putative chimeric sequences were filtered from the data set using the software package Bellerophon, preventing them from being misconstrued as novel organisms (19). The filtered sequences are considered to be the set of putative 16S rRNA gene amplicons. Sequences were clustered to enable each sequence of a cluster to be complementary to a set of perfectly matching (PM) probes. Putative amplicons were placed in the same cluster as a result of common 17-mers found in the sequence. The resulting 8,935 clusters, each containing approximately 0 to 3% sequence divergence, were considered OTUs representing all 121 demarcated prokaryotic orders. The taxonomic family of each OTU was assigned according to the placement of its member organisms in Bergey's Taxonomic Outline (16). The taxonomic outline maintained by Philip Hugenholtz (20) was consulted for phylogenetic classes containing uncultured environmental organisms or unclassified families belonging to named higher taxa. The OTUs comprising each family were clustered into subfamilies by transitive sequence identity according to a previously described method (10). Altogether, 842 subfamilies were found. The taxonomic position of each OTU and the accompanying NCBI accession numbers of the sequences composing each OTU can be viewed at http://greengenes.lbl.gov/Download/Taxonomic_Outlines/G2_chip_SeqDescByOTU_tax_outline.txt.

The objective of the probe selection strategy was to obtain an effective set of probes capable of correctly categorizing mixed amplicons into their proper OTU. For each OTU, a set of 11 or more specific 25-mers (probes) were sought that were prevalent in members of a given OTU but were dissimilar from sequences outside the given OTU. In the first step of probe selection for a particular OTU, each of the sequences in the OTU was separated into overlapping 25-mers, the potential targets. Then each potential target was matched to as many sequences of the OTU as possible. It was not adequate to use a text pattern search to match potential targets and sequences, since partial gene sequences were included in the reference set. Therefore, the multiple-sequence alignment provided by Greengenes was necessary to provide a discrete measurement of group size at each potential probe site. For example, if an OTU containing seven sequences possessed a probe site where one member was missing data, then the site-specific OTU size was only six. In ranking the possible targets, those having data for all members of that OTU were preferred over those found only in a fraction of the OTU members. In the second step, a subset of the prevalent targets was selected and reverse complemented into probe orientation, avoiding those capable of mis-hybridization to an unintended amplicon. Probes presumed to have the capacity to mis-hybridize were those 25-mers that contained a central 17-mer matching sequences in more than one OTU (34, 56). Thus, probes that were unique to an OTU solely due to a distinctive base in one of the outer four bases were avoided. Also, probes with mis-hybridization potential to sequences having a common tree node near the root were favored over those with a common node near the terminal branch. Probes complementary to target sequences that were selected for fabrication are termed PM probes. As each PM probe was chosen, it was paired with a control 25-mer (mismatching [MM] probe), identical in all positions except the 13th base (34). The MM probe did not contain a central 17-mer complementary to sequences in any OTU. The target probe and MM probes constitute a probe pair analyzed together. PM probes for each OTU can be viewed at http://greengenes.lbl.gov/cgi-bin/nph-show_probes_2_otu_alignments.cgi.

The chosen oligonucleotides were synthesized by a photolithographic method, at Affymetrix, Inc. (Santa Clara, CA), directly onto a 1.28-cm by 1.28-cm glass surface at an approximate density of 10,000 probes per μm^2 (6). Each unique probe sequence on the array had a copy number of roughly 3.2×10^6 (Affymetrix, personal communication). The entire array of 506,944 features was arranged as a square grid of 712 rows and columns. Of these features, 297,851 were oligonucleotide PM or MM probes targeting 16S rRNA gene sequences, and the

remaining were used for image orientation, normalization controls, or other unrelated analyses.

16S microarray sample preparation. From each of the 9 pools of 16S rRNA gene amplicons (3 each for Area 2, reduction, and oxidation sampling time points), 2 μg was DNaseI fragmented and biotin labeled, and an aliquot (1.85 μg) was hybridized to custom-made Affymetrix GeneChips (16S PhyloChip) according to the manufacturer's recommended protocol in the same way as cDNA is processed for expression arrays. Hybridization was performed at 60 rpm at 48°C overnight. PhyloChip washing and staining were performed according to standard Affymetrix protocols as described previously (32). In total, 9 microarrays were analyzed. Each PhyloChip was scanned and recorded as a pixel image, and initial data acquisition and intensity determination were performed using standard Affymetrix software (GeneChip microarray analysis suite, version 5.1).

Background subtraction. Background probes were identified as those producing intensities in the lowest 2% of all intensities. The average intensity of the background probes was subtracted from the fluorescence intensity of all probes. The noise value (N) was the variation in pixel intensity signals observed by the scanner as it read the array surface. The standard deviation of the pixel intensities within each of the identified background cells was divided by the square root of the number of pixels comprising that cell. The average of the resulting quotients was used for N in the calculations described below.

Detection and quantification criteria. Probe pairs scored as positive were those that met two criteria: (i) the intensity of fluorescence from the PM probe was greater than 1.3 times the intensity from the MM control and (ii) the difference in intensity, PM minus MM, was at least 500 times greater than the squared noise value ($>500\ N^2$). These two criteria were chosen empirically to provide stringency while maintaining sensitivity to the amplicons known to be present from sequencing of cloned Area 2 16S rRNA gene amplicons. The positive fraction (PosFrac) was calculated for each probe set as the number of positive probe pairs divided by the total number of probe pairs in a probe set. An OTU was considered present in the sample when over 90% of its assigned probe pairs were positive (PosFrac > 0.90). A hybridization intensity score (HybScore) was calculated in arbitrary units for each probe set as the trimmed average (maximum and minimum values removed before averaging) of the PM minus MM intensity differences across the probe pairs in a given probe set. All HybScores of <1 were shifted to 1 to avoid errors in subsequent logarithmic transformations. Two groupings are used to describe bacteria detected by the 16S microarray, the OTU and the subfamily. An OTU consists of a group of one or more sequences with typically 97 to 100% sequence homology, while a subfamily consists of a group of OTUs with typically no less than 94% sequence homology. Most comparisons are presented in terms of subfamilies as a conservative estimate of array specificity.

Array normalization. To account for variation from array to array, internal standards were added to each experiment. The internal standards were a set of 15 amplicons generated from yeast and bacterial metabolic genes, spiked into each amplicon pool prior to fragmentation. The known concentrations of the amplicons ranged from 4 pM to 605 pM in the final hybridization mix. HybScores resulting from the 15 corresponding probe sets were natural log transformed. Adjustment factors for each array were calculated by fitting a linear model using the least-squares method. The calculated array-specific adjustment factor was then subtracted from each probe set's ln(HybScore) for each array. When summarizing array results to the subfamily, the probe set (OTU) with the highest normalized array intensity (HybScore) was used as a representative.

Clone library comparison with microarray analyses. To compare the detection of bacteria using the 16S microarray with a more commonly used clone library approach, aliquots from the same 16S rRNA gene amplicon pools generated from Area 2 soil for array analysis were combined prior to ligation to pCR2.1 vectors. Following ligation, plasmids were transformed into TOP10 cells according to the manufacturer's instructions (Invitrogen, Carlsbad, Calif.). Clones (742) were randomly sequenced at the Department of Energy, Joint Genome Institute. Individual cloned rRNA genes were sequenced from each terminus, assembled using PHRED (14) and PHRAP (17), and required to pass quality tests of Phred 20 (base call error probability $< 10^{-2.0}$) to be included in the comparison. Chimeric sequences were identified using BELLEROPHON (19) and removed. Nonchimeric sequences of >1,300 bp in length were retained for further study and have been submitted to GenBank (see "Nucleotide sequence accession number" below). Representative 16S rRNA genes from each array-detected OTU were aligned with quality-assured sequences from the clone library. For phylogenetic placement, hypervariable regions of the alignment were masked, leaving 1,287 columns for generating a distance matrix. Phylogenetic trees were created from a neighbor-joining distance matrix using the ARB maximum parsimony algorithm (29). Rarefaction analysis of clone library sampling and Chao1 (5) estimates of richness were carried out using the software package DOTUR (46). Coverage was calculated according to the method described by Singleton et al. (49).

Identifying dynamic OTUs and subfamilies. Triplicate PhyloChip arrays were analyzed for each sample type, and normalized hybridization scores of representative OTUs within each subfamily were used to track changes in specific populations. A change in HybScore of 1,000 units is approximate to a 10-fold change in gene copy number (data not shown). To identify the most dynamic 16S rRNA gene amplicons, hierarchical clustering was performed on a prefiltered subfamily list containing only the 100 most variable subfamilies (based on standard deviation). This was performed within the R statistical programming environment (42) using the function "heatmap" within the package "made4" (9). Both samples and subfamilies were clustered with 1 Pearson's correlation as the distance metric and UPGMA (unweighted pair group method with arithmetic mean) as the linkage method. Heatmaps were reordered by dendrogram clustering to relocate covarying subfamilies and samples. The HybScore pattern over time for each cluster was plotted by averaging the intensities of all subfamilies within a cluster.

Real-time quantitative PCR confirmation of *Geothrix fermentans* population dynamics. Previously described primers (7) were used for quantitative real-time PCR (qPCR) (Gx.193F, 5′-GACCTTCGGCTGGGATGCTG-3′; Gx.448R, 5′-AGTCGTGCCACCTTCGT-3′). qPCR was performed using an iCycler iQ real-time detection system with the iQ Sybr Green Supermix kit (Bio-Rad). Reaction mixtures (final volume, 20 μl) contained 1X iQ SYBR green Supermix, 6 pmol of each primer, 10 μg bovine serum albumin, 0.5 μl DNA extract, and DNase/RNase-free water. Following enzyme activation (95°C, 3 min), 40 cycles of 95°C for 20 s, 51°C for 20 s, and 72°C for 30 s were performed. A specific data acquisition step at 87°C for 10 s was set above the melting temperature of potential primer dimers to minimize any nonamplicon SYBR green fluorescence. The copy number of *Geothrix fermentans* 16S rRNA gene molecules was quantified by comparing cycle thresholds to a standard curve (in the range of 10^2 to 10^8 copies), run in parallel, using *Geothrix fermentans* 16S rRNA gene amplicons generated by PCR (using universal bacterial primers 27F and 1492R) (62). Regression coefficients for the standard curves were consistently greater than 0.99, and postamplification melt curve analyses displayed a single peak at 90°C, indicative of specific *Geothrix fermentans* 16S rRNA gene amplification (data not shown).

Nucleotide sequence accession numbers. The nonchimeric sequences of >1,300 bp in length have been submitted to GenBank and assigned accession numbers DQ125500 to DQ125935.

RESULTS

Microbial biomass and activity. Unsurprisingly, column infusion with lactate stimulated microbial biomass, with DNA concentrations in soil samples taken during net U reduction being almost 10 times greater than in the original soil (Fig. 1A). However, biomass decreased between the net reduction and net oxidation phases. Lactate infusion also stimulated microbial activity (Fig. 1B), which increased from barely detectable levels in the original soil (1.2 μg iodonitrotetrazolium formazan [INF] $g^{-1}\ h^{-1}$) to over 690 μg INF $g^{-1}\ h^{-1}$ in samples from the net U reduction phase. Significantly, despite decreased biomass, microbial activity more than doubled to over 1,540 μg INF $g^{-1}\ h^{-1}$ in the net oxidation phase.

T-RFLP analysis of bacterial population shifts. Principal component analysis (PCA) of 16S rRNA gene-based T-RFLP profiles showed bacterial community divergence from the original starting material (Fig. 2), with separation occurring along both axes. The net U reduction phase and net oxidation phase samples also diverged but only along axis 2 (Fig. 2), indicating closer association to each other than to the original Area 2 soil community.

Clone library comparison with microarray analyses. Relatively large clone libraries (742 clones) were sequenced, assembled, and trimmed for quality before being screened for chimeras. After this quality control process, 429 high-quality clones remained, with almost 24% being detected as chimeric,

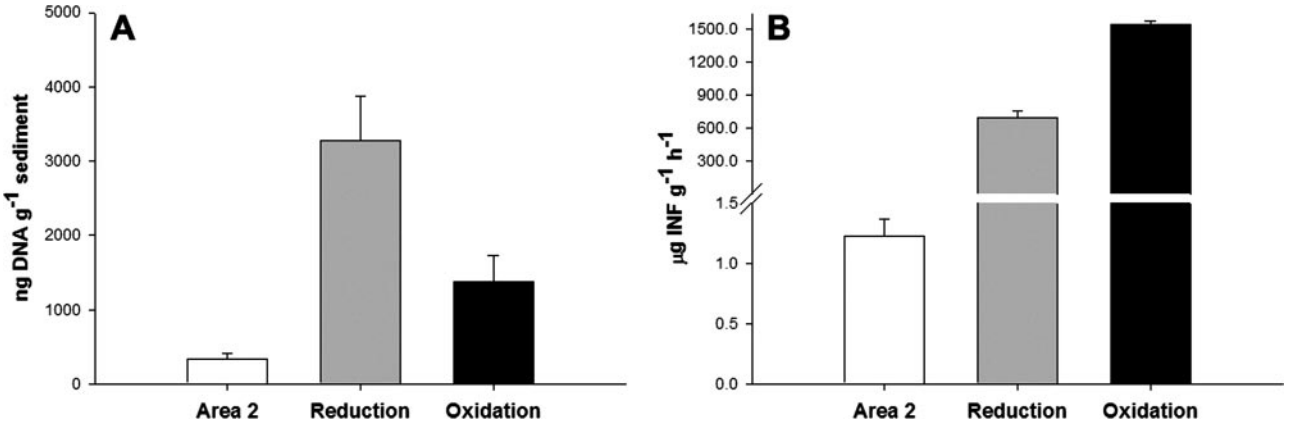


FIG. 1. Plots show DNA concentration (A) and microbial activity (dehydrogenase) (B) in soil samples taken from Area 2 soil before column packing and during net U reduction and oxidation phases. Bars show means ± standard errors ($n = 3$).

possibly due to the harsh nature of bead beating. Figure 3 shows a rarefaction curve with a sampling efficiency determined at 100% sequence homology (i.e., unique clones) and 99% sequence homology (i.e., the array defined OTU level). Neither curve reached an asymptote, indicating incomplete sampling of the Area 2 soil bacterial community after 429 high-quality clone sequences were analyzed. Coverage rates were 0.49 at 100% homology (unique), 0.61 at 99% homology (OTU), and 0.84 at 94% homology (subfamily).

Phylogenetic analysis of the composition of the Area 2 clone library (Fig. 4) shows an amplicon pool dominated by the *Actinobacteria* phylum accounting for almost 60% of the clones sampled. The families *Micrococcaceae* and *Microbacteriaceae* dominated the *Actinobacteria* clones, with the dominant species having their closest cultured homologs in *Arthrobacter* spp. and *Microbacterium* spp., respectively. The *Firmicutes* and the *Proteobacteria* represented the next most abundant phyla at 16% of clones each, with the latter being dominated by *Alphaproteobacteria*. Only one clone within the *Deltaproteobacteria* was detected.

In total, all clones detected could be assigned to 63 groups, most at the family or subfamily level. Of the 63 clone groups detected, the 16S microarray accurately identified a member of the same group in 59 cases, the other four cases consisted of two novel singleton clones (AKAU3608 and AKAU3516) not placed in any recognized bacterial grouping, a group of 11 clones within the family *Nocardiaceae* (subfamily 3) and a group of 13 novel *Chloroflexi* clones (which we have termed "uranium-contaminated soil clones") with low homology to previously sequenced *Chloroflexi* (3). In addition to the 59 groups of sequences in common between the clone and array approach, the array also detected 194 additional groups at the subfamily level. A phylogenetic tree comparing array and clone library detection of bacterial groups is available (see Fig. S1 in the supplemental material). Among the groups neglected by the cloning approach but detected by the array were organisms of significance to uranium bioremediation such as the

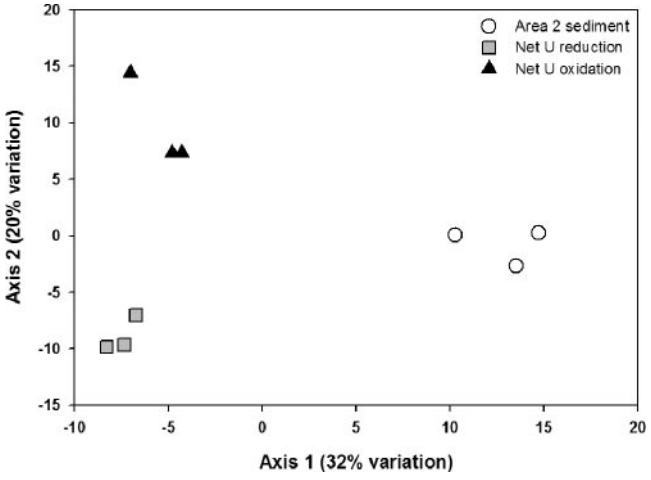


FIG. 2. PCA of 16S T-RFLP profiles showing bacterial community divergence from original FRC Area 2 soil prior to column packing and through net uranium reduction and reoxidation phases.

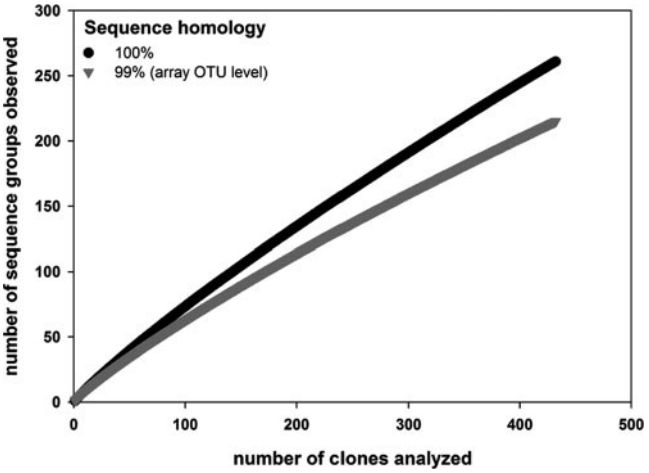


FIG. 3. Rarefaction analysis of clone library coverage of microbial diversity in uranium-contaminated Area 2 soil samples. Two sequence homology levels are shown: 100% represents the coverage estimate when clones are placed in groups using this homology cutoff and 99% represents the coverage estimate based on clones grouped using a 99% sequence homology cutoff, which corresponds to the typical divergence of sequences within array-defined OTUs.

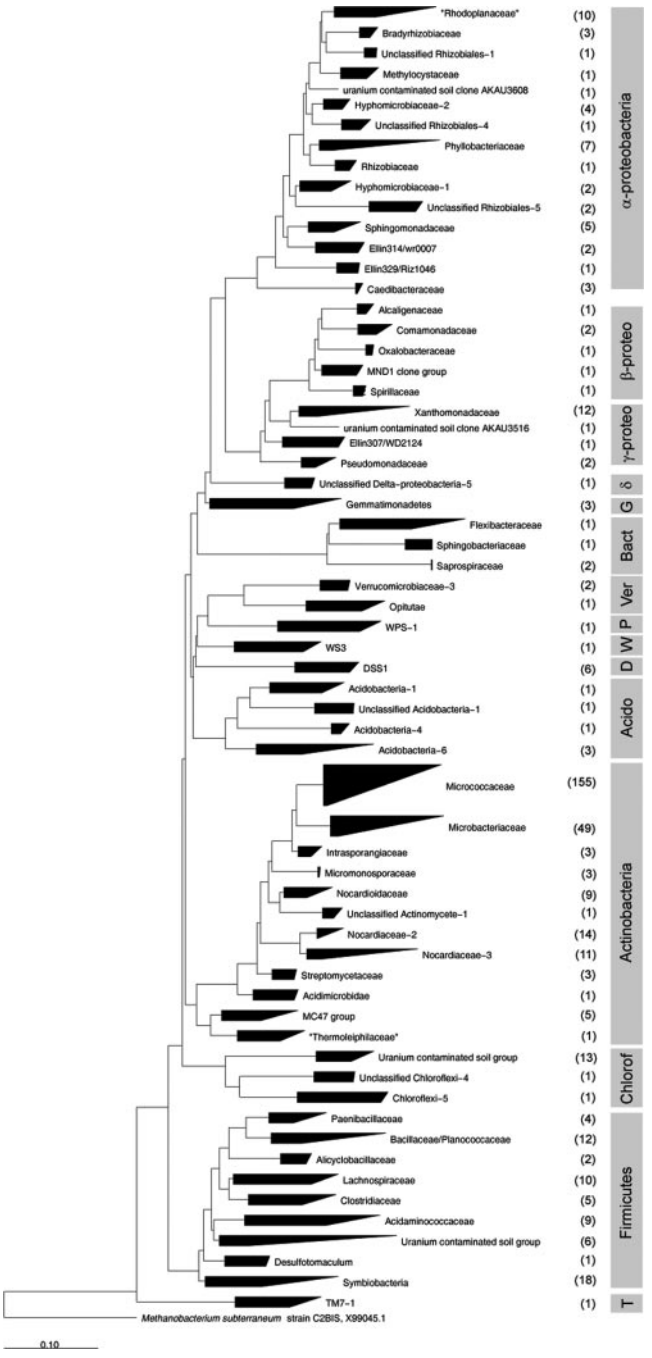


FIG. 4. Neighbor-joining tree showing phylogenetic position of 16S rRNA gene clones generated from Area 2 soil prior to column packing. Sequences were inserted by parsimony into a tree containing all bacteria and archaea in the "greengenes" database (as of 3 May 2005). Abbreviations are as follows: β-proteo, *Betaproteobacteria*; γ-proteo, *Gammaproteobacteria*; δ, *Deltaproteobacteria*; G, *Gemmatimonadetes*; Bact, *Bacteroidetes*; Ver, *Verrucomicrobia*; P, *Planctomycetes*; W, candidate division WS3; D, candidate division DSS1; Acido, *Acidobacteria*; Chlorof, *Chloroflexi*; T, candidate division TM7. Values in parentheses indicate numbers of clones in each group. *Methanobacterium subterraneum* strain C2BIS was used to root the tree. The scale bar represents 10 inferred nucleotide changes per 100 nucleotides analyzed.

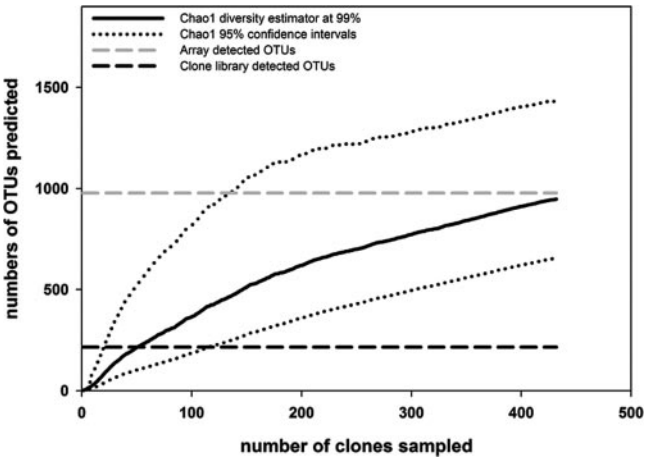


FIG. 5. Chao1 richness estimates of Area 2 soil bacterial communities based on clone library sampling. Richness estimates were made using a 99% sequence homology group classification equivalent to array-defined OTUs. The dashed gray line denotes the mean number of OTUs detected by array in Area 2 samples. The dashed black line denotes the number of OTUs detected by clone sequence analysis in Area 2 samples.

Geobacteraceae, which we have previously shown by primer-specific PCR to be present in this Area 2 sediment genomic DNA sample (60).

As rarefaction curves did not reach an asymptote, a nonparametric richness estimator, Chao1 (5), was used to predict richness (Fig. 5). The predicted richness of 948 OTUs corresponds well with the array-detected richness of 978 OTUs. While the curve has not yet leveled off, it is also possible that richness is being underestimated.

Identifying dynamic OTUs and subfamilies. Dynamic OTUs and subfamilies were detected by confining the analysis to the 100 subfamilies exhibiting the greatest standard deviation between the phases sampled. Hierarchical cluster analysis allowed detection of correlations between samples and also between subfamilies (Fig. 6). Array analysis indicated that Area 2 soil sample communities formed a separate cluster from those sampled during the net uranium reduction or oxidation phases, with the latter two arising from a common node. Overall, five groups of dynamic subfamilies were detected by cluster analysis (Fig. 6; see Table S2 in the supplemental material).

Cluster group 1 sequences increased following lactate amendment to reach a peak during the reduction phase and subsequently decline during the oxidation phase. This group contained mostly *Proteobacteria* with some *Actinobacteria*, *Firmicutes*, and members of the *Deinococcus-Thermus* and *Planctomyces* phyla. The *Proteobacteria* which responded in this manner were mostly orders within the α (*Azospirillales*, *Caulobacterales*, *Sphingomonadales*) or β (*Burkholderiales*, *Neisseriales*, *Rhodocyclales*) subphyla and are typically capable of nitrate reduction. Sequences within the *Cellulomonadaceae* family (order *Actinomycetales*) also increased following lactate addition and subsequently declined. Two orders of *Deltaproteobacteria*, *Bdellovibrionales* and *Desulfobacterales*, and a *Firmicute* of the homoacetogenic genus *Sporomusa* (44) also responded in a similar way.

