



Applying GeoChip Analysis to Disparate Microbial Communities

These high-throughput gene arrays, used to analyze microbial communities, link their structures to how they affect ecosystems

Jizhong Zhou, Zhili He, Joy D. Van Nostrand, Liyou Wu, Ye Deng

Microorganisms, the most diverse group of life, play integral roles in ecosystem functions and biogeochemical cycling of carbon, nitrogen, sulfur, phosphorus, and various metals. Microorganisms typically form complex communities whose structure, functions, interactions, and dynamics are critical to our lives. For instance, dramatic changes involving microbial communities from epidemics, plant or animal disease outbreaks, biological terrorist attacks, or atmospheric alterations from global climate change could prove a disaster for the environment and us. Although it is critical to understand which microorganisms are present and what they are doing, counting them is difficult, making it a major challenge to address such issues in quantitative ways and, thus, to make accurate predictions.

However, the recent development of large-scale, high-throughput metagenomics technologies along with GeoChip functional gene arrays are making it possible to obtain community-wide spatial and temporal information on microbial communities. With the advent of these new tools, microbial ecologists will no longer be frustrated by the shortage of data to address ecological questions. Large-scale metagenomic sequencing and high-throughput functional gene array technologies instead enable scientists to address research questions that formerly they could not. These new technologies are significantly reshaping the field of microbial ecology.

Why GeoChip?

Researchers who are detecting, identifying, characterizing, and quantifying microbial communities face several grand challenges. First, microbial communities are extremely diverse, whether in soil, water, food, or our bodies. One gram of a typical soil contains as many as 40,000 microbial species. Characterizing such a vast diversity and understanding the mechanisms shaping it presents numerous obstacles. Also, the majority of these microorganisms, more than 99%, have not been cultured. In addition, although microorganisms mediate many ecosystem processes, the relationships between microbial diversity and ecosystem functions

Jizhong Zhou is Presidential Professor of the Department of Botany and Microbiology and Director of the Institute for Environmental genomics, University of Oklahoma, Norman; Zhili He and Liyou Wu are research scientists at the Institute for Environmental Genomics; and Joy D. Van Nostrand and Ye Deng are postdoctoral research associates at the Institute for Environmental Genomics.

Summary

- GeoChip is a powerful, high-throughput metagenomic tool for analyzing microbial communities, including their structure, metabolic potential, diversity, and their impact on ecosystem functions.
- GeoChip was a recipient in 2009 of an R&D100 Award from *R&D Magazine*, which recognizes the 100 most innovative scientific and technical breakthroughs of the year.
- Recent developments helped to improve GeoChip in terms of specificity, sensitivity, probe design, computer program development, and target amplification of DNA and RNA molecules.
- These developments are making it possible to use functional gene arrays to analyze environmental samples, despite bottlenecks in processing and interpreting hybridization data.

remain elusive. This problem is due in part to a lack of sufficient information on microbial community-wide spatial and temporal dynamics.

To assess microbial community structure, functions, and dynamics in natural settings, measurement tools and techniques need to be: (i) simple, rapid, and robust; (ii) specific and sensitive with a broad comprehensive coverage of target microorganisms; (iii) quantitative and accurate with wide dynamic ranges; (iv) capable of detecting functions with high resolution; (v) capable of high throughput and in-parallel analysis; (vi) capable of making reliable comparisons across different sites, experiments, laboratories and times; and (vii) cost-effective. However, no such approaches are available.

Traditional culturing techniques provide only an extremely limited view of microbial diversity and functions. Meanwhile, although they are valuable, conventional nucleic acid detection approaches, such as 16S rRNA gene-based cloning methods, denaturing gradient gel electrophoresis, terminal-restriction fragment length polymorphism, quantitative PCR, and in situ hybridization also do not meet these analytic requirements. Each of these gene-based molecular approaches is slow and laborious or is of low throughput, has low resolution, is not quantitative, or provides limited functional information.

Although microarray technology is a powerful tool for analyzing gene expression, applying this technology to environmental studies presents numerous challenges in terms of probe design, the coverage of gene sequences, specificity, sensitivity and quantitation. To overcome such obstacles, a particular type of microarray, called functional gene arrays (FGAs), was developed and is now in widening use. FGAs contain probes targeting genes involved in various microbially mediated processes, and they are particularly useful in linking microbial diversity to community functions because the arrays target functionally known genes. Among various types of functional gene arrays, GeoChip is the most comprehensive array developed for biogeochemical, ecological, and environmental analyses.

