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

J. Zhou (PI)[1], S.-K. Rhee[1], C. Schadt[1], T. Gentry[1], Z. He[1],X. Li[2], X. Liu[1], J. Liebich[1], S.C. Chong[1], L. Wu[1]

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

Abstract

To effectively monitor microbial populations involved in various important processes, a 50-mer-based oligonucleotide microarray was developed based on known genes and pathways involved in: biodegradation, metal resistance and reduction, denitrification, nitrification, nitrogen fixation, methane oxidation, methanogenesis, carbon polymer decomposition, and sulfate reduction. This array contains approximately 2000 unique and group-specific probes with <85% similarity to their non-target sequences. Based on artificial probes, our results showed that at hybridization conditions of 50°C and 50% formamide, the 50-mer microarray hybridization can differentiate sequences having <88% similarity. Specificity tests with representative pure cultures indicated that the designed probes on the arrays appeared to be specific to their corresponding target genes. Detection limits were about 5–10ng genomic DNA in the absence of background DNA, and 50–100ng (~1.3 °ø10⁷ cells) in the presence background DNA. Strong linear relationships between signal intensity and target DNA and RNA concentration were observed (r^2 = 0.95–0.99). Application of this microarray to naphthalene-amended enrichments and soil microcosms demonstrated that composition of the microflora varied depending on incubation conditions. While the naphthalene-degrading genes from *Rhodococcus*-type microorganisms were dominant in enrichments, the genes involved in naphthalene degradation from Gram-negative microorganisms such as *Ralstonia*, *Comamonas*, and *Burkholderia* were most abundant in the soil microcosms (as well as those for polyaromatic hydrocarbon and nitrotoluene degradation). Although naphthalene degradation is widely known and studied in *Pseudomonas*, *Pseudomonas* genes were not detected in either system. Real-time PCR analysis of 4 representative genes was consistent with microarray-based quantification (r^2 = 0.95). Currently, we are also applying this microarray to the study of several different microbial communities and processes at the NABIR-FRC in Oak Ridge, TN. One project involves the monitoring of the development and dynamics of the microbial community of a fluidized bed reactor (FBR) used for reducing nitrate and the other project monitors microbial community responses to stimulation of uranium reducing populations via ethanol donor additions in situ and in a model system. Additionally, we are developing novel strategies for increasing microarray hybridization sensitivity. Finally, great improvements to our methods of probe design were made by the development of a new computer program, *CommOligo*. *CommOligo* designs unique and group-specific oligo probes for whole-genomes, meta-genomes, and groups of environmental sequences and uses a new global alignment algorithm to design single or multiple probes for each gene or group. We are now using this program to design a more comprehensive functional gene array for environmental studies. Overall, our results indicate that the 50mer-based microarray technology has potential as a specific and quantitative tool to reveal the composition of microbial communities and their dynamics important to processes within contaminated environments.

Introduction

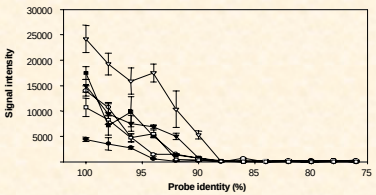
The recent development of microarrays as powerful, high-throughput genomic technology has spurred investigators toward their use for the study of various biological processes. However, adapting microarrays for use in environmental studies presents great challenges in terms of design, use and data analysis. 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 a 50mer FGA that indicates 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. Based on these results a much more comprehensive 50mer FGA of several thousand gene probes is being designed and tested in our laboratory.

Design and Analysis of 50mer FGA

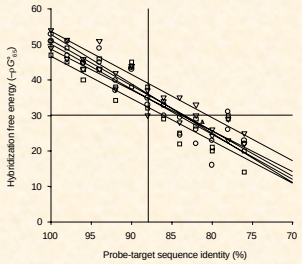
- A comprehensive 50-mer-based oligonucleotide microarray was developed based on most of the known genes (2,402) and pathways involved in biodegradation and metal resistance.
- The FGA contained 1,662 unique and group-specific probes with < 85% similarity to their non-target sequences.
- The FGA was tested for specificity, sensitivity, and linearity of signal intensity using serial dilutions of *Thauera aromatica* K-172 genomic DNA against a background of 1 µg *S. oneidensis* MR-1 genomic DNA as a negative control.
- Hybridization conditions were 50°C and 50% formamide.
- The FGA was used to analyze three different soils: 1) TFD – contaminated with 27.5 mg of polyaromatic hydrocarbons (PAHs) kg^{-1} and also various metals ; 2) PCT18 – contaminated with 3.3 mg PAHs kg^{-1}; and 3) IDT - 3.0 mg of BTEX (benzene, toluene, ethylbenzene, and xylene), 17 mg of Pb, and 6.3 mg of Cr kg^{-1}. A non-contaminated forest soil was used as the control.

