

ORIGINAL ARTICLE

GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes

Zhili He^{1,2}, Terry J Gentry^{2,3}, Christopher W Schadt², Liyou Wu^{1,2}, Jost Liebich^{2,5}, Song C Chong², Zhijian Huang^{2,6}, Weimin Wu⁴, Baohua Gu², Phil Jardine², Craig Criddle⁴ and Jizhong Zhou^{1,2}

¹Department of Botany and Microbiology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA; ²Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; ³Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA and ⁴Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA

Owing to their vast diversity and as-yet uncultivated status, detection, characterization and quantification of microorganisms in natural settings are very challenging, and linking microbial diversity to ecosystem processes and functions is even more difficult. Microarray-based genomic technology for detecting functional genes and processes has a great promise of overcoming such obstacles. Here, a novel comprehensive microarray, termed GeoChip, has been developed, containing 24 243 oligonucleotide (50 mer) probes and covering >10 000 genes in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. The developed GeoChip was successfully used for tracking the dynamics of metal-reducing bacteria and associated communities for an *in situ* bioremediation study. This is the first comprehensive microarray currently available for studying biogeochemical processes and functional activities of microbial communities important to human health, agriculture, energy, global climate change, ecosystem management, and environmental cleanup and restoration. It is particularly useful for providing direct linkages of microbial genes/populations to ecosystem processes and functions.

The ISME Journal (2007) 1, 67–77; doi:10.1038/ismej.2007.2

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: microarray; functional genes; microbial community; bioremediation

Introduction

Microorganisms are the foundation of the Earth's biosphere, and play integral and unique roles in ecosystem functions and biogeochemical cycling of carbon, nitrogen, sulfur, phosphorus and various metals. But the precise role of many microorganisms in these cycles is unknown (Fitter *et al.*, 2005). Understanding the structure, functions, stability and adaptations of microbial populations/communities is critical for basic science discovery, biotechnology,

agriculture, energy, environment and human health. However, the majority of microorganisms in natural environments are not cultivated yet (Amann *et al.*, 1995). Owing to their extremely high diversity and their as-yet uncultivated status, microbial detection, characterization and quantification in natural systems remain challenging, especially on a large scale and in a parallel and high-throughput fashion. Also establishing linkages between microbial diversity to ecosystem functions represents even more challenges (Fitter *et al.*, 2005; Levin, 2006).

Microarrays are a recently developed, powerful genomic technology and are widely used to study gene expression (Schena *et al.*, 1995; Lockhart *et al.*, 1996; DeRisi *et al.*, 1997; Liu *et al.*, 2003; Gao *et al.*, 2004), monitor environmental processes (Loy *et al.*, 2002; Taroncher-Oldenburg *et al.*, 2003; Zhou, 2003; Bodrossy and Sessitsch, 2004; Rhee *et al.*, 2004; Steward *et al.*, 2004; Tiquia *et al.*, 2004; Zhou *et al.*, 2004; Wu *et al.*, 2006a), and potentially apply to clinic diagnosis (Lesko *et al.*, 2003). Similar to the situation in which microprocessors have

Correspondence: Dr J Zhou, Department of Botany and Microbiology, Institute for Environmental Genomics, University of Oklahoma, 101 David L. Boren Blvd., Norman, OK 73019, USA. E-mail: jzhou@ou.edu

⁵Current address: Agrosphere Institute (ICG-IV), Forschungszentrum Juelich GmbH, 52425 Juelich, Germany.

⁶Current address: State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, PR China.

Received 15 November 2006; revised and accepted 8 February 2007

increased the speed of computation, microarray-based genomic technologies have revolutionized genetic analyses of biological systems. Although microarray technology has been used successfully to analyze global gene expression in pure culture studies (Schena *et al.*, 1995; Lockhart *et al.*, 1996; DeRisi *et al.*, 1997; Liu *et al.*, 2003; Gao *et al.*, 2004; Mukhopadhyay *et al.*, 2006), adapting microarray technology for use in environmental studies presents numerous challenges in terms of probe design, the coverage of gene sequences, specificity, sensitivity and quantitation (Loy *et al.*, 2002; Taroncher-Oldenburg *et al.*, 2003; Rhee *et al.*, 2004; Steward *et al.*, 2004; Tiquia *et al.*, 2004; Wu *et al.*, 2006a).

To overcome such obstacles for studying microbial communities in natural settings, a particular type of microarrays, called functional gene arrays (FGAs), has been developed and used (Taroncher-Oldenburg *et al.*, 2003; Rhee *et al.*, 2004; Steward *et al.*, 2004; Tiquia *et al.*, 2004). This type of arrays contains probes from the genes involved in key microbially mediated biogeochemical processes, such as C, N and S cycling, phosphorus utilization, organic contaminant degradation and metal resistance and reduction, and is particularly powerful for studying various biogeochemical processes. Because the arrays contain probes from the genes with known biological functions, they will be useful in linking microbial diversity to ecosystem processes and functions. Several systematic experimental evaluations indicated that FGA-based microarrays can be used as specific, sensitive and potentially quantitative tools for detecting microbial populations and functional activities in natural settings (Wu *et al.*, 2001, 2004, 2006a; Taroncher-Oldenburg *et al.*, 2003; Rhee *et al.*, 2004; Steward *et al.*, 2004; Tiquia *et al.*, 2004). However, one of the greatest challenges in using FGAs for detecting functional genes and/or microorganisms in the environment is to design oligonucleotide probes specific to the target genes/microorganisms of interest because sequences of a particular functional gene are highly homologous and/or incomplete, especially sequences derived from laboratory cloning of environmental samples. Another challenge for using FGAs for studying the microbial communities in natural systems is the lack of arrays containing comprehensive probe sets. To tackle those challenges, in this paper, we report the design, construction, evaluation and application of a comprehensive FGA, termed GeoChip, which contains more than 24 000 probes from all of the known genes involved in various biogeochemical, ecological and environmental processes. As this is the second generation of FGAs, we refer to this array as GeoChip 2.0. The developed GeoChip is highly specific, and was successfully used for tracking the dynamics of metal-reducing bacteria and associated communities for an *in situ* bioremediation study. To our knowledge, this is the most comprehensive microarrays currently available for studying the

functional processes and activities of microbial communities associated with human health, agriculture, energy, global changes, ecosystem management and environmental cleanup and restoration.