Cluster group 2 sequences increased following lactate stim-

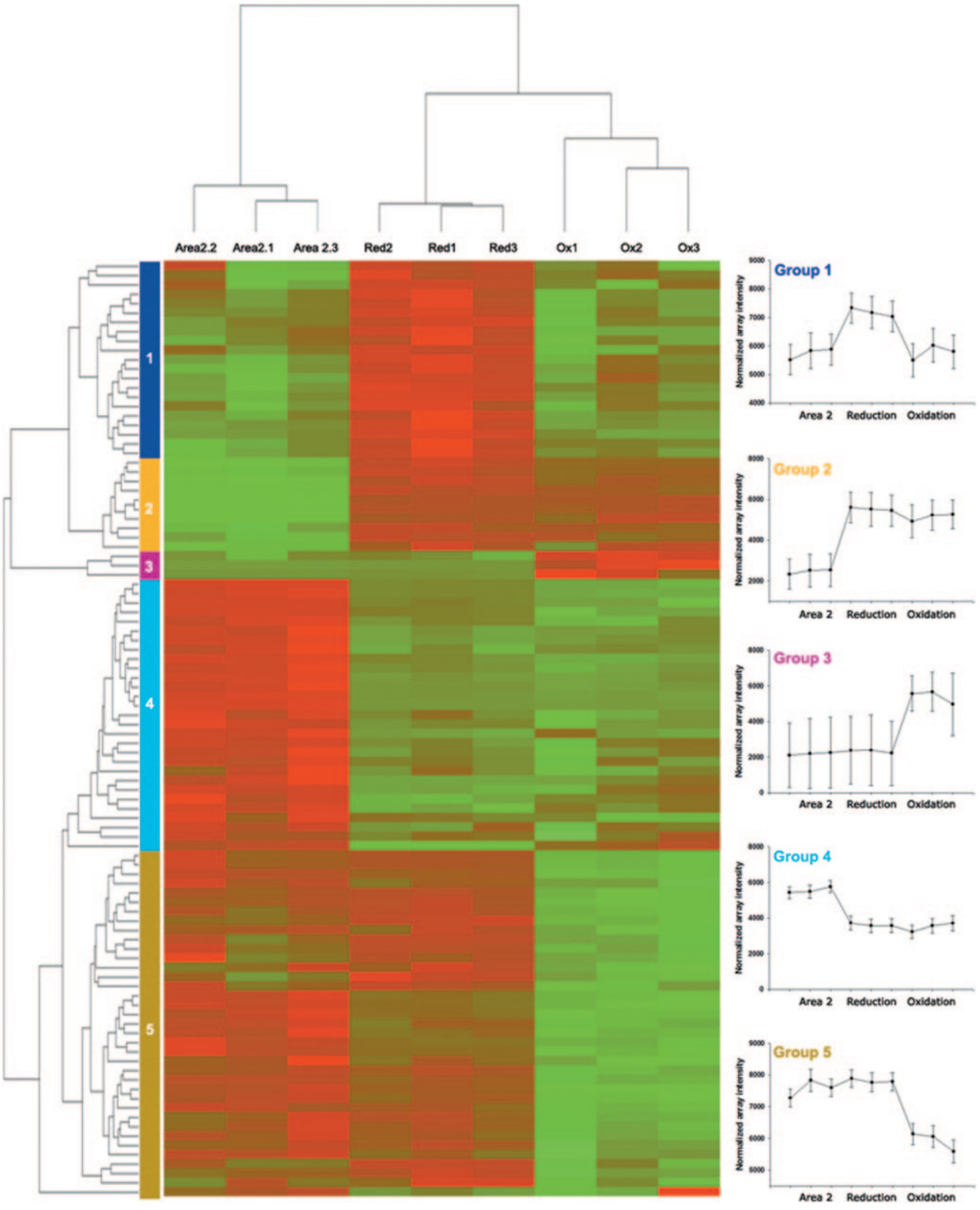


FIG. 6. Heatmap and dendrograms showing the response of 100 bacterial subfamilies (shown on *y* axis) exhibiting the highest standard deviation between samples (shown on *x* axis) taken pre-lactate stimulation (Area2.1, Area2.2, and Area2.3), during net uranium reduction (Red1, Red2, and Red3), and during net uranium reoxidation (Ox1, Ox2, and Ox3). The color gradient from green to red represents increasing array hybridization intensity. Five main response groups were detected, and the average intensity (HybScore) of the cluster group response is presented in line plots to the right of the heatmap. Error bars represent standard errors in HybScore differences between all subfamilies in a cluster.

ulation and remained at elevated concentrations in the amplicon pool throughout the net uranium reduction and oxidation phases. This group contained three families of *Deltaproteobacteria*, the *Nitrospinaceae*, *Geobacteraceae*, and *Anaeromyxobacter*. In addition to *Deltaproteobacteria*, members of the *Acidobacteriaceae* were also detected, specifically the species *Geothrix fermentans*. This cluster also contained *Rhodocyclaceae*, *Comamonadaceae* (*Malikia spinosa*), and an *Actinomycete* of the genus *Intrasporangium*.

Cluster group 3 sequences comprised three subfamilies which were of lower abundance in the original Area 2 soil and reduction-phase soil amplicon pools but increased during the net uranium oxidation phase. Two of the subfamilies were from the *Acidobacteria*, with both representative clones coming from halogenated organic carbon-degrading consortia (47, 59). The other member of cluster group 3 was a bacterium within the *Desulfovibrionaceae* family.

Cluster group 4 was a large cluster containing sequences from 28 bacterial subfamilies whose amplicons declined in abundance or disappeared following lactate addition. The groups consisted of four families of *Actinomycetes*, including *Microbacteriaceae*, one of the dominant families in the original Area 2 soil amplicon pool. Bacteria within the *Chloroflexi*, *Cyanobacteria*, and *Acidobacteria* also followed this pattern. Five families within the order *Bacillales* also declined following lactate addition, as did 8 families within the *Alphaproteobacteria*. The *Xanthomonadaceae*, which initially was the most abundant *Gammaproteobacteria* family in the Area 2 soil clone library, also declined.

The final cluster group observed, group 5, consisted mainly of *Actinobacteria*, *Firmicutes*, *Cyanobacteria*, and *Alphaproteobacteria* sequences, representing most of the initially dominant clones present in the original Area 2 soil, including the family *Micrococcaceae*. Following lactate addition, these amplicons appeared to maintain similar abundance through the uranium reduction phase but showed a sharp decline during the uranium reoxidation phase.

Analysis of the dynamics of all subfamilies detected indicated that, between the net U reduction and net U oxidation phases, the amplicon signatures of many (206 subfamilies) did not change, while a large number (139 subfamilies) declined, with only a small proportion (13 subfamilies) increasing in quantity (see Table S1 in the supplemental material). However, compared with other bacterial groups, a disproportionate amount of *Actinobacteria* and α-, β-, and γ-proteobacteria subfamilies showed a decline in amplicon concentration between the reduction and oxidation phases.

Real-time quantitative PCR confirmation of *Geothrix fermentans* population dynamics. We have previously demonstrated via 16S microarray analysis and confirmed by qPCR that bacteria within the family *Geobacteraceae* increased in abundance following lactate addition and maintained an increased abundance through both uranium reduction and reoxidation phases (60). To confirm array observations that the *Acidobacteria* species *Geothrix fermentans* showed this pattern of increased abundance, as detected by the 16S microarray, we used primers specific for this species together with a SYBR green-based quantitative PCR approach. This showed that initial populations of *G. fermentans* in Area 2 soil were below detection limits (10^3 copies g soil^{-1} for qPCR) (Table 1); however, following lactate stimulation, *G. fermentans* abundance increased to over 10^6 copies g soil^{-1} during the uranium reduction phase. Samples from the uranium reoxidation phase also showed increased *G. fermentans* populations over 10^6 copies g^{-1}, correlating well with 16S microarray observations (Table 1).

TABLE 1. Real-time quantitative PCR analysis of changes in *Geothrix fermentans* gene copy number compared with array intensity

Sample source[a]	Replicate no.	Normalized array intensity[d]	No. of 16S rRNA gene copies g^{-1}
Area 2	1	133[b]	ND[c]
	2	119[b]	ND
	3	135[b]	ND
Reduction	1	6,309	1.7×10^6
	2	6,295	3.3×10^6
	3	6,122	4.7×10^6
Oxidation	1	5,497	2.7×10^6
	2	6,054	4.0×10^6
	3	6,433	9.5×10^6

[a] Area 2 indicates the original sediment prior to column packing, reduction indicates samples taken during the net uranium reduction phase, and oxidation indicates samples taken during the net uranium oxidation phase.
[b] The *Geothrix fermentans* probe set was below the positive fraction threshold of 0.90 (90% of probe pairs must pass) in these samples and therefore is considered not detected by the array.
[c] ND, none detected.
[d] Array intensities are expressed in arbitrary fluorescence units.

DISCUSSION

The observed reoxidation of bioreduced U(IV) reported by Wan et al. (60) is of significant concern, considering the limited options available for uranium remediation. From this work, it now appears that even under highly reducing (methanogenic) conditions, the stability of bioreduced U(IV) can be transient. A by-product of microbial metabolism, dissolved inorganic carbon, coupled with high Ca^{2+} concentrations was shown by Wan et al. (60) to drive the redox potential of the U(VI):U(IV) oxidation couple more negative, thus accounting for the thermodynamic feasibility of U(IV) oxidation under reducing conditions. The terminal electron acceptor(s) for U(IV) oxidation have not yet been identified for this process, but it has been hypothesized that either Fe(III) or Mn(IV) is the likely candidate (60). What remains unclear is the role of alterations in microbial community composition coincident to the changing process dynamics. Is there an active microbially catalyzed oxidation process? Or is the microbial role indirect, simply through the production of CO_2? If U(VI) reduction and U(IV) reoxidation are thermodynamically favorable under similar redox conditions, then the possibility remains that a balance exists between the two processes and that during the net reoxidation phase, reoxidation rates exceed those of reduction. In terms of microbial community composition, a possible explanation for this would be the loss of microbial species or functional groups capable of U(VI) reduction. It is conceivable, for example, that *Geobacter* species, having reduced all available iron and uranium within their immediate vicinity or range of motility, are faced with no terminal electron acceptor with which to couple the oxidation of organic carbon, resulting in population decline as other organisms, possibly sulfate re-

ducers, fermenters, or methanogens, are able to out-compete the metal reducer. To test for functionally significant alterations in bacterial community composition, we analyzed samples from both the net U reduction and net U reoxidation phases of a column incubation using historically contaminated soil from Oak Ridge, Tenn.

Using DNA as a proxy of microbial biomass, it was clear that lactate infusion into contaminated soils resulted in significant biomass stimulation, which reached a peak during the net U reduction phase. However, biomass was observed to decline from the net U reduction to the net U oxidation phase despite a nonlimiting supply (60) of OC. A decrease in suitable terminal electron acceptors (TEAs), as indicated by methane production, low effluent concentrations of iron, uranium, and manganese, and low redox potential (60), may have resulted in decreased biomass as microorganisms become unable to fuel cellular metabolism. This effectively would result in a decline in organisms solely reliant on oxygen, nitrate, iron, uranium, manganese, or sulfate as TEAs for oxidation of OC. Despite the significant reduction in bacterial biomass, microbial activity (using dehydrogenase activity as a proxy) more than doubled between net reducing and oxidizing phases. However, the INT reduction assay used to estimate dehydrogenase activity may also be considered an assay of total reducing potential. Nonviable cells may possess residual reduction capacity, and nonbiological reduction of INT may be possible in environments with redox potentials below that of INT (−90 mV) (50), as found in the soil columns during this study (−250 to −150 mV) (60). Nevertheless, as the system redox potential was below that of INT in columns during both U reduction and oxidation phases, the increase in INT reduction during the oxidation phase cannot be solely due to abiotic effects. Studies using *Escherichia coli* have demonstrated that dehydrogenase activity, measured by INT reduction, is greater in cells growing fermentatively than in cells exhibiting anaerobic respiration (50). Such a shift to fermentation as the dominant process following TEA exhaustion may explain the observed increase in INT reduction in the net U oxidation phase despite reduced biomass.

To investigate shifts in the composition and dynamics of bacterial populations during the remediation process, we employed a high-density DNA microarray approach together with a traditional clone library as validation. Both methods sample a gene amplicon pool following PCR amplification; however, clone libraries are known to underestimate microbial diversity either due to insufficient number of clones sequenced (46) or cloning bias (58). DNA microarray approaches, on the other hand, are not subject to cloning bias and have the capacity to sample a far greater number of molecules, which represents a significant advantage, particularly when highly abundant sequences mask clone sampling efforts. However, it must be noted that, in its current form, the DNA microarray approach described here is subject to end-point PCR bias in the same way as the accepted 16S rRNA gene clone libraries. This bias is primarily due to primer degeneracies resulting in preferential amplification of specific sequences and potential misrepresentation of sequence abundance within genomic DNA extracts (40); however, misrepresentations based on these studies of noncomplex mixtures appear to be on the order of two- to fourfold (36). This bias is thought to be magnified by increasing the number of PCR cycles and has only recently been tested in complex environmental mixtures (1, 30). Acinas et al. (1) observed no effect of cycle number on PCR bias, where clone distribution was generally in good agreement with previously observed distributions for bacterioplankton. In this study, we used a conservative approach by confining analyses to the 100 most dynamic subfamilies detected, all of which exhibited changes 10-fold or greater. PCR bias is also thought to be most significant in the early cycles where random priming effects may result in misrepresentation of phylotype abundances. To minimize this effect, we combined triplicate PCRs for each sample (40), and although PCR bias due to template reannealing is thought to be minimal in complex environmental samples (51, 52), we attempted to reduce any bias by using the fastest possible temperature ramp times during PCR cycling (25). Like all PCR-based clone library analyses, we cannot rule out bias in altering the relative abundance of phylotypes or OTUs; however, we have taken several steps to minimize its influence.

The design of the probes contained in the 16S microarray was based on approximately 30,000 sequences of 600 bp or greater which were available in public databases in March 2002. Therefore, it was possible that recently published novel sequences would not be detectable using the 16S microarray. To confirm the ability of the 16S microarray to accurately classify bacterial sequences in the uranium-contaminated soil from Area 2 at the Oak Ridge Field Research Center (60), 16S amplicon pools from three independent soil extractions were processed and hybridized to separate 16S microarrays. For clone libraries, an aliquot of the three amplicon pools was combined and used in the ligation reaction. Assembled clone sequences were obtained for 742 clones, which after quality screening, yielded 429 high quality sequences.

Amplicons from the initial Area 2 soil bacterial community were dominated by high-G+C gram-positive bacteria, particularly the families *Micrococcaceae* (*Arthrobacter* spp.) and *Microbacteriaceae* (*Microbacterium* spp.). *Arthrobacter* species have been shown to be resistant to high concentrations of uranium while also capable of intracellular heavy metal accumulation (53) and uranium adsorption (55); as such, they may play a role in natural uranium immobilization at the Oak Ridge site. Although *Microbacterium* spp. are known to be capable of heavy metal (chromate) reduction (21) and have been reported to accumulate actinides (plutonium) intracellularly (23), their potential role in uranium immobilization is not clear. Other initially abundant 16S sequence types included α-proteobacteria within the *Rhizobiales* (*Hyphomicrobiaceae*, "*Candidatus* Rhodoplanaceae," *Phyllobacteriaceae*) and the *Sphingomonadales*, organisms typically considered oligotrophic. *Hyphomicrobium* has previously been detected at the Oak Ridge site as the dominant *nirK* type denitrifier (63).

Comparison of phylogenetic composition of the Area 2 soil by clone library and 16S microarray analysis showed that most clone groups (59 of 63) were detected by the array. As might be expected, two of the clone groups not identified by the array were singletons which could not be placed in any recognized bacterial phylum. The other two groups consisted of 11 clones within the *Nocardiaceae* (subfamily 3) and 13 clones representing novel *Chloroflexi*. While the 16S microarray was capable of accurately detecting most bacteria sampled by the clone library analysis, the 16S microarray detected many more (253 groups)

bacteria than the clone library (63 groups). One explanation for this observation may be inadequate clone library sampling, resulting in an underestimation of bacterial richness. Rarefaction analysis of clone library sampling demonstrated that the number of clones sequenced was indeed insufficient to adequately determine diversity within the Area 2 soil sample. As is typical in analysis of soil microbial communities, the predicted richness (948 OTUs), by Chao1 estimates in this case, was far greater than that observed by cloning (215 OTUs). Array analysis, however, detected a number (978 OTUs) similar to the Chao1 diversity estimate; therefore, it appears the array approach is capable of detecting substantially more sequence types than a relatively large sample clone library.

Another possible explanation for the discrepancy between array and clone library observed bacterial richness is the influence of nonspecific hybridization leading to false positives in array analysis. In designing this 16S microarray,we specifically chose the Affymetrix platform due to its probe capacity and the added confidence provided by the probe-pair strategy. Here, each perfectly matching probe is paired with a corresponding mismatch probe containing a single-base mismatch at the central nucleotide position, providing a nonspecific hybridization control for each probe on the array. In addition, we have previously confirmed by primer-specific PCR that families of bacteria that are detected by the 16S microarray and overlooked by cloning are indeed present (60).

A major strength of using high-density arrays is the ability to perform replicated analyses in a routine and reproducible manner. While clone libraries may also be replicated, the cost is generally prohibitive, and additional bias may occur (52). In this study, we required an approach that allowed statistical analysis of the changes in bacterial communities and their members over time to determine if the observed uranium reoxidation was a result of shifts in microbial groups of interest. T-RFLP analysis indicated that the bacterial community composition had changed following lactate addition but that the communities during the U reduction and U oxidation phases were only separated along one PCA axis. This is a useful approach, but 16S rRNA gene-based T-RFLP of diverse communities with universal primers does not provide reliable information regarding community composition. To mine the large quantities of data produced by this phylogenetic array and identify groups of covarying bacterial sequences, we used clustering approaches in a similar manner to those used for more standard gene expression arrays (9). Hierarchical clustering of the 100 most variable subfamilies detected by the 16S microarray demonstrated that variable sequences fell into 5 primary response groups from which it is possible to suggest common population dynamics. Cluster group 4 exhibited a distinct profile, declining rapidly following lactate addition through the U reduction and U reoxidation phases. This cluster consisted mostly of sequences from the actinobacteria, bacilli, and α-proteobacteria, which presumably declined as bacteria with more efficient lactate utilization flourished. Cluster group 1 sequences increased initially following the lactate amendment but declined during the U reoxidation phase. This group was primarily composed of α- and β-proteobacteria typically capable of nitrate reduction in addition to some bacteria recognized as potential uranium reducers, such as the family *Cellulomonadaceae* (45). While some *Deltaproteobacteria* sequences were also in this cluster, neither group has been reported to reduce U(VI). Interestingly *Bdellovibrio* spp. are bacterial parasites, and their increase may simply be a response to the increased biomass of their prey. The significant increase of this cluster of sequences followed by a significant decline may indicate a rapid oxidation of added carbon coupled with reduction of available TEAs and a subsequent decline as suitable TEAs are exhausted. Cluster group 5 sequences were initially at a higher density in the original Area 2 sediment amplicon pool prior to lactate amendment, but unlike cluster group 4, they did not immediately decline following carbon addition. Again, like cluster group 4, this sequence group contained many families of *Actinobacteria*, including the family *Micrococcaceae*, the dominant 16S rRNA clone type. This may be significant in explaining some of the observed mobilization of uranium in these column studies, as any intracellular uranium accumulated by these organisms would be released upon cell lysis. Cluster group 4 also contained many *Alpha-* and *Gammaproteobacteria* capable of nitrate reduction, which may have retained competitiveness until the in situ nitrate (only 1.1 mg kg^{-1}) was depleted.

Cluster group 2 sequences increased following lactate addition but showed no decline during the observed reoxidation phase, and significantly, this group contained many typical U(VI) reducing bacteria such as the family *Geobacteraceae* and the genera *Geothrix* and *Anaeromyxobacter*. These bacteria have been detected before in either iron- or uranium-reducing enrichments using other NABIR FRC sediments (36, 38, 48). We confirmed this observation for *Geothrix fermentans* using qPCR in this study and for *Geobacteraceae* in a previous study (60). The fact that these bacterial sequences did not decline during the oxidation phase strongly suggests that a shift in community composition away from metal reducers was not a factor in reoxidation of U(IV).

However, the questions remain, how would these bacterial sequences remain at elevated densities despite low concentrations of U(VI) and other suitable TEAs in highly reducing (methanogenic) conditions, and why did uranium reduction rates decline? One possible strategy for survival in the absence of a suitable abiotic TEA would be a syntrophic association with an organism acting as a biological electron acceptor for excess electrons generated during fermentation. The significant increase observed in dehydrogenase activity between the U reduction and U reoxidation phase may support such a shift to fermentative pathways. A closer look at other bacterial sequences in this cluster also indicates bacteria capable of survival under highly reducing conditions (i.e., reductive dechlorination, TCE-contaminated site clone FTLM5 [accession no. AF529125]). Cluster group 3 contained bacteria whose amplicons increased in abundance between the U reduction and U oxidation phases; significantly, two of the organisms detected (*Acidobacteria*) matched clones previously detected in reductive dechlorinating enrichments (47, 59), while the third organism was a *Desulfovibrio*-related isolate. This covariation of a sulfate reducer and reductive dechlorinating organisms may indicate an interaction such as interspecies hydrogen transfer (13). Drzyzga et al. (12) demonstrated the maintenance of a *Desulfovibrio* sp. in a chemostat in coculture with a dehalorespiring bacterium, which occurred despite the absence of sulfate. The authors proposed that the two organisms were

in syntrophy with *Desulfovibrio*, fermenting lactate and using the dehalogenating bacterium as a "biological electron acceptor." Another possibility may be a syntrophic association between the *Desulfovibrio*-like organism and methanogens (41), as we observed methane production (60) during both the U reduction and oxidation phases. Significantly, *Geobacter* spp. may also benefit from syntrophic relationships with biological electron acceptors. *Geobacter sulfurreducens* was shown to be capable of sustaining growth by electron transfer, following acetate oxidation, to an anaerobic partner bacterium in the absence of Fe(III) or other TEAs (8).

While the observed increase in a *Desulfovibrio* species deserves further investigation, the observation that other metal reducing bacteria, *Geothrix* and *Geobacter* spp., showed no decline indicates that a community shift away from more-efficient U(VI) reducers was not a primary factor in the net oxidation of U(IV), as was perhaps the case in the Colorado field study of Anderson et al. (2). The reoxidation of bioreduced U(IV) has been shown to be thermodynamically possible under the reducing conditions in these column studies (60), with Fe(III) and Mn(IV) being identified as possible electron acceptors. While the bacterial dynamics documented here by the array data do not support direct microbial involvement in this process, neither do the data eliminate this possibility. This is a focus of future studies. The high-density oligonucleotide-based microarray approach used here constitutes a powerful tool which holds significant promise for application in microbial ecology. The ability to detect and monitor a substantial portion of the microbial population in a routine, accurate, and reproducible manner will aid our understanding of the functioning of microbial communities, ranging from natural ecosystems to managed microbial consortia.

ACKNOWLEDGMENTS

We thank Phil Hugenholtz for assistance with phylogenetic analysis, chimera identification, and access to high-quality sequence alignments. We also thank Thomas Huber (BELLEROPHON), Pat Schloss (DOTUR), and Aedín Culhane (MADE4) for assistance with software and scripts. We thank John Coates for the generous gift of *Geothrix fermentans* cells, and we are grateful for helpful comments provided by Eric Roden and Sue Lynch.

We are grateful for the support of the U.S. DOE Natural and Accelerated Bioremediation Research (NABIR) Program in the Office of Biological and Environmental Research. This work was supported in part by contract no. DE-AC02-05CH11231.

