

# GeoChip 3.0: Further Development and Applications of Functional Gene Arrays (FGAs) for Analysis of Microbial Communities

Zhili He<sup>1,4</sup>, Ye Deng<sup>1</sup>, Liyou Wu<sup>1</sup>, Joy Van Nostrand<sup>1,6</sup>, Christopher Hemme<sup>1,6</sup>, Christopher Schadt<sup>2</sup>, Terry Gentry<sup>3</sup>, Weimin Wu<sup>4</sup>, Laurie Kellogg<sup>1</sup>, Baohua Gu<sup>1</sup>, David Watson<sup>3</sup>, Terry C. Hazen<sup>1,6</sup>, Phil Jardine<sup>2</sup>, Craig S. Criddle<sup>1</sup>, and Jizhong Zhou<sup>1,6</sup>

<sup>1</sup>University of Oklahoma, Norman, OK; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, TN; <sup>3</sup>Texas A&M University, College Station, TX; <sup>4</sup>Stanford University, Stanford, CA; <sup>5</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>6</sup>Virtual Institute for Microbial Stress and Survival



## ABSTRACT

Microarray technology provides the opportunity to identify thousands of microbial genes or populations simultaneously. Recently, a comprehensive functional gene array, called GeoChip 2.0, has been developed, evaluated and applied for characterizing microbial communities in natural systems. GeoChip 2.0 contains 24,243 oligonucleotide (50mer) probes and covers > 10,000 genes in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. It is a powerful generic tool, and can be used for: (i) profiling various environmental samples, such as soil, groundwater, sediments, oil fields, deep sea, animal guts, and etc; (ii) 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; (iii) exploring direct linkages of microbial genes/populations to ecosystem processes and functions; and (iv) detecting functional genes and/or organisms in a particular environment. Here, we present an application example on the dynamics and stability of microbial genes and associated communities during a bioremediation period at the Oak Ridge Field Research Center (FRC). Due to exponential increases in the number of genes and the number of sequences for each gene, a new generation of such an array (GeoChip 3.0) is in development. GeoChip 3.0 is expected to have more features, such as (i) it is more comprehensive, covering >37,700 gene sequences of 290 gene families; (ii) it includes the phylogenetic marker, *gyrB*; (iii) it is more automatic for sequence retrieval and selection, probe design and verification, array construction and data analysis, information storage, and automatic update, which greatly facilitate the management of such a complicated array, especially for future updates.

## Current version: GeoChip 2.0

**Table 1** List of major functional markers on the GeoChip 2.0

Gene category	Example of key enzyme (gene)	Total probe
<b>Nitrogen cycling</b>		<b>5310</b>
Nitrogen fixation	Nitrogenase ( <i>nifH</i> )	1225
Nitrification	Nitrite oxidoreductase ( <i>nxrA</i> ), nitrate reductase ( <i>narX</i> ), nitrate reductase ( <i>narX</i> ), nitrite reductase ( <i>nr1</i> ), nitrite reductase ( <i>nr2</i> )	306
Nitrogen assimilation	Glutamine synthetase ( <i>glnA</i> ), glutamate dehydrogenase ( <i>gdh</i> )	1432
<b>Carbon cycling</b>		<b>4599</b>
Carbon fixation	Rubisco ( <i>rbcL</i> , <i>rbcS</i> , <i>rbcX</i> , <i>rbcY</i> ), CDMC/PP2C	1019
Cellulose degradation	Cellulase, endoglucanase	1269
Lignin degradation	Laccase, ligninase	513
Chitin degradation	Chitinase ( <i>chiA</i> ), chitinase ( <i>chiB</i> )	744
Methane production	Methanococcus ( <i>mcrA</i> ), methanobrevibacter ( <i>mcrB</i> )	433
Methane oxidation	Methane monooxygenase ( <i>pmoA</i> )	330
Others	Lignin peroxidase ( <i>lign</i> ), peroxidase, cellulase	266
<b>Sulfate reduction</b>		<b>1615</b>
Phosphorus utilization	Phosphatase ( <i>phoA</i> , <i>phoX</i> )	145
<b>Metal reduction and resistance</b>		<b>4546</b>
Acidic resistance	Acid shock protein ( <i>aspA</i> ), <i>aspB</i> , <i>aspC</i>	873
Cadmium resistance	Cadmium transport ( <i>cadA</i> , <i>cadB</i> )	282
Chromium resistance	Chromium transport ( <i>chrA</i> )	319
Metal resistance/resistance	Metal ion reductase ( <i>mirA</i> ), metal reductase ( <i>mirB</i> )	548
Nickel resistance	Nickel transport ( <i>nick</i> ), <i>nickA</i> , <i>nickB</i>	180
Zinc resistance	Zinc resistance protein ( <i>znr</i> )	128
Other metal resistance/resistance	Other metal resistance/resistance ( <i>zinc</i> , <i>nick</i> , <i>zinc</i> , etc.)	2263
<b>Contaminant degradation</b>		<b>8028</b>
Biogenic amine, dihydroxybenzoyl, and tyrosine (BTEK) & related aromatic	Diarylesterase ( <i>dehA</i> ), dihydroxybenzoyl synthase ( <i>dhbA</i> ), tyrosine decarboxylase ( <i>tyrD</i> ), benzoyl-CoA reductase ( <i>bcrA</i> ), and aromatic 1,2-dioxygenase ( <i>aroA</i> )	4174
Chlorinated aromatic	Chlorobenzene reductase ( <i>chrA</i> )	90
Nitroaromatic	Nitrobenzene reductase ( <i>nirA</i> ), 4-nitrobenzoyl reductase ( <i>nirB</i> )	152
Polycyclic aromatic hydrocarbon (PAH)	Polycyclic aromatic hydrocarbon reductase ( <i>phrA</i> )	268
Chlorinated hydrocarbon (CHC)	Chlorinated hydrocarbon reductase ( <i>chrA</i> )	743
Chlorinated carbon (CC)	Polycyclic aromatic hydrocarbon reductase ( <i>phrA</i> )	268
Other organic compounds/by-product	Aromatic hydrolyase ( <i>ahp</i> ), homocysteine 1,2-dioxygenase ( <i>hcd</i> ), vanillin 1,6-dioxygenase ( <i>vanA</i> )	239
<b>Total</b>		<b>34243</b>