GeoChip Features Probes for Genes Involved in Biogeochemical Cycling

GeoChip contains oligonucleotide probes for genes involved in biogeochemical cycling of carbon, nitrogen, phosphorus, sulfur and various metals, antibiotic resistance, metal resistance, organic contaminant degradation, and energy processing, as well as *gyrB*-based phylogenetic markers. The current version of GeoChip contains 50-nucleotide probes with high resolution, specificity, and sensitivity for distinguishing about 70,000 functional gene variants in more than 400 func-

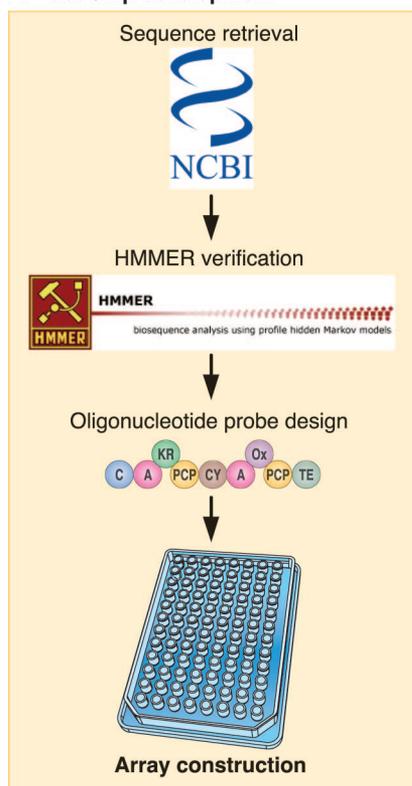
Summary of probe and covered sequence information on the current version of GeoChip

Gene category	Gene or enzymes number	Probe number	Sequence-specific probes	Group-specific probes	Covered CDS
Carbon cycling	41	5,196	1,765	3,431	10,573
Carbon fixation	5	1165	270	895	2794
Cellulose degradation	4	305	151	154	484
Lignin degradation	4	330	274	56	419
Other carbon compound degradation	25	3142	998	2243	6434
Methane oxidation	2	118	87	31	193
Methanogenesis	1	136	84	52	249
Nitrogen cycling	16	3,763	2,148	1,615	7,353
Nitrogen fixation	1	1224	764	460	2277
Nitrification	2	111	69	42	257
Denitrification	5	1543	1061	482	2995
Other nitrogen processes	8	885	254	631	1824
Phosphorus cycling	3	599	183	416	1,220
Sulphur cycling	4	1,504	1,083	421	2,042
Sulfite reduction	3	1309	1017	292	1703
Sulphur oxidation	1	195	66	129	339
Energy processing	2	508	410	98	671
c-type cytochromes	1	384	370	14	398
Hydrogenases	1	124	40	84	273
Metal Resistance	41	4,870	603	4,267	10,962
Arsenic	3	392	59	333	814
Mercury	6	291	77	214	620
Chromium	1	543	65	478	1292
Other metals	31	3644	402	3239	8236
Organic contaminant degradation	172	8,609	2,160	6,449	16,948
Aromatics	124	6192	1654	4538	12416
Chlorinated solvents	5	203	41	162	449
Herbicides and Pesticides	16	1104	172	932	2286
Other organic contaminants	27	1110	293	817	1797
Antibiotic resistance	11	1,539	263	1,276	2,784
Beta-lactamases	4	353	146	207	554
Transporters	5	1182	103	1021	2103
Others	2	62	14	48	127
Phylogenetic marker (<i>gyrB</i>)	1	1,318	947	371	1,777
Pathways for human microbiomes	136	7,763	6,958	805	12,601
Total	438	35,669	16,520	19,149	66,931