REFERENCES

1. **Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz.** 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl. Environ. Microbiol. **71:**8966–8969.
2. **Anderson, R. T., H. A. Vrionis, I. Ortiz-Bernad, C. T. Resch, P. E. Long, R. Dayvault, K. Karp, S. Marutzky, D. R. Metzler, A. Peacock, D. C. White, M. Lowe, and D. R. Lovley.** 2003. Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. Appl. Environ. Microbiol. **69:**5884–5891.
3. **Asami, H., M. Aida, and K. Watanabe.** 2005. Accelerated sulfur cycle in coastal marine sediment beneath areas of intensive shellfish aquaculture. Appl. Environ. Microbiol. **71:**2925–2933.
4. **Castiglioni, B., E. Rizzi, A. Frosini, K. Sivonen, P. Rajaniemi, A. Rantala, M. A. Mugnai, S. Ventura, A. Wilmotte, C. Boutte, S. Grubisic, P. Balthasart, C. Consolandi, R. Bordoni, A. Mezzelani, C. Battaglia, and G. De Bellis.** 2004. Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. Appl. Environ. Microbiol. **70:**7161–7172.
5. **Chao, A.** 1984. Non-parametric estimation of the number of classes in a population. Scand. J. Stat. **11:**265–270.
6. **Chee, M., R. Yang, E. Hubbell, A. Berno, X. C. Huang, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris, and S. P. A. Fodor.** 1996. Accessing genetic information with high-density DNA arrays. Science **274:**610–614.
7. **Coates, J. D., and L. A. Achenbach.** 2002. The biogeochemistry of aquifer systems, p. 719–727. *In* C. J. Hurst, R. L. Crawford, G. R. Knudsen, M. J. McInerney, and L. D. Stetzenbach (ed.), Manual of environmental microbiology. ASM Press, Washington, D.C.
8. **Cord-Ruwisch, R., D. R. Lovley, and B. Schink.** 1998. Growth of *Geobacter sulfurreducens* with acetate in syntrophic cooperation with hydrogen-oxidizing anaerobic partners. Appl. Environ. Microbiol. **64:**2232–2236.
9. **Culhane, A. C., J. Thioulouse, G. Perriere, and D. G. Higgins.** 2005. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics **21:**2789–2790.
10. **DeSantis, T. Z., I. Dubosarskiy, S. R. Murray, and G. L. Andersen.** 2003. Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics **19:**1461–1468.
11. **Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace.** 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl. Environ. Microbiol. **64:**3869–3877.
12. **Drzyzga, O., J. Gerritse, J. A. Dijk, H. Elissen, and J. C. Gottschal.** 2001. Coexistence of a sulphate-reducing *Desulfovibrio* species and the dehalorespiring *Desulfitobacterium frappieri* TCE1 in defined chemostat cultures grown with various combinations of sulphate and tetrachloroethene. Environ. Microbiol. **3:**92–99.
13. **Drzyzga, O., and J. C. Gottschal.** 2002. Tetrachloroethene dehalorespiration and growth of *Desulfitobacterium frappieri* TCE1 in strict dependence on the activity of *Desulfovibrio fructosivorans*. Appl. Environ. Microbiol. **68:**642–649.
14. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
15. **Finneran, K. T., M. E. Housewright, and D. R. Lovley.** 2002. Multiple influences of nitrate on uranium solubility during bioremediation of uranium-contaminated subsurface sediments. Environ. Microbiol. **4:**510–516.
16. **Garrity, G. M.** 2001. Bergey's manual of systematic bacteriology. Springer-Verlag, New York, N.Y.
17. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
18. **Holmes, D. E., K. T. Finneran, R. A. O'Neil, and D. R. Lovley.** 2002. Enrichment of members of the family *Geobacteraceae* associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. Appl. Environ. Microbiol. **68:**2300–2306.
19. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20:**2317–2319.
20. **Hugenholtz, P.** 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. **3:**reviews0003.1–reviews0003.8.
21. **Humphries, A. C., and L. E. Macaskie.** 2002. Reduction of Cr(VI) by *Desulfovibrio vulgaris* and *Microbacterium* sp. Biotechnol. Lett. **24:**1261–1267.
22. **Istok, J. D., J. M. Senko, L. R. Krumholz, D. Watson, M. A. Bogle, A. Peacock, Y. J. Chang, and D. C. White.** 2004. *In situ* bioreduction of technetium and uranium in a nitrate-contaminated aquifer. Environ. Sci. Technol. **38:**468–475.
23. **John, S. G., C. E. Ruggiero, L. E. Hersman, C. S. Tung, and M. P. Neu.** 2001. Siderophore mediated plutonium accumulation by *Microbacterium flavescens* (JG-9). Environ. Sci. Technol. **35:**2942–2948.
24. **Kelly, J. J., S. Siripong, J. McCormack, L. R. Janus, H. Urakawa, S. El Fantroussi, P. A. Noble, L. Sappelsa, B. E. Rittmann, and D. A. Stahl.** 2005. DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. Water Res. **39:**3229–3238.
25. **Kurata, S., T. Kanagawa, Y. Magariyama, K. Takatsu, K. Yamada, T. Yokomaku, and Y. Kamagata.** 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. Appl. Environ. Microbiol. **70:**7545–7549.
26. **Lehner, A., A. Loy, T. Behr, H. Gaenge, W. Ludwig, M. Wagner, and K. H. Schleifer.** 2005. Oligonucleotide microarray for identification of *Enterococcus* species. FEMS Microbiol. Letts. **246:**133–142.
27. **Lloyd, J. R., J. Chesnes, S. Glasauer, D. J. Bunker, F. R. Livens, and D. R. Lovley.** 2002. Reduction of actinides and fission products by Fe(III)-reducing bacteria. Geomicrobiol. J. **19:**103–120.
28. **Lovley, D. R., and R. T. Anderson.** 2000. Influence of dissimilatory metal reduction on fate of organic and metal contaminants in the subsurface. Hydrogeol. J. **8:**77–88.
29. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32:**1363–1371.
30. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-

subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Appl. Environ. Microbiol. **69:**320–326.
31. **Macaskie, L. E., J. R. Lloyd, R. A. P. Thomas, and M. R. Tolley.** 1996. The use of micro-organisms for the remediation of solutions contaminated with actinide elements, other radionuclides, and organic contaminants generated by nuclear fuel cycle activities. Nucl. Energy J. Br. Nucl. Energy Soc. **35:** 257–271.
32. **Masuda, N., and G. M. Church.** 2002. *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. J. Bacteriol. **184:**6225–6234.
33. **McCullough, J., T. C. Hazen, S. M. Benson, F. B. Metting, and A. C. Palmisano.** 2004. Bioremediation of metals and radionuclides: what it is and how it works. Lawrence Berkeley National Laboratory report 42595, 2nd ed. Lawrence Berkeley National Laboratory, Berkeley, Calif.
34. **Mei, R., E. Hubbell, S. Bekiranov, M. Mittmann, F. C. Christians, M. M. Shen, G. Lu, J. Fang, W. M. Liu, T. Ryder, P. Kaplan, D. Kulp, and T. A. Webster.** 2003. Probe selection for high-density oligonucleotide arrays. Proc. Natl. Acad. Sci. USA **100:**11237–11242.
35. **Nevin, K. P., K. T. Finneran, and D. R. Lovley.** 2003. Microorganisms associated with uranium bioremediation in a high-salinity subsurface sediment. Appl. Environ. Microbiol. **69:**3672–3675.
36. **North, N. N., S. L. Dollhopf, L. Petrie, J. D. Istok, D. L. Balkwill, and J. E. Kostka.** 2004. Change in bacterial community structure during in situ biostimulation of subsurface sediment cocontaminated with uranium and nitrate. Appl. Environ. Microbiol. **70:**4911–4920.
37. **Palmer, C., E. M. Bik, M. B. Eisen, P. B. Eckburg, T. R. Sana, P. K. Wolber, D. A. Relman, and P. O. Brown.** 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res. **34:**e5.
38. **Petrie, L., N. N. North, S. L. Dollhopf, D. L. Balkwill, and J. E. Kostka.** 2003. Enumeration and characterization of iron(III)-reducing microbial communities from acidic subsurface sediments contaminated with uranium(VI). Appl. Environ. Microbiol. **69:**7467–7479.
39. **Phillips, E. J. P., E. R. Landa, and D. R. Lovley.** 1995. Remediation of uranium contaminated soils with bicarbonate extraction and microbial U(VI) reduction. J. Ind. Microbiol. **14:**203–207.
40. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. **64:**3724–3730.
41. **Raskin, L., B. E. Rittmann, and D. A. Stahl.** 1996. Competition and coexistence of sulfate-reducing and methanogenic populations in anaerobic biofilms. Appl. Environ. Microbiol. **62:**3847–3857.
42. **R Development Core Team.** 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
43. **Rhee, S. K., X. D. Liu, L. Y. Wu, S. C. Chong, X. F. Wan, and J. Z. Zhou.** 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl. Environ. Microbiol. **70:**4303–4317.
44. **Rosencrantz, D., F. A. Rainey, and P. H. Janssen.** 1999. Culturable populations of *Sporomusa* spp. and *Desulfovibrio* spp. in the anoxic bulk soil of flooded rice microcosms. Appl. Environ. Microbiol. **65:**3526–3533.
45. **Sani, R. K., B. M. Peyton, W. A. Smith, W. A. Apel, and J. N. Petersen.** 2002. Dissimilatory reduction of Cr(VI), Fe(III), and U(VI) by Cellulomonas isolates. Appl. Microbiol. Biotechnol. **60:**192–199.
46. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501–1506.
47. **Schlotelburg, C., F. von Wintzingerode, R. Hauck, W. Hegemann, and U. B. Gobel.** 2000. Bacteria of an anaerobic 1,2-dichloropropane-dechlorinating mixed culture are phylogenetically related to those of other anaerobic dechlorinating consortia. Int. J. Syst. Evol. Microbiol. **50:**1505–1511.
48. **Shelobolina, E. S., K. O'Neill, K. T. Finneran, L. A. Hayes, and D. R. Lovley.** 2003. Potential for *in situ* bioremediation of a low-pH, high-nitrate uranium-contaminated groundwater. Soil Sediment Contam. **12:**865–884.
49. **Singleton, D. R., M. A. Furlong, S. L. Rathbun, and W. B. Whitman.** 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. Appl. Environ. Microbiol. **67:**4374–4376.
50. **Smith, J. J., and G. A. McFeters.** 1997. Mechanisms of INT (2-(4-iodophenyl)-3-(4-nitrophenyl)-5-phenyl tetrazolium chloride), and CTC (5-cyano-2,3-ditolyl tetrazolium chloride) reduction in *Escherichia coli* K-12. J. Microbiol. Methods **29:**161–175.
51. **Suzuki, M., and S. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
52. **Suzuki, M., M. Rappe, and S. Giovannoni.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl. Environ. Microbiol. **64:**4522–4529.
53. **Suzuki, Y., and J. F. Banfield.** 2004. Resistance to, and accumulation of, uranium by bacteria from a uranium-contaminated site. Geomicrobiol. J. **21:**113–121.
54. **Suzuki, Y., S. D. Kelly, K. A. Kemner, and J. F. Banfield.** 2003. Microbial populations stimulated for hexavalent uranium reduction in uranium mine sediment. Appl. Environ. Microbiol. **69:**1337–1346.
55. **Tsuruta, T.** 2002. Removal and recovery of uranyl ion using various microorganisms. J. Biosci. Bioeng. **94:**23–28.
56. **Urakawa, H., P. A. Noble, S. El Fantroussi, J. J. Kelly, and D. A. Stahl.** 2002. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Appl. Environ. Microbiol. **68:**235–244.
57. **Von Mersi, W., and F. Schinner.** 1991. An improved and accurate method for determining the dehydrogenase activity of soils with iodonitrotetrazolium chloride. Biol. Fertil. Soils **11:**216–220.
58. **von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21:**213–229.
59. **von Wintzingerode, F., B. Selent, W. Hegemann, and U. B. Gobel.** 1999. Phylogenetic analysis of an anaerobic, trichlorobenzene-transforming microbial consortium. Appl. Environ. Microbiol. **65:**283–286.
60. **Wan, J. M., T. K. Tokunaga, E. Brodie, Z. M. Wang, Z. P. Zheng, D. Herman, T. C. Hazen, M. K. Firestone, and S. R. Sutton.** 2005. Reoxidation of bioreduced uranium under reducing conditions. Environ. Sci. Technol. **39:**6162–6169.
61. **Warsen, A. E., M. J. Krug, S. LaFrentz, D. R. Stanek, F. J. Loge, and D. R. Call.** 2004. Simultaneous discrimination between 15 fish pathogens by using 16S ribosomal DNA PCR and DNA microarrays. Appl. Environ. Microbiol. **70:**4216–4221.
62. **Wilson, K. H., R. B. Blitchington, and R. C. Greene.** 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. J. Clin. Microbiol. **28:**1942–1946.
63. **Yan, T. F., M. W. Fields, L. Y. Wu, Y. G. Zu, J. M. Tiedje, and J. Z. Zhou.** 2003. Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate- and uranium-contaminated groundwater. Environ. Microbiol. **5:**13–24.

Microbial Ecology

High-Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library When Sampling the Environment

Todd Z. DeSantis[1,2], Eoin L. Brodie[1], Jordan P. Moberg[1], Ingrid X. Zubieta[1], Yvette M. Piceno[1] and Gary L. Andersen[1,2]

(1) Lawrence Berkeley National Laboratory, Center for Environmental Biotechnology, 1 Cyclotron Road, Mail Stop 70A-3317, Berkeley, CA 94720, USA
(2) Lawrence Berkeley National Laboratory, Virtual Institute for Microbial Stress and Survival, Berkeley, CA 94720, USA

Received: 15 June 2006 / Accepted: 20 June 2006

Abstract

Molecular approaches aimed at detection of a broad-range of prokaryotes in the environment routinely rely on classifying heterogeneous 16S rRNA genes amplified by polymerase chain reaction (PCR) using primers with broad specificity. The general method of sampling and categorizing DNA has been to clone then sequence the PCR products. However, the number of clones required to adequately catalog the majority of taxa in a sample is unwieldy. Alternatively, hybridizing target sequences to a universal 16S rRNA gene microarray may provide a more rapid and comprehensive view of prokaryotic community composition. This study investigated the breadth and accuracy of a microarray in detecting diverse 16S rRNA gene sequence types compared to clone-and-sequencing using three environmental samples: urban aerosol, subsurface soil, and subsurface water. PCR products generated from universal 16S rRNA gene-targeted primers were classified by using either the clone-and-sequence method or by hybridization to a novel high-density microarray of 297,851 probes complementary to 842 prokaryotic subfamilies. The three clone libraries comprised 1391 high-quality sequences. Approximately 8% of the clones could not be placed into a known subfamily and were considered novel. The microarray results confirmed the majority of clone-detected subfamilies and additionally demonstrated greater amplicon diversity extending into phyla not observed by the cloning method. Sequences matching operational taxonomic units within the phyla Nitrospira, Planctomycetes, and TM7, which were uniquely detected by the array, were verified with specific primers and subsequent amplicon sequencing. Subfamily richness detected by the array corresponded well with nonparametric richness predictions extrapolated from clone libraries except in the water community where clone-based richness predictions were greatly exceeded. It was concluded that although the microarray is unreliable in identifying novel prokaryotic taxa, it reveals greater diversity in environmental samples than sequencing a typically sized clone library. Furthermore, the microarray allowed samples to be rapidly evaluated with replication, a significant advantage in studies of microbial ecology.

Introduction

Molecular approaches designed to describe prokaryotic diversity routinely rely on classifying heterogeneous nucleic acids amplified via universal 16S rRNA gene polymerase chain reaction (PCR). The resulting mixed amplicons can be quickly, but coarsely, categorized using terminal-restriction fragment length polymorphism (t-RFLP), single-strand conformation polymorphism (SSCP), or temperature/denaturing gradient gel electrophoresis (T/DGGE) [39]. Association of taxonomic nomenclature to each group may be accomplished through sequencing, but this requires additional labor to physically isolate each gene and does not scale well for large comparative studies such as environmental monitoring. Species richness may be predicted from a few hundred sequences, but reproducible discovery of species composition may require $>10^4$ sequencing reactions per sample [14].

To increase the throughput of detection of microorganisms within complex samples, multiple DNA probes have been arrayed onto solid surfaces to allow for parallel, multispecies detection. Successful differen-

Correspondence to: Gary L. Andersen; E-mail: GLAndersen@lbl.gov

DOI: 10.1007/s00248-006-9134-9

tiation of specific collections of bacteria has been achieved by using 16S rRNA gene microarrays containing tens to hundreds of probes for Enterococcus [37], Cyanobacteria [2], nitrifying bacteria [33], fish pathogens [58], and other bacterial groups. Alternatively, protein-encoding genes have been targeted to survey environments using microarrays [63, 64]. Recently, high-density 16S rRNA gene microarrays have emerged in efforts to detect any bacterial type without *a priori* knowledge of the community structure. Two major challenges have impeded this goal: Probes must be designed that are sensitive to only a specified branch of the prokaryotic tree, and hybridization scoring algorithms are required to interpret probe responses into reliable identifications. If a single unique probe for a taxon cannot be found, several probes can be utilized in combination with rule-based scoring. By vastly increasing the total number of probes within a microarray, more taxa can be queried and detection confidence can be improved. Using this approach, it was shown that organisms from environmental samples were accurately classified into their respective orders using an array with 62,358 probes [12, 61].

In this study, a novel microarray containing 297,851 probes targeted to 16S rRNA genes was tested by using amplicons derived from soil, water, and aerosols. The community profiles derived from the hybridizations were compared to the results from cloning-and-sequencing the same amplicons. A fraction of the clones (8%) were sufficiently divergent from database sequences to be considered novel and were not identified by the array. The microarray results confirmed the majority of clone-detected subfamilies, but additionally showed greater amplicon diversity. Importantly, the microarray detected phyla that would have otherwise been overlooked if relying solely on the clone library. Three of these phyla have been confirmed with specific PCR amplification. The results illustrate the consequence of relying only on clone libraries or high-density 16S rRNA gene microarrays when profiling a microbial community.

Methods

Microarray Design. The microarray probe design approach previously described for differentiating Staphylococcaceae [9] was applied to all known 16S rRNA gene sequences containing at least 600 nucleotides. Briefly, sequences (*Escherichia coli* bp positions 47 to 1473) were extracted from a multiple sequence alignment composed of more than 30,000 records within the 15 March 2002 release of the 16S rRNA gene database, greengenes.lbl.gov [11]. This region was selected because it is bounded on both ends by universally conserved segments that can be used as PCR priming sites to amplify bacterial or archaeal [13] genomic material using only two to four primers. Putative chimeric sequences were filtered from the data set by using the software package Bellerophon [25], preventing them from being misconstrued as novel organisms [28]. Filtered sequences were clustered to enable each sequence of a cluster to be complementary to a set of perfectly matching (PM) probes. Putative amplicons were placed in the same cluster as a result of common 17-mers found in the sequence. The resulting 8935 clusters, each containing approximately 3% sequence divergence, were considered operational taxonomic units (OTUs) representing all 121 demarcated prokaryotic orders. The taxonomic family of each OTU was assigned according to the placement of its member organisms in Bergey's Taxonomic Outline [22]. The taxonomic outline as maintained by Hugenholtz [26] was consulted for phylogenetic classes containing uncultured environmental organisms or unclassified families belonging to named higher taxa. The OTUs comprising each family were clustered into subfamilies by transitive sequence identity according to a previously described method [9]. Altogether, 842 subfamilies were found. The taxonomic position of each OTU as well as the accompanying NCBI accession numbers of the sequences composing each OTU can be viewed at http://greengenes.lbl.gov/Download/Clones_v_Array/.

The objective of the probe selection strategy was to obtain an effective set of probes capable of correctly categorizing mixed amplicons into their proper OTU. For each OTU, a set of 11 or more specific 25-mers (probes) was sought to be prevalent in members of a given OTU but dissimilar from sequences outside the given OTU. The average number of probes chosen for each OTU was 24. In the first step of probe selection for a particular OTU, each of the sequences in the OTU was separated into overlapping 25-mers, the potential targets. Then each potential target was matched to as many sequences of the OTU as possible. It was not adequate to use a text pattern search to match potential targets and sequences because partial gene sequences were included in the reference set. Therefore, the multiple sequence alignment provided by Greengenes was necessary to provide a discrete measurement of group size at each potential probe site. For example, if an OTU containing seven sequences possessed a probe site where one member was missing data, then the site-specific OTU size was only six. In ranking the possible targets, those having data for all members of that OTU were preferred over those found only in a fraction of the OTU members. In the second step, a subset of the prevalent targets was selected and reverse-complimented into probe orientation, avoiding those capable of mishybridization to an unintended amplicon. Probes presumed to have the capacity to mishybridize were those 25-mers that contained a central 17-mer matching sequences in more than one OTU [56]. Thus, probes that were unique to an

OTU solely due to a distinctive base in one of the outer four bases were avoided. Also, probes with mishybridization potential to sequences having a common tree node near the root were favored over those with a common node near the terminal branch. Probes complementary to target sequences that were selected for fabrication are termed PM probes. As each PM probe was chosen, it was paired with a control 25-mer [mismatching probe (MM)], identical in all positions except the thirteenth base. The MM probe did not contain a central 17-mer complimentary to sequences in any OTU. The probe complementing the target (PM) and MM probes constitute a probe pair analyzed together. Sets of probes for each OTU can be viewed at: http://greengenes.lbl.gov/cgi-bin/nph-show_probes_2_otu_alignments.cgi.

The chosen oligonucleotides were synthesized by a photolithographic method at Affymetrix Inc. (Santa Clara, CA, USA) directly onto a 1.28 × 1.28 cm glass surface at an approximate density of 10,000 molecules/μm^2 [6]. The entire array of 506,944 probe features was arranged as a grid of 712 rows and columns. Thus, each unique probe sequence (feature) on the array occupied a square with an 18-μm side and had a copy number of roughly 3.2×10^6. Of these features, 297,851 were oligonucleotide PM or MM probes with exact or inexact complementarity, respectively, to 16S rRNA genes. The remaining were used for image orientation, normalization controls, or for pathogen-specific signature amplicon detection using additional targeted regions of the chromosome [62]. Each high-density 16S rRNA gene microarray was designed with additional probes that: (1) target amplicons of prokaryotic metabolic genes spiked into the 16S rRNA gene amplicon mix in defined quantities just before fragmentation and (2) are complimentary to prelabeled oligonucleotides added into the hybridization mix. The first control collectively tests the fragmentation, biotinylation, hybridization, staining, and scanning efficiency. It also allows the overall fluorescent intensity to be normalized across all the arrays in an experiment. The second control directly assays the hybridization, staining, and scanning.

Environmental Sampling and DNA Extraction. Air samples were collected at a flow rate of approximately 10 L/min onto 1.0 μm polyethylene terephthalate (Celanex) filters (Hoechst-Celanese, Dallas, TX) over a 24-h period simultaneously from six locations in San Antonio, TX, USA. Sample filters were washed in 10 mL buffer (0.1 M sodium phosphate, 10 mM EDTA, pH 7.4, 0.01% Tween 20), and the suspension was stored frozen until needed. One 0.6-mL aliquot of wash was taken from each thawed filter wash and combined in a "day pool". DNA was extracted by using a modification of a soil technique [46]. After centrifugation of the day pool at 16,000 × *g* for 25 min, the pellets were resuspended in 400 μL sodium phosphate buffer (100 mM NaH_2PO_4, pH 8) and transferred into two 2-mL silica bead lysis tubes containing 0.9 g of zirconia/silica lysis bead mix (0.3 g of 0.5 mm and 0.6 g of 0.1 mm). For each lysis tube, 300 μL buffered SDS [100 mM NaCl, 500 mM Tris pH 8, 10% (wt/vol) SDS] and 300 μL phenol/chloroform/isoamyl alcohol (25:24:1) were added. Lysis tubes were inverted and finger-flicked three times to mix the buffers before bead mill homogenization with a Bio101 Fast Prep 120 machine (Qbiogene, Irvine, CA, USA) at 6.5 m/s for 45 s. The bead-beating duration was selected for its ability to release DNA from spores while not overfragmenting genomes [12]. Following lysate centrifugation at 16,000 × *g* for 5 min, the aqueous supernatant was removed to a new 2-mL tube and maintained at –20°C for 1 h to overnight. An equal volume of chloroform was added to the thawed supernatant prior to vortexing for 5 s and centrifugation at 16,000 × *g* for 3 min. The supernatant was then combined with two volumes of binding Solution 3 (MoBio, Carlsbad, CA, USA). Genomic DNA (gDNA) from the mixture was isolated on a MoBio spin column, washed with Solution 4, and eluted in 60 μL of 1× TE according to the manufacturer's instructions. The gDNA was further purified by passage through a Sephacryl S-200 HR spin column (Amersham, Piscataway, NJ, USA) and stored at 4°C. Each of the gDNA preparations from four different "day pools" from the week of July 14, 2003 was independently PCR-amplified. PCR products were combined to constitute the sample for the week.

Subsurface water was collected during polylactate-stimulated bioremediation of a chromate-contaminated aquifer at the Hanford 100H site, WA (http://www-esd.lbl.gov/ERT/hanford100h/). Water, approximately 150 mL, was filtered through sterile 0.22-μm anodisc filters (Whatman, Florham Park, NJ, USA) and DNA was extracted by using a modification of the procedure described for air samples. Anodisc filters were manually fragmented in a sterile whirlpak bag, and 1 mL of phosphate buffer was added. Filter fragments in buffer were transferred to a bead lysis tube. Tubes were centrifuged at 16,000 × *g* for 5 min and 700 μL of buffer was removed. Next, 300 μL of buffered SDS solution and 300 μL of phenol/chloroform/isoamyl alcohol (25:24:1) were added and bead beating was performed at 5.5 m/s for 30 s. After centrifugation, the aqueous phase was mixed with an equal volume of chloroform in a phase-lock gel tube (Eppendorf, Westbury, NY, USA) and further extracted. The top phase containing nucleic acids was purified as for the air samples without the need for additional Sephacryl purification. DNA was eluted in 50 μL sterile water and stored at –20°C until needed.

Subsurface soils were obtained from a uranium-contaminated soil (area 2) at the NABIR Field Research

Center at Oak Ridge, TN, USA (http://www.esd.ornl.gov/nabirfrc/). More information about the soil characteristics is available at http://public.ornl.gov/nabirfrc/other/FRCSummary.pdf DNA was extracted from triplicate 500-mg (wet weight) subsamples of soil using a BIO101 soil DNA extraction kit (Qbiogene) according to the manufacturer's protocol.

16S rRNA Gene Amplification. The 16S rRNA gene was amplified from the DNA extracts using universal primers 27f.1 (5′-AGRGTTTGATCMTGGCTCAG) and 1492R (5′-GGTTACCTTGTTACGACTT). PCR for air and soil samples was carried out by using the TaKaRa *Ex Taq* system (Takara Bio Inc., Japan) as follows, with at least three replicate PCR reactions performed per sample and pooled before analysis. Each PCR reaction mix contained 1× buffer, 0.8 mM *TaKaRa* dNTP mixture, 0.02 U/μL *Ex Taq* polymerase, 0.4 mg/mL bovine serum albumin (BSA), and 1.0 μM of each primer. PCR conditions were 1 cycle of 3 min at 95°C, followed by 35 cycles of 95°C (30 s), 53°C (30 s), 72°C (60 s), and a final extension at 72°C for 7 min. DNA extracts from water samples were amplified by using a slightly different protocol using a range of eight different annealing temperatures between 48°C and 58°C. Only 30 cycles were performed for amplification from water samples and amplicons from the eight different annealing temperatures were combined.

Cloning-and-Sequencing. Amplicon pools from the three environments were subjected to cloning as follows: Amplicons were ligated and cloned by using the TOPO-TA pCR2.1 kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Individual clones containing organism-specific 16S rRNA gene fragments were purified by using magnetic beads [54], and sequenced from each terminus using an ABI3700 (Applied Biosystems, Foster City, CA), assembled using Phred and Phrap [16, 17], and were required to pass quality tests of Phred 20 (base call error probability $< 10^{-2.0}$) to be included in the analysis. Sequencing was performed at the DOE Joint Genome Institute (JGI; http://www.jgi.doe.gov/). Putative chimeric sequences were obtained by using Bellerophon [25]. Sequences were aligned to the Greengenes 7682-character format by using the NAST [10] web server (http://greengenes.lbl.gov/NAST). Similarity to public database records was calculated with DNADIST [19], by using the DNAML-F84 option assuming a transition/transversion ratio of 2.0, and an A, C, G, and T 16S rRNA gene base frequency of 0.2537, 0.2317, 0.3167, and 0.1979, respectively. This was empirically calculated from all records of the Greengenes 16S rRNA gene multiple sequence alignment over 1250 nucleotides in length. The lane mask [35] was used to restrict similarity observations to 1287 conserved columns (lanes) of aligned characters. Cloned sequences from this study were rejected from further analysis when less than 1000 characters could be compared to a lane-masked reference sequence. Sequences were assigned to a taxonomic node by using a sliding scale of similarity thresholds [52]. Phylum, class, order, family, subfamily, or OTU placement was accepted when a clone surpassed similarity thresholds of 80%, 85%, 90%, 92%, 94%, or 97%, respectively. When similarity to nearest database sequence was below 94%, the clone was considered to represent a novel subfamily and a novel class was denoted when similarity was less than 85%.