FGA Specificity



Effect of probe-target sequence identities on hybridization signal intensity. Symbols correspond to different naphthalene pathway genes.

- The FGA differentiated sequences having < 88% similarity.



Hybridization free energy between probes of varying similarity to target genes.

- The intersecting vertical line indicates the threshold value of cross hybridization signal at 88% sequence homology.
- The intersecting horizontal line indicates the suggested -ρG °65 value (30 in this case) under which little cross-hybridization occurred.

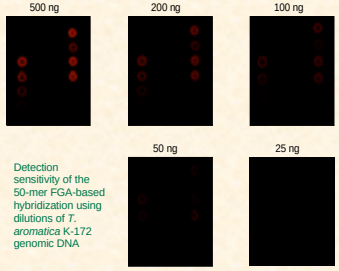
Table 1. Sequence similarities of *dsr*AB, *nirS*, *nirK*, *nifH*, *amoA* and *pmoA* genes in GenBank.

Phylogenetic hierarchies	Sequence similarities (%) ± standard deviation†					
	*dsr*AB	*nirS*	*nirK*	*nifH*	*amoA*	*pmoA*
Strain	0.93 ± 0.03	0.93 ± 0.04	0.91 ± 0.05	0.93 ± 0.06	0.99 ± 0.01	0.95 ± 0.03
Species	0.73 ± 0.13	0.70 ± 0.18	0.76 ± 0.00	0.82 ± 0.06	0.75 ± 0.11	0.79 ± 0.55
Genus	0.70 ± 0.09	0.67 ± 0.16	0.71 ± 0.07	0.70 ± 0.07	0.71 ± 0.07	0.75 ± 0.15
Family or higher	0.66 ± 0.13	0.57 ± 0.14	0.62 ± 0.14	0.66 ± 0.10	0.65 ± 0.13	-

†Average and standard deviation of similarity values of six gene groups at different phylogenetic levels.

- Based on the FGA's ability to differentiate sequences that are < 88% similar, it should be possible to detect genes at the species level (~70-80% similarity).

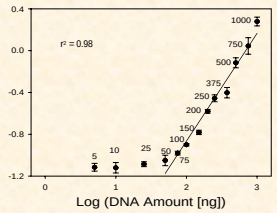
FGA Sensitivity



Detection sensitivity of the 50-mer FGA-based hybridization using dilutions of *T. aromatica* K-172 genomic DNA

- The FGA detection limit was about 50–100 ng of pure culture genomic DNA or 1.3 _ 10⁷ cells in the presence background DNA.

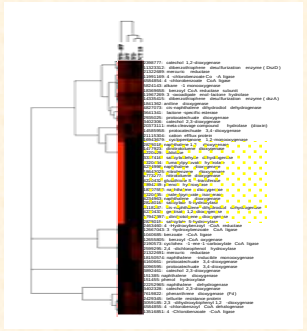
FGA Linearity of Detection



Quantitative relationship of the FGA-based hybridization with dilutions of *T. aromatica* K-172 genomic DNA in presence of background DNA.

- There was a strong linear relationship between FGA signal intensity and target DNA concentration (r^2 = 0.95–0.99).

FGA Analysis of Contaminated Soils



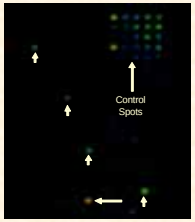
Hierarchical cluster analysis of soil microbial communities in four different contaminated soils based on hybridization signal intensity ratios for the genes showing SNR > 3. Soil TFD-NP represents TFD soil after exposure to naphthalene vapor.

- Several genes related to naphthalene degradation were clustered together and they were abundant in the naphthalene-amended soil.
- These genes also appeared to be present at very low levels in all three soils which is not surprising since at least two of these soils contained low levels of PAHs.

FGA Analysis of FRC Groundwater after Biostimulation with Ethanol

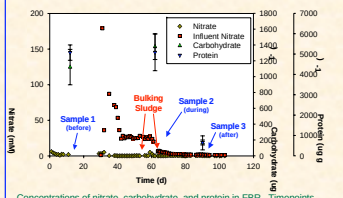
- FRC groundwater contaminated with high levels of uranium (~ 50 mg l^{-1}) and nitrate (~ 4000 mg l^{-1}) was biostimulated by the addition of ethanol as an electron donor.
- Samples were collected and analyzed using the FGA.