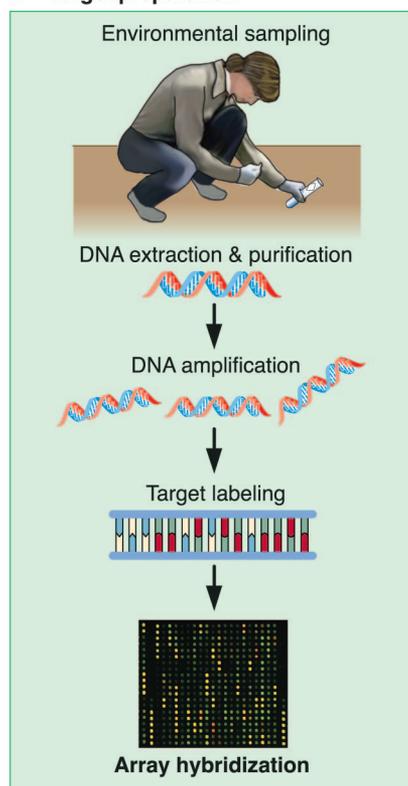


FIGURE 1

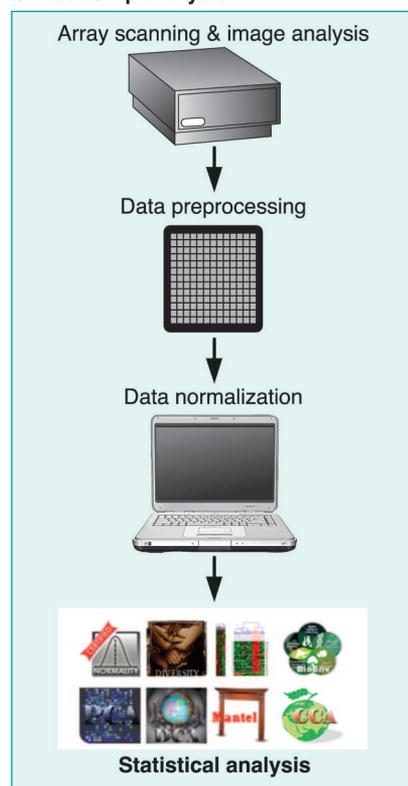
A GeoChip development



B Target preparation



C GeoChip analysis



GeoChip operation. (A) GeoChip development. (B) Target preparation. (C) GeoChip.

tional categories (Table 1). Its control probes include positive probes for ribosomal genes and negative probes for human, plant, and/or hyperthermophile genes.

Compared to other types of high-throughput technologies such as barcode-based pyrosequencing and 16S rRNA gene-based phylogenetic arrays, GeoChip has several unique characteristics in terms of content, coverage, design, and application. First, it provides functional information on microbial communities related to biogeochemical, ecological, and environmental processes. Second, GeoChip has higher resolution because functional genes are targeted, enabling it to differentiate microorganisms at the species and strain level. Reasons for this superiority include the fact that functional gene markers are generally more divergent than are phylogenetic gene markers. Third, GeoChip hybridization is quantitative because no PCR amplification is involved. This is critical for ecosystem-level studies. Fourth, PCR amplification-

based approaches cannot be used to detect and quantify many functional genes of interest because either appropriate probes are unavailable or functional genes prove difficult to amplify from environmental samples. In addition, GeoChip is designed to assay both DNA and RNA, and thus provide information on both community structure and activities. Finally, GeoChip also contains a universal standard consisting of an artificial sequence that is specific only to a complementary oligonucleotide that is labeled with a fluorescent dye and added to all samples. The universal standard hybridizes to each spot on an array and provides a way to normalize signal intensity for comparisons of different microbial communities across different sites, experiments, and laboratories.

Gauging How Well GeoChips Work

The current version of GeoChip contains approximately 30,000 probes whose sequences are