Accumulation curves, diversity estimates (Shannon–Weaver index [43]), and nonparametric richness estimations (Chao1 and ACE [4, 5]) were calculated by using the software DOTUR [51] with the clone distance matrix as input and a nearest-neighbor clustering algorithm. Dominance in clone libraries was calculated as 1 – Shannon evenness index ($1 - E$), where evenness (E) is represented as follows: $E = H/\ln S$ (H = Shannon–Weaver diversity index; S = total richness in a sample).

Accession Numbers. Sequences generated in this study have been deposited in Genbank as accession numbers DQ125500–DQ125935 (soil), DQ129237–DQ129656 (air), and DQ264398–DQ264650 (water). Fasta formatted records can also be obtained at http://greengenes.lbl.gov/Download/Clones_v_Array/.

Microarray Processing. Identical amplicon pools used for cloning were also used for array analysis. For air samples, 2 μg amplicons was concentrated to a volume less than 40 μL with a Microcon YM100 spin filter (Millipore, Billerica, MA, USA). For soil samples and water samples, 2 μg ($\sim 10^{12}$ gene copies) and 500 ng ($\sim 3 \times 10^{11}$ gene copies) of amplicons, respectively, were concentrated using a PCR clean up kit (MoBio). The PCR products were spiked with known concentrations of amplicons derived from prokaryotic metabolic genes. This mix was fragmented to 50–200 bp using DNase I (0.02 U/μg DNA; Invitrogen) and One-Phor-All buffer per the Affymetrix protocol. The complete mixture was incubated at 25°C for 10 min, 98°C for 10 min, and then labeled. Biotin labeling was accomplished using an Enzo Bioarray Terminal Labeling Kit (Affymetrix) as per the manufacturer's instructions. Next, labeled DNA was denatured (99°C for 5 min) and hybridized to the DNA microarray at 48°C overnight (>16 h) at 60 rpm. The arrays were subsequently washed and stained. Reagents, conditions, and equipment are detailed elsewhere [44].

Scanning and Probe Set Scoring. Arrays were scanned using a GeneArray Scanner (Affymetrix). The scan was captured as a pixel image using standard

Affymetrix software (GeneChip Microarray Analysis Suite, version 5.1) that reduced the data to an individual signal value for each probe. Background probes were identified as those producing intensities in the lowest 2% of all intensities. The average intensity of the background probes was subtracted from the fluorescence intensity of all probes. The noise value (N) was the variation in pixel intensity signals observed by the scanner as it read the array surface. The standard deviation of the pixel intensities within each of the identified background cells was divided by the square root of the number of pixels comprising that cell. The average of the resulting quotients was used for N in the calculations described below.

Probe pairs scored as positive were those that met two criteria: (1) the fluorescence intensity from the perfectly matched probe (PM) was at least 1.3 times greater than the intensity from the mismatched control (MM), and (2) the difference in intensity, PM minus MM, was at least 130 times greater than the squared noise value ($>130N^2$). The positive fraction (PosFrac) was calculated for each probe set as the number of positive probe pairs divided by the total number of probe pairs in a probe set. An OTU was considered "present" when its PosFrac was greater than 0.92 in all three replicates. Present calls were propagated upwards through the taxonomic hierarchy by considering any node (subfamily, family, order, etc.) as "present" if at least one of its subordinate OTUs was present.

Validation of Array-Detected OTUs Not Detected by the Clone Library. PCR primers targeting specific OTUs within the Nitrospiraceae and Planctomycetaceae (Table 3) were generated by ARB's probe design feature [41] and Primer3 [50]. Melting temperatures were constrained from 45°C to 65°C and G + C content between 40% and 70% was preferred. The primers were chosen to contain 3′ bases non-complimentary to sequences outside of the sub-family. TM7 phylum-specific primers [29] were obtained from the literature. DNA sequences were generated by PCR as described above with the necessary adjustments in annealing temperatures. Amplicons were purified (PureLink PCR Purification Kit, Invitrogen), sequenced, and matched to an OTU in the same manner as described in the cloning-and-sequencing method except the minimum count of base comparisons was not used to exclude data.

Results

Cloning. Sequences from the clone libraries of the three environments were assembled into contigs from two sequencing reactions initiated at the 5′ and 3′ termini of the 16S rRNA gene. Initially, 1391 contigs with low base call error probability were accepted for analysis. After removing contigs that did not contain sufficient data to allow at least 1000 characters to be compared to a lane-masked reference sequence and filtering the sequences for putative chimeras, 1155 remained. The vetted libraries from air, soil, and water contained 417, 485, and 253 clones, respectively. Figure 1 shows the class level (85% database sequence homology) distributions of clones within each of the three ecosystem samples; not shown are 1, 8, and 0 clones (air, soil, water) considered novel at the class level compared to existing database entries. For air, soil, and water clone libraries, 9.8%, 14.4%, and 1.6% of clones, respectively, were outside the subfamily level (94%) assignment threshold and were considered novel sequences.

Clone library analysis indicated that Firmicutes dominated the air sample mostly within the class Bacilli, whereas Actinobacteria dominated the soil sample. The water sample consisted solely of three classes: flavobacteria (Bacteroidetes), β-proteobacteria, and γ-proteobacteria, each with similar distribution. Figure 2 shows the accumulation curves for the three samples where the cumulative number of subfamilies observed is plotted against the sampling effort. All communities were incompletely sampled as evidenced by the nonasymptotic curves [30]; however, the water community appeared relatively well sampled compared to air and soil.

Comparison of Cloning with Microarray Analysis. A comparison between clone library and microarray assessment of community composition is shown for each taxonomic level in Table 1. The breadth of 16S rRNA gene sequence types was expressed as a count of distinct groups detected in each environment, by each method, at six levels of taxonomic resolution. It is clear that even at the phylum level cloning underestimated richness compared with the microarray. Among all three environments, the amplicons categorized by cloning were in general concordance with a subset of taxonomic categories reported from the arrays. This trend continued as the resolution of the comparisons increased from phylum to subfamily. Hybridization of aerosol amplicons produced "present" calls in 238 subfamilies, 178 of which were not found in the corresponding clone library. Subfamily richness appeared to be greater in the soil sample with 279 subfamilies detected; of these, 239 were not encountered in the clone library. The water sample was shown to have the lowest richness by both methods, but of the 99 subfamilies detected by microarray, only 6 were detected by cloning. For all sample types, very few subfamilies deemed present by the clone libraries were not reflected in the microarray hybridization results.

A synopsis of the phylum-level community composition is assembled in Table 2, allowing comparison of the 34 phyla reported in at least one of the environments

by at least one of the methods. By cloning, 10 phyla were detected in the aerosol and soil samples, whereas only 2 phyla were detected in the water sample. In contrast, the array method detected all of the phyla detected by cloning and an additional 17, 23, and 17 phyla in the air, soil, and water samples, respectively.

PCR and Sequencing Confirmation of Additional Phyla Detected by Microarray. To determine whether the additional phyla detected by microarray were true positives, and not the product of unforeseen cross-hybridization, three OTUs from diverse phyla, detected in aerosols by microarray only, were chosen for further investigation (Table 3). OTU 864 (OTU numbers correspond to 16S rRNA microarray probe sets) within the phylum Nitrospira comprised sequences discovered in sludge, soil, reservoirs, and in an aquifer. All 13 probe pairs in the probe set for OTU 864 were positive in 3 of 3 arrays. The sample was interrogated with primers designed from known sequences in the OTU and the resulting amplicons were sequenced, revealing similarity to OTU 864. Similarly, all 11 probe pairs of a Planctomycetes OTU (OTU 4948) were consistently positive despite this OTU being unrepresented among the 417 aerosol clones. Primers were designed from the five 16S rRNA gene sequences generated from a municipal wastewater plant [7] that defined OTU 4948. Taxon-specific PCR and sequencing confirmed that a sequence matching this OTU was present in the sample. Evidence for the presence of phylum TM7 came from the probe set complementary to OTU 8155. General TM7 phylum-specific primers [29] produced sequences attributed to a related TM7 OTU identified as 3664.

Diversity Estimates. Table 4 lists diversity estimates and richness predictions based on the clones sampled and also compares predicted richness values to those observed by both cloning and array methods. Shannon–Weaver diversity estimates for the clone libraries indicate that sample diversity is of the order air > soil >>> water. Both Chao1 and ACE nonparametric richness estimators predicted that the subfamily level richness of the air and soil samples is far greater than that observed through clone sampling and these estimates were in strong agreement with the subfamily counts reported by the array analysis. The water sample produced a large discrepancy in subfamily richness between the cloning and array methods regardless of whether direct clone observations or nonparametric predictors were used for the comparison. In Table 4, a trend was observed—the greater the dominance encountered in the clone library, the greater the difference between the cloning and array observed counts of subfamilies.

Discussion

The postulated complexity of each microbial community that has been isolated from the environment, combined with the number of potentially unique ecosystems, has hindered efforts to sufficiently catalog microbial biodiversity. There are estimates of thousands [14, 55] to millions [21] of unique bacterial genomes present in a gram of soil. Outdoor aerosols may be equally complex, composed of organisms released from multiple habitats, both locally and over long distances. Furthermore, microbial communities can change temporally as environmental conditions vary. The need for high-throughput accurate biological monitoring is clear.

Community fingerprinting has provided methods for rapidly profiling microbial communities with replication. These approaches, based on heterogeneity in amplicon length, endonuclease cleavage sites, melting profiles, or single strand secondary structure, have allowed high-throughput inference of species richness and evenness [39]. However, fingerprinting methods are

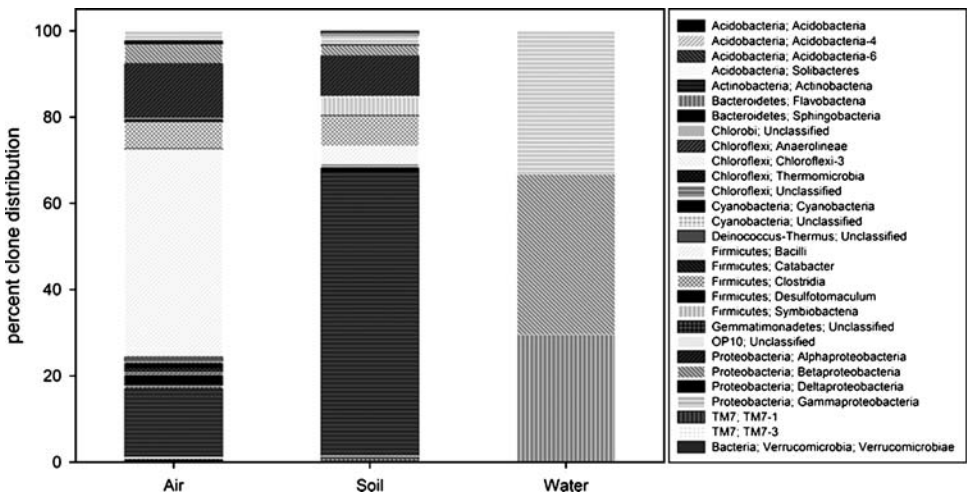


Figure 1. Class-level distribution of clones within libraries from air, soil, and water samples.

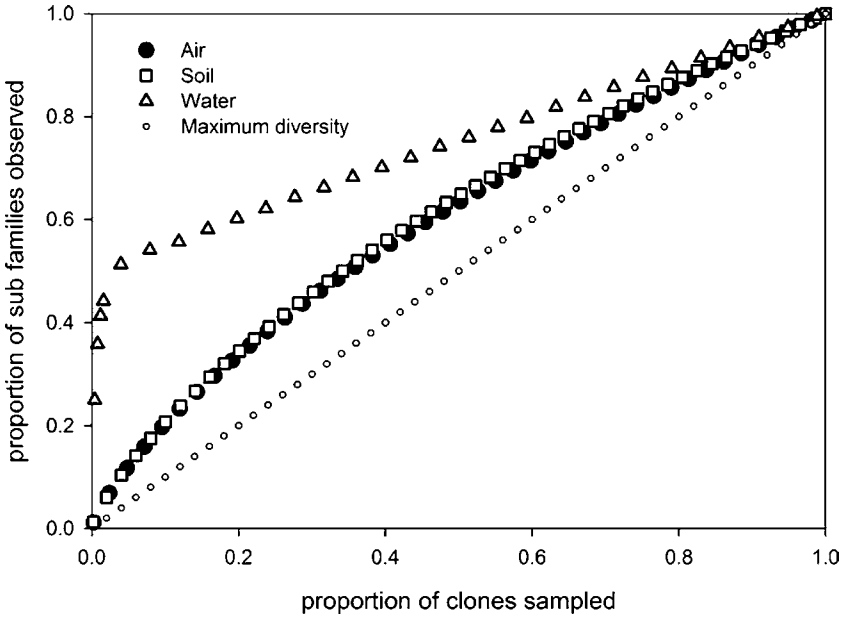


Figure 2. Accumulation curves for air (•; n = 417), soil (□; n = 485) and water (△; n = 253) bacterial communities analyzed by clone library. The maximum diversity scenario where every sample is a new observation (○). Sampling efficiency at subfamily level is presented: subfamily is defined as 94% sequence homology. Data for curves represent an average of 1000 simulations performed using the software DOTUR.

generally deficient in providing taxonomic microbial identity and typically yield less than 100 clearly defined bands, peaks, or products for analysis. T-RFLP offers the greatest taxonomic resolution of the rapid fingerprinting methods, potentially capable of class identification [34] when DNA sequences in the sample are nearly identical to database reference sequences, or when over 10 restriction enzymes are used in parallel for each sample [49]. However, the reliability of T-RFLP for taxonomic assignment is unclear, because the presence of classes reported by T-RFLP—but not found in a corresponding clone library—have been left unverified.

PCR has made it possible to easily obtain composite samples of mixed rRNA genes from natural environments [53, 60]. Although amplification biases have been demonstrated in defined communities due to primer selection, number of cycles, and template concentration [18, 48, 59], this technique has been valuable in increasing our understanding of the complexity of individual communities [27]. Regardless of the method used to limit biases from environmental samples, most of our knowledge on microbial composition of specific communities comes from isolating individual, amplified 16S rRNA genes for cloning and sequencing. The sequences are compared to references in large databases, allowing either specific phylogenetic classification or proposal of novel taxa when a clone is sufficiently divergent from known groups. The limitation becomes the number of clones or PCR products requiring sequencing and analysis. It has been suggested that environmental samples may require over 40,000 sequencing reactions to document 50% of the richness [14]. This approach is laborious, costly, and time-consuming, often taking weeks to complete the analysis of one clone library. Thus, performing studies with sample replication becomes rapidly overwhelming. Despite the effort required, clone libraries are often the chosen method, and the current "gold standard" for obtaining the greatest estimate of diversity. Typical libraries of cloned 16S rRNA gene fragments include fewer than 1000 sequences [15, 32, 45], well below the suggested quantity.

By making it possible to conduct sequence analyses on the complete pool of 16S rRNA gene fragments at once, high-density photolithography microarrays have the ability to provide microbial identification within complex environmental samples in a high-throughput manner. Hybridization (with replicates) requires only

Table 1. Count of distinct taxonomic groups detected by microarrays and/or clone libraries from three sample matrices

Matrix	*Taxonomic level*	*Array only*	*Array and clone*[a]	*Clone only*[a]
Air	Phylum	17	10	0
	Class	34	21	1
	Order	82	35	2
	Family	157	62	5
	Subfamily	178	60	6
	OTU	961	54	76
Soil	Phylum	23	10	0
	Class	49	19	1
	Order	105	29	1
	Family	194	46	3
	Subfamily	239	40	3
	OTU	1237	37	28
Water	Phylum	17	2	0
	Class	31	3	0
	Order	60	4	0
	Family	86	6	0
	Subfamily	93	6	0
	OTU	250	17	5

[a] The DNAML-F84 homology cutoff used for clone assignment to each level of taxonomic resolution is as follows: Phylum 80%, Class 85%, Order 90%, Family 92%, Subfamily 94%, OTU 97%.

Table 2. Phyla detected in different sample matrices by cloning and sequencing or by high-density DNA microarray analysis

	Air		*Soil*		*Water*	
Phylum	*Clone*	*Array*	*Clone*	*Array*	*Clone*	*Array*
Acidobacteria	×	×	×	×		×
Actinobacteria	×	×	×	×		×
AD3		×		×		
Aquificae		×		×		
Bacteroidetes	×	×	×	×	×	×
BRC1		×		×		×
Caldithrix		×		×		
Chlamydiae		×		×		
Chlorobi	×	×		×		
Chloroflexi	×	×	×	×		×
Coprothermobacteria		×		×		
Cyanobacteria	×	×		×		×
Deinococcus–Thermus	×	×		×		
DSS1		×		×		
Firmicutes	×	×	×	×		×
Gemmatimonadetes	×	×		×		
Lentisphaerae				×		×
marine group A		×		×		×
Natronoanaerobium		×		×		
NC10				×		×
Nitrospira		×[a]		×		×
OD1		×				
OP10			×	×		×
OP8		×		×		
OP9/JS1				×		×
Planctomycetes		×[a]		×		×
Proteobacteria	×	×	×	×	×	×
SPAM				×		
Spirochaetes		×		×		
Termite group 1				×		
TM7		×[a]	×	×		×
Verrucomicrobia		×	×	×		×
WS3		×	×	×		×
WS5				×		×

[a]Phylum confirmed by taxa-specific PCR.

Table 3. Phyla detected only by DNA microarray analysis and subsequently verified by PCR

Phyla verified	*Confirming amplicon GenBank acc. #*	*Closest matching database sequence (GenBank acc. #)*	*Closest matching array OTU*	*Primer sequences (5′ to 3′)*	*PCR annealing temperature (°C)*
Nitrospira	DQ129656	Sludge clone GC86 (Y14644)	*Nitrospiraceae*; OTU 864[a]	For-TCGAAAAG CGTGGGG Rev-CTTCCTC CCCCGTTC	47.0
Planctomycetes	DQ129666	anoxic basin clone CY0ARA027D12 (BX294763)	*Planctomycetaceae*; OTU 4948[a]	For-GAAACTGC CCAGACAC Rev-AGTAACG TTCGCACAG	60.0
TM7	DQ236249	Sludge clone SBR1071 (AF268996)	TM7-1; OTU 3664[b]	For-AYTGGGCGT AAAGAGTTGC Rev-TACGGYTA CCTTGTTACGACTT	66.3

[a]Closest match of confirming amplicon is contained within array-detected OTU.
[b]Closest match of confirming amplicon is contained within alternate TM7 OTU.
Amplicons confirming the phyla detected by microarray were sequenced and identified by comparison to database sequences.

Table 4. Clone library based estimates of diversity and predicted richness compared with observed richness determined by cloning and array approaches

	Estimated diversity		*Predicted richness*[a]		*Observed richness*		
Matrix	*Shannon*[b]	*Dominance*[c]	*Chao1*	*ACE*	*Clone*	*Array*	*Difference*[d]
Air	2.74	0.38	155–396	165–354	66	238	3.61
Soil	2.18	0.46	119–276	123–250	43	279	6.48
Water	0.65	0.64	4–17	2–17	6	99	16.5

[a]95% confidence intervals are shown both Chao 1 and ACE richness estimators.
[b]Shannon–Weaver index of diversity.
[c]Dominance is expressed as 1 − Shannon's equitability (evenness) index.
[d]Fold difference in richness observation between clone library and array approaches.

1.5 days compared to the clone-and-sequence method, which necessitates 3 days or longer to sequence and analyze a typical library of several hundred clones. It has previously been demonstrated that organisms from complex environmental samples can be accurately classified into their respective orders by using an array with 62,358 probes [12, 61]. The present study investigated the response of a novel microarray containing 297,851 probes when hybridized to 16S rRNA gene amplicons generated from aerosols, soil, and water. The microarray design approach was based on the anticipation that the tool would be used to characterize samples without prior knowledge of their microbial composition. For this reason, more than 30,000 diverse 16S rRNA gene sequences were clustered into 8935 OTUs consisting of 842 subfamilies. Every OTU was interrogated by 24 probes, on average, each adjacent to a control probe used to subtract the effects of nonspecific hybridization. The requirement of a sequence-specific interaction from multiple unique probes to identify the presence of each OTU was implemented to increase the confidence of detection over single probe per OTU methods.

This high-density microarray targets the most unique portions of the 16S rRNA gene for a given cluster, and the results are summarized at higher phylogenetic levels. This is in contrast to scoring probes at each node in a hierarchical tree [40]. Our approach accommodates OTUs that may be divergent from other members of the same encapsulating node (which is often the case with environmental sequences) by not requiring that a single probe solution must be found for the entire node. Although the microbial census from every environment is far from complete [3], the key questions for suitability of this approach are: "Is the sequence variation from all extant prokaryotes unlimited, encompassing every possible nucleotide variation within the 16S rRNA gene?", or conversely, "Can a majority of the organisms be classified on the basis of similarity to identified sequences in the databases?" Unlike estimates of microbial genomic variability [21], sequence variability of the 16S rRNA gene appears more constrained [52], most likely because of the functional necessity of the ribosome. Thus, to some degree, sequences not yet in the database may share some homology with targets used for probe analysis. This basic assumption of probe design, which has enabled the identification of one or a few OTUs using Southern analysis [20, 36], fluorescence *in situ* hybridization (FISH) [8, 23], and quantitative PCR (QPCR) [1, 24], was extended to design probes for a substantially greater number of OTUs. To allow for the detection of environmental sequences slightly divergent from those represented on our array, we do not require a sequence specific interaction of 100% of probe pairs defining an OTU.

It was hypothesized that a phylogenetic profile calculated from an array analysis should reflect the composition of sequence types obtained by cloning the same amplicon pool. Specifically, we tested the effectiveness of the novel microarray in detecting and categorizing environmental 16S rRNA genes into taxa with defined nomenclature. Three environments were selected for bacterial community evaluation by means of DNA extraction and universal 16S rRNA gene PCR amplification. Products were split for sampling by cloning-and-sequencing or microarray hybridization.

The three clone libraries produced 253 – 485 sequences each and varied in composition relative to each other. The soil and air were dominated by Actinobacteria and Bacilli, respectively (Fig. 1), and contained over 40 subfamilies each. The water library possessed considerably less richness with only six subfamilies. As predicted, accumulation curves demonstrated that hundreds of clones were insufficient to catalog all subfamilies putatively present, but that the water appeared more thoroughly sampled than the others. The divergent characteristics of the three clone libraries were considered beneficial for testing the array against dissimilar 16S rRNA gene amplicon community structures.

After each amplicon pool was hybridized to replicate arrays, probe responses were matched to OTUs in the database. Detection of an OTU required more than 92% of the probe pairs assigned to the particular probe set to hybridize such that the PM probe had a greater intensity than the MM probe partner. This threshold was chosen to allow sequences with minor divergence from database entries, from which the array was designed, to be

detectable by the array. The OTUs found by the array and/or the cloning method were summarized to the subfamily, family, order, class, and, lastly, phylum to discern the resolution at which the results deviate. Regardless of the resolution considered, the array consistently revealed greater richness than the corresponding clone library (Table 1). This result was expected because nonasymptotic accumulation curves demonstrated that the clone libraries were only a partial sample of the total sequence diversity. The array predicted the presence of every phylum represented in the clone library. The same concurrence held for nearly all classes and orders in all three environmental samples. The atmospheric bacterial samples offered the most relevant example of a previously uncharacterized environment, because very few aerosol-derived 16S rRNA gene sequences are publicly available. In this relatively unstudied environment, OTU-level matching of cloned sequences and array positives was in poor agreement. Yet, the two methods concurred at the subfamily ranking with some exceptions. For example, in air samples, subfamilies within the Myxococcaceae (δ-proteobacteria) and Williamsiaceae (Actinobacteria) were overlooked by the array. Similarly, in the soil, cloned subfamilies within the Opitutaceae (Verrucomicrobia) and Propionibacteriaceae (Actinobacteria) were not found by the arrays using a PosFrac threshold of 0.92. Although an explanation for the reduced PosFrac within Myxococcaceae and Opitutaceae could not be attributed to mismatches between the clone and probe sequences, it was clear that divergence at the loci targeted by the probes would prohibit a sequence-specific response for the Williamsiaceae and Propionibacteriaceae.

The greater number of phyla reported by the array, but not represented in the clone libraries, was unanticipated (Table 2). There are two main factors which, individually or combined, may help explain this anomaly: (1) either the array approach overestimates richness as a result of nonspecific hybridization leading to false-positives or (2) cloning does not truly represent sequence distribution because of insufficient sampling or perhaps cloning bias. Three phyla detected only by array analysis (Nitrospira, Planctomycetes, and TM7) in air samples were chosen for further investigation. Amplification using specific PCR primers and sequencing of amplicons confirmed the presence of the phyla and in two cases the exact OTU detected by the array was also confirmed. This demonstrated not only that the microarrays revealed broader diversity than a typical clone library, but also that the additional components could be identified and subsequently verified with a confirmatory third method. It is significant that entire phyla would have been overlooked if the clone library were the sole source of taxonomic sampling.

It was impractical to determine if sequencing to extinction (asymptotic accumulation/rarefaction curves) would have revealed the additional phyla, because it has been estimated that $>10^4$ sequences may only be sufficient to encounter half of an environment's microbial richness [14]. However, the clone libraries presented in this work reflect the method as it is typically practiced rather than how it would be statistically complete. Nonetheless, it is possible to predict richness within microbial communities by using rarefaction and statistical estimators [4, 5, 51]. As expected for environmental bacterial community sampling efforts, accumulation curves and nonparametric richness estimators demonstrated that no community in this study was sampled to completion. Importantly, the predicted richness extrapolated from cloning observations from the air and soil was quite similar to that enumerated by array analysis. The richness detected by the array for the water sample considerably exceeded the predicted richness due to high dominance.