- Genes for dissimilatory sulfate reductase (*dsrAB*), nitrate reductase (*narG*) and nitrite reductase (*nirS*) were detected (indicated by arrows).
- Spot color scales with hybridization signal intensity (blue=low, green=med, red=high).



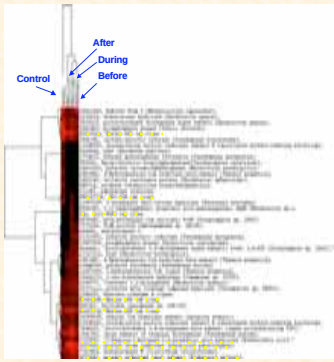
FGA Analysis of a Fluidized Bed Reactor Community

- A 1400-L fluidized bed reactor (FBR) is being used to denitrify FRC groundwater.
- Ethanol added as electron donor.
- Occurrences of bulking sludge interrupted operation of FBR.
- Samples from before, during, and after bulking event were analyzed using the FGA.
- Samples were also assayed for total carbohydrate and protein content.
- Microorganisms were visualized with staining and confocal microscopy.



Concentrations of nitrate, carbohydrate, and protein in FBR. Timepoints tested by FGA and microbial analyses are indicated along with two occurrences of bulking sludge.

- Carbohydrate and protein levels decreased following the bulking sludge event but this may have been due to other factors such as reduced levels of nitrate in the influent.



Hierarchical cluster analysis of FBR microbial communities before, during, and after bulking sludge occurrence. An uncontaminated forest soil is included for comparison.

- FGA analysis did not detect significant changes in the FBR microbial community during the bulking occurrence.
- However, current methods may not detect shifts in subdominant populations (< 1% of total community).
- Denitrification genes were dominated at all time points by a nitrate reductase (*nasA*) similar to *Klebsiella oxytoca* and a nitrous oxide reductase (*nosZ*) similar to *Paracoccus pantotrophus*.
- Several sulfur reduction genes were also detected.

- Visual and microscopic analyses suggested the bulking was due to Zoogloeal clustering.

- This research suggests the bulking was probably caused by chemical factors, possibly micronutrient deficiencies, and not changes in microbial community structure.

Development of New FGA

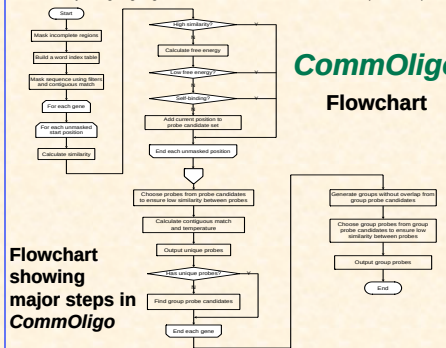
- Based on the results highlighted in this poster, we are currently designing a more comprehensive 50mer FGA that contains several thousand probes.
- A new software program, *CommOligo*, has also been developed that greatly improves the quality of designed probes.

***CommOligo* Probe Design Software**

Features

- Using novel global alignment algorithms.
- Designing unique probes for a single target.
- Designing group-specific probes for a group of highly homologous sequences.
- Considering probe mismatch position with non-targets.
- Choosing desirable regions for probe design.
- Considering all possible probe candidates and selecting optimal oligonucleotides.
- Potentially designing oligonucleotides for environmental studies (FGA, etc.).

Flowchart showing major steps in *CommOligo*

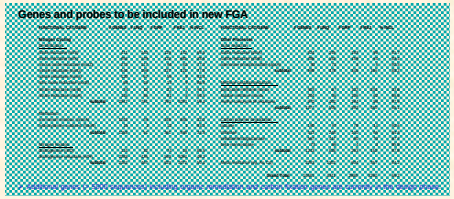


***CommOligo* Flowchart**

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 probes	Non-specific	Group-specific	Total ORFs	ORFs rejected	Probes designed	Specific probes	Non-specific	Group-specific
ArrayOligoSelector	1766	1	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**

- In comparison to other probe design programs, *CommOligo* designed more gene-specific and less non-specific probes for tested gene sets.
- *CommOligo* also designed group-specific probes for highly homologous sequences.



Conclusions

- These results indicate that the 50mer FGA has potential as specific, sensitive, and potentially quantitative parallel tools for characterizing the composition, structure, activities 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.

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

UT-BATTELLE