Development and Use of Integrated Microarray-Based Genomic Technologies for Assessing Microbial Community Composition and Dynamics

J. Zhou (PI)[1], L. Wu[1], T. Gentry[1], C. Schadt[1], Z. He[1], and X. Li[2]

[1]Oak Ridge National Laboratory and [2]Perkin Elmer Life and Analytical Sciences

Abstract

The recent development of microarray technology provides great opportunity for the simultaneous identification of thousands of microbial genes/populations, but low microbial biomass often prevents application of this technology to microbial communities in natural settings. We have developed a whole community genome amplification (WCGA)-assisted microarray-based detection approach for analysis of microbial communities whose members can not be studied using conventional technology. With optimized buffer systems, as few as two bacterial cells could be detected. Whole genome microarray hybridization showed that representative detection of individual genes or genomes was obtained within the DNA concentrations of 1 to 100 ng from individual or mixed genomes. Significant linear relationships were observed between signal intensity and initial DNA concentration ranging from: (a) 40pg to 125ng for the majority of *Shewanella* genes (r^2=0.65-0.99) and other organisms as detected by whole genome arrays, (b) genomes in constructed communities from 0.1 to 1000 ng (r^2=0.91) using community genome arrays and (c) community DNAs diluted from a stimulated natural groundwater ranging from 0.01 ng to 250 ng (r^2=0.96-0.98) using functional gene arrays. We applied this technology to investigate microbial communities in five groundwater samples contaminated with uranium and other metals using functional gene arrays (~2,000 probes). The results indicated that microbial populations containing important genes involved in contaminant degradation and immobilization have locally heterogeneous distributions and that microbial diversity is greatly decreased in contaminated environments. This is the first time that microarrays have been successfully used to analyze low biomass communities, such as those commonly found in settings important to human health, industry, and environmental management. We have also developed a software program, CommOligo, for designing probes from whole-genomes, meta-genomes or a group of sequences. The program uses a new global alignment algorithm to design single or multiple unique probes for each gene with default settings for maximal similarity of 85%, maximal number of continuous match of 15 bases, and free energy of -30 kcal/mol. The program is also able to design single or multiple group specific probes for a group of genes with minimal similarity of 96% within a group and the same parameters as unique probes outside a group. The program was evaluated using both whole-genome and highly homologous sequence data and compared with other probe design software. The results clearly demonstrate that CommOligo performed better and can be used for oligonucleotide probe design from various types of sequence data. Using this program, a more comprehensive functional gene array containing ~24,000 probes for important biogeochemical cycling (C, N, S, & P), metal resistance, and contaminant degradation genes has been designed and constructed. This is the most comprehensive array currently available for environmental studies. We applied this microarray to the study of microbial communities at the NABIR-FRC during ethanol biostimulated uranium reduction. The array revealed that the different sampled wells initially contained heteogenous microbial populations that became more similar to each other over the 1 year treatment process. Trends in levels of homologous functional genes, including those for nitrite reductase, correlated with biochemical changes, while the FGA revealed that genes from individual species were more varied in their response to biostimulation. These results indicate the potential for microarray-based characterization of microbial community structures and dynamics in environmental samples.

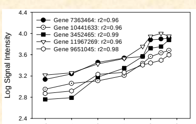
Introduction

The recent development of microarrays as powerful, high-throughput genomic technology has spurred investigators toward their use for the study of various biological processes. However, adapting microarrays for use in environmental studies presents great challenges in terms of design, use and data analysis. The genes encoding functional enzymes involved in various biogeochemical cycling (e.g. nitrogen, carbon and sulfur) and bioremediation processes, are very useful as signatures for monitoring the potential activities and physiological status of the microbial populations that drive these environmental processes. Both oligonucleotides and DNA fragments derived from functional genes can be used for fabricating functional gene arrays (FGAs). However, microarrays containing large DNA fragments as probes are generally constructed from polymerase chain reaction (PCR)-amplified DNA. Obtaining all the diverse environmental clones and bacterial strains required as templates for this amplification from their various sources is virtually impossible.

To circumvent this problem, FGAs containing synthetic oligonucleotides (oligos) have been developed for use. The main advantage of oligo FGAs is that construction is much easier than DNA-based FGAs, because the probes can be directly designed and synthesized based on sequence information from public databases. Therefore, comprehensive arrays representing the extreme diversity of known environmental sequences can be constructed. This poster details results from use of 50mer FGAs that indicate the array has potential as specific, sensitive, and potentially quantitative parallel tools for characterizing the composition, structure, activities and dynamics of microbial communities in natural environments.

Whole Community Genome Amplification

- A whole community genome amplification (WCGA)-assisted microarray-based detection approach was developed for the analysis of microbial communities whose members can not be studied using conventional technology.
- Amplified DNAs were analyzed using a 50-mer oligonucleotide functional gene array (FGA) with ~2,000 probes for C, N, and S cycling, biodegradation, and metal resistance.
- WCGA was tested on pure culture DNAs and groundwater samples from the NABIR Field Research Center (FRC) in Oak Ridge, TN.