Materials and methods

Oligonucleotide probe design, synthesis and fabrication

A new version of *CommOligo* (Li *et al.*, 2005) with group-specific probe design features was used to design both gene- and group-specific oligonucleotide probes based on the following criteria: (i) gene-specific probes: $\leq 90\%$ sequence identity, ≤ 20 -base continuous stretch and ≥ -35 kcal/mol free energy (Liebich *et al.*, 2006); (ii) group-specific probes: a group-specific probe has to meet the above requirements for nontarget groups, and it also has to have $\geq 96\%$ of sequence identity, ≥ 35 -base continuous stretch and ≤ -60 kcal/mol of free energy within the group (He *et al.*, 2005). The information about the probes and their targets is available on our web site: <http://ieg.ou.edu/GeoChip2.0/GeoChip2.0-info.xls>. All designed probes were subsequently verified by *ProbeChecker*, and synthesized by MWG Biotech Inc. (High Point, NC, USA). The concentration of all oligonucleotides was adjusted to 100 pmol/ μ l. All oligonucleotide probes and controls were arrayed onto Corning UltraGAPS (Corning, NY, USA) slides using a Microgrid II Arrayer (Genomic Solutions, Ann Arbor, MI, USA) as described previously (He *et al.*, 2005).

Preparations of synthesized oligonucleotide and polymerase chain reaction PCR amplicon targets

Oligonucleotides that are complementary to probes spotted on the array were synthesized and labeled at the 5' end with Cy5 or Cy3 dye during synthesis (Table S1 in Supplementary Data 1). 17 (11 for gene-specific probes and six for group-specific probes) sequences were selected as templates from pure cultures or environmental clones available in our laboratory. The desired gene fragments were amplified by PCR (Tables S2 and S3 in Supplementary Data 1). To effectively evaluate GeoChip, most of the selected probes for generating targets have sequence identity, continuous stretch and free energy values close to the design criteria. Each PCR product had a minimal length to cover all available probes. Normally, 50 pg of each synthesized oligonucleotide or PCR-amplicon was used alone (single-target experiments) or in a mixture (multiple-target experiments) of multiple targets.

DNA extraction, labeling, GeoChip processing and chemical analysis

Community DNAs were extracted from groundwater samples as described previously (Zhou *et al.*, 1996). Labeling, array hybridization and scanning were

conducted as described previously (He *et al.*, 2005). The analysis of uranium concentration in the groundwater samples was carried out as described by Wu *et al.* (2006c).

Data normalization and analysis

Scanned images were quantified using the software ImaGene 6.0 (Biodiscovery Inc., El Segundo, CA, USA). A Perl script was developed to analyze digital array data. This script included the following key steps:

- (i) Poor-quality spots were removed.
- (ii) Signal intensity of each spot was normalized by mean.
- (iii) Spots with low signal intensities were removed based on the signal-to-noise ratio (SNR) (Wu *et al.*, 2006a). A commonly accepted SNR cutoff value of 3.0 (Verdick *et al.*, 2002) was used for synthesized oligonucleotide and PCR-amplicon targets, and 2.0 for environmental samples.
- (iv) For outlier removal, if any of replicates (slides) had (signal–mean) more than three times the standard deviation, this replicate was moved. This process continued until no such replicates were identified. The Mantel test was used to examine the correlations between the differences of uranium concentrations and those of various functional gene abundances (Legendre and Legendre, 1998).

Results

GeoChip design strategies and construction

Owing to the nature of functional gene sequences (highly similar and incomplete), it can be extremely challenging to design specific oligonucleotide probes for some functional genes using routine probe design strategies. Thus, in this study, four strategies were implemented. First, retrieved sequences were aligned using a multiple sequence alignment (MSA) program. Only the shared regions of the functional genes were used for probe design. Second, experimentally established oligonucleotide design criteria and a novel software tool specifically developed to deal with highly similar sequences were used to select 50-mer oligonucleotide probes. Third, to detect both divergent and closely related sequences, both gene- and group-specific probes were designed. Finally, to increase the confidence of detection, multiple probes for each sequence or each group of sequences were designed. The major steps for GeoChip construction are shown in Figure 1, including sequence retrieval, oligonucleotide design, probe verification and output. The Supplementary Data 2 provides more detailed information for those individual steps.

With the above considerations and major steps, we have developed the GeoChip containing 24 243 oligonucleotide probes targeting >150 functional

groups of >10 000 genes essential to the biogeochemical cyclings of carbon, nitrogen, phosphorus and sulfur along with metal resistance, metal reduction and organic contaminant degradation. Among them, 19 959 (82.3%) probes are gene-specific, whereas 4 284 (17.7%) probes are group-specific (Table 1). About 11.6% of the probes target the genes involved in carbon degradation, 4.2% for carbon fixation, 5.1% for nitrogen fixation, 1.4% for nitrification, 9.5% for denitrification, 5.9% for nitrogen mineralization, 6.7% for sulfate reduction, 3.2% for methane reduction and oxidation, 18.8% for metal reduction and resistance, and 33.1% for degradation/transformation of a variety of organic chemical compounds such as acrylonitrile (1.8%), benzoate (3.7%), biphenyl (1.1%), catechol compounds (4.0%), naphthalene (1.0%), phenol (1.2%) and protocatechuate (1.5%) (Table 1, Supplementary Data 3). Almost all (~98.2%) of the gene sequences were from bacteria whereas the rest (~1.8%) were from fungi. In addition, the following controls were spotted to check hybridization, printing, gridding, quality and data analysis: (i) 16S rRNA gene probes as positive controls (192 probes), (ii) quantitative and negative controls with 10 probes from 10 human genes (960 spots), and (iii) blanks.

