

Microarray-Based Analysis of Subnanogram Quantities of Microbial Community DNAs by Using Whole-Community Genome Amplification†

Liyou Wu,^{1,‡§} Xueduan Liu,^{1,2,3,‡§} Christopher W. Schadt,¹ and Jizhong Zhou^{1,*‡}

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6038,¹ and Department of Plant Pathology, Hunan Agricultural University,² and School of Minerals Processing & Bioengineering, Central South University,³ Changsha 410083 Hunan, People's Republic of China

Received 18 November 2005/Accepted 15 April 2006

Microarray technology provides the opportunity to identify thousands of microbial genes or populations simultaneously, but low microbial biomass often prevents application of this technology to many natural microbial communities. We developed a whole-community genome amplification-assisted microarray detection approach based on multiple displacement amplification. The representativeness of amplification was evaluated using several types of microarrays and quantitative indexes. Representative detection of individual genes or genomes was obtained with 1 to 100 ng DNA from individual or mixed genomes, in equal or unequal abundance, and with 1 to 500 ng community DNAs from groundwater. Lower concentrations of DNA (as low as 10 fg) could be detected, but the lower template concentrations affected the representativeness of amplification. Robust quantitative detection was also observed by significant linear relationships between signal intensities and initial DNA concentrations ranging from (i) 0.04 to 125 ng ($r^2 = 0.65$ to 0.99) for DNA from pure cultures as detected by whole-genome open reading frame arrays, (ii) 0.1 to 1,000 ng ($r^2 = 0.91$) for genomic DNA using community genome arrays, and (iii) 0.01 to 250 ng ($r^2 = 0.96$ to 0.98) for community DNAs from ethanol-amended groundwater using 50-mer functional gene arrays. This method allowed us to investigate the oligotrophic microbial communities in groundwater contaminated with uranium and other metals. The results indicated that microorganisms containing genes involved in contaminant degradation and immobilization are present in these communities, that their spatial distribution is heterogeneous, and that microbial diversity is greatly reduced in the highly contaminated environment.

Microorganisms play integral and often unique roles in ecosystem functions, yet we often know little about dominant populations that presumably perform these functions, nor do we know much about how these populations differ with habitat. Understanding the structure and composition of microbial communities and their responses to environmental perturbations, such as toxic contamination, climate change, and land use changes, is critical for prediction, maintenance, and restoration of desirable ecosystem functions. Due to the extremely high diversity of environmental samples, microbial detection, characterization, and quantification are great challenges. The development and application of nucleic acid-based techniques have largely eliminated the reliance on cultivation-dependent methods for microbial detection and consequently have greatly advanced characterization of microorganisms in natural habitats (2).

Compared to nucleic acid hybridization with porous membranes, real-time PCR, and other molecular approaches, microarray-based hybridization has the advantages of high

throughput and parallel detection. Although microarray technology has been used successfully to analyze global gene expression in pure-culture studies (8, 11, 15, 16, 23, 24, 31), adapting microarray hybridization for use in environmental studies presents numerous challenges in terms of specificity, sensitivity, and quantitation (21, 22, 25, 28, 32, 35). Various environmental microarray formats, such as functional gene arrays (FGA) (21, 22, 25, 28), community genome arrays (CGA) (29), and phylogenetic oligonucleotide arrays (10, 17), have been developed for microbial community analyses of environmental samples and evaluated. Because of their high-throughput capacity, it is expected that microarray-based genomic technologies will revolutionize analyses of microbial community structure, functions, and dynamics (32). However, one of the main challenges of successful application is the fact that current detection sensitivities are often not sufficient for detecting the less dominant microbial populations in an environmental sample (5, 21, 22). Currently, for single-copy genes, genomic DNA from approximately 10^7 cells is required to obtain reasonably strong hybridization using 50-mer-based oligonucleotide microarrays (21). However, individual populations in any particular environmental sample, even surface soils in which biomass is typically high, generally consist of less than 10^7 cells/g. This leads to great difficulties in analyses of natural microbial communities. Appropriate manipulation (e.g., amplification) of community DNAs prior to hybridization is needed, but it is challenging to amplify these DNAs in a representative and quantitative fashion (27, 35). Traditional PCR-based amplification methods suffer from inherent problems associated with biases and artifacts (27, 35), and their gene-by-

* Corresponding author. Present address: Institute for Environmental Genomics, University of Oklahoma, Stephenson Research and Technology Center, 101 David L. Boren Blvd., Norman, OK 73019. Phone: (405) 325-6073. Fax: (405) 325-3442. E-mail: zhou-aem@rccc.ou.edu.

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

‡ L.W., X.L., and J.Z. contributed equally to this work.

§ Present address: Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019.

gene nature makes application of these methods to comprehensive, high-throughput microarray analyses impractical. Thus, we evaluated and optimized multiple displacement amplification (MDA) (3, 6, 12, 14, 18, 19, 20, 27) for a whole-community genome amplification (WCGA)-assisted microarray detection approach to analyze microbial community structure and demonstrated its application to low-biomass groundwater microbial communities.

MATERIALS AND METHODS

Environmental samples, cultures, and isolation of genomic DNA. *Shewanella oneidensis* MR-1 from our laboratory culture collection and *Rhodospseudomonas palustris* CGA009 and *Nitrosomonas europaea* ATCC 19718 provided by Caroline Harwood, Department of Microbiology, University of Washington, and Daniel J. Arp, Department of Botany and Plant Pathology, Oregon State University, Corvallis, respectively, were used to construct whole-genome cDNA microarrays and also to construct community genomic DNA arrays in this study. The following 13 other distantly related bacteria were also used to construct community genomic DNA arrays: α -proteobacterium C1-4, *Bacillus methanolicus* F6-2, *Marinobacter* sp. strain D5-10, *Halomonas variabilis* B9-12, *Pseudomonas* sp. strain G179, and *Azoarcus tayloriticus* Td1, which were obtained from our collection or were marine isolates; and *Thauera aromatica*, *Paracoccus denitrificans*, *Achromobacter xylosoxidans*, *Rhizobium meliloti*, *Ochrobactrum anthropi*, *Azospirillum brasilense*, and *Pseudomonas mendocina*, which were obtained from the American Type Culture Collection (Manassas, VA). Most of the bacteria were grown in Luria-Bertani broth; the exceptions were *N. europaea*, which was grown in *N. europaea* medium, and *R. palustris*, which was grown in nutrient broth. Cells were harvested at the exponential phase and frozen at -80°C .

To evaluate the performance of whole-community genome amplification for microbial community analysis, groundwater samples obtained from the Field Research Center (FRC) site of the U.S. Department of Energy Environmental Remediation Science Program at Oak Ridge Reservation, Oak Ridge, Tenn., were used. The FRC site includes three areas in which the soil and groundwater are contaminated and an uncontaminated background area in which the soils are similar to those found in the contaminated areas. In the past, the site contained four unlined ponds that received approximately 10^6 liters of liquid nitric acid and uranium-bearing wastes per year for approximately 30 years until it was closed in 1984. The waste ponds contribute to the contamination by nitrate, uranium, heavy metals, and a variety of low-level organic contaminants of the surrounding sediment and groundwater. A full description of the site can be found at the FRC website (<http://www.esd.ornl.gov/nabirfrc>). Groundwater samples were obtained from five wells. Wells FW010 and FW024, located in area 3, are 32.5 m apart and are approximately 20 m from a former waste pond. Well FW021 is 27 m from the waste pond embankment in area 1 and is approximately 130 m from the wells in area 3. Well FW003 is located in area 2, which is approximately 275 m down-gradient from the waste ponds in area 1 and area 3. Well FW300 is located in the uncontaminated background area, approximately 6 km northwest of the source ponds. Water was collected from a screened interval below the water table at each of the six wells on the same day (2 April 2003). Another groundwater sample was collected from well FW029 located in area 1 that had been experimentally amended with ethanol to stimulate the anaerobic microbial community (13).

The genomic DNAs of the pure cultures were isolated using previously described protocols (34). All genomic DNA samples were treated with RNase A (Sigma, St. Louis, MO) and analyzed on agarose gels stained with ethidium bromide prior to microarray hybridization. Groundwater samples were collected and transported to the laboratory in amber glass bottles. Bacteria were harvested by centrifugation ($10,000 \times g$, 4°C , 30 min), and the pellets were stored at -80°C until DNA was extracted. Each cell pellet was resuspended in lysis buffer, and the cells were disrupted using a previously described grinding method (33) and were purified by gel electrophoresis plus a minicolumn purification (Wizard DNA clean-up system; Promega, Madison, WI). DNA concentrations were determined in the presence of ethidium bromide by fluorometric measurement of the excitation at 360 nm and emission at 595 nm using an HTS700 BioAssay reader (Perkin Elmer, Norwalk, CT).