In a previous study on environmental amplicon sampling, we demonstrated a lack of correlation between the numbers of clones from a phylogenetic taxon and the corresponding hybridization intensity by using a 62,358 probe array [61]. Analogous discrepancies have been documented when sampling environmental PCR products by cloning *versus* SSCP [31], or *versus* T-RFLP [42]—suggesting a cloning bias. It is possible that the cloning process is limited because of nonrandom selection from a heterogeneous pool when amplicons are nonuniform in length [47] or form variable secondary structures [38]. Conversely, the 16S rRNA microarray has the potential advantage of increased sensitivity. Where typical clone libraries must be pruned of sequencing aberrations (including chimeras) usually resulting in only hundreds of amplicons graduating to the final taxonomic assessment, the array accepts the entire mass of PCR products to be exposed to the probes. Using the described method of data analysis, the microarray requires $>10^7$ gene copies for detection (manuscript in preparation). In this study, between 10^{11} and 10^{12} molecules were sampled by the array, whereas only hundreds were analyzed by cloning. Therefore, minority amplicon types, with concentrations 4 orders of magnitude less than those in the majority, will have an increased probability of being detected by the microarray. In fact, high dominance within clone libraries correlated with large differences between the richness detected by array and cloning approaches (Table 4). The trend may predict that underestimation of the true richness can occur when cloning efforts produce only a limited pool of diversity. We acknowledge that further investigation of this trend is necessary, especially to exclude the possibility that simply small sample sizes are the sole cause of underestimation.

The described high-density universal 16S rRNA microarray has been successfully used to monitor metal-reducing bacteria during uranium bioremediation [57] and flux in airborne prokaryote populations in urban

settings (Andersen *et al.*, unpublished data). In this study, we presented the results of applying PCR products to the arrays; however, by interrogating nonamplified rRNA, a significant source of bias can be alleviated. This is the focus of ongoing studies.

In summary, although this microarray is unreliable in classifying novel taxa it, was capable of confirming the majority of clone-detected subfamilies in addition to revealing greater richness, even at the phylum level. Furthermore, richness observations from the array analysis corresponded well with nonparametric richness predictions calculated from clone sampling, indicating a more complete inventory of the sampled ecosystems. A subset of taxa uniquely identified by the array was verified, illustrating the consequence of relying solely on clone libraries when profiling a microbial community. The laborious, costly, and time-consuming nature of clone library analysis diminishes its utility in studies requiring replication and temporal monitoring. The responsiveness of the 16S rRNA microarray to nucleic acids from diverse phyla in complex mixtures and its suitability for investigations requiring replication, demonstrated a necessary advance toward the goal of high-throughput ecological monitoring. For these reasons, we believe the high-density DNA microarray offers a promising approach for studies of microbial ecology.

Acknowledgments

The authors wish to thank Dr. Phil Hugenholtz and Dr. Paul Richardson of the Joint Genome Institute for assistance in constructing the clone libraries. We acknowledge the expert technical assistance of Sonya Murray, Richard Phan, Mark Rojas, and Ramine Cromartie-Thornton. Computational support was provided through the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov). This work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, and was funded in part by the Department of Homeland Security under grant number HSSCHQ04X00037 and in part by the Department of Energy, Environmental Remediation Sciences Program (ERSP) and the Genomics: GTL program.

References

1. Bach, HJ, Tomanova, J, Schloter, M, Munch, JC (2002) Enumeration of total bacteria and bacteria with genes for proteolytic activity in pure cultures and in environmental samples by quantitative PCR mediated amplification. J Microbiol Methods 49: 235–245
2. Castiglioni, B, Rizzi, E, Frosini, A, Sivonen, K, Rajaniemi, P, Rantala, A, Mugnai, MA, Ventura, S, Wilmotte, A, Boutte, C, Grubisic, S, Balthasart, P, Consolandi, C, Bordoni, R, Mezzelani, A, Battaglia, C, De Bellis, G (2004) Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. Appl Environ Microbiol 70: 7161–7172
3. Chandler, DP, Jarrell, AE (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. BioTechniques 38: 591–600
4. Chao, A (1984) Non-parametric estimation of the number of classes in a population. Scand J Statist 11: 265–270
5. Chao, S, Lee, AM (1992) Estimating the number of classes via sample coverage. J Am Stat Assoc 87: 210–217
6. Chee, M, Yang, R, Hubbell, E, Berno, A, Huang, XC, Stern, D, Winkler, J, Lockhart, DJ, Morris, MS, Fodor, SP (1996) Accessing genetic information with high-density DNA arrays. Science 274: 610–614
7. Chouari, R, Le Paslier, D, Dauga, C, Daegelen, P, Weissenbach, J, Sghir, A (2005) Novel major bacterial candidate division within a municipal anaerobic sludge digester. Appl Environ Microbiol 71: 2145–2153
8. Crocetti, GR, Hugenholtz, P, Bond, PL, Schuler, A, Keller, J, Jenkins, D, Blackall, LL (2000) Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. Appl Environ Microbiol 66: 1175–1182
9. DeSantis, TZ, Dubosarskiy, I, Murray, SR, Andersen, GL (2003) Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics 19: 1461–1468
10. DeSantis, TZ, Hugenholtz, P, Keller, K, Brodie, EL, Larsen, N, Piceno, YM, Phan, R, Andersen, GL (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 34: W394–399
11. DeSantis, TZ, Hugenholtz, P, Larsen, N, Rojas, M, Brodie, EL, Keller, K, Huber, T, Dalevi, D, Hu, P, Andersen, GL (2006) Greengenes, a chimera-checked 16S rRNA genes database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072
12. DeSantis, TZ, Stone, CE, Murray, SR, Moberg, JP, Andersen, GL (2005) Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiol Lett 245: 271–278
13. Dojka, MA, Hugenholtz, P, Haack, SK, Pace, NR (1998) Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl Environ Microbiol 64: 3869–3877
14. Dunbar, J, Barns, SM, Ticknor, LO, Kuske, CR (2002) Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol 68: 3035–3045
15. Elshahed, MS, Senko, JM, Najar, FZ, Kenton, SM, Roe, BA, Dewers, TA, Spear, JR, Krumholz, LR (2003) Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. Appl Environ Microbiol 69: 5609–5621
16. Ewing, B, Green, P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186–194
17. Ewing, B, Hillier, L, Wendl, MC, Green, P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8: 175–185
18. Farrelly, V, Rainey, FA, Stackebrandt, E (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl Environ Microbiol 61: 2798–2801

19. Felsenstein, J (1989) PHYLIP—Phylogeny Inference Package (Version 3.65). Cladistics 5: 164–166
20. Felske, A, Akkermans, AD (1998) Prominent occurrence of ribosomes from an uncultured bacterium of the Verrucomicrobiales cluster in grassland soils. Lett Appl Microbiol 26: 219–223
21. Gans, J, Wolinsky, M, Dunbar, J (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309: 1387–1390
22. Garrity, GM (2001) Bergey's Manual of Systematic Bacteriology, 2nd edn. Springer-Verlag, New York
23. Ginige, MP, Hugenholtz, P, Daims, H, Wagner, M, Keller, J, Blackall, LL (2004) Use of stable-isotope probing, full-cycle rRNA analysis, and fluorescence *in situ* hybridization-microautoradiography to study a methanol-fed denitrifying microbial community. Appl Environ Microbiol 70: 588–596
24. Holmes, DE, Finneran, KT, O'Neil, RA, Lovley, DR (2002) Enrichment of members of the family Geobacteraceae associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. Appl Environ Microbiol 68: 2300–2306
25. Huber, T, Faulkner, G, Hugenholtz, P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20: 2317–2319
26. Hugenholtz, P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3: 1–8
27. Hugenholtz, P, Goebel, BM, Pace, NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180: 4765–4774
28. Hugenholtz, P, Huber, T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. Int J Syst Evol Microbiol 53: 289–293
29. Hugenholtz, P, Tyson, GW, Webb, RI, Wagner, AM, Blackall, LL (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Appl Environ Microbiol 67: 411–419
30. Hughes, JB, Hellmann, JJ, Ricketts, TH, Bohannan, BJ (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol 67: 4399–4406
31. Junca, H, Pieper, DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. Environ Microbiol 6: 95–110
32. Kazor, CE, Mitchell, PM, Lee, AM, Stokes, LN, Loesche, WJ, Dewhirst, FE, Paster, BJ (2003) Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients. J Clin Microbiol 41: 558–563
33. Kelly, JJ, Siripong, S, McCormack, J, Janus, LR, Urakawa, H, El Fantroussi, S, Noble, PA, Sappelsa, L, Rittmann, BE, Stahl, DA (2005) DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. Water Res 39: 3229–3238
34. Kent, AD, Smith, DJ, Benson, BJ, Triplett, EW (2003) Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. Appl Environ Microbiol 69: 6768–6776
35. Lane, DJ (1991) Nucleic acid techniques in bacterial systematics. In: Stackebrandt, E, Goodfellow, M (Eds.) Nucleic Acid Techniques in Bacterial Systematics, Vol 1. Wiley, New York, pp 115–175
36. Lanser, JA, Adams, M, Doyle, R, Sangster, N, Steele, TW (1990) Genetic relatedness of *Legionella longbeachae* isolates from human and environmental sources in Australia. Appl Environ Microbiol 56: 2784–2790
37. Lehner, A, Loy, A, Behr, T, Gaenge, H, Ludwig, W, Wagner, M, Schleifer, KH (2005) Oligonucleotide microarray for identification of *Enterococcus* species. FEMS Microbiol Lett 246: 133–142
38. Lindstrom, UM, Kool, ET (2002) An orthogonal oligonucleotide protecting group strategy that enables assembly of repetitive or highly structured DNAs. Nucleic Acids Res 30: e101
39. Liu, WT, Stahl, DA (2002) Molecular approaches for the measurement of density, diversity, and phylogeny. In: Hurst, CJ, Crawford, RL, Knudsen, GR, McInerney, MJ, Stetzenbach, LD (Eds.) Manual of Environmental Microbiology, 2nd edn., pp 114–134
40. Loy, A, Lehner, A, Lee, N, Adamczyk, J, Meier, H, Ernst, J, Schleifer, KH, Wagner, M (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. Appl Environ Microbiol 68: 5064–5081
41. Ludwig, W, Strunk, O, Westram, R, Richter, L, Meier, H, Yadhukumar, Buchner, A, Lai, T, Steppi, S, Jobb, G, Forster, W, Brettske, I, Gerber, S, Ginhart, AW, Gross, O, Grumann, S, Hermann, S, Jost, R, Konig, A, Liss, T, Lussmann, R, May, M, Nonhoff, B, Reichel, B, Strehlow, R, Stamatakis, A, Stuckmann, N, Vilbig, A, Lenke, M, Ludwig, T, Bode, A, Schleifer, KH (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371
42. Lueders, T, Friedrich, M (2000) Archaeal population dynamics during sequential reduction processes in rice field soil. Appl Environ Microbiol 66: 2732–2742
43. Magurran, AE (1988) Ecological Diversity and Its Measurement. Princeton University Press, Princeton, NJ
44. Masuda, N, Church, GM (2002) Escherichia coli gene expression responsive to levels of the response regulator EvgA. J Bacteriol 184: 6225–6234
45. McCaig, AE, Glover, LA, Prosser, JI (1999) Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Appl Environ Microbiol 65: 1721–1730
46. Miller, DN, Bryant, JE, Madsen, EL, Ghiorse, WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl Environ Microbiol 65: 4715–4724
47. Ohara, O, Temple, G (2001) Directional cDNA library construction assisted by the *in vitro* recombination reaction. Nucleic Acids Res 29: E22
48. Polz, MF, Cavanaugh, CM (1998) Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol 64: 3724–3730
49. Rosch, C, Bothe, H (2005) Improved assessment of denitrifying, N_2-fixing, and total-community bacteria by terminal restriction fragment length polymorphism analysis using multiple restriction enzymes. Appl Environ Microbiol 71: 2026–2035
50. Rozen, S, Skaletsky, H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–386
51. Schloss, PD, Handelsman, J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71: 1501–1506
52. Schloss, PD, Handelsman, J (2004) Status of the microbial census. Microbiol Mol Biol Rev 68: 686–691
53. Schmidt, TM, DeLong, EF, Pace, NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 173: 4371–4378
54. Skowronski, EW, Armstrong, N, Andersen, G, Macht, M, McCready, PM (2000) Magnetic, microplate-format plasmid isolation protocol for high-yield, sequencing-grade DNA. BioTechniques 29: 786–788, 790, 792
55. Torsvik, V, Daae, FL, Sandaa, RA, Ovreas, L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. J Biotechnol 64: 53–62

56. Urakawa, H, Noble, PA, El Fantroussi, S, Kelly, JJ, Stahl, DA (2002) Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Appl Environ Microbiol 68: 235–244
57. Wan, J, Tokunaga, TK, Brodie, E, Wang, Z, Zheng, Z, Herman, D, Hazen, TC, Firestone, MK, Sutton, SR (2005) Reoxidation of bioreduced uranium under reducing conditions. Environ Sci Technol 39: 6162–6169
58. Warsen, AE, Krug, MJ, LaFrentz, S, Stanek, DR, Loge, FJ, Call, DR (2004) Simultaneous discrimination between 15 fish pathogens by using 16S ribosomal DNA PCR and DNA microarrays. Appl Environ Microbiol 70: 4216–4221
59. Webster, G, Newberry, CJ, Fry, JC, Weightman, AJ (2003) Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. J Microbiol Methods 55: 155–164
60. Wilson, KH, Blitchington, RB, Greene, RC (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction [published erratum appears in J Clin Microbiol 1991 Mar; 29(3): 666]. J Clin Microbiol 28: 1942–1946
61. Wilson, KH, Wilson, WJ, Radosevich, JL, DeSantis, TZ, Viswanathan, VS, Kuczmarski, TA, Andersen, GL (2002) High-density microarray of small-subunit ribosomal DNA probes. Appl Environ Microbiol 68: 2535–2541
62. Wilson, WJ, Strout, CL, DeSantis, TZ, Stilwell, JL, Carrano, AV, Andersen, GL (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol Cell Probes 16: 119–127
63. Wu, L, Thompson, DK, Li, G, Hurt, RA, Tiedje, JM, Zhou, J (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol 67: 5780–5790
64. Zhou, J (2003) Microarrays for bacterial detection and microbial community analysis. Curr Opin Microbiol 6: 288–294

Applied and Environmental Microbiology, July 2006, p. 5069–5072 Vol. 72, No. 7
0099-2240/06/$08.00+0 doi:10.1128/AEM.03006-05

Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB

T. Z. DeSantis,[1] P. Hugenholtz,[2] N. Larsen,[3] M. Rojas,[4] E. L. Brodie,[1] K. Keller,[5] T. Huber,[6] D. Dalevi,[7] P. Hu,[1] and G. L. Andersen[1]*

Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 70A-3317, Berkeley, California 94720[1]; Microbial Ecology Program, DOE Joint Genome Institute, 2800 Mitchell Drive, Bldg. 400-404, Walnut Creek, California 94598[2]; Danish Genome Institute, Gustav Wieds vej 10 C, DK-8000 Aarhus C, Denmark[3]; Department of Bioinformatics, Baylor University, P.O. Box 97356, 1311 S. 5th St., Waco, Texas 76798-7356[4]; Department of Bioengineering, University of California, Berkeley, California 94720[5]; Departments of Biochemistry and Mathematics, The University of Queensland, Brisbane, Queensland 4072, Australia[6]; and Department of Computer Science, Chalmers University of Technology, SE-412 96 Göteborg, Sweden[7]

Received 20 December 2005/Accepted 15 April 2006

A 16S rRNA gene database (http://greengenes.lbl.gov) addresses limitations of public repositories by providing chimera screening, standard alignment, and taxonomic classification using multiple published taxonomies. It was found that there is incongruent taxonomic nomenclature among curators even at the phylum level. Putative chimeras were identified in 3% of environmental sequences and in 0.2% of records derived from isolates. Environmental sequences were classified into 100 phylum-level lineages in the *Archaea* and *Bacteria*.

Comparative analysis of 16S small-subunit rRNA genes is commonly used to survey the constituents of microbial communities (4, 13, 23, 24), to infer bacterial and archaeal evolution (14, 19), and to design monitoring and analysis tools, such as microarrays (5, 10, 17, 20, 29, 30). Because the rate of production of 16S small-subunit rRNA gene sequence records for uncultured organisms now exceeds the rate of production for their cultured counterparts, taxonomic placement of sequences lags behind. In fact, 43% of full-length 16S small-subunit rRNA gene records in the GenBank database are amalgamated into the pseudodivisions "environmental samples" and "unclassified." Annotation styles are inconsistent, creating barriers for computational categorization of biological sources. Furthermore, since rRNA genes from environmental DNA are usually PCR amplified, it is suspected that many clandestine chimeric sequences are intercalated into the public databases. For a small sample of 1,399 sequence records from known phyla, it was estimated that 3% of the public data might contain chimeras (2). The effect of these poor-quality data, exacerbated by barriers in exchanging nomenclature, has led to several conflicting taxonomies. The probability of mistakenly adopting a chimeric sequence in a phylogenetic inference or as a reference for probe/primer design is increasing noticeably. Finally, ARB (21) database administration needs to be streamlined for workers who maintain 16S small-subunit rRNA gene collections on their local computers.

Greengenes addresses these concerns by providing four features: a standardized set of descriptive fields, taxonomic assignment, chimera screening, and ARB compatibility. Heuristics are used to consider the author's annotations and categorize each source as a named or unnamed isolate, an unnamed symbiont, or an uncultured organism. Other standard descriptors include sequence quality measurements, authors, and a "study_id" that links all the records associated with a project. Greengenes maintains a consistent multiple-sequence alignment (MSA) of both archaeal and bacterial 16S small-subunit rRNA genes to facilitate taxonomic placement. Taxonomy proposed by independent curators, including the NCBI, the Ribosomal Database Project (RDP) (Bergey's) (7), Wolfgang Ludwig (21), Phil Hugenholtz (16), and Norman Pace (23), is tracked to promote user awareness of several estimations of phylogenetic descent, allowing a balanced approach to node nomenclature when dendrograms are generated. Comprehensive chimera assessment is a distinguishing characteristic of the Greengenes data assembly process. Each sequence is scored for chimeric potential, a breakpoint is estimated, and parent sequences are identified. Furthermore, since biologists often collect and visualize 16S small-subunit rRNA gene relationships using the freely available ARB software, Greengenes simplifies the chore of keeping a research group's private ARB database current by providing standardized alignments and an import filter (greengenes.ift) that imports the alignment and other standardized fields from 16S small-subunit rRNA gene records vetted weekly from GenBank.

To illustrate the utility of the Greengenes data assembly process and to examine the validity of prokaryotic candidate phyla, we aligned and chimera checked more than 90,000 public 16S small-subunit rRNA gene sequences. Taxonomic classifications from the major curators were used when such classifications were available. Sequence data were imported from NCBI for complete or nearly complete gene sequences (length, >1,250 nucleotides) deposited as of 2 April 2006. Alignment of both archaeal and bacterial sequences was performed with the NAST aligner (8) against a "Core Set" of templates selected from a phylogenetically broad collection (16). The resulting MSA was formatted so that each sequence occupied a consistent 7,682 characters or 4,182 characters; the latter allowed

* Corresponding author. Mailing address: Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 70A-3317, Berkeley, CA 94720. Phone: (510) 495-2795. Fax: (510) 486-7152. E-mail: GLAndersen@lbl.gov.

compatibility with RDP v8.1 (22) alignments. Both these formats were concise enough for browsing in common MSA graphical interfaces, such as ClustalX (28), MEGA (18), and the platform-independent interface Jalview (6), as well as ARB. Other standard expansions, such as the >20,000-character Ludwig alignment, are alternate formats that will be available in future releases to give maximum flexibility to researchers.

For high-throughput chimera screening of the aligned sequences, the program Bellerophon (15) was used with two modifications. First, the algorithm was modified to reduce the number of potential parents considered in the partial trees, which allowed run time to scale linearly rather than logarithmically with the count of candidate sequences in a collection. Second, a new metric was implemented, which weighted the likelihood of a sequence being chimeric according to the similarity of the parent sequences. The more distantly related the parent sequences were to each other relative to their divergence from the candidate chimeric sequence, the greater the likelihood that the inferred chimera was real. This metric, called the divergence ratio, used the average sequence identity between the two fragments of the candidate and the corresponding parent sequences as the numerator and the sequence identity between the parent sequences as the denominator. All calculations were restricted to 1,287 conserved columns of aligned characters using a 300-bp window on either side of the most likely breakpoint. A divergence ratio of >1.1 and fragment-to-parent levels of similarity of >90% were required for classifying sequences as putatively chimeric.

Taxonomy was linked to each record by various methods. NCBI taxonomic nomenclature and RDP taxonomic nomenclature were extracted directly from the corresponding GenBank-formatted records. The Pace and Ludwig annotations were exported from curated ARB databases. The Hugenholtz taxonomy was also derived from a curated ARB database in which tree topologies had been verified using RAxML-VI (27) for maximum likelihood inference. The general time-reversible model of evolution was applied together with optimization of substitution rates and site-specific rates according to a gamma distribution. Different search algorithms were considered depending on the run time of the standard hill climb (SHC) search method. If the running time was less than 8 h, simulating annealing (SA) was processed with the default starting temperature and a termination time set at approximately 24 h. If simulating annealing was not used and SHC terminated within 24 h, SHC was used. Furthermore, rapid hill climb was used in all other cases when the running time was less than 24 h. If rapid hill climb did not terminate within the set limit, the number of taxa was reduced. After 100 bootstrap replications, a consensus tree was calculated using Consense (12) and imported into ARB. This database (greengenes.arb) is available for download through Greengenes and is updated periodically.

Of the 90,000 NCBI records analyzed, 54% were derived from uncultured organisms, the majority of which were deposited in the last 5 years (Fig. 1). Only three studies have submitted more than 1,000 full-length clones; however, we expect the number of large 16S small-subunit rRNA gene surveys to increase due to the availability and falling cost of high-throughput sequencing. Bellerophon detection of putative chimeras in

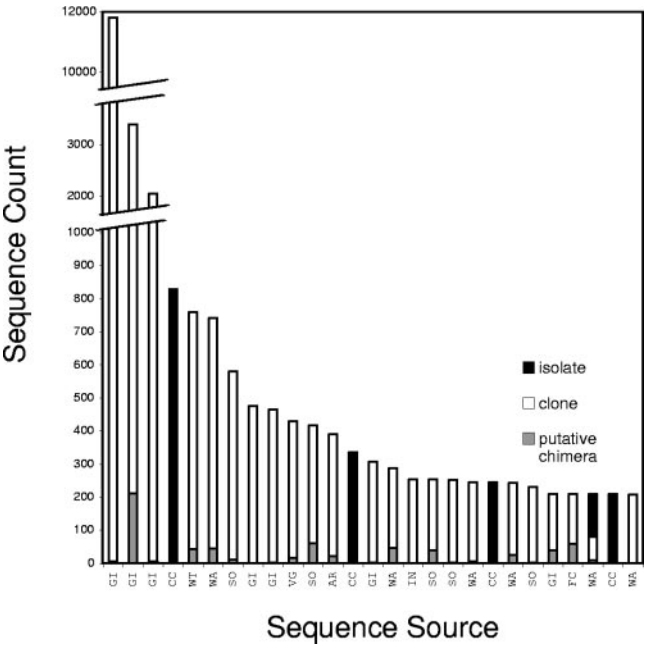


FIG. 1. 16S rRNA gene sequencing projects that produced more than 200 full-length records. All projects were submitted to GenBank between October 2000 and February 2006. Sequences were generated from gastrointestinal (GI), soil (SO), vaginal (VG), aerosol (AR), culture collection (CC), insect (IN), water (WA), waste treatment (WT), and fecal (FC) sources as indicated on the *x* axis. The projects are ordered by sequence count.

3% of the sequences from uncultured organisms was not unexpected considering the initial estimates (2). Surprisingly, 0.2% of sequences derived from pure cultures were also determined to be putative chimeras. Multiple distinct 16S rRNA genes have been encountered when clone libraries have been created from colonies assumed to be pure cultures prepared from numerous third-party sources (Colleen Cavanaugh, personal communication). It is possible that isolated colonies contain symbiotic bacteria which increase PCR template complexity, enabling chimera formation. In addition, thousands of full-length 16S small-subunit rRNA gene-annotated GenBank records were only partially aligned using NAST. Future versions of NAST could be altered to allow alignment extensions across regions having low template similarity or to allow candidates to be aligned in sections using divergent templates. Both of these options may allow a greater abundance of chimeric data to be imported into Greengenes but perhaps would capture novel phyla from the public repositories. Alternately, manually aligned sequences from novel phyla can be offered from the user community for recruitment to the Core Set advocating periodic reevaluation of the partially aligned set.

Discovery of chimeras in 16S small-subunit rRNA gene data collections is crucial if the data set is going to be a foundation for applied bioinformatics. Chimeras are a fundamental problem when they are used as templates with probe selection software, a growing concern with the recent increase in 16S small-subunit rRNA gene microarray probe development (3, 8, 11). The 15 to 30 bases surrounding the chimeric breakpoint can appear to be sufficiently different from all other records in

a database to cause a probe selection algorithm to justifiably identify the region as a target's signature and suggest complementary probes that can be synthesized. These probes could appear to be very valuable considering their minimal mishybridization potential, but in fact, they would rarely be useful since they target nonexistent organisms. Chimera test results from Greengenes allow greater control over input to probe selection software, should aid in avoiding artificial terminal restriction fragment length polymorphism pattern predictions from ARB-compatible TRF-CUT (25), and can increase the accuracy of sampling rarefaction curves (26).