Computational evaluation of GeoChip specificity

To assess the specificity of the designed probes, the distributions of the maximum sequence identity,

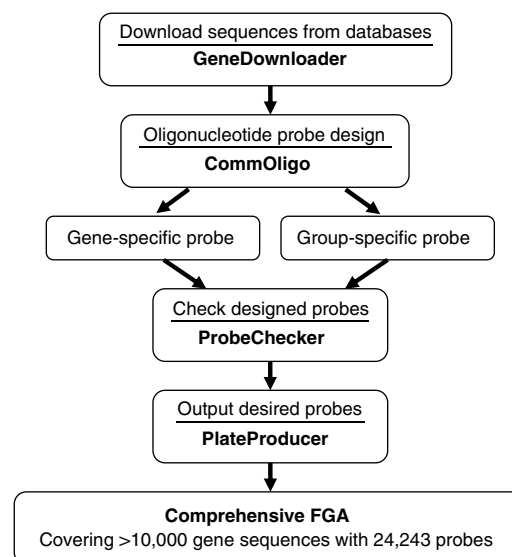


Figure 1 Major steps for construction of the 50-mer GeoChip. *CommOligo* is the core program to select gene- and group-specific oligonucleotide probes for each functional gene sequence based on criteria: identity ≤ 90%, stretch ≤ 20 bases and free energy ≥ −35 kcal/mol for gene-specific probes, and identity ≥ 96%, stretch ≥ 35 bases and free energy ≤ −60 kcal/mol for group-specific probes. *GeneDownloader*, *ProbeChecker* and *PlateProducer* were Perl scripts to pre-process gene sequences or post-process oligonucleotide probes.

maximum stretch length and minimal free energy to their nontargets were examined computationally. The majority of the designed probes fall in the ranges of sequence identity, stretch or free energy far away from the thresholds of probe design criteria, and only a very small portion of the designed probes were very close to the thresholds (3.4% of the gene-specific probes from the identity range of 86–90%, 7.0% from the maximal stretch lengths of 16–20 bases, and <1.7% with the minimal free energy within the range from –30 to –35 kcal/mol (Figure 2). Similar results were observed for group-specific probes (Figure 3). Approximately 93.0% of group-specific probes had 100% sequence identities with their target members in the same group, and the other 4.7 and 2.3% of group-specific probes had 98 and 96% identities with their group targets, respectively (Figure 3a). For the stretch length, 94.3% of group-specific probes had 45–50-base stretches with

their group targets, and the other 2.9 and 2.8% of group-specific probes had 40–44-base, and 35–39-base stretches with their group targets, respectively (Figure 3b). Most group-specific probes (86.0%) had the maximum free energy from –65 to –85 kcal/mol, and 8.1 and 5.9% of group-specific probes had the maximum free energy –60 to –65, and <–85 kcal/mol, respectively (Figure 3c). On the basis of our previous results on specificity evaluation (Rhee *et al.*, 2004; Tiquia *et al.*, 2004; He *et al.*, 2005; Liebich *et al.*, 2006), all probes are expected to be highly specific to their corresponding targets.

Experimental evaluation of GeoChip specificity

The specificity of the designed GeoChip was further evaluated experimentally. Representatives of the probes whose values of sequence identity, stretch

Table 1 Summary of the numbers of probes by functional gene category on the GeoChip

Gene category	Unique probes	Group probes	Total
Nitrogen cycling	3896	1414	5310
Nitrogen fixation	1225	0	1225
Denitrification	1805	501	2306
Nitrification	251	96	347
Nitrogen mineralization	615	817	1432
Carbon cycling	3708	891	4599
Carbon fixation	739	279	1018
Cellulose degradation	1213	72	1285
Lignin degradation	423	90	513
Chitin degradation	651	93	744
Methane production	194	243	437
Methane oxidation	243	93	336
Others	245	21	266
Sulfate reduction	1286	329	1615
Phosphorus utilization	89	56	145
Metal reduction and resistance	4039	507	4546
Arsenic resistance	775	102	877
Cadmium resistance	237	45	282
Chromium resistance	303	16	319
Mercury resistance/reduction	439	109	548
Nickel resistance	135	5	140
Zinc resistance	110	18	128
Other metals and metalloids	1450	182	1632
Other metal resistance/reduction	590	30	620
Contaminant degradation	6941	1087	8028
BTEX and related aromatics	3652	524	4176
Chlorinated aromatics	79	11	90
Nitoaromatics	134	18	152
Heterocyclic aromatics	176	37	213
Hydrocarbons (e.g. PAHs)	597	144	741
Polychlorinated biphenyls (PCBs)	303	85	388
Chlorinated solvents	178	54	232
Pesticides	504	54	558
Other chemicals & by-products	1297	160	1457
Others	21	0	21
Total	19 959	4284	24 243

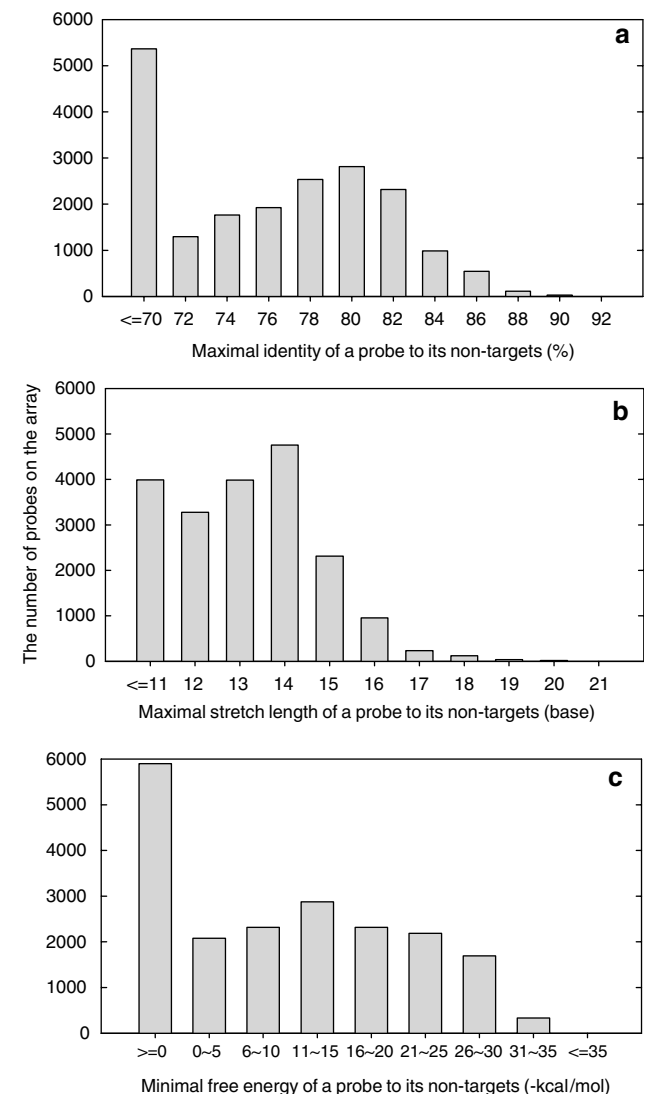


Figure 2 Distribution of 19 959 gene-specific probes at their (a) maximal sequence identities, (b) maximal stretch lengths or (c) minimal free energy with their non-targets.