Whole-genome DNA amplification using phi 29 DNA polymerase. A Templiphi 500 amplification kit (Amersham Biosciences, Piscataway, NJ) was used for whole-genome amplification or whole-community amplification with incubation 2 to 6 h at 30°C with a modified buffer. Appropriate amounts of genomic DNA (10 fg to 100 ng) were mixed thoroughly with 50 μl of reaction buffer

containing random hexamers, deoxynucleotides, and 2 μl of an enzyme mixture. Reactions were stopped by heating the mixtures at 65°C for 10 min, and the amplified products were quantified as described above and visualized on 1% agarose gels. The effects of *Escherichia coli* single-strand binding protein (SSB) (267 ng/ μl), spermidine (0.1 mM), betaine (1 M), RecA protein (260 ng/ μl), and dimethyl sulfoxide (DMSO) (1%) individually and in combination on amplification biases and yields were examined. The effects of amplification time and DNA template concentration on amplification were also assessed based on the optimized buffer.

Microarray construction. Whole-genome microarrays for *S. oneidensis* MR-1 (~4.9 Mb), a metal-reducing bacterium, *R. palustris* (4.8 Mb), a photosynthetic bacterium, and *N. europaea*, an ammonium-oxidizing bacterium (2.7 Mb), were constructed as described previously (11, 15) in order to evaluate the representation of the whole-community genome amplification procedure. The total gene coverage for the three whole-genome arrays ranged from 95 to 99%. PCR products or 50-mer oligonucleotide probes in 50% DMSO were spotted in duplicate onto aminopropyl silane-coated Ultra GAPS glass slides (Corning, Corning, NJ) or Superamine glass slides (TeleChem International, Inc., Sunnyvale, CA). The printing quality was evaluated by direct scanning of the slides, PicoGreen (Molecular Probes Inc., Eugene, OR) staining, and direct genomic DNA hybridization. Arrays were postprocessed by following the instructions of the slide manufacturers.

A community genome array consisting of whole genomic DNAs from 16 bacterial strains was constructed as described by Wu et al. (29) in order to determine the representation and quantitation of whole-community genome amplification for an artificial microbial community. The CGA was composed of multiple microbial species whose G+C contents ranged from 43 to 68%. Five *Saccharomyces cerevisiae* genes were included in this array as negative controls. All 21 probes (including negative controls) were arranged as a matrix consisting of 15 rows and two columns (designated columns *a* and *b*). Genomic DNA samples were prepared for deposition, printed, and postprocessed as described above. Each glass slide contained three replicates of genomic DNA from individual strains.

An oligonucleotide (50-mer) functional gene array for monitoring bioremediation and nutrient cycling was constructed using the methods described previously (21, 25), and it was used to evaluate whole-community genome amplification. This FGA contained probes from various groups of genes involved in degradation of organic contaminants, metal resistance, and nutrient cycling. A total of 2,006 oligonucleotide probes were printed in duplicate on each slide. Information concerning the probe sequences, predicted melting temperatures, organismal origins, and gene functions is available at <http://www.esd.ornl.gov/facilities/genomics/index.html>. Probes from six human genes and four plant genes were included on the microarrays as negative or quantitative controls. In addition, two highly conserved 16S rRNA gene probes were included as positive controls. Probes were prepared for microarray deposition, printed, and postprocessed as described previously (21, 25).

DNA labeling and hybridization. Genomic DNA or DNA amplified from a small amount of genomic DNA by WCGA was fluorescently labeled using the random priming method and was purified as described previously (21, 25, 28). All microarray experiments were performed in triplicate, unless indicated otherwise, so that statistical analyses could be performed. Each hybridization solution (total standard volume, 30 μl) contained denatured fluorescently labeled genomic DNA, 50% formamide, $3 \times \text{SSC}$ ($1 \times \text{SSC}$ is 150 mM NaCl plus 15 mM sodium citrate), 2 μg of unlabeled herring sperm DNA (Promega, Madison, WI), and 0.3% sodium dodecyl sulfate (SDS). The hybridization solutions were heated at 95°C for 3 min and were kept warm in a 50°C incubator. Microarray slides, coverslips, and pipette tips were warmed and were also kept warm in an incubator prior to the hybridization. Microarrays were placed into self-contained flow cells (Telechem International) in a 50°C water bath immediately for overnight hybridization. Following hybridization, coverslips were removed in prewarmed washing buffer ($1 \times \text{SSC}$ -0.2% SDS) and then washed sequentially for 5 min in $1 \times \text{SSC}$ -0.2% SDS and $0.1 \times \text{SSC}$ -0.2% SDS and for 30 s in $0.1 \times \text{SSC}$ before they were air dried in the dark.

Microarray scanning and data processing. A ScanArray 5000 microarray analysis system (Perkin-Elmer, Wellesley, MA) was used to scan microarrays. A quick scan at a resolution of 50 μm was performed prior to the real scanning at a resolution of 10 μm , and the laser power and photomultiplier tube gain were adjusted to avoid saturation of spots and to make the two fluorescence channels comparable. Scanned image displays were saved as 16-bit TIFF files and were analyzed by quantifying the pixel density (intensity) of each spot using ImaGene, version 5.0 (Biodiscovery, Inc., Los Angeles, CA). The mean signal intensity was determined for each spot, and the local background signal was subtracted automatically from the hybridization signal of each spot. The fluorescence intensity

values for all replicates of the negative control genes, of 10 *Arabidopsis thaliana* genes for the three whole-genome arrays, of five yeast genes for the small community genomic DNA arrays, or of the human genes for the FGAs were averaged, and then the averages were subtracted from the background-corrected intensity values for the hybridization signals. The signal-to-noise ratio (SNR) was also calculated based on the following formula of Verdnik et al. (26): $SNR = (\text{signal intensity} - \text{background}) / \text{standard deviation of background}$. Spots with SNR that were less than 3 were defined as poor spots.

The outliers, represented by data points that were not consistently reproducible and had a disproportionately large effect on the statistical results, were detected and removed at a P value of <0.01 . When the absolute value of a data point minus the mean was greater than 2.90σ , the data point was considered an outlier and removed. To make sure that different treatments in the experiments for testing additives, different genomes, template concentrations, and mixtures were comparable, poor spots and outliers were removed based on hybridizations only with the nonamplified genomic DNA.

The signal intensities of the WCGA DNA (Cy5) and the nonamplified genomic DNA (Cy3) were normalized based on the mean signal intensity for all genes on the arrays. Briefly, the mean signal intensity for all of the genes on an array in each channel was calculated. Since the same amounts of amplified and nonamplified genomic DNAs were used for labeling and hybridization, we expected that the average signal intensities for all of the genes would be approximately equal. Thus, a coefficient was obtained by dividing the mean signal intensity from the Cy5 channel by the mean signal intensity from the Cy3 channel. Then the signal intensities of individual genes from the Cy3 channel were multiplied by this coefficient to obtain normalized signal intensities. For the microarray data for community genomic DNA arrays and 50-mer oligonucleotide arrays, normalization was performed using the mean for the spiked internal positive control genes. The normalized microarray data were then used for further analysis.

Data analysis. Three indexes were used to evaluate amplification representativeness. The first index was representational bias (D_j^{total}). $R_{i,j}$ was the ratio of the signal intensity with amplified DNA to the signal intensity with genomic DNA for the i th gene in the j th experiment, and $LR_{i,j} = \log_{10}R_{i,j}$. If the signal intensity with amplified DNA was equal to the signal intensity with genomic DNA for the i th gene in the j th experiment, then the ratio was 1 or the log ratio was 0. Similar to Euclidean distance, we defined D_j^{total} as the average distance of the log ratio from the reference point, 0, where no bias was introduced during amplification and hybridization. Then,

$$D_j^{\text{total}} = \sqrt{\sum_{i=1}^{N_j} (LR_{i,j} - 0)^2 / N_j} = \sqrt{\sum_{i=1}^{N_j} (LR_{i,j})^2 / N_j}$$

where N_j is the number of genes detected in unamplified genomic DNAs in the j th experiment. D_j^{total} describes the overall average representational bias for the j th experiment. D_j^{total} is equal to 0 if there is no bias. The smaller the D_j^{total} , the smaller the bias. However, there is no upper limit for the value of D_j^{total} . Thus, D_j^{total} is more meaningful for relative comparisons.