The fraction of putative chimeras in the deposited sequences from an individual study varies from none to more than 20% (Fig. 1), suggesting that chimera screening is still not being uniformly applied by sequence generators. The problem is exacerbated with sparsely populated candidate phyla. For instance, the bacterial phyla "SAM" and "5" and the class GN4 (*Proteobacteria*) may require reevaluation. Likewise, the genera *Tistrella*, *Caldotoga*, *Dehalobacterium*, and *Desulfovermiculus* are currently anchored by sequences with evidence of chimeric composition. Additional sequences could lead to empirical rejection of certain classifications or may aid in defining the true breadth of sequence variation for these taxa.

Comparison of five different taxonomies uncovered surprisingly great disparity between expert curators. Loosely interpreting a "phylum" to be any labeled group or division immediately subordinate to the domain *Archaea* or *Bacteria*, we compared the five curations in a Venn diagram (Fig. 2). The main source of the disparity is the discordant naming of novel candidate phyla or the absence of names for candidate phyla. For example, Pace and Hugenholtz have independently named more than 12 phylum-level lineages, many of which are the same lineages, and RDP has not named any of these lineages. This is a consequence of the huge number of environmental sequences in the public databases and the frequent redundant naming of environmental lineages in the literature. We hope that making multiple taxonomic classifications available through Greengenes will aid in standardizing classification, particularly classification of environmental lineages.

Greengenes is also a functional workbench to assist in analysis of user-generated 16S rRNA gene sequences. Batches of sequencing reads can be uploaded for quality-based trimming and creation of multiple-sequence alignments (9). Three types of non-MSA similarity searches are also available, seed extension by BLAST (1), similarity based on shared 7-mers by a tool called "Simrank," and a direct degenerative pattern match for probe/primer evaluation. Results are displayed using user-preferred taxonomic nomenclature and can be saved between sessions.

In summary, Greengenes offers annotated, chimera-checked, full-length 16S rRNA gene sequences in standard alignment formats. The relational database links taxonomies from multiple curators and multiple sequences from a single study. We found that there is incongruent taxonomic nomenclature among curators even at the phylum level. Bellerophon found putative chimeras in sequences derived from both uncultured and isolated organisms. The data set can be compared to user-provided sequences via a web interface or can be imported directly into ARB for advanced analyses. We anticipate

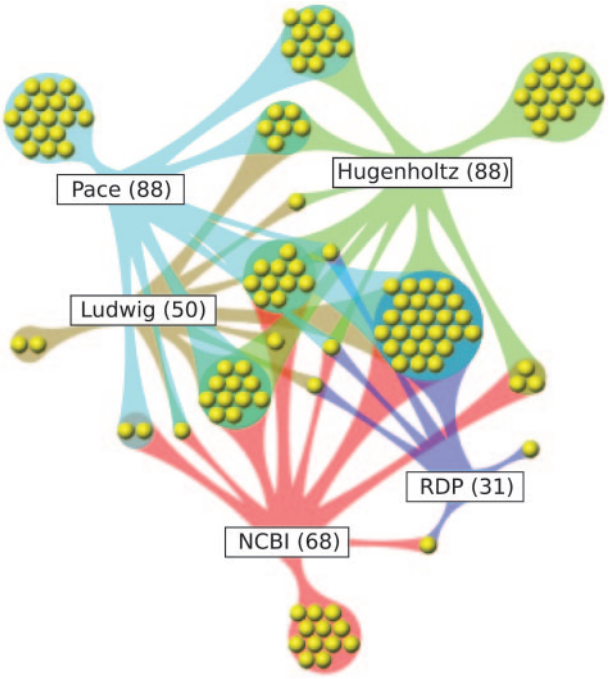


FIG. 2. Phylum-level nomenclature shared by independent curators represented as a five-way Venn diagram. Yellow spheres represent the 126 phylum or candidate division names encountered in at least one of the five taxonomy systems (Pace, Hugenholtz, Ludwig, RDP, or NCBI). The numbers in parentheses are the counts for phylum or candidate division names recognized by an individual curator. Clusters of yellow spheres connected by more than one colored web symbolize names recognized by multiple curators. The image was rendered by the AutoFocus software (Aduna B.V., The Netherlands). A complete table of phylum-level nomenclature comparisons is available at http://greengenes.lbl.gov/TaxCompare.

that Greengenes will be valuable to researchers conducting environmental surveys and for 16S rRNA microarray design.

In the immediate future, we plan to develop and implement a number of community curation tools. This should allow the user community to actively participate in improving the quality of the Greengenes database and should ensure that time-consuming manual improvements of sequence and sequence-associated data, including taxonomic corrections, are propagated for the benefit of the whole community. Specifically, five curation tools that should capture manual improvements are in development: (i) improvements in individual sequence alignments, (ii) manual verification of putative chimeras, (iii) recruitment of novel lineages to the Core Set, (iv) corrections in the Greengenes description (the abbreviated description of the record usually has the form [habitat] clone [clone name] for environmental sequences), and (v) updating taxonomic group names. One of the main challenges in the implementation of these tools is to ensure that only high-quality manual edits are incorporated into Greengenes. For example, for a suggested alignment alteration, the submitted sequence must (i) match the existing sequence, (ii) preserve the location of highly conserved positions in the 16S rRNA gene, and (iii) record the curator information as part of the update transaction. We recognize the desire of many users to contribute to a distrib-

uted curation effort, and we hope that Greengenes will become a resource to facilitate this desire.

We thank Kirk Harris and Norman Pace for sharing their ARB database and Richard Phan and Yvette Piceno for assistance with the web interface.

The computational infrastructure was provided in part by the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the U.S. Department of Energy Office of Science Office of Biological and Environmental Research Genomics:GTL Program and the Natural and Accelerated Bioremediation Research Program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. Web application development was funded in part by the Department of Homeland Security under grant HSSCHQ04X00037.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman.** 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. **71:**7724–7736.
3. **Ashelford, K. E., A. J. Weightman, and J. C. Fry.** 2002. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. Nucleic Acids Res. **30:**3481–3489.
4. **Brodie, E., S. Edwards, and N. Clipson.** 2002. Bacterial community dynamics across a floristic gradient in a temperate upland grassland ecosystem. Microb. Ecol. **44:**260–270.
5. **Castiglioni, B., E. Rizzi, A. Frosini, K. Sivonen, P. Rajaniemi, A. Rantala, M. A. Mugnai, S. Ventura, A. Wilmotte, C. Boutte, S. Grubisic, P. Balthasart, C. Consolandi, R. Bordoni, A. Mezzelani, C. Battaglia, and G. De Bellis.** 2004. Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. Appl. Environ. Microbiol. **70:**7161–7172.
6. **Clamp, M., J. Cuff, S. M. Searle, and G. J. Barton.** 2004. The Jalview Java alignment editor. Bioinformatics **20:**426–427.
7. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. **33:**D294–D296.
8. **DeSantis, T. Z., I. Dubosarskiy, S. R. Murray, and G. L. Andersen.** 2003. Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics **19:**1461–1468.
9. **DeSantis, T. Z., P. Hugenholtz, K. Keller, E. L. Brodie, N. Larsen, Y. M. Piceno, R. Phan, and G. L. Andersen.** NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res., in press.
10. **DeSantis, T. Z., C. E. Stone, S. R. Murray, J. P. Moberg, and G. L. Andersen.** 2005. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiol. Lett. **245:**271–278.
11. **Emrich, S. J., M. Lowe, and A. L. Delcher.** 2003. PROBEmer: a web-based software tool for selecting optimal DNA oligos. Nucleic Acids Res. **31:**3746–3750.
12. **Felsenstein, J.** 1989. PHYLIP—Phylogeny Inference Package (version 3.65). Cladistics **5:**164–166.
13. **Harris, J. K., S. T. Kelley, and N. R. Pace.** 2004. New perspective on uncultured bacterial phylogenetic division OP11. Appl. Environ Microbiol. **70:**845–849.
14. **Harris, J. K., S. T. Kelley, G. B. Spiegelman, and N. R. Pace.** 2003. The genetic core of the universal ancestor. Genome Res. **13:**407–412.
15. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20:**2317–2319.
16. **Hugenholtz, P.** 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. **3:**1–8.
17. **Kelly, J. J., S. Siripong, J. McCormack, L. R. Janus, H. Urakawa, S. El Fantroussi, P. A. Noble, L. Sappelsa, B. E. Rittmann, and D. A. Stahl.** 2005. DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. Water Res. **39:**3229–3238.
18. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. **5:**150–163.
19. **Lane, D. J., A. P. Harrison, Jr., D. Stahl, B. Pace, S. J. Giovannoni, G. J. Olsen, and N. R. Pace.** 1992. Evolutionary relationships among sulfur- and iron-oxidizing eubacteria. J. Bacteriol. **174:**269–278.
20. **Lehner, A., A. Loy, T. Behr, H. Gaenge, W. Ludwig, M. Wagner, and K. H. Schleifer.** 2005. Oligonucleotide microarray for identification of *Enterococcus* species. FEMS Microbiol. Lett. **246:**133–142.
21. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32:**1363–1371.
22. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. **29:**173–174.
23. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734–740.
24. **Radosevich, J. L., W. J. Wilson, J. H. Shinn, T. Z. DeSantis, and G. L. Andersen.** 2002. Development of a high-volume aerosol collection system for the identification of air-borne micro-organisms. Lett. Appl. Microbiol. **34:**162–167.
25. **Ricke, P., S. Kolb, and G. Braker.** 2005. Application of a newly developed ARB software-integrated tool for in silico terminal restriction fragment length polymorphism analysis reveals the dominance of a novel *pmoA* cluster in a forest soil. Appl. Environ. Microbiol. **71:**1671–1673.
26. **Schloss, P. D., and J. Handelsman.** 2004. Status of the microbial census. Microbiol. Mol. Biol. Rev. **68:**686–691.
27. **Stamatakis, A., T. Ludwig, and H. Meier.** 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics **21:**456–463.
28. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
29. **Webster, G., C. J. Newberry, J. C. Fry, and A. J. Weightman.** 2003. Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. J. Microbiol. Methods **55:**155–164.
30. **Wilson, K. H., W. J. Wilson, J. L. Radosevich, T. Z. DeSantis, V. S. Viswanathan, T. A. Kuczmarski, and G. L. Andersen.** 2002. High-density microarray of small-subunit ribosomal DNA probes. Appl. Environ. Microbiol. **68:**2535–2541.

Long-Term Sustainability of a High-Energy, Low-Diversity Crustal Biome

Li-Hung Lin,[1,2]* Pei-Ling Wang,[3] Douglas Rumble,[4] Johanna Lippmann-Pipke,[5] Erik Boice,[6] Lisa M. Pratt,[6] Barbara Sherwood Lollar,[7] Eoin L. Brodie,[8] Terry C. Hazen,[8] Gary L. Andersen,[8] Todd Z. DeSantis,[8] Duane P. Moser,[9] Dave Kershaw,[10] T. C. Onstott[1]

Geochemical, microbiological, and molecular analyses of alkaline saline groundwater at 2.8 kilometers depth in Archaean metabasalt revealed a microbial biome dominated by a single phylotype affiliated with thermophilic sulfate reducers belonging to *Firmicutes*. These sulfate reducers were sustained by geologically produced sulfate and hydrogen at concentrations sufficient to maintain activities for millions of years with no apparent reliance on photosynthetically derived substrates.

Most subsurface microbial ecosystems examined to date (including subseafloor sediments, deep-sea hydrothermal vents, terrestrial sedimentary aquifers, and petroleum reservoirs) ultimately depend on sunlight. These studies have been mostly confined to depths of less than 1 km, and the ecosystems are either supported by photosynthetically produced electron donors and acceptors transported by groundwater or seawater with ages much less than a million years, or are in constant contact with oxygenated seawater migrating through the underlying fractured basaltic aquifer (*1*–*4*). Although two occurrences of autotrophic microbial communities have been reported to exist in <300-m-deep volcanic aquifers flushed with fresh meteoric water (*5*, *6*), their long-term sustainability on H_2 and isolation from photosynthetically produced substrates have not been demonstrated. Although the existence of subsurface microorganisms at depths greater than 1 km in pristine environments is well established (*7*), much is still unknown regarding the abundance, diversity, and sustainability of these microbial communities over geological time scales.

To determine the long-term sustainability of a deep terrestrial environment, we examined the microbial diversity and metabolic activity of a 3-to 4-km-deep fracture in the 2.7-billion-year-old Ventersdorp Supergroup metabasalt, in which fracture water ages of tens of millions of years (*8*), abundant abiogenic hydrocarbons (*9*), and radiolytically produced H_2 (*10*) have been reported. To characterize the indigenous microbial composition and its principal respiratory pathway, we analyzed 16*S* ribosomal DNA (rDNA) sequences, aqueous and gas geochemistry, and stable and noble gas isotopic signatures of moderately saline groundwater emanating from a fracture zone 2.825 km below the land surface (kmbls) in the Mponeng gold mine, South Africa. This high-pressure water-bearing fracture was intersected during exploratory drilling ahead of a tunnel advancing into an unmined zone ~100 m above the Ventersdorp Contact Reef (VCR) ore zone. The fracture water was initially sampled as soon as it was safe to do so (4 days after the fracture intersection), and three subsequent samples were obtained over a 54-day interval (Table 1) to monitor drilling contamination and possible changes of community structure and geochemistry as the fracture was dewatered and before being sealed by the mine's cementation team.

Fracture water samples yielded a uniform community structure dominated by a single phylotype (MP104-0916-b1) that constituted

Table 1. Geochemical and microbiological characteristics of fracture water and mining water. NA, not available; <d.l., below detection limit, which is 1 μM for O_2 and 0.5 fg ml^{-1} for archaeal DNA; Ma, million years ago.

Sample no. (number of days since intersection)	1 (4)	2 (7)	3 (15)	4 (58)	5
Sample name	MP104E65X C-091602	MP104E65X C-091902	MP104E65X C-092702	MP104E65X C-110902	MP104E65XC-SW-091602
Origin	Fracture water	Fracture water	Fracture water	Fracture water	Mining water
Water and gas flow rates (liters min^{-1}) (borehole volumes)	40/2.4 (45)	10.9/0.8 (55)	8.2/1.2 (64)	2.3/1.7 (96)	NA
pH	9.3	9.3	9.3	9.2	NA
E_h (mV)	−330	−350	−340	−263	NA
Temperature (°C)	>60	52	52	52	20
Formate (μM)‖	7.6	NA	7.1	8.9	1.4
Acetate (μM)‖	24.6	NA	22.5	35.7	5.9
Cl^- (mM)*‖	54.1	NA	71.9	84.9	0.42
Br^- (μM)‖	125	NA	177	218	3.8
SO_4^{2-} (μM)‖	529	NA	900	1860	171
HS^- (μM)†‖	NA	1390	NA	1060	NA
H_2 (μM)†‖	1940	2600	2090	3715	NA
CH_4 (μM)†‖	8580	11800	9320	16600	NA
O_2 (μM)‡‖	6	<d.l.	<d.l.	<d.l.	285
δ^2H-H_2 (‰ VSMOW)‖	−684	−684	−688	−695	NA
δ^{13}C-CH_4 (‰ PDB)§‖	−31.6	−31.7	−32.8	−33.2	NA
δ^2H-CH_4 (‰ VSMOW)‖	−364	−367	−366	−390	NA
^{4}He model age (Ma)	20.9 ± 10.5	20.4 ± 10.2	15.8 ± 7.9	NA	NA
^{40}Ar model age (Ma)	16.3 ± 8.2	21.3 ± 10.6	16.9 ± 8.4	NA	NA
^{134}Xe model age (Ma)	18.7 ± 7.0	19.4 ± 3.8	21.0 ± 6.0	NA	NA
^{136}Xe model age (Ma)	21.6 ± 6.0	25.0 ± 3.8	23.8 ± 4.6	NA	NA
Bacterial DNA (pg ml^{-1})	15 ± 8	16 ± 8	30 ± 15	30 ± 15	3 ± 1.5 × 10^5
Archaeal DNA (pg ml^{-1})	~5 × 10^{-4}	<d.l.	~5 × 10^{-4}	~5 × 10^{-4}	206 ± 100
Cell density (cells ml^{-1})	5.1 ± 0.5 × 10^4	NA	NA	3.3 ± 0.3× 10^4	NA

*The concentrations derived from charged balance were 10.2 mM for sample 1, 27.6 mM for sample 3 and 50.6 mM for sample 4. †The concentrations of dissolved gases were reported as concentrations corrected for diffusive loss (*8*). Diffusive correction was not applied to sample 4 because of the lack of noble gas analysis. ‡The positive O_2 content for sample 1 may be derived from incomplete isolation of fracture water from the mining environment caused by the extremely high water pressure and flow rate. §The carbon isotopic value was referenced to Pee Dee belemnite (PDB). ‖The uncertainties for aqueous and gas chemistry are ±10% and ±20%, respectively. The uncertainties for isotopic measurements are ±0.5‰ for δ^{13}C-hydrocarbon and ±5‰ for δ^2H-hydrocarbon and δ^2H-H_2.

[1]Department of Geosciences, Princeton University, Princeton, NJ, USA. [2]Department of Geosciences, [3]Institute of Oceanography, National Taiwan University, Taipei, Taiwan. [4]Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC, USA. [5]GeoForschungsZentrum Potsdam, Telegrafenberg, Potsdam, Germany. [6]Department of Geological Sciences, Indiana University, Bloomington, IN, USA. [7]Department of Geology, University of Toronto, Toronto, ON, Canada. [8]Ecology Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [9]Division of Earth and Ecosystems Sciences, Desert Research Institute, Las Vegas, NV, USA. [10]Mponeng Mine, Anglo Gold, Johannesburg, South Africa.

*To whom correspondence should be addressed. E-mail: lhlin@ntu.edu.tw

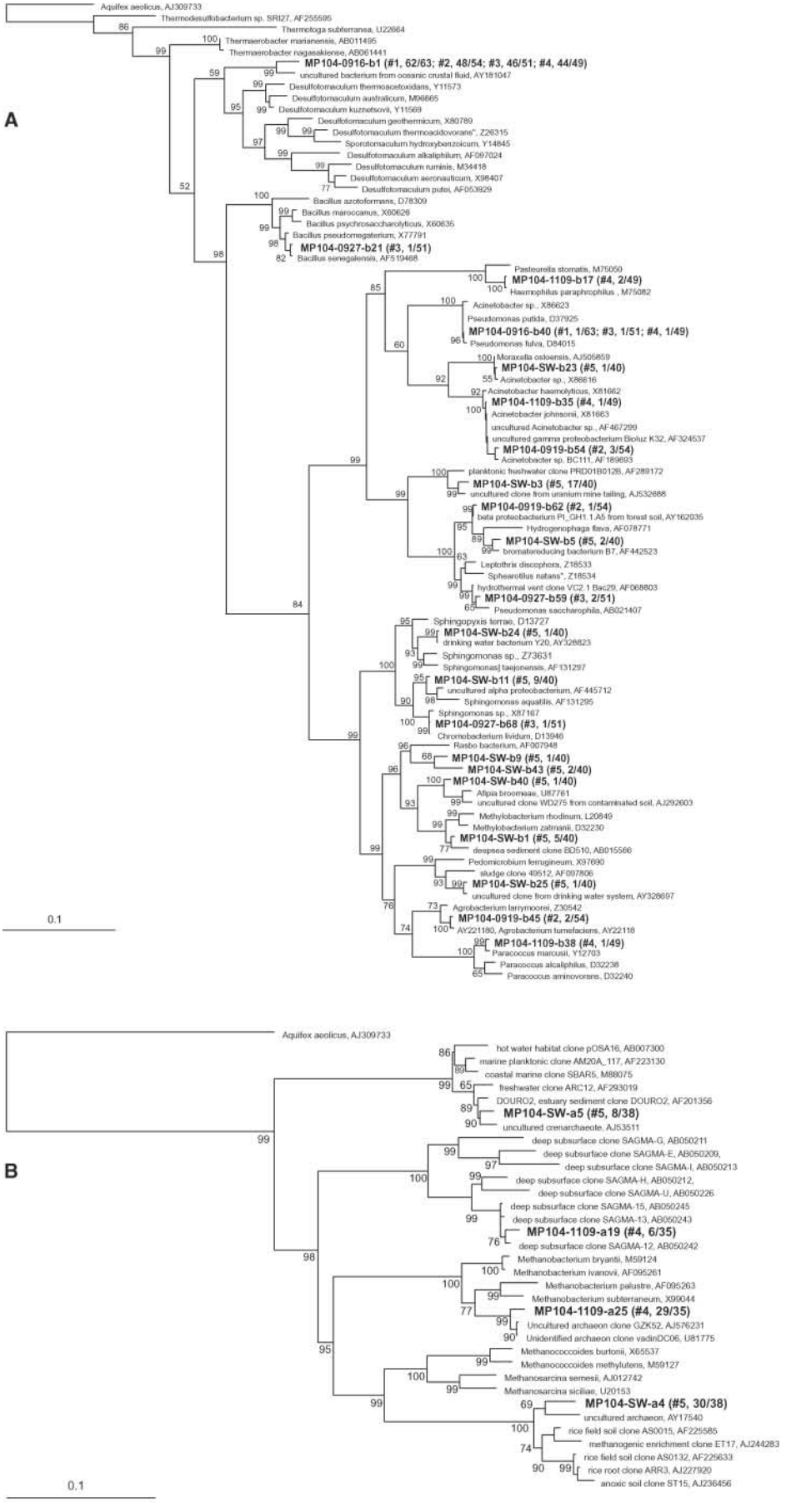


Fig. 1. Phylogenetic trees for bacteria (**A**) and archaea (**B**) based on 16*S* rDNA sequences recovered from fracture and mining water in the Mponeng gold mine, South Africa. The fraction number in parentheses represents the number of clones for each phylotype versus the total number of clones analyzed in each sample. The scale bar is equivalent to the sequence variation of 10%.

>88% of the clones in the 16*S* rDNA libraries we generated but did not appear in the water used for mining (Fig. 1A). This phylotype was related (96% similarity) to an uncultured clone recovered from thermal fluids of oceanic crust (*11*) or to *Desulfotomaculum kuznetsovii* (91% similarity) (*12*), a sulfate reducer growing at moderately thermophilic conditions. Of the other minor bacterial and archaeal phylotypes associated with the fracture water, MP104-1109-a19 resembled (98% similarity) environmental clones (the SAGMEG-2 group) obtained from fracture water in an adjacent mine (*13*) (Fig. 1B). Other phylotypes were closely related (98 to 99% similarity) to environmental clones recovered from surface environments (such as soils and sludge) or to various mesophiles distributed within *Proteobacteria*, *Firmicutes*, and *Methanobacteria* (Fig. 1). These minor phylotypes did not resemble those from the mining water [see supporting online material (SOM)].

High-density 16*S* rDNA microarray analysis was also used as a more sensitive approach to assess microbial diversity. Array hybridization results supported the finding that microbial diversity was much less (table S2) and the *Firmicutes* were of greater relative abundance (table S3) in the fracture water than in the mining water. Some overlap between the sequences found in the fracture water and those in the mining water is expected because mining water is a mixture of recycled fracture water and surface water.

The geochemistry of the fracture water was characterized to identify its origin and to assess the principal metabolic pathways (Table 1 and tables S1 and S4). Over the observation period, the temperature decreased from >60° to 52°C, while the pH was constant and the redox potential (E_h) increased from –330 to –263 mV. The Cl^-, Br^-, and SO_4^{2-} concentrations increased significantly, whereas other anion concentrations fluctuated or decreased slightly. Reduced gases (H_2 and CH_4) were abundant (>1 mM), and their concentrations increased as water flow rates declined and gas flow rates remained constant. Formate and acetate concentrations ranged from 7 to 9 μM and 22 to 36 μM, respectively. The δ^2H and $\delta^{13}C$ of the abundant dissolved C_{1-4} hydrocarbons indicated an abiogenic origin, based on their similarity to abiogenic hydrocarbon gases identified at other Precambrian Shield sites (*9*). The δ^2H and $\delta^{18}O$ values of the fracture water, –29.1 per mil (‰) and –6.9‰ VSMOW (Vienna standard mean ocean water), respectively, plotted slightly above the global and local meteoric line (*14*). The δ^2H-H_2 values ranged between –680 and –690‰ VSMOW, which when compared with the δ^2H-H_2O value yielded isotopic equilibrium temperatures of 62°C for the first sample, declining to 49°C for the last sample. Based on the thermal conductivity model (*15*) and the heat flow data (*16*), the initial temperature equates to that at 4.2 kmbls and the final temperature to that at 2.9 kmbls.

Noble gas analyses yielded elevated radiogenic and fissiogenic concentrations of ^4He, ^{40}Ar, ^{134}Xe, and ^{136}Xe (table S1) but exhibited no systematic trend during the observation period. The bulk model age of the fracture water ranged from 15.8 ± 7.9 million years to 25.0 ± 3.8 million years (Table 1). This model age represents either the true subsurface residence time or the mixing between a hydrothermal fluid that is ancient (0.8 to 2.5 billion years old); saline; and H_2-, C_{1-4} hydrocarbon–, and SO_4^{2-}-rich; and paleometeoric water that is ~3 to 4 million years old; moderately saline; and H_2-, C_{1-4} hydrocarbon– and SO_4^{2-}-poor (SOM).

The exceptional dominance of the *Firmicutes* clones; δ^2H and $\delta^{18}O$ values of the fracture water that are distinct from those of the mining water (*17*); and increasing H_2, CH_4, He, Cl^-, Br^-, and SO_4^{2-} concentrations during the observation period suggest that the fracture environment favors the survival of these *Firmicutes*-related microorganisms and that mining operations had minimal impact on this environment during depressurization and dewatering. *Desulfotomaculum*-related environmental clones and isolates have been detected in other subsurface environments, including an oil reservoir (*18*), deep sedimentary strata (*19*, *20*), an ore deposit (*21*), and saline water emanating from a 3.2-kmbls borehole located 13 km east of this borehole (*22*). The evident success of these microorganisms may not be surprising because their ability to form endospores would facilitate their survival during periods of low water activity, nutrient deprivation, and suboptimal temperature (*23*). Whether the less abundant members of the clone libraries represent indigenous microorganisms that are capable of acquiring energy through metabolisms that are distinct from those of the strains they most resemble phylogenetically, or alternatively represent moribund relics of shallower, less saline paleometeoric water that has mixed with this fracture water, is not known.