length and free energy are close to the probe design criteria thresholds were selected for experimental evaluation of specificity using both synthesized oligonucleotide targets and PCR-amplified targets. First, the specificity of the GeoChip was evaluated using an equal mixture (50 pg for each target) of 15 synthesized oligonucleotides (Table S1 in Supplementary Data 1), nine (T1–T9) targeted by gene-specific probes and six (T10–T15) targeted by group-specific probes under different hybridization temperatures (42, 45, 50 and 60°C). Our results suggested that the optimal hybridization temperature was between 45 and 50°C in the presence of 50% formamide (Figure 4), which is consistent with our previous results with the arrays containing fewer probes (Rhee *et al.*, 2004; Tiquia *et al.*, 2004). Second, a mixture of 25 synthesized oligonucleotide targets (Table S1 in Supplementary Data 1) was hybridized to the GeoChip at 50°C and 50% formamide. All of the 25 probes corresponding to

their targets showed positive signals but three unrelated probes showed positive hybridization, and no false negatives were detected (Table 2). Finally, a mixture of 17 PCR-amplicons (Table S2 and S3 in Supplementary Data 1) were obtained using gene-specific primers, labeled and used as targets to evaluate GeoChip specificity. It is expected that 35 oligonucleotide probes on the array would hybridize with the targets. The results showed that all 35 expected probes had positive hybridization with an average signal of 9265 ± 5270 ($n = 5$), and an average of $\text{SNR} = 67.6 \pm 38.72$ ($n = 5$). In addition, four probes showed false positives, and no false negatives were observed (Table 2). By considering the numbers of probes on the arrays, the percentage

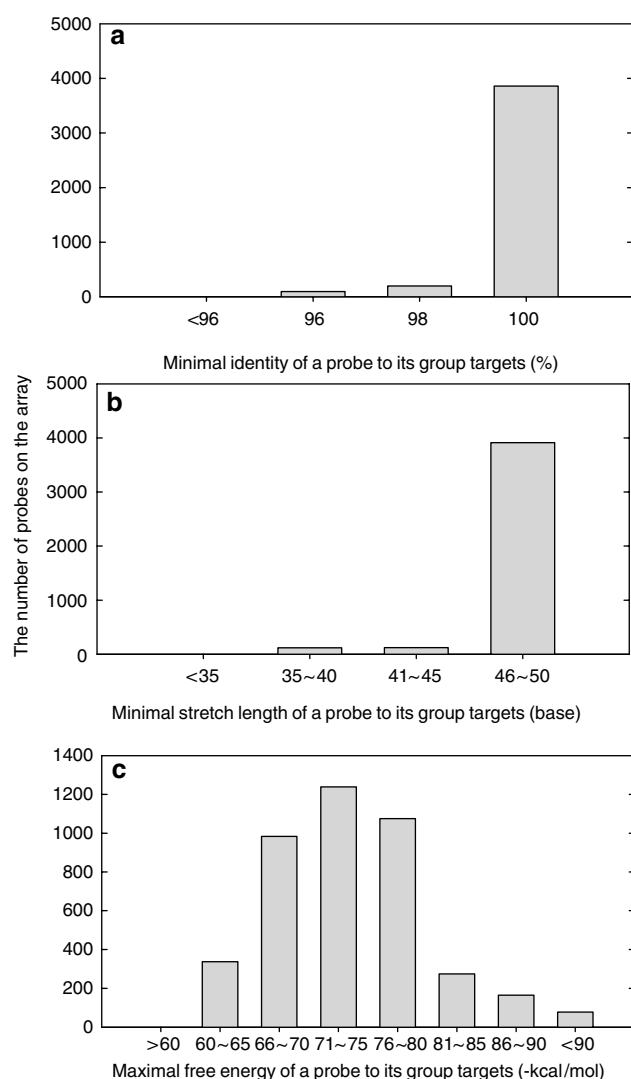


Figure 3 Distribution of 4284 group-specific probes at their (a) minimal sequence identities, (b) minimal stretch lengths or (c) maximal free energy with their group targets.

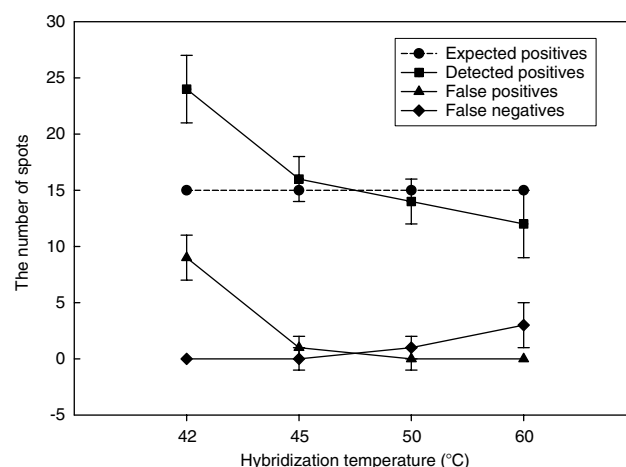


Figure 4 The GeoChip was hybridized with a mixture of 15 synthesized oligonucleotide targets at 42°C, 45°C, 50°C and 60°C. The numbers of detected spots, expected spots, false positives and false negatives were shown with five replicates for each condition.

Table 2 Summary of the GeoChip hybridization with different targets (oligonucleotides or PCR-amplicons)

Targets	Oligonucleotide	PCR-amplicon
No. of targets	25	17
Expected no. of probes detected	25	35
No. of probes hybridized	28	39
Average signal intensity \pm s.d.	5678 ± 3372	9265 ± 5270
Average SNR \pm s.d.	55.9 ± 27.56	67.6 ± 38.72
No. of false positives	3	4
Percentage of false positives (%) ^a	4.9×10^{-4}	9.7×10^{-4}
No. of false negatives	0	0
Percentage of false negatives (%)	0	0

Abbreviations: PCR, polymerase chain reaction; SNR, signal-to-noise ratio; s.d., standard deviation.