The second index was the percentage of genes whose ratios of amplified DNA to nonamplified genomic DNA were significantly different from the reference ratio, 1, at a P value of 0.01. This index described the percentage of the genes in an amplified sample that were significantly different from genes in the nonamplified genomic DNA sample. The smaller the value, the less the bias contributed by the amplification.

The third index used was the percentage of genes for which the hybridization ratio of amplified DNA to nonamplified genomic DNA was larger than the indicated fold change (i.e., 1.5-, 2-, 3-, and 4-fold).

For the analysis of microarray data for FRC samples, cluster analysis was performed using the pairwise average-linkage hierarchical clustering algorithm (9) provided in the CLUSTER software (<http://rana.stanford.edu>), and the results of hierarchical clustering were visualized using the TREEVIEW software (<http://rana.stanford.edu/>). A standard t test was used to test the significance of array data for different treatments. Principal-component analysis and canonical analysis were also performed using SYSTAT 10.0 and SAS for comparing the microarray data for the FRC samples and the chemical data for the sampling sites.

RESULTS AND DISCUSSION

Optimization of amplification conditions. It is generally believed that certain additive reagents, such as SSB, spermidine,

the RecA protein, betaine, and DMSO, might increase enzymatic amplification of DNA by various mechanisms, such as removing inhibitors, breaking up GC-rich regions, protecting single-stranded DNA, or increasing the local concentrations of macromolecules, and hence lead to higher enzyme reaction efficiency (1). To improve the amplification efficiency, reaction buffers containing various additive reagents were evaluated. One nanogram of *R. palustris* genomic DNA was amplified in triplicate using the commercial buffer with or without additive reagents. Compared to the DNA yield with the commercial buffer without additive reagents, the modified buffer containing SSB or spermidine substantially improved the DNA yield (49 to 66%), whereas the buffer containing DMSO or betaine decreased the amplification efficiency (11 to 14%). Only a slight increase (16%) in the DNA yield was observed with the buffer containing the RecA protein (see Fig. S1 in the supplemental material). The buffer containing both SSB and spermidine resulted in slightly higher yields, and the amplification reactions reached plateau phases earlier (about 4 h) than the amplification reactions with the buffer containing betaine, DMSO, or no additive reagent reached plateau phases (approximately 5 h) (see Fig. S1 in the supplemental material).

Considerable improvements in sequence representation in the amplified DNAs were obtained with the buffers containing additive agents when they were tested with the whole-genome open reading frame (ORF) arrays (11, 15) (see Table S1 in the supplemental material). For instance, the overall average representational bias for the buffer containing both SSB and spermidine (0.107/0.045) was more than twofold lower than the overall average representational bias for the commercial buffer without additive reagents and was comparable to or even lower (0.078/0.045) than the overall average representational bias observed with nonamplified genomic DNAs (see Table S1 in the supplemental material). The proportions of the genes whose hybridization ratios were significantly different from the reference point, 1, were considerably less for the buffer containing both SSB and spermidine (0.4%) than for the commercial buffer (8.9%) (see Table S1 in the supplemental material). In addition, the proportions of the genes whose hybridization signal ratios (amplified DNA/genomic DNA) exhibited >2 -fold changes were substantially lower for the modified buffer containing both SSB and spermidine (0.2%) than for the commercial buffer (1.5%). These results suggest that use of the additive reagents could substantially improve sequence representation in amplified samples. However, compared to the commercial buffer, little effect on the sequence representation was observed for the buffer containing the RecA protein. Therefore, the buffer containing both SSB and spermidine was used in all subsequent experiments.

Amplification sensitivity. The amplification sensitivity with the modified buffer containing both SSB and spermidine was determined using a series of 10-fold genomic DNA dilutions that resulted in amounts ranging from 1 fg to 1 ng. Very robust amplification (more than 7 μg) was obtained with amounts of genomic DNA as small as 10 fg, whereas no DNA amplification was observed with 1 fg of template DNA (Fig. 1A), suggesting that the amplification sensitivity is between 1 and 10 fg DNA. This is equivalent to the average DNA content of approximately one or two bacterial cells (assuming that the DNA content is 5 fg per cell, like the DNA content of *E. coli*), and

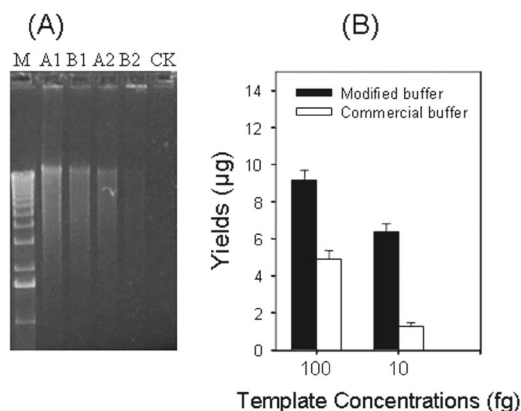


FIG. 1. Improved amplification sensitivity with modified reaction buffer. (A) Comparison of MDA amplification products obtained by using an optimized buffer containing SSB and spermidine (lanes A1 and A2) and an unmodified commercially available buffer (lanes B1 and B2) with 100 fg (lanes A1 and B1) and 10 fg (lanes A2 and B2) of template DNA and a control with no DNA (lane CK). (B) Comparison of DNA yields. The error bars are based on three replicate samples.

the sensitivity is up to 10-fold higher than the sensitivity of the commercial buffer (Fig. 1B). The sensitivity which we obtained is very similar to that obtained by Raghunathan et al. (20) but is 7 orders of magnitude higher than that obtained by Vora et al. (27).

Representative amplification with different species. The representativeness of amplification from three different genomes (*R. palustris*, *S. oneidensis* MR-1, and *N. europaea*) was determined using whole-genome ORF microarrays (11, 15). As expected, the hybridization signal ratios were aligned along a line corresponding to ratios close to 1:1 for the genomes examined (see Fig. S2 in the supplemental material). The representative bias in the amplified DNA was very similar to or slightly lower than that in the cohybridized nonamplified DNA (Table 1) for all organisms examined except *N. europaea*, whose representative bias was about six times higher in amplified DNA than in nonamplified genomic DNA. Also, the proportions of the genes whose hybridization ratios were significantly different

TABLE 2. Effect of DNA template concentration on representative amplification with *R. palustris* DNA

Parameter	Amt of DNA template ^a			
	10 ng	1 ng	100 pg	10 pg
Total no. of genes	4,867	4,838	4,840	4,724
Representational bias	0.045	0.227	0.545	0.592
SDG _{0.01} ^b	0.4	13.9	29.4	32.2
F _{1.5} ^c	1.0	40.1	81.8	76.2
F _{2.0}	0.2	17.6	68.9	76.9
F _{3.0}	0	4.4	48.9	52.9
F _{4.0}	0	1.4	32.5	34.5

^a Different amounts of genomic DNA from *R. palustris* were amplified for 4 h in triplicate. The average yields were ~1.5, 8, 21, and 25 µg, corresponding to original template amounts of 0.01, 0.1, 1, and 10 ng, respectively. Amplified and genomic DNAs (2 µg) were labeled with Cy3 and Cy5, respectively, in triplicate and cohybridized with a whole-genome ORF array.

^b See Table 1, footnote d.

^c See Table 1, footnote e.

from the reference point, 1 ($P < 0.01$), were comparable for the amplified DNA and the nonamplified genomic DNA for all species (Table 1). In addition, the proportions of the genes whose hybridization signal ratios (amplified DNA/genomic DNA) showed >2-fold differences were <0.5% for all genomes examined, and no genes showed >3-fold changes (Table 1). These results indicated that the WCGA-assisted microarray hybridization was highly representative. Additionally, this improved method showed considerably less representational bias than the methods used for MDA-amplified human, yeast, and *E. coli* DNAs (6, 14, 19).