Sulfur isotopic analyses yielded $\delta^{34}S$ values of HS^- ranging from 11 to 13‰ Vienna Canon Diablo meteorite (VCDT) and those of SO_4^{2-} from 19 to 26‰ VCDT (table S4). The depleted $\delta^{34}S$ value of HS^- relative to that of SO_4^{2-} (–7.7‰ for sample 4 and –12.4‰ for sample 7) is consistent with microbial sulfate reduction. The fact that SO_4^{2-} concentrations in the fracture water were higher than that of the overlying dolomitic aquifer (*13*), and that SO_4^{2-} has been reported in analyses of fluid inclusions from hydrothermal quartz in the VCR (*24*) and the reservoir mixing-fractionation model (SOM), all suggest that the SO_4^{2-} originated from the ancient hydrothermal fluid, not the paleometeoric end member. The $\delta^{34}S$ values of fracture pyrite (0 to 2‰ VCDT) were less than that of the coexisting barite (10.1‰ VCDT) (table S4) and were consistent with those of hydrothermal 2.0-billion-year-old pyrite from the VCR in the same mine (*25*). The $\Delta^{33}S$ values for all samples clustered around 0‰ (table S4). To reproduce the $\Delta^{33}S$ of total dissolved S species (–0.022‰) for sample 7, a mixing of SO_4^{2-} derived from

Table 2. Gibbs free energy (ΔG), substrate consumption rate and steady-state free-energy flux for various microbial redox reactions.

Sample no.	1			3			4		
	ΔG (kJ mol^{-1})	Substrate rate* (μM year^{-1})	Energy flux* (kJ cell^{-1} s^{-1})	ΔG (kJ mol^{-1})	Substrate rate* (μM year^{-1})	Energy flux* (kJ cell^{-1} s^{-1})	ΔG (kJ mol^{-1})	Substrate rate* (μM year^{-1})	Energy flux* (kJ cell^{-1} s^{-1})
H_2-sulfate reduction†‡	–148	5.9×10^6	-9.2×10^{-13}	–155	8.6×10^6	-1.4×10^{-12}	–146	1.8×10^7	-2.7×10^{-12}
Acetate-sulfate reduction†‡	–86	1.9×10^5	-1.7×10^{-14}	–94	1.5×10^5	-1.5×10^{-14}	–82	2.4×10^5	-2.1×10^{-14}
H_2-methanogenesis‡	–94	2.1×10^5	-2.1×10^{-14}	–95	2.9×10^4	-2.9×10^{-15}	–93	3.8×10^5	-3.7×10^{-14}
Acetate-methanogenesis‡	–33	1.9×10^5	-6.6×10^{-15}	–34	1.5×10^5	-5.5×10^{-15}	–28	2.4×10^5	-7.1×10^{-15}
Formate-methanogenesis‡	–74	1.8×10^4	-1.4×10^{-15}	–86	1.5×10^4	-1.3×10^{-15}	–42	1.8×10^3	-8.1×10^{-17}
H_2-acetogenesis‡	–62	1.4×10^5	-9.3×10^{-15}	–61	2.0×10^4	-1.2×10^{-15}	–64	2.5×10^5	-1.7×10^{-14}

*The calculation for maximum substrate consumption rate and steady-state free-energy flux is shown in SOM. †The HS^- concentration used in the calculation for sulfate reduction for samples 1 and 3 were assumed to be 1.3 mM. The free energy for sulfate reduction only varied less than 5% when HS^- concentration was changed between 1.1 to 1.5 mM. ‡The reactions for free energy calculations were as follows: H_2-sulfate reduction: $4\ H_2 + H^+ + SO_4^{2-} \rightarrow HS^- + 4\ H_2O$; acetate-sulfate reduction: $CH_3COO^- + SO_4^{2-} \rightarrow 2\ HCO_3^- + HS^-$; H_2-methanogenesis: $4\ H_2 + H^+ + HCO_3^- \rightarrow CH_4 + 3\ H_2O$; acetate-methanogenesis: $CH_3COO^- + H_2O \rightarrow CH_4 + HCO_3^-$; formate-methanogenesis: $4\ HCOO^- + H^+ + H_2O \rightarrow CH_4 + 3\ HCO_3^-$; and H_2-acetogenesis: $4\ H_2 + H^+ + 2\ HCO_3^- \rightarrow CH_3COO^- + 4\ H_2O$. The uncertainties are ±10%.

pyrite oxidation and barite dissolution with a Rayleigh isotopic fractionation by microbial sulfate reduction was required (SOM). Such a mixture, composed of 30% SO_4^{2-} derived from the oxidation of pyrite by radiolytically produced oxidants (*10*) and 70% SO_4^{2-} derived from the dissolution of barite, produced an initial $\delta^{34}S$ value of 8.13‰ VCDT and a $\Delta^{33}S$ value of –0.0087‰ (SOM). The isotopic evolution from the initial 8.13‰ VCDT to the present observation (18.63‰ VCDT for total dissolved S species) requires a Rayleigh fractionation process in which 70% of the initial SO_4^{2-} was removed as microbially precipitated pyrite (SOM). By this model, the total HS^- formed by microbial sulfate reduction would be 4.35 mM [1.52 mM × 70%/30% + 0.80 mM HS^- (SOM)].

The Gibbs free energy for microbial redox reactions was calculated to provide additional constraints on the dominant respiratory pathway occurring in this fracture water. Sulfate reduction dominated the free-energy yields for a wide range of electron donor and acceptor combinations (Table 2). The free-energy yields for these reactions were much greater than the minimum requirement for synthesis of 1/3 of an adenosine triphosphate molecule by pumping one proton across the cell membrane (~ –20 kJ) (*26*). Among electron donors available for sulfate reduction and methanogenesis, H_2 yielded more free energy than acetate and formate, and H_2-utilizing sulfate reduction yielded the highest free energy and energy flux (Table 2), suggesting that CO_2-utilizing methanogens and acetogens cannot sustain as high a population density as sulfate reducers and therefore would be minor constituents of the community, as is observed in the clone libraries. The prominence of the free-energy flux for H_2-utilizing sulfate reduction over other metabolic reactions is consistent with the depletion of $\delta^{34}S$ values of HS^- relative to SO_4^{2-} (*27*) and the physiological characteristics inferred from the dominant phylotype.

The in situ rate of microbial sulfate reduction was estimated to range from 0.22 to 1.45 nM $year^{-1}$, or from 5.5×10^{-18} to 3.6×10^{-17} moles per cell $year^{-1}$ (assuming that all observed cells, 4×10^7 cells $liter^{-1}$, actively reduce SO_4^{2-}), based on the potential microbially produced sulfide [4.35 mM (SOM)] and fracture water residence time of ~3 or ~20 million years inferred from noble gas analysis. Such a rate is comparable to rates reported in subseafloor sediments (*28*) and to the estimated rate of radiolytic H_2 generation (*10*), but far less (4 to 5 orders of magnitude) than the maintenance energy demand of mesophilic sulfate reducers as determined in laboratory experiments (*29*) (SOM). The estimated in situ sulfate reduction rate, when combined with the experimental maintenance energy demand [48 mol (g dry weight cell)$^{-1}$ $year^{-1}$] (*29*) and assuming 20 fg per cell, would support only 200 to 1600 cells $liter^{-1}$, which is far below the observed $\sim 4 \times 10^7$ cells $liter^{-1}$. If cells were constantly growing and dying, then the experimental sulfate reducer yield of 12.2 g of dry weight cell mol^{-1} (*29*), when combined with the in situ sulfate reduction rate, would correspond to cell turnover times of 45 to 300 years. The isotopic estimate of the long-term in situ microbial activity, however, is 10^9 to 10^{10} times less than the maximum substrate consumption rate (Table 2), which would seem to indicate that as-yet-unidentified factors play a role in restraining the microbial respiration.

The hot, reducing, gaseous water emanating from a fracture at 2.8 to 4.2 kmbls harbored a microbial community dominated by a single *Firmicutes* phylotype. The *Firmicutes* probably penetrated the Mponeng fracture zone at current depths during infiltration of paleometeoric water between 3 and 25 million years ago and since then have relied on nonphotosynthetically derived H_2 and SO_4^{2-} converted from Archaean/Proterozoic pyrite/barite. Nutrient concentrations have remained much higher than observed in shallower crustal environments, suggesting that the deep crustal biosphere may be energy-rich, is not approaching entropic death, and is capable of sustaining microbial communities indefinitely by geological processes.

References and Notes

1. K. O. Stetter *et al.*, *Nature* **365**, 743 (1993).
2. D. M. Karl, in *The Microbiology of Deep-Sea Hydrothermal Vents*, D. M. Karl, Ed. (CRC Lewis, Boca Raton, FL, 1995), pp. 35–124.
3. E. M. Murphy *et al.*, *Water Res. Res.* **28**, 723 (1992).
4. S. D'Hondt *et al.*, *Science* **306**, 2216 (2004).
5. F. H. Chapelle *et al.*, *Nature* **415**, 312 (2002).
6. T. O. Stevens, J. P. McKinley, *Science* **270**, 450 (1995).
7. T. C. Onstott *et al.*, in *Enigmatic Microorganisms and Life in Extreme Environments*, J. Seckbach, Ed. (Kluwer Academic, Dordrecht, Netherlands, 1999), pp. 489–500.
8. J. Lippmann *et al.*, *Geochim. Cosmochim. Acta* **67**, 4597 (2003).
9. B. Sherwood Lollar *et al.*, *Chem. Geol.* **226**, 328 (2006).
10. L. H. Lin *et al.*, *Geochem. Geophys. Geosyst.* **6**, Q07003 10.1029/2004GC000907 (2005).
11. J. P. Cowen *et al.*, *Science* **299**, 120 (2003).
12. T. N. Nazina, A. E. Ivanova, L. P. Kanchaveli, E. P. Rozanova, *Mikrobiologiya* **57**, 823 (1988).
13. K. Takai, D. P. Moser, M. DeFlaun, T. C. Onstott, J. K. Fredrickson, *Appl. Environ. Microbiol.* **67**, 5750 (2001).
14. H. Craig, *Science* **133**, 1702 (1961).
15. G. Omar, T. C. Onstott, J. Hoek, *Geofluids* **3**, 69 (2003).
16. M. Q. W. Jones, *J. Geophys. Res.* **93**, 3243 (1988).
17. G. R. Drennan, M. C. Boiron, M. Cthelineau, L. J. Robb, *Mineral. Petrol.* **66**, 83 (1999).
18. C. Tardy-Jacquenod, M. Magot, B. K. Patel, R. Matheron, P. Caumette, *Int. J. Syst. Bacteriol.* **48**, 333 (1998).
19. H. Saas, H. Cypionka, *Syst. Appl. Mcirobiol.* **27**, 541 (2004).
20. Y. Liu *et al.*, *Int. J. Syst. Bacteriol.* **47**, 615 (1997).
21. T. Nakagawa *et al.*, *FEMS Microbiol. Ecol.* **41**, 199 (2002).
22. D. P. Moser *et al.*, *Gemicrobiol. J.* **20**, 517 (2003).
23. L. L. Campbell, R. Singleton, in *Bergey's Manual of Systematic Bacteriology*, P. H. A. Sneath, N. S. Mair, M. E. Sharpe, J. G. Holt, Eds. (Williams & Wilkins, Baltimore, MD, 1986), vol. 2, pp. 1200–1202.
24. H. E. Frimmel, D. K. Hallbauer, V. H. Gartz, *Mineral. Petrol.* **66**, 55 (1999).
25. B. Zhao, C. Harris, L. J. Jordaan, L. J. Robb, *Geosci. Afr. 2004* **2**, 732 (2004).
26. B. Schink, *Microbiol. Mol. Biol. Rev.* **61**, 262 (1997).
27. A. L. W. Kemp, H. G. Thode, *Geochim. Cosmochim. Acta* **32**, 71 (1968).
28. R. J. Parkes *et al.*, *Nature* **436**, 390 (2005).
29. W. Badziong, R. K. Thauer, *Arch. Microbiol.* **117**, 209 (1978).

30. This work was supported by grants from NSF Life in Extreme Environments and NASA to T.C.O.; the National Science Council of Taiwan to L.-H.L. and P.-L.W.; the Jet Propulsion Laboratory, NASA Cosmochemistry, and NSF Earth Science to D.R.; the Deutsche Forschungsgemeinschaft (German Research Foundation) to J.L.-P.; the NASA Indiana-Princeton-Tennessee Astrobiology Initiative to L.P.; the National Science and Engineering Research Council of Canada and Killam Fellowship funds to B.S.L.; and the U.S. Department of Energy to T.C.H. We thank J. Hall, G. Southam, G. Wagner, B. Tipple, T. Gihring, E. Van Heerden, and D. Litthauer of the Witwatersrand Microbiology Project, R. Wilson of SRK-Turgis Technology, and geologists of the Mponeng mine for their assistance in the setup of the field laboratory, field sampling, and the coordination of logistic supply. We also acknowledge D. Chivian for discussion of DNA data and S. Ono for assisting with sulfur isotope analysis. GenBank accession numbers for sequences are from DQ088779 to DQ088817.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5798/479/DC1
SOM Text
Figs. S1 and S2
Tables S1 to S4
References

13 March 2006; accepted 21 September 2006
10.1126/science.1127376

2008 R&D 100 AWARDS

Berkeley Lab PhyloChip

Appendix C

Selected Press Coverage

"Microbiology's Air Force," *Science*, December 20, 2006.
http://sciencenow.sciencemag.org/cgi/content/full/2006/1220/1

"Microbe Census Reveals Air Crawling with Bacteria," *Scientific American*, December 19, 2006.
http://www.sciam.com/article.cfm?chanID=sa003&articleID=9B426131-E7F2-99DF-31803CBBE807C66A

"Surprise in Airborne Bacteria Census," *United Press International*, December 19, 2006.
http://www.upi.com/NewsTrack/Science/2006/12/19/surprise_in_airborne_bacteria_census/8350/

"A Nationwide Census of Airborne Bacteria," *Environmental Science and Technology News*, August 18, 2004.
http://pubs.acs.org/subscribe/journals/esthag-w/2004/aug/tech/pt_nationwide.html

"Berkeley Lab Technology Could Help Areas Flooded by Katrina and Rita," *Berkeley Lab Research News*, September 27, 2005.
http://www.lbl.gov/Science-Articles/Archive/ESD-Katrina-phylochip.html

"Researchers Find Pathogens in Unlikely Places Using High-Density Microarrays," *Microarray Bulletin*, AMB Community, January 18, 2006.
www.microarraybulletin.com/community/article.php?p=71

Also, to hear a PBS interview with Todd DeSantis on studies of airborne bacteria, (aired by KCPW Salt Lake City-Park City, December 27, 2006), go to:
http://greengenes.lbl.gov/Download/Podcasts/PhyloChip_Aerosols_DeSantis_KCPW_06-12-27.mp3

Next Article

Article Tools

- Add to Personal Folders
- E-Mail This Page
- Print This Page
- Share this article:
  - Del.icio.us
  - Digg it!
  - Submit to Slashdot

Related Content
Search Google Scholar for:

- Articles by Bohannon, J.

Search PubMed for:

- Articles by Bohannon, J.

Search *Science*NOW
Keyword Search Enter Keyword

GO ›

- Advanced Search ›
- *Science* Magazine ›

My Science

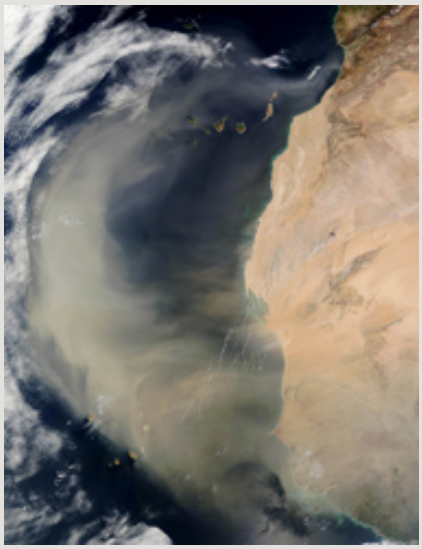
- My Folders
- My Alerts
- My Saved Searches
- Sign In

About Access
*Science*NOW articles are free for four weeks after publication. Older articles from the archive may require AAAS membership or separate subscription.

Alerts & Feeds

- Email Alerts
- RSS Feeds

Science Signaling
New on *Science Signaling*



Microbes galore.
This African sandstorm will deposit millions of tons of dust over the Caribbean and Florida. A study of airborne microbes indicates that these storms may be delivering more than just hazy skies.

Credit: NASA

Microbiology's Air Force

By John Bohannon
*Science*NOW Daily News
20 December 2006

Air is filthy with microbes, at least in the city. That's the conclusion from a genetic study of airborne bacteria in two U.S. cities that suggests the atmosphere may be a more important part of global microbial ecology than was assumed. It also provides the first baseline for monitoring the air for bioterrorist attacks.

Because airborne bacteria are exposed to extremes of temperature, dryness, and radiation, they are thought to be far less diverse and abundant than their peers down on the ground. It has been difficult to test that idea. The traditional method of surveying microbial populations--growing colonies in the lab--doesn't work well for studying the atmosphere's flora. Not only are airborne populations relatively sparse, but the cells are often also in a quiescent state, requiring special conditions to start growing. Finding better methods would be useful, not only for studying basic microbial ecology but also for improving a 3-year-old U.S. government effort called BioWatch, which aims to put sniffing machines in hundreds of public places to sound the alarm in case of a bioterrorism attack.

To get a better fix on flying microbes, a team led by Gary Andersen, a microbial ecologist at Lawrence Berkeley National Laboratory in Berkeley, California, used a recently developed microbe-detector called the Phylochip. Rather than relying on microbes to grow into visible colonies, the device detects individual cells by grabbing their ribosomal RNA (rRNA) genes out of a solution of broken-up cells. Fluorescent dye then gloms onto the microbial rRNA. To see who's who, the rRNA is then washed over a target array of more than 8000 different matching rRNA strands that represent the 121 taxonomic orders of prokaryotic microbes. A computer records positive hits by checking for bright spots on the array.

- Novel Roles for the NF-κB in Regulating Neuronal Function
- Exosomes From Infected Macrophages Are Proinflammatory
- More from *Science Signaling*

For their atmospheric samples, the researchers used three sites each in San Antonio and Austin, Texas, both part of the BioWatch monitoring network. The team pumped air samples through fine-meshed filters at the rate of 10 liters per hour and collected the filters weekly for 17 weeks in the summer of 2003.

A surprising discovery was that several nonpathogenic microbes closely related to bioterror weapons of choice, including anthrax and tularemia, were common in city air, which may explain some of BioWatch's false alarms in recent years. Another surprise was the sheer diversity of the airborne microbes, which included 1800 types of bacteria, as diverse an assemblage as typically found in soil, the team reports online today in *Proceedings of the National Academies of Science*. The diversity and concentration varied dramatically from week to week, with much of the variation explained by weather conditions.

The bottom line? "We humans are embedded in a microbial world that we barely acknowledge, and this study is one that is beginning to shed light on what kinds of organisms are out there in the environment," says microbiologist Norman Pace of the University of Colorado at Boulder. "The only problem I see is that the numbers of samples and locales are very limited. We need a lot more of this in a lot of environments."

Related sites

- The BioWatch program
- More about the Phylochip

You have reached the bottom of the page. Back to top

News - December 19, 2006

Microbe Census Reveals Air Crawling with Bacteria

Thousands of microbes that change with the weather drift in the atmosphere

By David Biello

AIRBORNE MICROBES: Despite the difficulties of living in thin air, more than 1,800 microbes showed up in a recent census of the atmosphere in Austin and San Antonio, Tex.

The air is a hostile place for a microbe. Often dry, lacking in nutrients and filled with deadly ultraviolet radiation, the atmosphere would seem to be the last place a microbe would want to find itself. Yet, a new genetic census of some air samples from Austin and San Antonio, Tex., finds that as many as 2,000 different kinds of microbes may be present in the air we breathe on any given day.

Microbial ecologist Gary Andersen of Lawrence Berkeley National Laboratory and his colleagues collected air samples in the two Texas cities over a period of 17 weeks, starting in 2003. They then used a specially designed microarray--a small chip roughly the size of a quarter that carries probes to detect specific genetic information--to search for a gene involved in the making of a protein (16S) that is found in many microbes. "We designed a 500,000-probe array to identify up to about 9,000 different groups of bacterial and archaean organisms," Andersen explains. "It looks at the differences in the 16S sequence to identify a specific type of prokaryotic organism."

In the air samples, the researchers uncovered at least 1,800 different types of microbes, including those such as the diarrhea-causing *Arcobacter* and ulcer-inducing *Heliobacter* genera that can be dangerous to human health. Previous efforts to determine microbe counts in the atmosphere had relied on culturing the air to see what grew. "Over 90 percent you can't recover even though it was not only present but viable," Andersen notes. "It's just something about the physiological state it gets in; when it's not in rich media, it has a different physiology."

This puts the diversity of microbes in the air on par with the diversity of microbes in the soil, a fertile environment for such life-forms. In fact, there is a large crossover between the microbes in the air of a city and the microbes in its soil. The ecologists found that airborne microbes were broadly the same in Austin and San Antonio as well, and varied more depending on the weather than any other factor.

The most common microbes included those that thrive in hay fields and deteriorating exterior paint, according to the paper published online December 18 in *Proceedings of the National Academy of Sciences USA*. "We were surprised at how many different types of sequences we were seeing," Andersen adds. "Obscure phyla, like TM7, which have been seen in soil and gum tissue; hot springs type organisms; and microbes from sewage treatment plants."

The census provides a background for ongoing efforts by the U.S. Department of Homeland Security to monitor city air for potential bioterrorism attacks as well as fills in a gaping hole in the understanding of where microbes persist. For example, storms of Saharan dust--and their attendant microbes--have been linked to local meningitis outbreaks and tracked across the Pacific to coral deaths and an increase in childhood asthma in the Caribbean. "It's important to do a microbial census to see what's in the air we breathe. I believe it's going to change as the climate changes," Andersen says. "We may see very different populations of microbes in the air and that may have some health implications."

The air is now being tested in at least 30 U.S. cities and Andersen hopes the effort will be broadened. "How is it in the middle of the oceans?" he asks. "Is it just what's coming up from the sea's surface or is it a long-range dispersal of organisms?" Whatever the case, fresh, clean, bacteria-free air is rarer than previously thought.

FURTHER READING

Shrinking to Enormity
Coral-Killing Dust
Anthrax Genome Spills Its Deadly Secrets

Science

RSS Feed

Surprise in airborne bacteria census

Published: Dec. 19, 2006 at 4:22 PM

Print story · Email to a friend · Font size: A A

BERKELEY, Calif., Dec. 19 (UPI) -- U.S. scientists, conducting a first-of-its-kind census, found Texas air teeming with more than 1,800 types of bacteria.

The researchers from the U.S. Department of Energy's Lawrence Berkeley National Laboratory used a DNA test to catalog bacteria in air samples taken from San Antonio and Austin, Texas.

They said some 1,800 types of bacteria showed up in the samples.

They said the widely varied bacterial population rivaled the diversity found in soil. They also found naturally occurring relatives of microbes that could be used in bioterrorist attacks.

"Before this study, no one had a sense of the diversity of the microbes in the air," said lead author Gary Andersen, who noted the findings pave the way for regional bacterial censuses that will help a Department of Homeland Security bioterrorism surveillance program differentiate between normal and suspicious fluctuations in airborne pathogens.

The study will also help scientists establish a baseline of airborne microbes, which they can use to track how climate change affects bacterial populations.

The research appears in the online early edition of the Proceedings of the National Academy of Sciences.

Print story · Order reprints · Email to a friend · Post comment

Share this story: · Digg · Del.icio.us

» Next in Science: **Scientific DNA test validated for accuracy**

Astr
labo
spac

Latest H
Ne
Co
Ca
Ru
Do
Ch
Ne
Bil

Research Section
A-Page Section
Meetings Calendar
Links
to environmental & funding sites.
Online News
- Policy News
- Science News
- Technology News
- Business & Education News

About *ES&T*
- How to Subscribe
- About *ES&T*
- Masthead
- Editors (pdf)
- Magazine Staff
- Sample Issue (Research pages)

For Advertisers
- Media Information
- Ad Rates - Print
- Ad Rates - Web

For Help
- Editorial Office
- Technical Support
- Contact Us

Site Map

Technology News - August 18, 2004

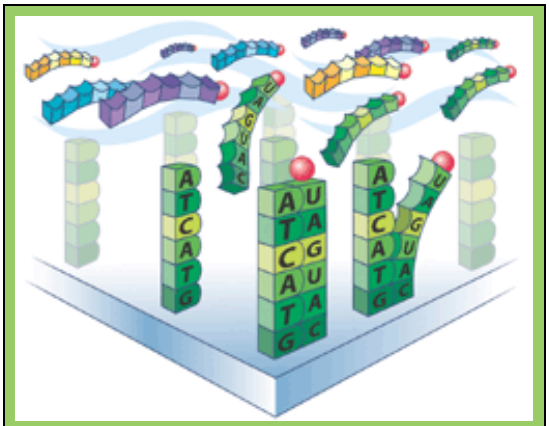
A nationwide census of airborne bacteria

Scientists at Lawrence Berkeley National Labs (LBNL) are using microarrays to compile a national database of airborne bacteria, which could be a boon for environmental researchers. Immediately after the September 11 attacks, American security officials began planning for a possible chemical or biological attack, and the Department of Homeland Security set up detection systems in major cities. But the detectors have to be calibrated to ignore background noise or, in the case of a biological attack, ambient airborne bacteria.