SNR > 3.0 was used to distinguish a positive spot from its background. A mix of all targets (25 for synthesized oligonucleotides or 17 PCR amplicons) was hybridized with the GeoChip with five ($n = 5$) replicates.

^aIt was calculated by dividing the number of false positives observed with the number of target genes tested and the total probes on the GeoChip.

of false positives ($4.9\text{--}9.7 \times 10^{-4}$) is negligible. These results suggest that the probes on the GeoChip are highly specific to their corresponding targets.

Application of the GeoChip to analysis of *in situ* uranium bioremediation

To demonstrate the power of the developed GeoChip, we used it to monitor microbial community dynamics in groundwater undergoing *in situ* biostimulation for uranium reduction at the Department of Energy (DOE) Field Research Center (FRC) in Oak Ridge (TN, USA). The groundwater and sediments at this field site has been heavily contaminated with high levels of uranium (up to 60 mg/l in groundwater and 800 mg/kg in sediments) and nitrate (up to 160 mM in groundwater), which presents a great challenge for environmental cleanup. In this field test, an above-ground treatment system was used to remove nitrate and other inhibitors to provide a favorable conditions for microbial growth and then ethanol was injected to subsurface from day 137 to 142 (1/7/2004 to 1/12/2004) and from day 163 to 166 to stimulate *in situ* microbial denitrification and subsequently microbial reduction of U(VI) to insoluble U(IV) (Wu *et al.*, 2006b,c). The ethanol injection was intermittently conducted. By the end of June 2005 (day 670), the uranium concentrations were reduced below the USA Environmental Protection Agency (EPA) maximum contaminant level (MCL) for drinking water ($<30 \mu\text{g/l}$) in monitoring wells (Figure 5). Sulfate consumption and sulfide formation were evident during uranium remediation. This is the first *in situ* demonstration that high level of uranium contamination in subsurface can be successfully bioremediated to the level below the US EPA MCL.

The microbial community dynamics from one of the four frequently sampled monitoring wells (FW 102-3) was intensively analyzed with the GeoChip. More than 2993 genes in >100 gene categories

showed statistically significant positive hybridization signals. As dissimilatory Fe(III)-reducing bacteria (FeRB) such as *Geobacter* spp. and sulfate-reducing bacteria (SRB), such as *Desulfovibrio* spp. are the two major groups of microorganisms capable of U(VI) reduction through both direct enzymatic (Lovley *et al.*, 1993a,b; Lovley, 1995; Truex *et al.*, 1996; Tebo and Obraztsova, 1998; Petrie *et al.*, 2003; Wu *et al.*, 2006c) and/or indirect chemical mechanisms (Mohagheghi *et al.*, 1985; Liger *et al.*, 1999), we have focused on the analysis of the dynamics of FeRB and SRB. During the uranium reduction period, both FeRB (Figure 5) and SRB (data not shown) populations reached their highest levels at day 212, followed by a gradual decrease over 500 days. Consequently, the uranium in groundwater and sediments was reduced and thus uranium concentrations in the groundwater decreased. Because *Geobacter*-type FeRB and some SRB can use U(VI) as electron acceptor by obtaining energy for growth (Lovley *et al.*, 1993a,b; Truex *et al.*, 1996; Tebo and Obraztsova, 1998), it is expected that these types of microbial populations would change with uranium concentrations. As expected, the uranium concentrations in the groundwater were significantly correlated with the total abundance of *c*-type cytochrome genes ($r=0.73$, $P<0.05$, Figure 5) from *Geobacter*-type FeRB and *Desulfovibrio*-type SRB, and with the total abundance of *dsrAB* (dissimilatory sulfite reductase) genes ($r=0.88$, $P<0.05$) (data not shown). Mantel test also indicated that there was significant correlation between the differences of uranium concentrations and those of total *c*-cytochrome gene abundance ($r_m=0.75$, $P<0.001$) or *dsrAB* gene abundance ($r_m=0.72$, $P<0.01$). These results suggested that *Geobacter*-type FeRB and SRB played significant roles in reducing uranium to a level below the drinking standard ($<30 \mu\text{g/l}$).

To examine what genes/populations of FeRB and SRB were the key players for uranium reduction, a Mantel test was also performed to determine whether the changes in the abundance of each gene were correlated with the changes in uranium concentrations. The changes of more than a dozen of *c*-type cytochrome genes from *Geobacter sulfurreducens* and *Desulfovibrio desulfuricans* showed significant correlations to the changes of uranium concentrations among different time points (Figure 6), further suggesting that *Geobacter* and *Desulfovibrio* species did play a significant role in the success of the *in situ* uranium bioremediation. Also the changes of more than 10 *dsrAB*-containing populations, including both cultured (e.g. *Desulfovibrio desulfuricans*, *Desulfovibrio termitidis*, *Desulfotomaculum kuznetsovii* and *Thermosedulfovibrio yellowstonii*) and noncultured SRB were significantly related to the changes in uranium concentrations, indicating their importance in uranium reduction (Figure 6). Interestingly, as expected, the changes of several *dsrAB*-containing sulfate-reducing populations previously recovered from this site

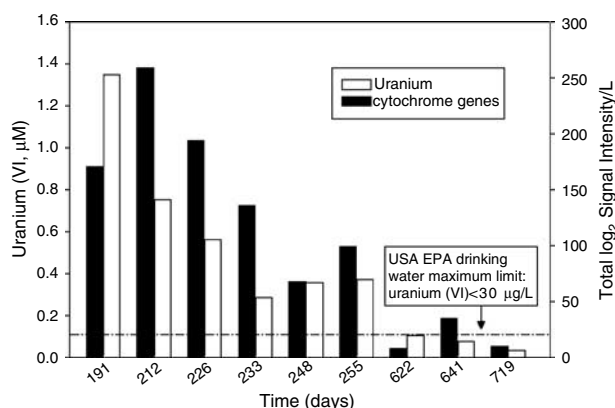


Figure 5 Relationships between uranium concentrations and the total abundance of *c*-type cytochrome genes detected by GeoChip for groundwater microbial communities in the monitoring well FW 102-3. Ethanol was injected for 2 days every week till day 711.



Figure 6 Hierarchical cluster analysis of gene relationships of groundwater microbial communities in the monitoring well FW 102-3 based on GeoChip hybridization signal intensity. Representative genes were from different samples of Well FW 102-3 at different days. These genes showed significant correlations ($P \leq 0.05$) with uranium concentrations based on the Mantel test.