Effects of DNA template concentrations on representational bias. Since the amplification process is random, representational bias could be dependent on the template DNA concentration. To evaluate the effects of DNA template concentration on representational biases, amounts of genomic DNA from *R. palustris* ranging from 10 pg to 10 ng were amplified (1.5 to 25 µg after 4 h), and the normalized ratios of the amplified DNA to the nonamplified DNA were analyzed as described above. The template DNA concentration had dramatic effects on the overall average representational bias. For instance, the overall

TABLE 1. Representative amplification for three microbial genomes

Parameter	<i>R. palustris</i>		<i>N. europaea</i>		<i>S. oneidensis</i>	
	Genomic DNA ^a	RCA ^b	Genomic DNA ^a	RCA ^b	Genomic DNA ^a	RCA ^b
Total no. of genes ^c	4,670	4,867	2,116	1,797	4,683	4,925
Representational bias	0.078	0.045	0.013	0.074	0.024	0.021
SDG _{0.01} ^d	2.4	0.4	0.2	2.0	0.4	0.2
F _{1.5} ^e	4.026	1.025	0.000	4.508	0.000	0.000
F _{2.0}	0.514	0.182	0.000	0.000	0.000	0.000
F _{3.0}	0.000	0	0.000	0.000	0.000	0.000
F _{4.0}	0.000	0	0.000	0.000	0.000	0.000

^a Genomic DNAs (2 µg) were labeled with both Cy3 and Cy5 in triplicate and cohybridized with whole-genome ORF arrays.

^b Genomic DNAs (10 ng) from individual genomes were amplified for 4 h in triplicate. The amplified DNA (2 µg) was labeled with Cy5, whereas the nonamplified genomic DNA (2 µg) was labeled with Cy3. Both Cy3- and Cy5-labeled DNAs were cohybridized with whole-genome ORF arrays.

^c Poor spots and outliers for hybridization with nonamplified genomic DNA were removed and not considered for data analysis. The number of genes is the total number of effective genes used for calculating various indexes.

^d SDG_{0.01}, percentage of genes whose ratios of amplified DNA to nonamplified genomic DNA are significantly different from the reference ratio, 1, at a P value of 0.01.

^e F_{1.5}, F_{2.0}, F_{3.0}, and F_{4.0}, percentages of genes whose hybridization ratios of amplified DNA to nonamplified genomic DNA are more than 1.5-, 2.0-, 3.0-, and 4.0-fold, respectively.

TABLE 3. Representative amplification and detection of equally mixed genomic DNAs^a

Parameter	Genomic DNAs (10 ng) from:			Genomic DNAs (1 ng) from:		
	<i>R. palustris</i>	<i>S. oneidensis</i>	<i>N. europaea</i>	<i>R. palustris</i>	<i>S. oneidensis</i>	<i>N. europaea</i>
Total no. of genes	4,445	4,942	1,861	4,733	5,015	1,555
Representational bias	0.088	0.045	0.161	0.187	0.272	0.243
SDG _{0.01} ^b	2.8	0.5	5.3	14.2	11.9	10.6
F _{1.5} ^c	5.5	0.6	29.2	32.1	50.5	48.4
F _{2.0}	0.3	0.1	6.1	9.2	26.4	23.1
F _{3.0}	0.02	0.02	0.0	1.7	7.7	4.3
F _{4.0}	0.02	0.02	0.0	0.6	3.1	0.3

^a Genomic DNAs (10 or 1 ng) from the species were equally mixed and amplified with RCA for 4 h. Then 2- μ g portions of amplified and nonamplified genomic DNAs were labeled with Cy3 and Cy5, respectively, in triplicate and cohybridized separately with each whole-genome ORF array.

^b See Table 1, footnote *d*.

^c See Table 1, footnote *e*.

average representational bias was more than threefold higher with 1 ng of DNA template than with 10 ng of DNA template (Table 2). While the overall average representational biases were similar when 10 and 100 pg of DNA were used, they were twofold higher than the representational bias when 1 ng of DNA was used. Also, the proportion of the genes whose hybridization ratios were significantly different from the reference point at a *P* value of 0.01 increased as the DNA template concentration decreased (Table 2). Around 30% of the ORFs were significantly different (*P* = 0.01) with 10 and 100 pg of template. The majority of the ORFs (69% to 77%) showed >2-fold differences with the lower concentrations (10 and 100 pg), whereas only a small portion of the ORFs (0.2 to 18%) showed >2-fold differences with 1 and 10 ng. About 33 to 34% of the genes showed >4-fold differences with 10 and 100 pg of template DNA, whereas less than 1.3% of the ORFs showed >4-fold differences with 1 and 10 ng of template DNA. These results suggest that depending on the level of precision required, various amounts of DNA can be used and that very representative detection can occur with as little as 1 ng DNA in the absence of other nontarget DNA templates.

Representational detection with artificial communities of mixed species. To determine whether representative detection can occur with mixed community samples, equal quantities of template DNAs (1 or 10 ng) from *Shewanella*, *Rhodopseudomonas*, and *Nitrosomonas* were mixed, amplified (~18 μ g after 4 h), and subjected to hybridization and analysis as described above. With 30 ng of DNA template (10 ng from each species), the overall average representational bias in *Nitrosomonas* (0.1609) was about twice the overall average representational bias in *Rhodopseudomonas* (0.0884), which was twice the overall average representational bias in *Shewanella* (0.0447). The overall average representational bias for individual genomes with mixed DNA was about 1.3 to 2.2 times the overall average representational bias in the absence of other genomic DNA, suggesting that the presence of DNA from other species has effects on the amplification of individual genomes. However, the proportions of the genes whose hybridization ratios were significantly different from the reference point at a *P* value of <0.01 in mixed DNA templates (0.5 to 5.3%) (Table 3) were comparable to the proportions observed in the absence of other DNA templates (0.2 to 3.5%) (Table 2). Also, very small proportions of the genes (0.1 to 6.3%) showed twofold differences, but none of the genes showed a threefold difference.

These results were comparable to the observed in the absence of other DNA templates. These results indicate that although the presence of other DNA templates has effects on representative amplification, the effects appear to be very small.

When the total amount of DNA template was 3 ng (1 ng from each species), the differences in the average overall representational biases among *Rhodopseudomonas* (0.1865), *Shewanella* (0.2717) and *Nitrosomonas* (0.2433) were considerably less than the differences when 10-ng portions of the DNAs of these organisms were mixed (Table 3), but the average overall representational biases were substantially higher (~1.5- to 6-fold) with 3 ng of mixed DNA than with 30 ng of mixed DNA. For about 10 to 14% of the genes the hybridization ratios were significantly different from the reference point, 1 (at *P* < 0.01), which is substantially higher than the percentage when a total of 30 ng of DNA (0.5 to 5.3%) was used (Table 3). Although about 9 to 26% of the genes showed >2-fold differences, the proportions of the genes that showed threefold differences (1.7 to 7.7%) were much lower. Again, the results indicate that the DNA concentration has significant effects on the overall performance of the WCGA-based microarray detection approach.

To understand the effects of mixed templates on the amplification performance at lower concentrations better, the hybridization ratio data for *Rhodopseudomonas* obtained with 1 ng of DNA in different experiments were compared further. The overall average representational bias, the proportions of the genes showing values significantly different from the reference point, and the proportions of the genes having >1.5-, >2-, >3-, or >4-fold differences were very similar or even less for mixed DNA templates (Table 3) than for templates lacking other DNAs (Table 2). The results indicated that the presence of nontarget DNA templates could improve the amplification performance with low DNA template concentrations.

In natural microbial communities, not all species are equally abundant. To determine the representational bias in the context of environmental applications, genomic DNAs from *Rhodopseudomonas* (10 ng), *Shewanella* (1 ng), and *Nitrosomonas* (0.1 ng) were mixed, amplified (~17 μ g after 4 h), and then hybridized in triplicate and analyzed as described above. Similar to the results described above, the overall average representational bias increased as the DNA concentration decreased (Table 4). Although the representational biases were higher than those observed in other experiments (Tables 2 and

TABLE 4. Representative amplification and detection of unequally mixed genomic DNAs^a

Parameter	<i>R. palustris</i> (10 ng)	<i>S. oneidensis</i> (1 ng)	<i>N. europaea</i> (100 pg)
Total no. of genes	4,630	4,067	2,057
Representational bias	0.109	0.314	0.333
SDG _{0.01} ^b	1.9	10.9	25.4
F _{1.5} ^c	9.8	59.4	55.2
F _{2.0}	0.9	34.1	33.4
F _{3.0}	0.1	13.1	13.4
F _{4.0}	0.02	5.9	6.9

^a Different concentrations of genomic DNAs from the species were mixed and amplified for 4 h. Then 2- μ g portions of the amplified and nonamplified genomic DNAs (2 μ g) were labeled with Cy3 and Cy5, respectively, in triplicate and cohybridized separately with each whole-genome ORF array.

^b See Table 1, footnote d.