## CONCLUSIONS

- GeoChip has been constructed with more than 24,000 oligos covering more than 10,000 gene sequences. To our knowledge, this is the most comprehensive functional gene array currently available for environmental studies.
- GeoChip has been evaluated, and demonstrates that it can be used as a powerful tool for a rapid, high-through-put and cost-effective analysis of microbial communities.
- Microbial activities and associated communities were successfully monitored for *in situ* bioremediation at the Oak Ridge FRC site.
- A new generation of GeoChip (version 3.0) with more features is in development, which is expected to provide a more comprehensive picture for a given microbial community.

## ACKNOWLEDGEMENTS

This research was funded by the U.S. Department of Energy (Office of Biological and Environmental Research, Office of Science) grants from the ERSP Program and the GTL:Genomics ESPP Program.

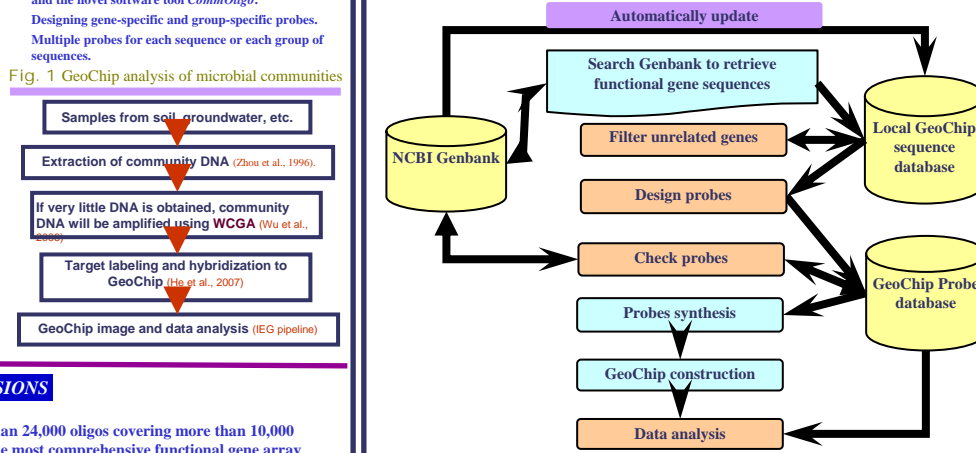


## Further Development of GeoChip 3.0

### New features for GeoChip 3.0

- GeoChip 3.0 is more comprehensive, and is expected to cover >37,700 (~10,000 for GeoChip 2.0) gene sequences of 290 gene families (~150 gene family on GeoChip 2.0) (Table 2). Thus, GeoChip 3.0 will be more representative.
- The homology of automatically retrieved sequences by key words is verified by HUMMER using seed sequences so that unrelated sequences can be removed.
- A software package (including databases) has been developed for sequence retrieval, probe and array design, probe verification, array construction, array data analysis, information storage, and automatic update, which greatly facilitate the management of such a complicated array, especially for future updates (Fig. 2).
- GeoChip has implemented a universal standard, which can compare different samples, and normalize data.
- GeoChip 3.0 implements a genomic control/standard, which can quantitatively analyze functional gene data.
- GeoChip 3.0 also includes GeoChip 2.0 probes, and those probes are checked against with new databases.