(e.g. FW003269B and FW300181B) showed significant correlations to the differences of uranium concentrations (Figure 6). All of the above results indicate that the GeoChip is able to reveal microbial community differences, and that it is a powerful tool for tracking bioremediation processes, and for linking microbial populations to functional processes.

Discussion

The development and application of microarray-based genomic technology for microbial detection and community analysis has received a great deal of attention. Because of its high-density and high-throughput capacity, it is expected that microarray-based genomic technologies will revolutionize the analysis of microbial community structure, function and dynamics. Therefore, we have developed a novel comprehensive microarray (GeoChip) containing > 24 000 gene probes and covering > 10 000 genes in > 150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. To our knowledge, this is the most comprehensive microarray currently available for studying various biogeochemical processes and functional activities of microbial communities. Our experimental results with uranium bioremediation experiments indicate that the developed GeoChip is able to reveal microbial community differences, and is a powerful tool for tracking bioremediation processes, and for linking microbial populations to functional processes.

Specificity is one of the critical issues in microarray assays, especially for environmental studies. To ensure microarray hybridization specificity, we have experimentally established probe design criteria based on sequence identity, continuous sequence stretches and free energy (He *et al.*, 2005; Liebich *et al.*, 2006) and developed a novel software tool, *CommOligo*, for designing microarrays probes (Li *et al.*, 2005). The developed GeoChip was designed using the newly developed software based on these experimentally established criteria (He *et al.*, 2005; Liebich *et al.*, 2006). Computational analysis showed that the majority (93–98%) of the probes on the GeoChip fall in the ranges of sequence identity, stretch length or free energy far away from the thresholds of probe design criteria, indicating that the designed probes should be specific to their corresponding targets. The specificity of the probe representatives whose values of sequence identity, stretch length and free energy are close to the probe design criteria thresholds were experimentally evaluated and only very small portions of false positive (0.002–0.004%) were observed. Further sequence analysis indicated that these false positive probes do not have high similarities, long stretches, or low free energy values with the targets used, suggesting that these false positives are due to random errors. Possible explanations for those false positives include high concentrations of targets used, errors in probe or/and target sequences, contaminations during probe/target preparation and array construction, and/or the lack of full understanding of the factors controlling probe-target kinetics. Finally, the comparison of the sequences from pure cultures indicated that the 50-mer oligonucleotide probes

could provide species-strain level resolution for analyzing microorganisms involved in nitrification, denitrification, nitrogen fixation, methane oxidation, and sulfite reduction (Tiquia *et al.*, 2004). Thus similar taxonomic resolution is expected for the developed GeoChip.

In contrast to the whole genome open reading frame (ORF) arrays for gene expression study of individual pure cultures, the developed GeoChip also contains group-specific probes (17.7%) and covers about 3000 gene sequences, which is important for environmental studies because many target sequences involved key biogeochemical, ecological and environmental processes are highly homologous. Computational analysis showed that the majority (93%) of the probes on the developed GeoChip have 100% sequence homology to their corresponding target sequences and have at least less than 90% of homology to the non-target sequences, suggesting that these group-specific probes will be able to detect various sequence groups in the environmental samples.

Sensitivity is another critical parameter that impacts the effectiveness of microarray-based approaches for detecting genes in environmental samples. When PCR-amplicon-based FGAs were used, the detection limit for *nirS* genes was approximately 1 ng of pure genomic DNA and 25 ng of soil community DNA without amplification of target templates (Wu *et al.*, 2001). Studies with a 50-mer FGA showed that the detection limit without target template amplification ranges from 5–10 ng of pure genomic DNA and 50–100 ng in a mixture of genomic DNA from different organisms (Rhee *et al.*, 2004; Tiquia *et al.*, 2004). By combining whole-community genome amplification (WCGA) approach, the 50-mer FGA can detect subnanogram quantities of microbial community DNAs as low as 10 pg (Wu *et al.*, 2006a). Therefore, it is expected that the developed GeoChip will have a similar level of sensitivity of the 50-mer FGAs because the hybridization conditions are identical.

The quantitative capability of microarray-based technology is another central issue for environmental applications. Several previous studies showed that very good linear relationships were obtained between hybridization signal intensity and target DNA or RNA concentration from pure cultures, mixed DNA templates and cells, and environmental samples (Wu *et al.*, 2001, 2004; Rhee *et al.*, 2004). Recently, we showed that reliable quantification could be obtained using 50-mer FGAs with randomly amplified DNAs (Wu *et al.*, 2006a) or randomly applied RNAs (Gao *et al.*, 2007). Thus, it is expected that the gene-specific probes on the GeoChip should be able to provide quantitative information for their corresponding target genes as demonstrated in our previous studies because the hybridization conditions are identical. One potential problem for quantifying population abundance with group-specific probes is that the target

sequences with one or two mismatches may result in lower hybridization signals than the target sequences with perfect matches, which may lead to inaccurate estimations of population abundance. However, since most (93%) of these group-specific groups on this designed GeoChip have ~100% homology to their sequences, the quantitative inaccuracy resulted from mismatched sequences should be a less concern.

As potential cross-hybridization is always a concern, especially when dealing with environmental samples of unknown composition, it is important to use the GeoChips for relative comparisons. In general, relative changes in microbial communities can be measured by the hybridization signal ratios of treatment samples to a common reference or control sample. The effects of cross-hybridization can be canceled out when the hybridization intensity signals from treatment samples are divided by the hybridization intensity signals from the common reference samples under the assumption that the community composition is similar between the treatment and reference samples. Thus, using hybridization ratios will help to minimize the effects of cross-hybridization on quantitative accuracy. Also, multiple hybridizations with replicate samples are always important for statistically assessing the reliability of the hybridization data and for obtaining reliable quantitative results with high confidence.