^c See Table 1, footnote e.

3), the percentages of genes whose ratios of amplified DNA to nonamplified genomic DNA are significantly different from the reference ratio, 1, at a *P* value of 0.01 and the proportions of the genes having >1.5-, >2-, >3-, or >4-fold differences were very similar. These results suggest that WCGA-assisted microarray hybridization with mixed DNAs whose concentrations are not equal can representatively detect the target genomes of interest.

Although individual genes in a genome could be amplified unequally with different template concentrations, the overall amplification at the whole-genome level could be equal because the genes which are under- or overrepresented could compensate for each other. To test this hypothesis, equal quantities of genomic DNAs from nine species were mixed (100 pg each) and amplified with MDA for 2 h in triplicate. The amplified genomic DNAs were labeled with Cy5. Then 1.111 μ g of the amplified genomic DNA was cohybridized with 1.111 μ g of nonamplified genomic DNA using CGA containing the whole-genome DNA as a probe. No significant differences (*P* = 0.05) between the hybridization signal intensity for individual genomes with the amplified genomic DNA and the hybridization signal intensity for individual genomes with the nonamplified genomic DNA were observed (Fig. 2). Significant correlations between the average signal intensity for the amplified DNA and the average signal intensity for the nonamplified DNA were obtained for the nine microbial genomes (r^2 = 0.64). These results indicate that the amplification at the whole-genome level is representative.

Quantitation of WCGA-assisted microarray hybridization.

The quantitative power of WCGA-assisted microarray hybridization was first determined using *S. oneidensis* MR-1 whole-genome ORF arrays (11). Genomic DNA of *S. oneidensis* MR-1 was diluted in fivefold series to obtain amounts ranging from 0.04 to 125 ng. To make sure that the amplified DNAs were in the exponential phase, the diluted DNAs were amplified for 2 h in triplicate. All of the amplified DNAs were labeled with Cy5 and cohybridized with Cy3-labeled nonamplified genomic DNA. At a *P* value of 0.05, for ~80% of the 4,173 effective genes there was a significant linear relationship (r^2 = 0.65 to 0.99) between signal intensity and the initial amount of DNA for amounts ranging from 0.04 to 125 ng, while for 86% of the genes there was a significant linear rela-

tionship at a *P* value of 0.1 (r^2 = 0.53 to 0.99) (Fig. 3A). These results suggest that WCGA-assisted microarray hybridization is quantitative for the vast majority of genes. The quantitative nature of microarray-based hybridization is consistent with the findings of microarray studies of gene expression (4, 7, 15, 24).

To determine whether WCGA-assisted microarray hybridization is quantitative for target organisms in the presence of other nontarget DNAs, the quantitative relationships between signal intensity and DNA concentration were examined further using a CGA containing the entire genomic DNAs from five bacteria representing different genera and species. A strong linear relationship (r^2 = 0.91) between the signal intensity for the amplified DNAs and the signal intensity for the nonamplified DNAs for amounts ranging from 0.1 to 1,000 ng was obtained (Fig. 4A), and this relationship was similar to that observed for self genomic DNA-genomic DNA hybridization (Fig. 4B). These results suggest that the overall WCGA-assisted, CGA-based microarray hybridization method is quantitative for mixed DNA templates.

Representative and quantitative detection of environmental samples. To further test the representative and quantitative nature of WCGA-assisted microarray detection with real environmental samples in which the microbial community structure was more complex than the structure of the mixed artificial communities used in the experiments described above, an ethanol-amended, uranium-contaminated groundwater was analyzed. Unlike the study described below, in this sample biomass was not limiting as the groundwater had been repeatedly fed ethanol (13) and high DNA yields were observed. This allowed us to use approaches similar to those utilized for the controlled evaluations described above with an actual environmental sample. The purified community DNA was diluted to

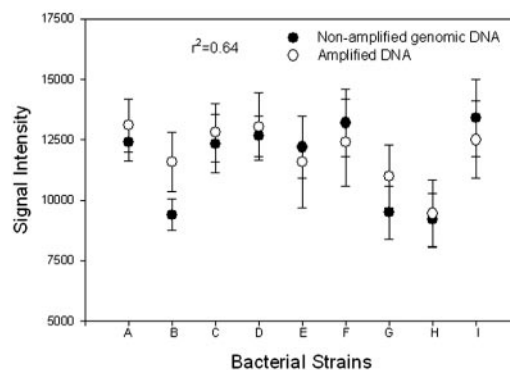


FIG. 2. Representational detection determined using community genome arrays. Equal quantities (100 pg) of DNAs from nine distinctly different bacteria, *P. denitrificans* (A), *T. aromatica* (B), *R. palustris* (C), *S. oneidensis* MR-1 (D), *O. anthropi* (E), *Marinobacter* sp. strain D5-10 (F), *Pseudomonas* sp. strain C179 (G), *P. mendocina* (H), and α -proteobacterium C1-4 (I), were mixed and amplified with MDA for 2 h in triplicate. Equal amounts (1.111 μ g) of amplified genomic DNA and nonamplified genomic DNA (0.1 μ g from each species) were labeled with Cy5 and Cy3, respectively, and cohybridized to the community genome arrays containing the whole-genome DNAs as probes. No significant differences (*P* = 0.05) between the hybridization signal intensity for the amplified DNA and the hybridization signal intensity for the nonamplified genomic DNA were observed for individual genomes.

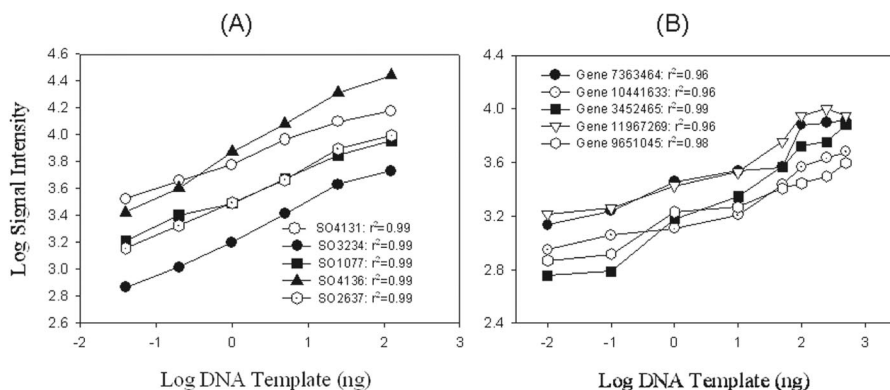


FIG. 3. Quantitation of WCGA-assisted microarray hybridization. (A) Quantitative relationship between the signal intensity and the concentration of DNA from a pure culture. Genomic DNA from *S. oneidensis* MR-1 was diluted fivefold to obtain concentrations ranging from 0.04 to 125 ng (125, 25, 5, 1, 0.2, and 0.04 ng). Diluted DNAs were obtained at the exponential phase (2 h) in triplicate. All amplified DNAs were labeled with Cy5 and cohybridized with Cy3-labeled, nonamplified genomic DNA (2 μg). The average signal intensities of individual genes at each dilution were determined, and a linear regression model was fitted for signal intensities and DNA concentrations for each gene. Overall, for 80% of the 4,173 genes there was a significant linear relationship between signal intensity and template DNA concentration ($r^2 = 0.65$ to 0.99 ; $P = 0.05$). The quantitative relationships of five representative genes are shown. (B) Quantitative relationship between the signal intensity and the concentration of DNA from a biostimulated groundwater sample. The purified community DNA was diluted to obtain amounts ranging from 0.01 to 500 ng, amplified, labeled, and hybridized to functional gene arrays. For all of the genes detected there was a significant linear relationship between signal intensity and DNA concentration. The quantitative relationships for five genes are shown.