- **Linear amplification of 5 different genes from dilutions of FRC groundwater.**
- **Individual genes can be representatively amplified from 1 - 100 ng of template.**

Development of New FGA

- A new software program, *CommOligo*, has also been developed that greatly improves the quality of designed probes.
- We used this software to design a more comprehensive 50mer FGA that contains ~24,000 probes.

***CommOligo* Probe Design Software**

Features

- Uses novel global alignment algorithms.
- Designs unique probes for a single target.
- Designs group probes for highly similar sequences.
- Considers probe mismatch position with non-targets.
- Chooses desirable regions for probe design.
- Selects optimal oligonucleotides.

Number of designed gene-specific and group-specific probes (50-mer) by different programs

Programs used	Whole-genome sequences of *M. maripaludis* (1766 ORFs)						Group sequences of *nirS* and *nirK* (842 gene sequences)					
	Total ORFs	ORFs rejected	Probes designed	**Specific probe**	Non-specific	**Group-specific**	Total ORFs	ORFs rejected	Probes designed	**Specific probe**	Non-specific	**Group-specific**
ArrayOligoSeector	1766	7	1759	**1415**	344	**N/A**	842	0	842	**117**	725	**N/A**
OligoArray	1766	68	1698	**1654**	44	**N/A**	842	35	807	**70**	737	**N/A**
OligoArray 2.0	1766	68	1698	**1464**	234	**N/A**	842	51	791	**35**	756	**N/A**
OligoPicker	1766	18	1748	**1745**	3	**N/A**	842	657	185	**141**	44	**N/A**
CommOligo	1766	9	1757	**1745**	0	**12**	842	512	330	**147**	0	**183**

- **Compared to other probe design programs, *CommOligo* designed more gene-specific and less non-specific probes for tested gene sets.**
- ***CommOligo* also designed group probes for very similar sequences.**

Overall Summary of Probes on New FGA

Gene Category	Number of Probes: Unique	Group	Total
Carbon Degradation	2,532	276	2,808
Carbon Fixation	584	215	799
Metal Resistance/Reduction	4,039	507	4,546
Methane/Methanogenesis	437	333	770
Nitrogen Fixation	1,225	0	1,225
Nitrogen Metabolism	865	902	1,767
Nitrogen Reduction	1,805	501	2,306
Organic Contaminant	6,920	1,087	8,007
Perchlorate Remediation	21	0	21
Sulfur Reduction	1,286	329	1,615
Total	**19,714**	**4,150**	**23,864**

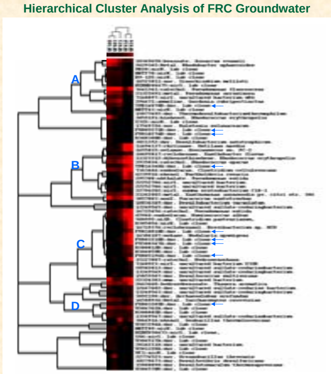
FRC Groundwater WCGA

Selected characteristics of analyzed FRC wells

Chemical Parameter	FRC Well: FW300	FW003	FW021	FW010	FW024
pH	6.7	6.0	3.4	3.5	3.6
Aluminum (mg/L)	0.2	0.4	398.0	1120.0	527.4
Chloride (mg/L)	2.4	124.7	220.2	686.4	281.4
Nitrate (mg/L)	2.6	1015	8823	43019	8481
Sodium (mg/L)	2.0	44.0	522.0	849.0	950.1
Sulfate (mg/L)	6.4	16.3	122.1	8.3	987.1
Uranium (mg/L)	0	0.10	12.2	0.17	44.8
Technetium (pCi/L)	0	141	30974	7190	36956
Spec. Conduct. (mS/cm)	0.3	1.6	11.4	36.0	15.8

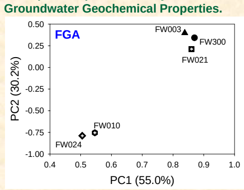
- **FW10, 21 & 24 are highly contaminated.**
- **FW300 is an uncontaminated background well.**

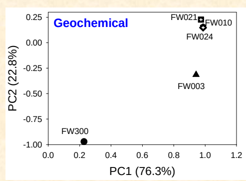
Hierarchical Cluster Analysis of FRC Groundwater



- **FGA analysis indicated that some genes (clusters B & C) were more prevalent at heavily contaminated sites (FW010 & FW024) while others were most common at less contaminated sites (cluster A) or were present at all sites (cluster D).**
- **Genes (*dsrA/B*) indicating the presence of sulfate reducing bacteria were found at all sites.**

Principal Components Analysis of Selected FRC Wells Based on FGA Results and Groundwater Geochemical Properties.