Microbial community sequencing presents a new age in biology, but one of the greatest challenges is how to link genomics, as well as microbial diversity, to ecosystem processes and functions (Zhou *et al.*, 2004; Fitter *et al.*, 2005; Oremland *et al.*, 2005). In contrast to small subunit ribosomal RNA gene-based PhyloChips (Loy *et al.*, 2002; Zhou, 2003; Zhou *et al.*, 2004), the developed GeoChip will be an ideal tool for providing direct functional linkages of genes/populations to ecosystem processes and functions because it contains probes from all functionally known geochemical, ecological and environmental processes. The developed GeoChip can be used to analyze microbial community structure of both heavy and light fractions from stable isotope probing approach (Leigh MB, Ostrom NE, J Zhou, Tiedje JM, the ASM General Meeting Abstract, 2006). By coupling the GeoChip-based hybridization with stable isotope probing analysis, we can rapidly know which functional groups/populations are active. In addition, the GeoChip can be used to measure community gene expression because all GeoChip probes were selected from coding sequences of functional genes. Thus, probing mRNAs with the developed GeoChip and/or stable isotope probing will provide valuable insights into functions of the genes/populations in critical geochemical and ecological processes. Such information will be particularly useful in establishing mechanistic linkages between the diversity of microbial genes/populations and ecosystem functions.

The developed GeoChip is expected to be a powerful tool in studying microbially mediated geochemical, ecological and environmental processes. Two major types of applications can be visualized for the developed GeoChip. One is to track microbial community dynamics under different environmental/treatment conditions as we described above. The developed GeoChip has been successfully used to track the changes of the responsible microbial populations during the bioremediation processes. We have also used the GeoChip to address specific questions and/or hypotheses related to microbial population/community dynamics at a particular site, such as whether a contaminant or a change in environmental conditions adversely affects certain key microbial populations (data not shown). The other is to use it as a generic tool for profiling the differences between microbial communities. For this purpose, we have used the developed GeoChip to analyze microbial communities from a variety of habitats, such as bioreactors, soils, marine sediments and animal guts (data not shown). All of these results suggest that the developed GeoChip is useful for studying various biogeochemical, ecological and environmental processes and associated microbial communities in natural settings in a rapid, high throughput and potentially quantitative fashion. With the developed GeoChips, it is possible to address many fundamental and applied research questions in microbial ecology important to human health, agriculture, energy, global climate changes, ecosystem management and environmental cleanup and restoration.

As a by-product of nuclear weapons production during the Cold War era, many DOE field sites are contaminated with mixtures of metals and radionuclide as well as nitrate, chlorinated solvents and hydrocarbons. Among the mixed contaminants, uranium, which has half-lives ranging from 247 000 to 4.5 billion years, is the most predominant radionuclide contaminant at DOE sites. Owing to the risks to liver damage and cancer, there is increasing concern about the fate of uranium in the contaminated areas. Microbially mediated reduction of highly soluble uranium (VI) to insoluble uranium (IV) is a promising strategy for the potential remediation of uranium-contaminated groundwaters. In the field plot experiments, the uranium concentrations were reduced below $<30 \mu\text{g/l}$. Experimental results from GeoChip analysis suggested that *Geobacter*-type FeRB and SRB played significant roles in uranium reduction, suggesting that uranium remediation using indigenous microorganisms could be a valid option in heavily uranium-contaminated sites.

Although GeoChips have the potential to be powerful tools in characterizing microbial community structure, a number of challenges will need to be addressed and overcome. First of all, it appears that the enzymes used for amplification and labeling are very sensitive to the residual contaminants (e.g.

humic substance) in purified community DNAs as well as freshness of reagents. High quality of community DNAs is critical to minimize experimental variations for improving microarray-based quantitative accuracy. Second, another difficulty in using GeoChips for addressing environmental questions is the lack of appropriate standards for data comparison. At this time, it is difficult to compare microarray data among different laboratories and even among different experiments in a single laboratory (Zhou, 2003). This could limit the power of this technology to address ecological and environmental questions. Further development of universal standards which allow quantitative comparisons across different conditions are urgently needed. Third, the target sequences in public database increase exponentially, and hence the GeoChip needs to be continuously updated. One of the challenges is that the current probe design program is difficult to handle many sequences. Rapid probe designing tools capable of handling 10 000 of homologous sequences are needed. In addition, the quantity of data generated by microarray-based studies of environmental samples will likely be enormous, but rapid processing, comparing, interpreting hybridization data still remain difficult endeavors. Bioinformatic tools developed for gene expression analysis can be used to analyze environmental samples to some extent, but they have difficulty in dealing with the complexity of environmental samples. Development of novel bioinformatics tools for data analysis and interpretation is urgently needed. Finally, as we have always emphasized (Zhou and Thompson, 2002), GeoChips are only tools, and as such, they should be integrated with studies to address ecological and environmental questions and hypotheses. Only in this way can the power of GeoChips for analyzing environmental samples be ascertained.

In summary, the developed GeoChip contains over 24 000 oligonucleotides covering more than 150 functional groups of 10 000 gene sequences involved in various biogeochemical and environmental processes. Computational and experimental evaluation indicates that the developed GeoChip is highly specific. Successful application of the GeoChip for monitoring bioremediation processes of uranium reduction in fields demonstrates that it can be used as a powerful tool for rapid, high-throughput, parallel and cost-effective analysis of microbial communities, and for providing mechanistic linkages between microbial populations and bioremediation processes. The GeoChip is the most comprehensive microarray currently available for microbial community studies. The developed GeoChip can be used as a generic high-throughput tool to address various biological questions in different systems such as bioreactors, soils, groundwaters, marine sediments and animal guts although further developments are needed in terms of quantitative data comparisons across different experiments,

laboratories and times, high-throughput probe design, rapid data processing, analysis and visualization, and interpretation within the context of environmental and ecological application.

Acknowledgements

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 (VIMSS; <http://vimss.lbl.gov>), 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. Oak Ridge National Laboratory is managed by University of Tennessee UT-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725. Jost Liebich was supported by a fellowship within the postdoctoral program of the German Academic Exchange Service (DAAD).