“The problem is that we really had to create the technology to do this,” says Gary Andersen of the Center for Environmental Biotechnology at LBNL in Berkeley, Ca. “Nobody else has ever done this before.” Created with the aid of gene-chip company Affymetrix, the current generation microarray chip now holds 500,000 probes for over 9000 taxa.

“You end up with a table of what’s present and their relative abundance,” says LBNL research associate Todd DeSantis. “So you can see from the sample when a taxa is declining or if it’s on the rise.” This is the first work that shows microarrays can detect not just the presence but also the quantity of taxa. The findings will be published this fall in the journal *Environmental Microbiology*.

The microarray chips uses 25 DNA-base long probes to bind selected sections of 16s ribosomal DNA. The ribosomal DNA is probed because it has sufficient variability to serve as a marker of biological taxa. DeSantis says designing the first chip for broad environmental sampling was difficult, because microarrays were developed for gene- expression work. “We initially wanted to handle this with off-the-shelf technology, using microarray techniques for expression analysis,” he says. “We found that this just wasn’t working.”



While the above drawing shows a microarray with binding sites that are only a six bases in length, the airborne bacteria study uses binding sites that are 25 bases long. This is the first microarray to broadly sample for different taxa.

After almost two years of work, the lab has already evaluated thousands of samples and is now awash in data. The researchers now have to figure out what to do with it all. DeSantis says they have just enlisted the help of a statistician and are building a website so that scientists can ask questions with different parameters. For instance, a wildlife biologist in Louisiana studying avian disease might want to identify which bacteria are present in the local environment, while an epidemiologist at a hospital in Minnesota might want to know the seasonal fluctuations of bacteria in a nearby city.

“We’re just squirreling away a lot of data and trying to put together an ftp site so that people can select cities and download information,” he says.

The LBNL researchers recently analyzed data collected during the summer of 2003 from two

cities in Texas. Separated by only 80 miles, Austin and San Antonio show some significant differences in their bacterial biota. Both cities have airborne bacteria from the genera *Pirellula*, *Pseudomonas*, *Rickettsia*, *Clostridium*, and *Bacillus*. But San Antonio has a much greater diversity of bacteria and also supports *Denitrovibrio*, *Legionella*, and *Actinomyces*. It is not known why two cities so close geographically differ so dramatically in their bacteria. “There must be some local reservoirs for the bacteria,” speculates Andersen. “We’re looking at weather patterns to see if that might explain it.”

Although the sampling phase of the project will end in the summer of 2005, analysis of the data could take years. In the future, the research might be expanded, according to Pete Pesenti, a project officer with the Science and Technology Directorate at the Department of Homeland Security. “We’ll take a look at the project next year and then decide if we might need to look at other things in the air like viruses and fungi,” he says. —PAUL D. THACKER

Return to Top | Technology News Home | *ES&T* Home

arrow up Return to Top

RESEARCH NEWS
BERKELEY LAB

September 27, 2005 news releases | receive our news releases by email | science@berkeley lab

Berkeley Lab Technology Could Help Areas Flooded by Katrina and Rita

DNA 'Phylochip' Scans for Thousands of Disease-Causing Microbes

Contact: Lynn Yarris (510) 486-5375, lcyarris@lbl.gov

BERKELEY, CA – The flood waters that filled the streets of New Orleans in the aftermath of Hurricanes Katrina and Rita are now heavily populated with bacteria, viruses and other disease-causing microbes that could pose a grave risk to residents long after the waters have been cleared. An experimental technology developed by scientists with the U.S. Department of Energy's Lawrence Berkeley National Laboratory (Berkeley Lab) holds promise for use in situations like this to provide a comprehensive picture of bacterial presence in the air, soil and water and enable authorities to track how that presence changes over time.

The floods that filled the streets of New Orleans and elsewhere in the aftermath of Hurricanes Katrina and Rita pose long-term health risks to residents, even after the waters have been cleared. (Photo by Kathy Anderson, LOLA)

"Warm and stagnant bodies of water frequently give rise to large numbers of different bacterial pathogens and opportunistic organisms," said Terry Hazen, a microbiologist who heads the Ecology Department of Berkeley Lab's Earth Sciences Division and is a leading authority on pathogens in warm environments. "A key challenge is a comprehensive identification of all potentially deleterious organisms without prior knowledge of the diversity of pathogens that may be present."

Originally developed for air sampling and still being tested by Berkeley Lab for the Department of Homeland Security, the Phylochip and its application to other media like those found on the Gulf Coast are pending further studies in soil and water.

Using a square-shaped DNA chip, the size of a quarter and packed with an array of 500,000 probes, Hazen and his colleagues can analyze a sample for the unique DNA signatures of 9,000 known species in the Phyla of bacteria and Archaea. The analysis can be completed within 24 hours from the start of the process.

"Our Phylochip makes its identification based on variations in the 16S rRNA gene, which is essential for protein synthesis and is present in all bacteria," said Gary Andersen, a microbiologist who now heads the Molecular Microbial Ecology Group within the Lab's Ecology Department. "Small regions of DNA base-pair sequence differences within the 16S rRNA gene can be used to distinguish different bacterial species, including pathogens. The advantage of this approach is that multiple pathogens can be identified simultaneously by targeting unique regions of the 16S rRNA gene sequence in the samples."

The bacterial problems facing the coastal areas of Louisiana, Mississippi and Alabama, which were ravaged by Hurricane Katrina, and areas in Texas and Louisiana impacted by Hurricane Rita, are likely to persist for an extended period of time, Hazen said. As flooded areas dry out, some of the pathogens in the contaminated water will become airborne. Others will be absorbed or adsorbed into the soil. Some of these will percolate down into the groundwater, and others will remain near the soil surface where they could be desorbed through precipitation.

“Every time it rains, we could be seeing a spike in bacterial pathogen activity,” said Hazen. “The microbial population currently in that flood environment is complex and dynamic.”

With the Phylochip and special software developed at Berkeley Lab, health and environmental officials monitoring the areas flooded by Katrina and Rita could be provided with what Hazen characterizes as “snapshots” over a period of time. These snapshots would tell officials whether specific populations of bacterial organisms are increasing or decreasing. Samples collected from critical locations would have to be transported to Berkeley Lab for DNA extraction, polymerase chain reaction (PCR) amplification, and array hybridization on the Phylochip.

The key to the Phylochip’s success is the 16S rRNA gene, which codes for a small substructure of the organelle in biological cells known as a ribosome. For bacteria, the DNA sequences that make up the 16s rRNA gene have been highly conserved throughout evolution, and can be used to positively identify individual species within the Phyla. These DNA sequences can also be used to infer natural relationships between different species.

In addition to bacterial pathogens, the 16s Phylochip would also reveal the presence in samples of bacterial species that are capable of degrading or transforming contaminants of concern. Under the right conditions, bioremediation and natural attenuation can be the safest and most cost-effective environmental remediation technique.

“We can validate the possibility of natural biodegradation and determine which species of biodegraders should be stimulated,” said Hazen. “With this information, appropriate treatability studies can be designed for determining the best and most cost-effective remediation strategy.”

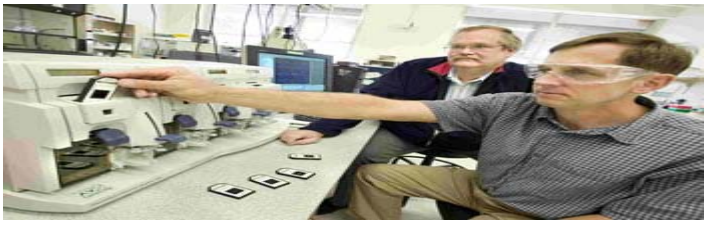
The 16s Phylochip has been extensively tested under a variety of conditions and locales and has proven to be a more sensitive and accurate means of detecting bacterial species, including pathogens, than the current standard of using 16S rRNA gene clone library sequencing.

Hazen and the other scientists in the Ecology Department are experienced at tracking bacterial pathogens in tropical and subtropical waters, which pose a special challenge to health and environment officials. The standard indicators used to assess whether water has been contaminated with human feces, the primary source of pathogens that threaten human health, are the bacteria *E. coli* and *Enterococcus*. However, according to Hazen, these standard indicators can survive indefinitely in waters where the temperature approaches that of the human body, which means they do not necessarily reveal whether the human fecal contamination was recent, a critical point for assessing the risk of disease.

“Our research has demonstrated that other bacteria, such as Enterococci, Bifidobacteria and Clostridia, provide better indirect evidence of recent human fecal contamination,” Hazen said. “We also have experience with direct detection of pathogens like *Vibrio vulnificus* and *Vibrio parahaemolyitcus*, common in warm coastal water and already causing infections, some fatal, in Katrina-affected areas.”

Hazen and his colleagues have developed a number of techniques for detecting both alternative indicators and opportunistic pathogens. These techniques can be used to identify contamination sources and help health and environment officials better determine the real risk involved with

contact and consumption of various water sources. In addition, they have shown how other technologies can also be brought to bear on the situation. For example, they have learned to use a micro-respirometer — a device designed to monitor real-time oxygen consumption, carbon dioxide or methane production — to determine within a few hours what would be the optimum conditions for achieving biodegradation of contaminants.



Terry Hazen (left) and Gary Andersen, microbiologists with Berkeley Lab, are shown here with the 16s Phylochip they developed that can analyze a sample for the unique DNA signatures of all known species of bacteria. (Photo by Roy Kaltschmidt, CSO)

“A combination of the technologies we are developing could provide functional design criteria for the remediation of contaminated soil and water at a number of sites in the hurricane-affected areas,” Hazen said.

Hazen and the scientists in his department are now working on new version of the Phylochip that could be used to detect pathogenic viruses and fungi. They are also developing chips that can be used to test for specific pathogens, such as *Yersinia pestis*, the bacterium that causes the plague. This work was principally funded by the Biological and Environmental Research program of the Department of Energy’s Office of Science.

Berkeley Lab is a U.S. Department of Energy national laboratory located in Berkeley, California. It conducts unclassified scientific research and is managed by the University of California. Visit our Website at www.lbl.gov.

Additional Information

- For further information about all the technologies described in this release, contact Terry Hazen by phone at (510) 486-6223 and by e-mail at TCHazen@lbl.gov, or Gary Andersen at (510)495-2795, GLAndersen@lbl.gov
- For more information about Berkeley Lab’s Earth Sciences Division, visit the Website at http://www-esd.lbl.gov/index.html
- For more information about Berkeley Lab’s Ecology Department, visit the Website at http://www-esd.lbl.gov/ECO/index.html

Top

Researchers Find Pathogens in Unlikely Places Using High-density Microarrays

Lawrence Berkeley's Gary Andersen and David Rasko of the University of Texas Southwestern Medical Center discuss the surprising organisms found using microarrays to detect pathogens in the environment.

By Stacey Ryder

Berkeley, January 18, 2006 — Scientists at Lawrence Berkeley National Laboratory are using two new microarrays to detect dangerous pathogens in the air, soil,

food and water, before they spread to human populations. In just nine hours, the group's 16S rRNA gene (16S phylogenetic) array identifies up to 8,900 distinctive organisms in a single experiment, while their multiple pathogen ID array (MPID) tests for more than 140 genetic regions that make bioterror agents particularly dangerous.

The team leader, Dr. Gary Andersen, said that his group typically discovers about twice the number of organisms found by conventional methods of cloning and sequencing. Using the array, the group is able to increase sampling size and can more accurately determine the distribution and complexity of bacterial species in the environment.

Andersen's research group has already used an earlier, prototype version of the phylogenetic array to identify more than 14 orders of prokaryotes andthree phyla of eukaryotes in British air samples. The array detects the organisms with probes complementary to variable regions of a universally conserved gene sequence—the small ribosomal subunit.

In a separate proof of concept study, Andersen's group used their MPID array to detect Category A biothreat agents responsible for anthrax, plague and tularemia, in complex environmental samples. The group designed the array to identify the presence of virulence and pathogenicity genes from 18 bioterrorist agents. They are now working on a multi-year project funded by the U.S. Department of Homeland Security to develop a comprehensive bacterial detection array that can simultaneously detect multiple organisms in complex environmental samples. Andersen's laboratory is currently examining the microbial composition of aero-sols obtained from pathogen detection systems to identify which microorganisms, including pathogens, are normally present in the environment.

"I guess what blew our socks off in the beginning were the pathogens we would see in, for instance, the air," said Andersen. "We'd often see things like *Clostridium botulinum*. That pathogen is anaerobic, so we were a little surprised."

Andersen recently spoke to David Rasko, a senior research scientist at the University of Texas Southwestern Medical Center at Dallas, about the advantages of using microarrays to look at microbial ecology and some of the surprising findings his team has made.

Rasko, who is also an adjunct collaborative investigator at The Institute for Genomic Research (TIGR), studies the genomic evolution of closely related pathogenic species. The two discussed:

- the sensitivity of microarray studies compared to plating and surveying studies of microbial ecosystems
- the role of 16S rRNA copy number in distinguishing microbial species from one another
- the potential for tracking the path of infectious diseases using microarray technology

Microarrays Vs. Plating and Surveying

Rasko: We've known for quite some time that using traditional plating and survey methods, you miss many species and see underrepresented diversity. Your chip gives a better idea of that diversity. What percentage do you think is still unresolved in those environments that you've tested?

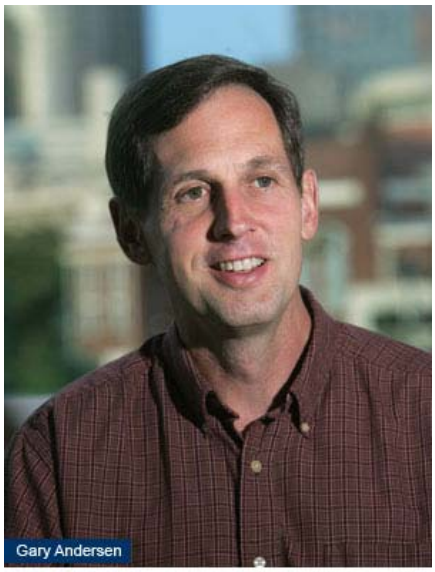
Andersen: Well, it's hard to say. Less than 1% of bacteria can be seen by culture. The route we've chosen is identification of bacteria by conserved genes, such as the 16S small subunit ribosomal gene. With this method, we can find greater diversity and species richness in the natural environment.

This has been going on for a number of years and it's allowed for the discovery of completely novel organisms that have never been observed in culture. But the number of species in different environments depends on your definition of a species.

For microbial ecologists like myself, the definition is 97% identity at the 16S gene level, or 70% DNA relatedness at the DNA/DNA hybridization level to be called a species. For organisms with important phenotypes, such as causing disease or catalyzing important reactions, we tend to use a much higher percent identity to call a species. As we sample new environments, we uncover new species. But as we sample many related environments, we are seeing less and less novel organisms being identified.

I think where the 16S phylochip helps is in determining the distribution and complexity of bacterial species in the environment. In our arrays we typically identify about twice the number of organisms found by the cloning and sequencing. Most of this is due to sampling size. Cloning is expensive and time consuming. Arrays are less expensive so many sampling points can be identified.

We typically use anywhere from 400 to 800 16S rRNA clones to compare with an array hybridization. Even though that's a large number, we're still missing a lot. If we increased the number, we'd probably see more, approaching the number we'd see on the array.a greater degree of sensitivity. Taking advantage of the sequence



Click for Bio

information from multiple laboratories helps us identify the most specific probes to identify a particular grouping of organisms we call a taxon.

The other is that we use the perfect match/mismatch probe pair to reduce the effects of cross-hybridization. Nonspecific hybridization has to be taken into account.

You have many very similar 16S molecules and you're trying to find the differences between them.

Rasko: Do you think the lack of identification using traditional cloning and plating type methods is based on low abundance of some members of the diverse sample?

Andersen: We noticed that it especially has to do with the way clone libraries are sequenced. Particularly when there's a high abundance member organism present, it seems to dominate clone libraries and crowds out the low abundance organisms.

One difference with the array is that instead of just pulling out individual 16S amplified molecules, we're putting the entire mixture on the array, allowing us to see the less abundant organisms.

We've split samples to compare our microarrays with the clone libraries, so we know we're not introducing any amplification bias when we compare microarray results with clone library results. But some probes just hybridize stronger than others. So, we've done Latin square experiments.

We take 16S targets of known organisms. In a rotating mixture we have different defined concentrations and we measure the hybridization intensity on the array. For an individual target, we see that as we increase or decrease the concentration, we increase or decrease the hybridization intensity. But if you compare several different targets at one concentration, you'll find that there will be quite a bit of difference in their sequence hybridization as measured by average difference hybridization score.

One thing that tells us is that we can't use our array for measuring absolute relative abundance of an organism unless we've already calibrated it to a known concentration. The strength of what we do, though, is we can measure the relative increase or decrease in popula-

tion size as measured by hybridization intensity over time or treatment.

The Role of 16S rRNA

Rasko: Different species have different numbers of 16S genes. Does the copy number of the 16S rRNA affect your hybridization or do you not see a huge bias based on the genomic copy number of this gene?

Click for Bio

Andersen: There's actually a significant advantage to the different copy numbers. We get an estimate of metabolic activity of the organisms in the environment by knowing how much rRNA, used for protein synthesis, is present. In addition to using conserved primers for a 16S gene amplification which will tell us which organisms are present, we can directly measure rRNA in a sample without the need for amplification to infer which organisms are most actively expressing proteins.

The typical active bacterium contains about 20,000 copies of rRNA. As a rule, organisms with more copies of the gene or with a higher metabolic rate make more rRNA molecules. Those with fewer copies of the gene or with a lower metabolic rate make fewer copies. There's no easier way of getting this information than by measuring the rRNA hybridization directly on our array.

The breakthrough that made this possible was the development of a protocol by Affymetrix to directly label RNA molecules without cDNA synthesis. The protocol was developed for mRNA, but we noticed it actually works even better for our rRNA.

We've used this method to measure the rise and fall of metabolic activity in multiple organisms in field experiments. For example, we were recently able to identify a syntrophic relationship between several Archaeal species and Proteobacterial metal reducers in a subsurface uranium-contaminated site. This is something we would only see using this high-density phylogenetic array and looking at the rRNA directly.

Rasko: Well, that's neat. That's all I'll say! That's just neat!

Andersen: This is really neat. It's something we've just recently done.

Rasko: Do you think that a version of your array could be used in a very similar way to identify different pathogens in respiratory diseases? Could you see the 16S and directed approaches working hand in hand or do you see them as competing technologies?

Andersen: I actually can see them working hand in hand. We've already experimented with a targeted approach for identifying pathogens in the environment. In a previous paper, we developed an array that targeted unique regions in the pathogen genomes to identify 18 specific pathogens. We had primers that were specific to each one of the pathogens. We used those primers to amplify 100 to 200 bp fragments. Multiple probe pairs were then used for what ended up being very highly specific and sensitive detection.

I think combining these two methods can give you a general overview of what's in a sample. If you're interested in specific pathogens, you can target them with a number of specific probe sets. As the number of probes that you can put on an array increases, I think we will be able to have everything on one array and run one sample to answer both those questions simultaneously.

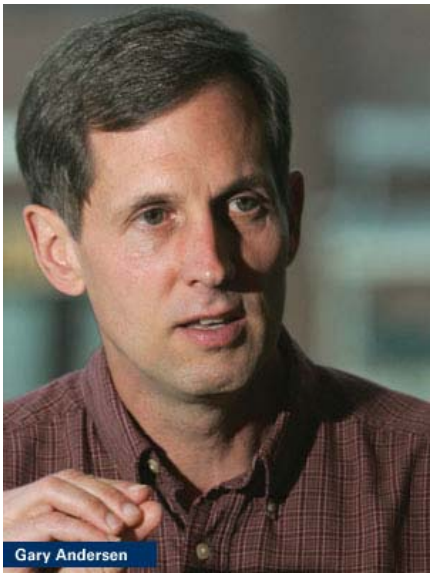
With our latest generation microarray, which we're using in the air, there are certain pathogenic organisms that we knew we couldn't identify by 16S because they are so similar to other environmental organisms. We put these pathogen-specific regions on the array so we could distinguish the pathogen from other environmental organisms.

For example, there's just one base pair difference, or in some cases even no base pair differences, between *Bacillus cereus* and *anthracis*. The 16S can tell you that there are some organisms from the *Bacillus cereus* group there, but we have more specific probes for the *anthracis* pathogen.

Tracking Pathogen Movement

Rasko: Do you think you can use this microarray technology to identify movement of bacteria, fungi or viruses on a global level? For example, if you have something that gets aerosolized in California after weeks, months or years, would you expect that novel species to show up in Washington, D.C., or Florida or New York?

Andersen: In a way, we're doing those experiments now. We're currently measuring the bacterial populations in urban aerosols in a number of cities across the United States and over different seasons, for a U.S. Department of Homeland Security project. In the process of doing this we've identified a number of local reservoirs of bacteria that are entrained into air. We've also observed long-range dispersal of certain organisms.



By looking at multiple locations over several years, we're for the first time getting an idea of the scope of variation of organisms at a given location and what factors influence the microbial composition.

For the Department of Homeland Security, this will be especially important in view of their BioWatch program for the biosurveillance of major cities for potential bioweapon release. One question that comes up is whether a pathogen detected by a sensitive detector could have occurred naturally.

That's what we're looking at now. What are the background organisms that would typically come up in a sample? We can see, for instance, different classes of organisms that seem to be more predominant based on prevailing wind patterns and other meteorological conditions. To some extent, that influences long-range dispersal as well.

Rasko: So, theoretically, you could identify the avian flu if it entered the United States in California and track its progress across the country through a methodology like this?

Andersen: You could, yes. You'd have to use much more extensive sampling than we're doing now and use probes for the specific organism you're looking for.

But yes, that would be an ideal way for measuring the spread and dispersal. One of the reasons it hasn't caught on rapidly is probably the type of arrays. We use very large arrays, so there's a cost component there. We can do hundreds of samples, but I think it would require thousands or tens of thousands of samples. More inexpensive versions of the array might be necessary for monitoring these specific pathogens.

Surprises in the Air

Rasko: From your papers, you've already identified some things that you wouldn't have thought would be in the air or the soil or the water samples. Is there anything that really surprised you?

Andersen: All the time, actually! We see many examples of organisms that were identified as deep-sea or from the Yellowstone Hot Springs or the Arctic ice, in the air or other environments. I think there are a number of reasons for that.

An organism could truly be ubiquitous in a number of locations, but I think a more likely explanation is that there are subtle differences in the 16S sequence. So, it's probably a variant of a species or a related species that has adapted to these different environments. It shares many of the properties, but not all. So, a deep subsurface vent organism, or a related organism with similar 16S, could also be in the middle of the cold Arctic Ocean.

I guess what blew our socks off in the beginning were the pathogens we would see in, for instance, the air. We'd often see things like *Clostridium botulinum*. That pathogen is anaerobic, so we were a little surprised, but it makes sense when you think about it. A lot of these organisms are spore-forming. They're probably not active in the air, but they're dispersing.

Limitations of Microarrays

Rasko: What do you think is the biggest limitation to chips in general?

Andersen: For the array that we're developing, I'd say one of the biggest limitations is the resolution of organism identification using the 16S rRNA. But for chips in general, we have a problem with data overload from the integration of our multiple experiments.

Let me first talk about our specific limitations using 16S rRNA. It gets back to our question of what is a species. The 16S gene is somewhat limited in the resolution and identification of particular organisms. It's about 1500 bp long, and the difference between closely related species is at best a few regions within the gene with a few base pairs of sequence variation.

In addition to the *Bacillus anthracis/Bacillus cereus* example, there's *Yersinia pestis/Yersinia pseudo-tuberculosis.* The organisms have very different disease lifestyles, but a one base pair difference in the 16S rRNA gene. Oftentimes, the resolution of this gene is not appropriate for individual pathogen species identification.

However, there are other candidate regions that can increase the resolution. For instance, there's the large subunit, or 23S rRNA, gene. It has all the advantages of the 16S in that it's present in every organism and there are conserved primers for amplification, but it's twice as long with a greater amount of sequence variation between species.

There are a number of other conserved molecules or genes you could use to increase resolution as well, so we need to develop a database. For 16S we havever 200,000 sequences in the database. Until we can get something approaching that for these other genes, they won't be as useful.

The other thing is that, although I'm really keen on using the large subunit at some point, it's 3,000 bases long. Until you can use one forward and one reverse reaction for sequencing a clone with a 3000 bp gene, I don't think making a database for it is going to be that popular. It's just going to be too expensive.

The other part of data overload is the informatics. We have 500,000 data points for each array and multiple arrays for each experiment. I have a small lab without much bioinformatics support, so we get overwhelmed quite quickly. We can see what's in a sample and compare one place to another at a gross level, but it takes us a lot of work to get into more detailed questions. I think if we had a better way of handling these large amounts of data, we would be able to get to these questions a lot quicker.

FOR MORE INFORMATION

Contacts

Gary L. Andersen, Ph.D.
Scientist
Center for Environmental Biotechnology
Lawrence Berkeley National Laboratory
1 Cyclotron Road, Mail Stop 70A-3317
Berkeley, CA 94720

David A. Rasko, Ph.D.
Senior Research Scientist
University of Texas Southwestern Medical Center at Dallas
Department of Microbiology
6000 Harry Hines Blvd. NA6.138
Dallas, TX 75235

Companies

Affymetrix Inc.

Further Reading

DeSantis TZ, Stone CE, Murray SR, Moberg JP, Andersen GL. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiol Lett.* 2005 Apr 15; 245(2):271-8.

Hu P, Brodie EL, Suzuki Y, McAdams HH, Andersen GL. Whole-genome transcriptional analysis of heavy metal stresses in *Caulobacter crescentus. J Bacteriol.* 2005 Dec; 187(24):8437-49.

Radnedge L, Agron PG, Hill KK, Jackson PJ, Ticknor LO, Keim P, Andersen GL. Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis. Appl Environ Microbiol.* 2003 May; 69(5):2755-64.

Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes. 2002* Apr; 16(2):119-27.