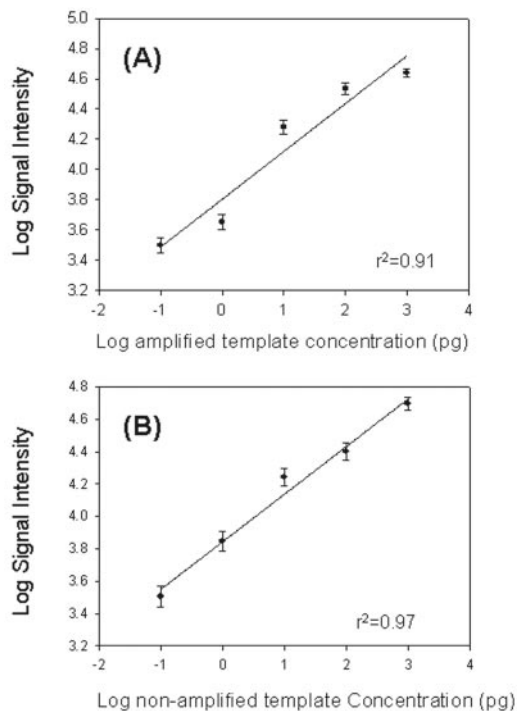


FIG. 4. Quantitation of WCGA-assisted CGA-based microarray hybridization. Unequal amounts of genomic DNAs (0.1, 1, 10, 100, and 1,000 ng) from five bacterial species representing different genera and species (*P. denitrificans*, *T. aromatica*, *R. palustris*, *S. oneidensis* MR-1, and *O. anthropi*) were mixed and amplified in triplicate for 2 h. Then the same amount of amplified DNA (total, 1,111.1 ng) was labeled with Cy3 or Cy5 and cohybridized with the same amount of Cy5- or Cy3-labeled nonamplified DNA mixtures. (A) Cy5-labeled amplified DNA. (B) Nonamplified genomic DNA.

obtain amounts ranging from 0.01 to 500 ng, amplified, labeled, and hybridized to an FGA. Forty-two of the 2,006 probes on the array were derived from previous studies of dissimilatory sulfite reductase (*dsrAB*) (3a) and nitrite reductase (*nirS* and *nirK*) (30) genes from the same site. Altogether, 61 genes were detected in nonamplified DNA samples (Table 5). When more than 1 ng of template DNA from the same sample was used for WCGA-assisted microarray analysis, 93 to 98% of the same genes were detected, and more than 50% of the genes were detected even with amounts of template DNA as small as 10 pg (Table 5). When the amount of initial community DNA used for WCGA was greater than 1 ng, the

TABLE 5. Representative detection of genes from an environmental sample

Parameter	Amt of DNA (ng) ^a							
	500	250	100	50	10	1	0.1	0.01
No. of genes detected ^b	59	60	60	60	59	57	40	31
% of genes detected ^c	96.7	98.3	98.3	98.3	96.7	93.4	65.5	50.8
Representational bias	0.164	0.127	0.150	0.125	0.076	0.098	0.465	0.452
SDG _{0.01} ^d	4.2	3.8	6.3	4.5	2.9	5.3	53.0	61.2
F _{1.5} ^e	3.4	8.3	18.3	6.7	1.7	5.3	92.5	58.1
F _{2.0}	3.3	3.3	5.0	3.3	1.7	0	80.0	48.4
F _{3.0}	3.4	1.7	1.7	1.7	0	0	42.5	29.0
F _{4.0}	3.4	1.7	1.7	1.7	0	0	17.5	19.4

^a Amount of community DNA used for amplification. Community DNA was isolated from a groundwater sample, diluted, amplified for 2 h in triplicate with RCA, and labeled with Cy3. Two micrograms of nonamplified community DNA was also labeled with Cy5 in triplicate, mixed with Cy3-labeled amplified DNA, and cohybridized on the 50-mer oligonucleotide arrays.

^b Number of genes detected whose SNR was greater than 3.

^c Percentage of genes detected in the amplified DNA samples compared to the total number of the genes detected in the nonamplified DNA sample. A total of 61 genes were detected in the nonamplified DNA sample.

^d See Table 1, footnote d.

^e See Table 1, footnote e.

TABLE 6. Proportion of unique genes in individual groundwater wells and matrix representation of the number of overlapping genes for wells

Parameter	FW300	FW003	FW021	FW010	FW024
No. (%) of unique or overlapping genes ^a					
FW300	61 (20) ^b	189 (36)	174 (35)	80 (21)	111 (23)
FW003		25 (11) ^b	144 (35)	61 (17)	84 (20)
FW021			10 (5) ^b	64 (20)	90 (24)
FW010				6 (5) ^b	118 (37)
FW024					30 (16) ^b
Total no. of genes detected	302	219	192	130	190
Simpson's genetic diversity index ^c	125.5	67.1	26.6	17.4	35.7

^a Unless indicated otherwise, the values are values for overlapping genes.

^b Values for unique genes.

^c Simpson's inverse diversity index was calculated using the program Estimates S (<http://vicroy.eeb.uconn.edu/estimates>) with the relative signal intensity ([measured/slide average] × 100) used as a measure of abundance. Thus, diversity in this case indicates the collective detected genetic diversity of each sample rather than species diversity.

representational bias was also low, and there were ≥ 2 -fold differences between nonamplified and amplified samples for less than 5% of the genes (Table 5). These results indicated that WCGA-assisted microarray hybridization-based detection was representative with real environmental samples containing as little as 1 ng of community DNA. When these results were compared to the results described above that were obtained with pure cultures and artificial mixed communities, it appeared that 1 ng of DNA was needed to obtain representative detection by WCGA-assisted microarray hybridization, whether the DNA source was homogeneous or heterogeneous.

The quantitative relationships between the hybridization signal intensity and the template genomic DNA used for amplification were analyzed further. The genes detected showed significant linear relationships for signal intensities with amounts of DNA ranging from 0.01 to 250 ng ($r^2 = 0.96$ to 0.98) (Fig. 3B). Thus, these results suggest that WCGA-assisted microarray hybridizations are quantitative for real environmental samples for a wide range of DNA concentrations.

Interestingly, although the results described above indicate that it might be difficult to obtain representative amplification when the total amount of DNA template is less than 1 ng, some genes can still be amplified when the total amount of DNA template is less than this amount, and there are still quantitative relationships for these amplified genes. In this example, quantitative relationships still occurred with as little as 0.01 ng of DNA. However, it should be noted that this was the smallest amount of template DNA examined and that the lower limit could vary with the complexity of the samples and the type of microarray format.

Application of WCGA to low-biomass groundwater communities. We used the optimized approach described above to evaluate the effects of contaminants on microbial community structure and the adaptation of microbial communities to environmental conditions in groundwater samples. Unlike the samples used in the evaluation described above, the samples originated from unstimulated communities with typical concentrations of $\sim 10^4$ cells ml⁻¹ and nanogram DNA yields.

Thus, conventional microarray analysis of these samples could not have been performed due to the very low biomass.

To analyze the microbial community structure, cells were collected from 2 liters of groundwater, the DNA was extracted, purified, and resuspended in 20 μ l of water, and then 1 μ l of the preparation was used as a template for amplification for 2 h. All of the amplified DNA was used for hybridization. At least three replicate amplifications and hybridizations were performed for each sample. As determined by the WCGA-assisted microarray hybridization approach, more than 400 genes showed statistically significant positive hybridization signals (Table 6) (see Fig. S3 in the supplemental material). As expected, the highest number of genes was detected for uncontaminated background samples (well FW300), while the lowest number of genes was detected for the highly contaminated sample (well FW010) (Table 6). The overall genetic diversity detected (Table 6) in each of the groundwater samples suggested that contaminants had strong effects on the microbial communities. Simpson's diversity index indicated that the levels of genetic diversity in the uncontaminated background well and less contaminated down-gradient well (wells FW300 and FW003, respectively) were much higher than the levels of genetic diversity in the more heavily contaminated wells in areas 1 and 3 (wells FW021, FW024, and FW010). The expected observation that the diversity at contaminated sites was substantially lower than the diversity in noncontaminated samples also suggested that the WCGA-assisted microarray

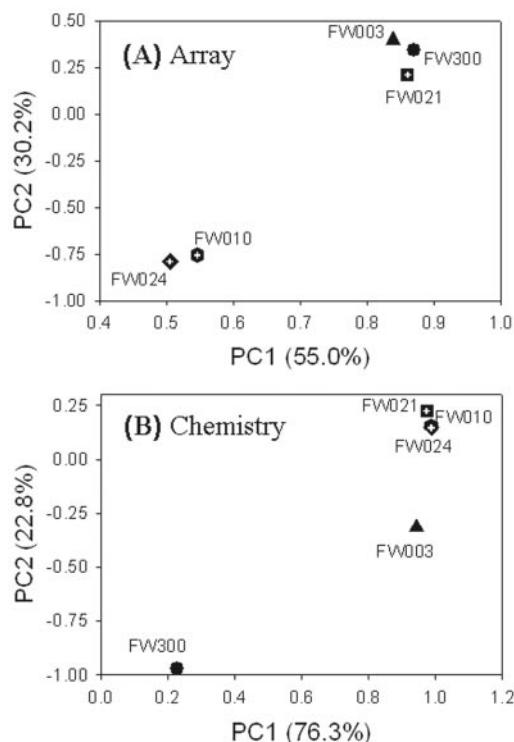


FIG. 5. Principal-component analysis of microarray hybridization signal intensity data (A) and groundwater chemistry and contaminant concentration data (B). Some of the chemical factors used in the analysis were uranium, nitrate, and aluminum. See Table S2 in the supplemental material for more information about the levels of contaminants in the groundwater wells.