References

- Amann RI, Ludwig W, Schleifer K. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Bodrossy L, Sessitsch A. (2004). Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol* **7**: 245–254.
- DeRisi JL, Iyer VR, Brown PO. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686.
- Fitter AH, Gilligan CA, Hollingworth K, Kleczkowski A, Twyman RM, Pitchford JW. (2005). Biodiversity and ecosystem function in soil. *Funct Ecol* **19**: 369–377.
- Gao H, Wang Y, Liu X, Yan T, Wu L, Alm E et al. (2004). Global transcriptome analysis of the heat shock response of *Shewanella oneidensis*. *J Bacteriol* **186**: 7796–7803.
- Gao H, Yang ZK, Gentry TJ, Wu L, Schadt CW, Zhou J. (2007). Microarray-based analysis of microbial community RNAs by whole community RNA amplification (WCRA). *Appl Environ Microbiol* **73**: 563–571.
- He Z, Wu L, Li X, Fields MW, Zhou J. (2005). Empirical establishment of oligonucleotide probe design criteria using perfect match and mismatch probes and artificial targets. *Appl Environ Microbiol* **71**: 3753–3760.
- Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd edn. Elsevier: New York, NY.
- Lesko LJ, Salerno RA, Spear BB, Anderson DC, Anderson T, Brazell C et al. (2003). Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: Report of the First FDA-PWG-PhRMA-DruSafe Workshop. *J Clin Pharmacol* **43**: 342–358.
- Levin SA. (2006). Fundamental questions in biology. *PLoS Biol* **9**: 1471–1472.
- Li X, He Z, Zhou J. (2005). Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* **33**: 6114–6123.
- Liebich J, Chong SC, Schadt CW, He Z, Rhee SK, Zhou J. (2006). Improvement of oligonucleotide probe design criteria for the development of functional gene microarrays for environmental applications. *Appl Environ Microbiol* **72**: 1688–1691.
- Liger E, Charlet L, van Chappellen P. (1999). Surface catalysis of uranium (VI) reduction by iron(II). *Geochim Cosmochim Acta* **63**: 2939–2955.
- Liu Y, Zhou J, Omelchenko M, Beliaev A, Venkateswaran A, Stair J et al. (2003). Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA* **100**: 4191–4196.
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**: 1675–1680.
- Lovley DR. (1995). Bioremediation of organic and metal contaminants with dissimilatory metal reduction. *J Ind Microbiol* **14**: 85–93.
- Lovley DR, Roden EE, Phillips EJP, Woodward JC. (1993a). Enzymatic iron and uranium reduction by sulfate-reducing bacteria. *Mar Geol* **113**: 41–53.
- Lovley DR, Widman PK, Woodward JC, Phillips EJP. (1993b). Reduction of uranium by cytochrome c_3 of *Desulfovibrio vulgare*. *Appl Environ Microbiol* **59**: 3572–3576.
- Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J et al. (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.
- Mohagheghi A, Updegraff DM, Goldhaber MB. (1985). The role of sulfate reducing bacteria in the deposition of sedimentary uranium ores. *Geomicrobiol J* **4**: 153–173.
- Mukhopadhyay A, He Z, Yen HC, Alm EJ, He Q, Huang K et al. (2006). Salt stress in *Desulfovibrio vulgaris* Hildenborough: an integrated genomics approach. *J Bacteriol* **188**: 4068–4078.
- Oremland RS, Capone DG, Stoltz JF, Fuhrman J. (2005). Whither or wither geomicrobiology in the era of ‘community metagenomics’. *Nat Rev Microbiol* **3**: 572–578.
- Petrie L, North NN, Dollhopf SL, Balkwill DL, Kostka JE. (2003). Enumeration and characterization of iron (III)-reducing microbial communities from acidic subsurface sediments contaminated with uranium (VI). *Appl Environ Microbiol* **69**: 7467–7479.
- Rhee SK, Liu X, Wu L, Chong SC, Wan X, Zhou J. (2004). Detection of biodegradation and biotransformation genes in microbial communities using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Schena M, Shalon D, Davis RW, Brown PO. (1995). Quantitative monitoring of gene-expression patterns with complementary-DNA microarray. *Science* **270**: 467–470.
- Steward GF, Jenkins BD, Ward BB, Zehr JP. (2004). Development and testing of a DNA microarray to assess nitrogenase (*nifH*) gene diversity. *Appl Environ Microbiol* **70**: 1455–1465.
- Taroncher-Oldenburg G, Griner EM, Francis CA, Ward BB. (2003). Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69**: 1159–1171.
- Tebo BM, Obraztsova AY. (1998). Sulfate-reducing bacterium grows with Cr(VI), U(VI), Mn(IV), and Fe(III) as electron acceptors. *FEMS Microbiol Lett* **162**: 193–198.

- Tiquia SM, Wu L, Chong SC, Passovets S, Xu D, Xu Y *et al.* (2004). Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664–675.
- Truex MJ, Peyton BM, Valentine NB, Gorby YA. (1996). Kinetics of U(VI) reduction by a dissimilatory Fe(III)-reducing bacterium under nongrowth conditions. *Biotechnol Bioeng* **55**: 490–496.
- Verdick D, Handran S, Pickett S. (2002). Key considerations for accurate microarray scanning and image analysis. In: Kamberova G (ed). *DNA Array Image Analysis: Nuts and Bolts*. DNA Press LLC: Salem, MA, pp 83–98.
- Wu L, Liu X, Schadt CW, Zhou J. (2006a). Microarray-based analysis of sub-nanogram quantities of microbial community DNAs using Whole Community Genome Amplification (WCGA). *Appl Environ Microbiol* **72**: 4931–4941.
- Wu L, Thompson DK, Li G, Hurt R, Huang H, Tiedje JM *et al.* (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67**: 5780–5790.
- Wu L, Thompson DK, Liu X, Fields MW, Bagwell CE, Tiedje JM *et al.* (2004). Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ Sci Tech* **38**: 6775–6782.
- Wu W, Carley J, Fienen M, Mehlhorn T, Lowe K, Nyman J *et al.* (2006b). Pilot-scale *in situ* bioremediation of uranium in a highly contaminated aquifer. 1. Conditioning of a treatment zone. *Environ Sci Technol* **40**: 3978–3985.
- Wu W, Carley J, Gentry T, Ginder-Vogel MA, Fienen M, Mehlhorn T *et al.* (2006c). Pilot-scale *in situ* bioremediation of uranium in a highly contaminated aquifer. 2. Reduction of U(VI) and geochemical control of U(VI) bioavailability. *Environ Sci Technol* **40**: 3986–3995.
- Zhou J. (2003). Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288–294.
- Zhou J, Bruns MA, Tiedje JM. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 461–468.
- Zhou J, Thompson DK. (2002). Challenges in applying microarrays to environmental studies. *Curr Opin Biotech* **13**: 204–207.
- Zhou J, Thompson DK, Xu Y, Tiedje JM. (2004). *Microbial Functional Genomics*. John Wiley & Sons: Hoboken, New Jersey, p 624.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)