- **Principal components analysis of the FGA data grouped two of the highly contaminated wells (FW010 & FW024) together, thus suggesting that the contaminants impacted the microbial community structure. However, other parameters (pH) may have also influenced these results.**
- **The presence of sulfate-reducing bacteria at all sites implies the potential for *in situ* metal (uranium) reduction/remediation.**

Microbial Characterization of Stanford/ORNL Groundwater Treatment System

Groundwater Chemistry

- **Uranium ~ 50 mg/L**
- **Nitrate ~ 4000 mg/L**
- **pH < 3.6**

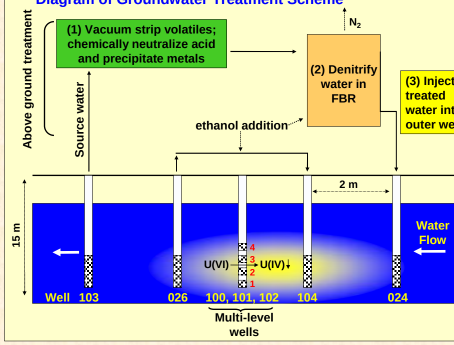
Treatment Strategy

- **Above ground denitrification and neutralization of groundwater**
- ***in situ* ethanol biostimulation and U(VI) reduction**

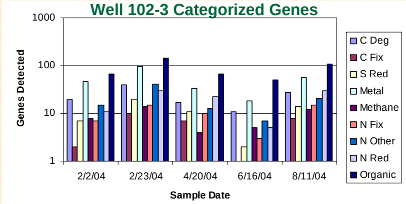
Sampling Strategy

- **Sample with peristaltic pump**
- **Harvest biomass by filtration**
- **Collect samples weekly to monthly**
- **Focus on wells 101-2, 102-3, & 026**
- **Analyze samples with new FGA**

Diagram of Groundwater Treatment Scheme

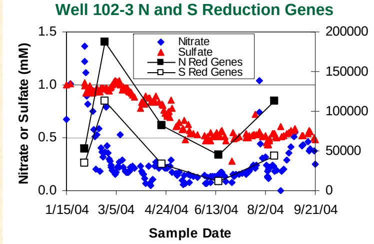


Well 102-3 Categorized Genes



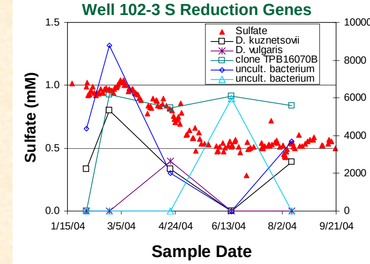
- **Largest number of genes detected in 2/23/04 sample (during denitrification)**
- **Organic and metal genes most commonly detected but represented the largest groups of probes on FGA**

Well 102-3 N and S Reduction Genes



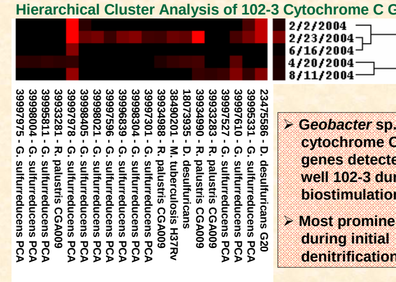
- **Total N & S reduction gene signals correlated with nitrate & sulfate levels**
- **Additional samples after 8/04 are being analyzed to see if the trend continued**

Well 102-3 S Reduction Genes



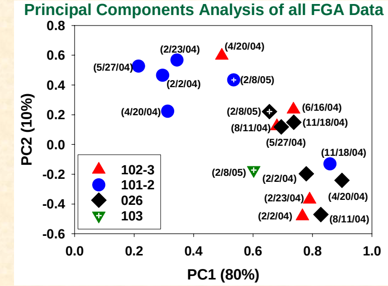
- **Overall, S reduction genes correlated with sulfate levels, but the dynamics of individual genes differed**

Hierarchical Cluster Analysis of 102-3 Cytochrome C Genes



- ***Geobacter* sp.-like cytochrome C genes detected in well 102-3 during biostimulation**
- **Most prominent during initial denitrification phase**

Principal Components Analysis of all FGA Data



- **Initial 026 & 102-3 samples (2/2/04) were similar but distinct from 101-2 indicating heterogeneity in the microbial populations**
- **Over time, the populations in the different wells became more similar to each other possibly due to continual influx of injected groundwater**

Conclusions

- These results indicate that the 50mer FGA has potential as specific, sensitive, and potentially quantitative parallel tools for characterizing the composition, structure, and dynamics of microbial communities in natural environments.
- Development of the new, expanded FGA should further enhance the application of this technology to the investigation of critical environmental issues.

Acknowledgments

This research was supported by The United States Department of Energy under the Natural and Accelerated Bioremediation Research Program of the Office of Biological and Environmental Research, Office of Science. Oak Ridge National Laboratory is managed by University of Tennessee-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.