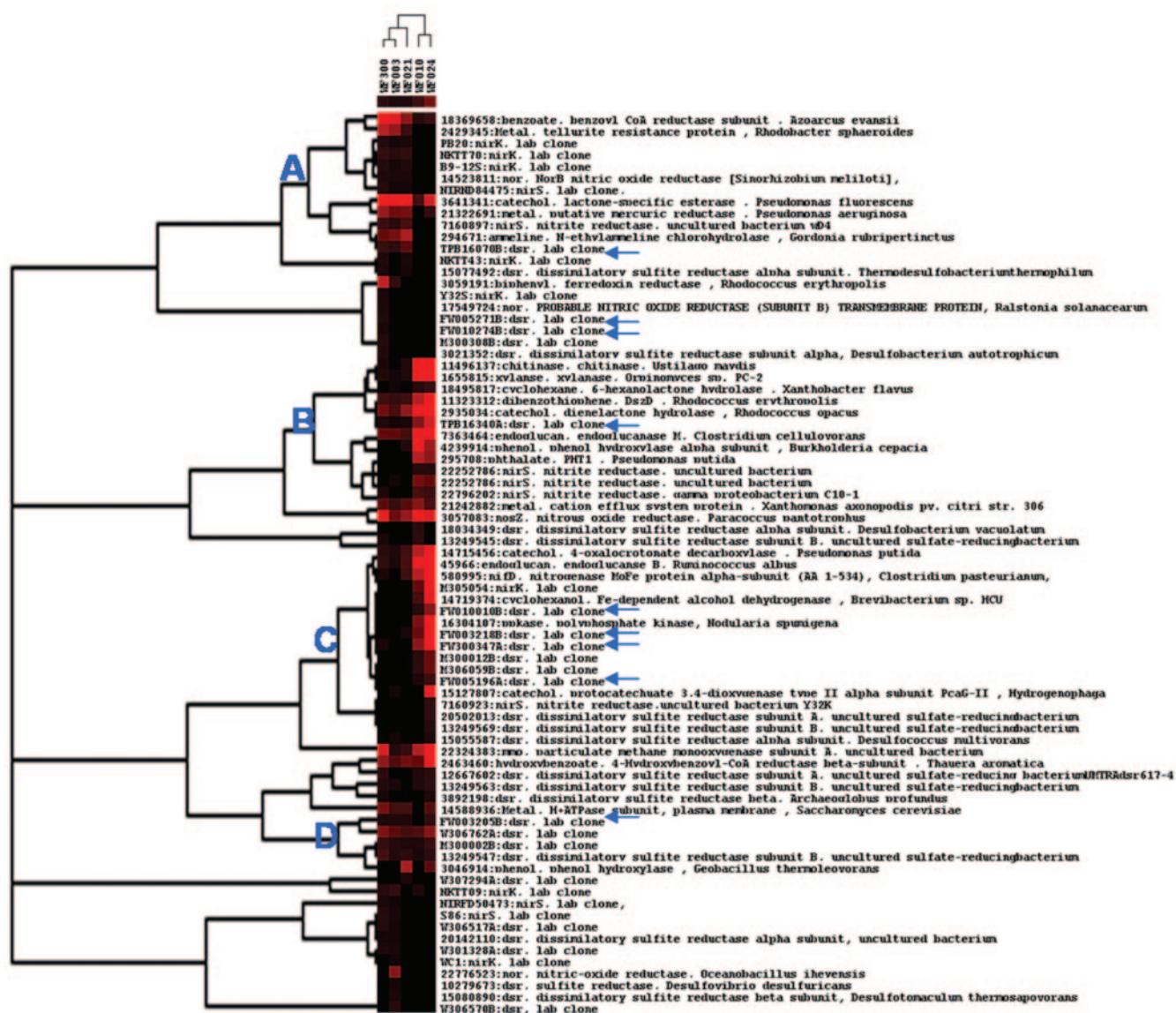


FIG. 6. Hierarchical cluster analysis of gene relationships based on hybridization signal intensity. Representative genes from five groundwater samples collected from contaminated and uncontaminated sites at the FRC are shown. The tree was generated using CLUSTER and was visualized with TREEVIEW. Black indicates no detectable hybridization above the background levels, while red indicates positive hybridization signals. The color intensities indicate differences in hybridization signal intensity. A, B, C, and D indicate similar detection patterns for samples. Genes in clusters A, B, and C showed especially high levels of cooccurrence and similar intensity patterns for the highly contaminated samples (wells FW024 and FW010) or for the background and less contaminated samples (wells FW300 and FW003). The blue arrows indicate *dsrA/B* genes that originated from clone libraries developed from previous FRC groundwater studies. The complete results of this analysis are shown in Fig. S3 in the supplemental material. CoA, coenzyme A; AA, amino acids.

hybridization method is capable of revealing real biological differences.

Principal-component analysis indicated that the highly contaminated samples from wells FW010 and FW024 were clustered together based on both geochemistry and array data (Fig. 5). However, while geochemical data clustered the heavily contaminated FW021 well with wells FW010 and FW024, the relationships based on array data were less clear. The uncontaminated FW300 well and the less contaminated FW003 well were well separated from the heavily contaminated groundwater by both array and geochemical data. These results sug-

gest that the overall community structures were different for these samples, and there was some correlation between contaminants and microbial community structure and composition. However, the relationships between microbial communities and geochemistry were also probably complicated by other factors, such as pH values; the pH values in wells FW003 and FW300 were similar, but they were much greater than the pH values in the three heavily contaminated wells.

Although the contaminant levels and the geochemistry were different for the different wells, the percentages of the genes detected that were shared by samples were significant (Table

6). The proportion of overlapping genes in different samples was consistent with the contaminant level and geochemistry. For instance, for the background sample without contamination less than 23% of the genes overlapped with the genes found in the highly contaminated samples from wells FW010 and FW024, but 36% of the genes detected overlapped with the genes in the less contaminated sample from well FW003 (Table 6). Some important genes involved in denitrification (e.g., *nosZ* [3057083] and *nirS* [7160897]), degradation of organic contaminants (e.g., diene lactone hydrolase [2935034] and lactone-specific esterase [3641341] genes), and metal resistance (e.g., mercuric reductase [21322691] gene) similar to genes of common genera, such as *Pseudomonas*, *Rhodococcus*, and *Paracoccus*, were observed in all samples (Fig. 6), suggesting that the microbial populations containing these genes are widespread. Dissimilatory sulfate-reducing bacteria are important in the reduction of uranium from soluble U(VI) to insoluble U(IV). In contrast to the results described above, while some dissimilatory sulfite-reducing organisms (*dsrAB*) were found in all of the samples (group D) (Fig. 6), the abundance and presence of most types (groups A, B, and C) seemed to vary with the origin of the sample. This suggests that the contaminants and geochemical conditions may have selected for or against certain populations. Additionally, a significant portion (5 to 20%) of all genes detected were unique to samples, even for the samples from wells FW010 and FW024, which are ~32.5 m apart (Table 6). Thus, important microbial populations appear to be highly heterogeneous at this site.

The Environmental Remediation Science Program FRC site is contaminated with nitrate, uranium, and technetium, as well as some residual organic compounds. Microarray analyses showed that microbial populations containing genes involved in sulfate reduction, denitrification, metal reduction, and degradation of organic contaminants are prominent at this site. Thus, strategies that are now being employed to stimulate the indigenous microbial populations for remediating these contaminants should be successful, if carried out carefully, without additional bioaugmentation of desired species (13). However, the great heterogeneity of the microbial populations in different samples implies that biostimulation could also be complicated and that a variety of optimized strategies to stimulate and maintain the desired populations may need to be considered for achieving the remediation goals at this site.