**Fig. 2** Work flow for GeoChip 3.0 design, construction and data analysis



**Fig. 1** GeoChip analysis of microbial communities

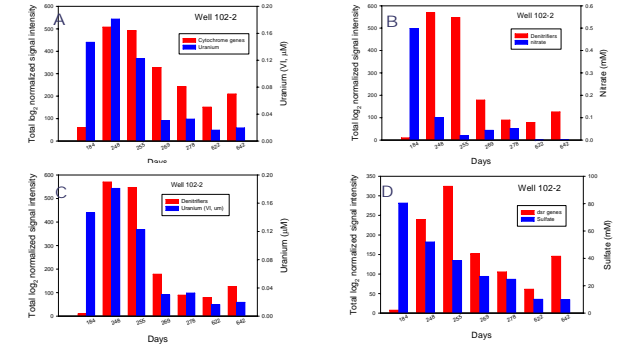
**Table 2.** The summary of GeoChip 3.0 probes and sequences information

Summary of GeoChip 3.0 probes and gene sequences (Up to May 2007)					
Gene category	Cat. No.	DL seq. no.	PD seq. no.	Total probe	Total CDS
Carbon degradation	24	18337	4092	1924	3192
Carbon fix	5	4682	2218	887	1614
Methane reduction/oxidation	3	4134	1853	447	752
Metal Resistance	43	28820	9625	3510	7021
Nitrogen	12	20800	19229	4006	7334
Organic Remediation	197	55598	18650	7093	12843
Phosphorus	2	1876	1441	471	1069
Sulphur	3	2523	2291	1464	1800
Others (e.g. <i>gyrB</i> )	1	8163	5252	1040	2089
<b>Total</b>	<b>290</b>	<b>144,933</b>	<b>64,651</b>	<b>20,842*</b>	<b>37,714</b>

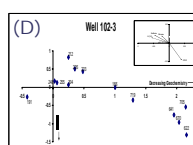
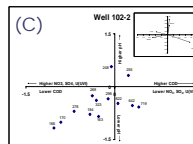
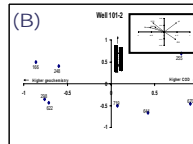
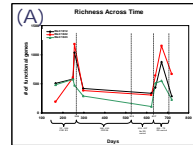
\*Only a single probe is counted for each sequence or each group of sequences.

## Monitoring microbial activities during *in situ* bioremediation of uranium at the FRC site in Oak Ridge using GeoChip 2.0

**Fig. 3** The microbial community dynamics from the monitoring well (102-2) during the bioremediation period from day 166 (3/1/2004) to day 719 (8/31/2005).



**Fig. 4** Statistical analysis of richness across operational time and correlations (A) between GeoChip 2.0 results, geochemistry data, time and microbial communities (B, C, D).



**(B) CCA analysis for Well 101-2**  
**Geochemistry:** (i) Nitrate, COD, and sulfate were most highly correlated to axis 1, and (ii) pH, U(VI) were positively correlated to axis 2 (pH more strongly than U(VI)).  
**Time/communities:** (i) Day 166, 248, 298, and 622 were all associated with increasing nitrate and sulfate, (ii) Day 166 and 248 were also associated with increasing U(VI), (iii) Day 255 was associated with pH and COD, and (iv) Days 641, 670, and 719 were strongly associated with COD.

**(C) CCA analysis for Well 102-2**  
**Geochemistry:** (i) Nitrate and U(VI) were both strongly correlated to the first axis (negatively). However, they diverged from each other along the gradient; (ii) COD again was highly positively correlated to the first axis; and (iii) pH and sulfate were both highly correlated to second axis.  
**Time/communities:** (i) The early time periods (with the exception of 248 and 255) were distributed along a gradient of nitrate and U(VI), (ii) Days 248 and 255 were more correlated to sulfate and pH (primarily pH), (iii) The later time periods were distributed along a gradient of COD; and (iv) there is a temporal gradient that is likely being strongly influenced by the decreasing nitrate and U(VI) and increasing COD.

**(D) Analysis for Well 102-3**  
**Geochemistry:** (i) All but the COD were negatively associated with the first axis, and (ii) COD was highly negatively correlated to the second axis.  
**Time/communities:** (i) The gradient of communities is moving away from the geochemistry and toward the COD, (ii) Day 191 is highly correlated to pH and nitrate, (iii) Day 166, a bit of an outlier, is highly correlated to axis 1 and is being pulled by very high nitrate, but low U(VI) and pH, and (iv) all of the other communities are along a changing gradient beginning with high geochemistry and low COD, to the later time points where geochemistry is much lower and COD much higher.