Our application of the new WCGA-assisted microarray-based detection approach to contaminated groundwater samples indicated that this technology is indeed very powerful for analyzing and monitoring the composition and structure of microbial communities. Although MDA has been used to amplify human and yeast DNAs, it has not been tested previously with complex natural communities. This new approach should permit systems-level analyses of microbial communities whose members cannot be detected using conventional microarray-based approaches. This study is also one of the first demonstrations that microarray-based technology can be used to successfully visualize the functional structure of microbial communities in real environmental samples in a high-throughput fashion. Although the power of this approach was demonstrated with microbial communities in contaminated groundwater, the principles and technologies could be applied to any situation in which microbial populations are analyzed. There-

fore, this approach could be useful for addressing questions concerning microbial communities associated with human health, plant and animal quarantine, pathogen epidemiology, rhizosphere ecology, animal productivity (e.g., intestinal and rumen populations), forestry, oceanography, fisheries, ecology, and biodiversity discovery (e.g., pharmaceutical discovery), as microbial communities play important roles in each of these areas and the available microbial biomass is often very restricted.

ACKNOWLEDGMENTS

This research was supported by the United States Department of Energy under the Environmental Remediation Science Program, Genomics:GTL Program, through the Virtual Institute of Microbial Stress and Survival (<http://vimss.lbl.gov>), Microbial Genome Program, Biotechnology Investigations—Ocean Margins Program, and Carbon Sequestration Program (as part of the consortium on research to enhance carbon sequestration in terrestrial ecosystems [CSiTE]) of the Office of Biological and Environmental Research, Office of Science. X.L. was supported in part by the Chinese National Science Foundation (NSFC) Joint Research Fund for Overseas Chinese Young Scholars (grant 30428014), the NSFC Fund for Creative Research (grant 50321402), and Chinese National Basic Research (grant 2004CB619201). Oak Ridge National Laboratory is managed by University of Tennessee—Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

REFERENCES

1. Abu Al-Soud, W., L. J. Jonsson, and P. Radstrom. 2000. Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. *J. Clin. Microbiol.* **38**:345–350.
2. Amann, R. L., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143–169.
3. Andras, S. C., J. B. Power, E. C. Cocking, and M. R. Davey. 2001. Strategies for signal amplification in nucleic acid detection. *Mol. Biotechnol.* **19**:29–44.
- 3a. Bagwell, C. E., X. Liu L. Wu, and J. Zhou. 2006. Effects of legacy nuclear waste on the compositional diversity and distributions of sulfate-reducing bacteria in a terrestrial subsurface aquifer. *FEMS Microbiol. Ecol.* **55**:424–431.
4. Bartosiewicz, M., M. Trounstein, D. Barker, R. Johnston, and A. Buckpitt. 2000. Development of a toxicological gene array and quantitative assessment of this technology. *Arch. Biochem. Biophys.* **376**:66–73.
5. Cho, J. C., and J. M. Tiedje. 2002. Quantitative detection of microbial genes by using DNA microarrays. *Appl. Environ. Microbiol.* **68**:1425–1430.
6. Dean, F. B., S. Hosono, L. H. Fang, X. H. Wu, A. F. Faruqi, P. Bray-Ward, Z. Y. Sun, Q. L. Zong, Y. F. Du, J. Du, M. Driscoll, W. M. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99**:5261–5266.
7. Denev, V. J., J. Park, T. V. Tsoi, J. M. Rouillard, H. Zhang, J. A. Wibbenmeyer, W. Verstraete, E. Gulari, S. A. Hashsham, and J. M. Tiedje. 2004. Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. *Appl. Environ. Microbiol.* **70**:4961–4970.
8. DeRisi, J. L., V. R. Iyer, and P. O. Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**:680–686.
9. Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**:14863–14868.
10. El Fantroussi, S., H. Urakawa, A. E. Bernhard, J. J. Kelly, P. A. Noble, H. Smidt, G. M. Yershov, and D. A. Stahl. 2003. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl. Environ. Microbiol.* **69**:2377–2382.
11. Gao, H. C., Y. Wang, X. D. Liu, T. F. Yan, L. Y. Wu, E. Alm, A. Arkin, D. K. Thompson, and J. Z. Zhou. 2004. Global transcriptome analysis of the heat shock response of *Shewanella oneidensis*. *J. Bacteriol.* **186**:7796–7803.
12. Hutchison, C. A., H. O. Smith, C. Pfannkoch, and J. C. Venter. 2005. Cell-free cloning using phi 29 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **102**:17332–17336.
13. 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.

14. Lage, J. M., J. H. Leamon, T. Pejovic, S. Hamann, M. Lacey, D. Dillon, R. Segraves, B. Vossbrinck, A. Gonzalez, D. Pinkel, D. G. Albertson, J. Costa, and P. M. Lizardi. 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.* **13**:294–307.
15. Liu, Y. Q., J. Z. Zhou, M. V. Omelchenko, A. S. Beliaev, A. Venkateswaran, J. Stair, L. Y. Wu, D. K. Thompson, D. Xu, I. B. Rogozin, E. K. Gaidamkova, M. Zhai, K. S. Makarova, E. V. Koonin, and M. J. Daly. 2003. Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc. Natl. Acad. Sci. USA* **100**:4191–4196.
16. Lockhart, D. J., H. L. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. W. Wang, M. Kobayashi, H. Horton, and E. L. Brown. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**:1675–1680.
17. Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K. H. Schleifer, and M. Wagner. 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.
18. Monstein, H. J., C. Olsson, I. Nilsson, N. Grahn, C. Benoni, and S. Ahnre. 2005. Multiple displacement amplification of DNA from human colon and rectum biopsies: bacterial profiling and identification of *Helicobacter pylori*-DNA by means of 16S rDNA-based TTGE and pyrosequencing analysis. *J. Microbiol. Methods* **63**:239–247.
19. Nallur, G., C. H. Luo, L. H. Fang, S. Cooley, V. Dave, J. Lambert, K. Kukanskis, S. Kingsmore, R. Lasken, and B. Schweitzer. 2001. Signal amplification by rolling circle amplification on DNA microarrays. *Nucleic Acids Res.* **29**:e118.
20. Raghunathan, A., H. R. Ferguson, C. J. Bornarth, W. M. Song, M. Driscoll, and R. S. Lasken. 2005. Genomic DNA amplification from a single bacterium. *Appl. Environ. Microbiol.* **71**:3342–3347.
21. 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.
22. Schadt, C. W., J. Liebich, S. C. Chong, T. J. Gentry, Z. L. He, H. B. Pan, and J. Z. Zhou. 2005. Design and use of functional gene microarrays (FGAs) for the characterization of microbial communities, p. 331–368. *In* T. Savidge and C. Pothubakis (ed.), *Methods in microbiology*, vol. 34. Microbial imaging. Elsevier, San Diego, Calif.
23. Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. 1995. Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* **270**:467–470.
24. Schena, M., D. Shalon, R. Heller, A. Chai, P. O. Brown, and R. W. Davis. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **93**:10614–10619.
25. Tiquia, S. M., L. Y. Wu, S. C. Chong, S. Passovets, D. Xu, Y. Xu, and J. Z. Zhou. 2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* **36**:664–670.
26. Verdnik, D., S. Handran, and S. Pickett. 2002. Key considerations for accurate microarray scanning and image analysis, p. 83–98. *In* G. Kamberova and S. Shah (ed.), *DNA array image analysis nuts & bolts*. DNA Press, Eagleville, Pa.
27. Vora, G. J., C. E. Meador, D. A. Stenger, and J. D. Andreadis. 2004. Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl. Environ. Microbiol.* **70**:3047–3054.
28. Wu, L. Y., D. K. Thompson, G. S. Li, R. A. Hurt, J. M. Tiedje, and J. Z. Zhou. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67**:5780–5790.
29. Wu, L. Y., D. K. Thompson, X. D. Liu, M. W. Fields, C. E. Bagwell, J. M. Tiedje, and J. Z. Zhou. 2004. Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ. Sci. Technol.* **38**:6775–6782.
30. 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.
31. Ye, R. W., W. Tao, L. Bedzyk, T. Young, M. Chen, and L. Li. 2000. Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J. Bacteriol.* **182**:4458–4465.
32. Zhou, J. H. 2003. Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol.* **6**:288–294.
33. Zhou, J. Z., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**:316–322.
34. Zhou, J. Z., M. R. Fries, J. C. Cheesanford, and J. M. Tiedje. 1995. Phylogenetic analyses of a new group of denitrifiers capable of anaerobic growth on toluene and description of *Azoarcus toluolyticus* sp. nov. *Int. J. Syst. Bacteriol.* **45**:500–506.
35. Zhou, J. Z., and D. K. Thompson. 2002. Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol.* **13**:204–207.