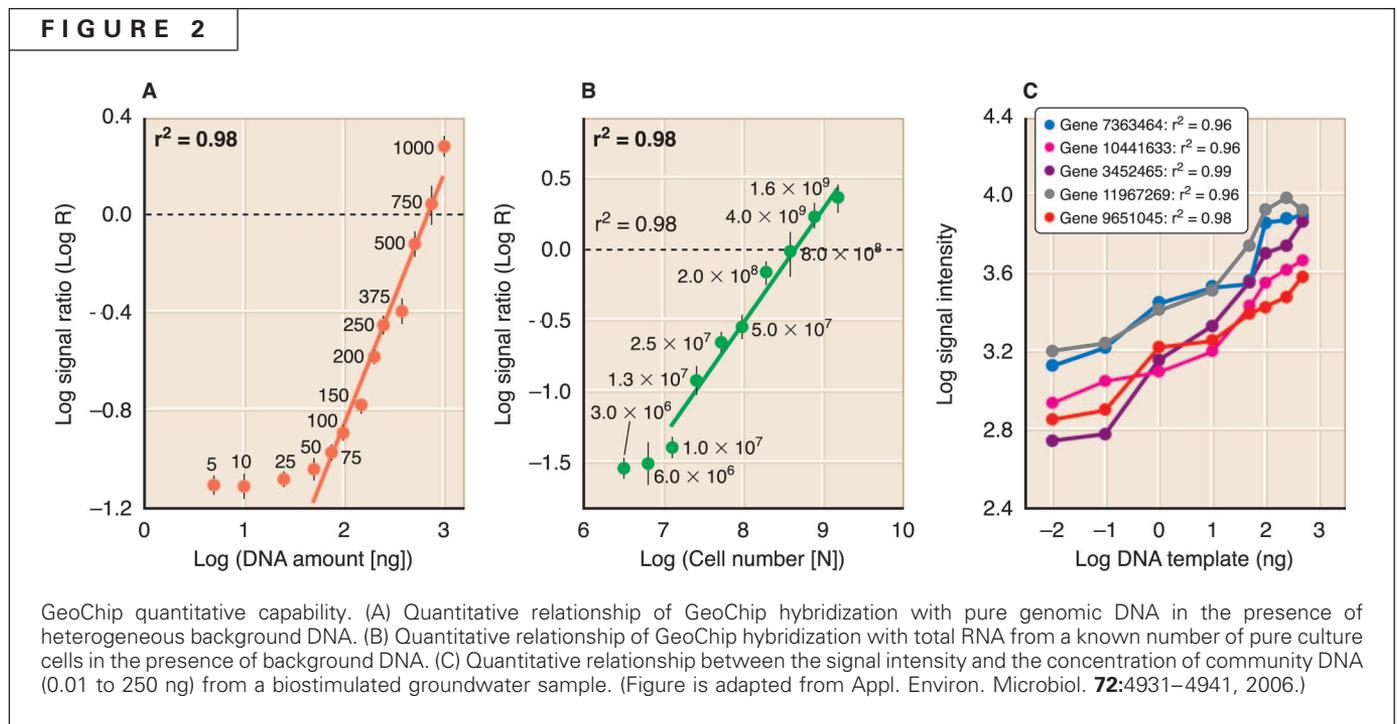
matched with those from public databases via key words. Once sequences are retrieved, a program called HMMER is used to identify correct sequences via seed sequences with validated functions. Specific probes are based on correctly identified sequences using the program *CommOligo*, and they are commercially synthesized for array construction (Fig. 1A). Fabricated arrays are stored at room temperature. Next, nucleic acids (DNA or RNA) are extracted and purified from environmental samples (Fig. 1B), amplified, and then labeled with fluorescent dyes before hybridization with GeoChip. After washing away free DNA/RNA, the fluorescently labeled and captured DNA/RNA molecules are digitally imaged by a laser scanner (Fig. 1C), and this image is used to assess the abundance of target DNA/RNA species. Statistical tools are available to rapidly analyze these hybridization data.

Specificity is a key parameter for determining the quality of microarray data, especially when those data are from environmental samples. Probe design and hybridization conditions help to control GeoChip specificity. For instance, hybridizing samples at 50°C in 50% formamide differentiates sequences with less than 90–92% identity. With such high resolution, GeoChip differentiates microorganisms at the species-strain level. Meanwhile, software packages such

as *CommOligo* are available for designing microarray probes, and they yield false positives in the range of 0.002–0.004% and no false negatives.

Sensitivity is another major concern, especially for samples from environments with many gene variants in low abundance. With GeoChips having 50-mer probes, the detection limits range from 5 to 10 ng of genomic DNA in the absence of background DNA, 50 to 100 ng of pure-culture genomic DNA in the presence of background DNA, or about 10⁷ cells. Such sensitivity is sufficient for analyzing environmental samples where microbial populations are abundant, such as those from bioreactors and wastewater treatment systems. But it may not be high enough for analyzing most environmental samples in natural settings such as soils, groundwater, and marine water columns because the populations in these samples are generally lower than the detection limits.

This challenge is being met through use of a whole-community genome amplification (WCGA)-assisted microarray, which can detect individual genes or genomes while using as little as 1 to 500 ng of community DNA. Even smaller amounts of DNA, as little as 10 fg from as few as two bacterial cells, can be detected by using a modified amplification buffer. Further, the WCGA approach yields significant linear rela-





tionships between signal intensity and initial DNA concentrations or cell numbers (Fig. 2).

Although detecting DNA tells whether particular genes are present in an environment, measures of microbial activity depend on detecting RNA molecules. Because WCGA amplifies DNA but not RNA, investigators developed whole-community RNA amplification (WCRA) methods to provide enough mRNA from environmental samples for microarray analysis. Current WCRA approaches work with as little as 10 ng of RNA to detect active populations in natural settings.

Insights from GeoChip Analysis of Environmental Sites

Here are several examples of how GeoChip is being used to profile microbial communities from different habitats:

- When microbial communities in grassland soils were exposed to higher-than-usual amounts of carbon dioxide, GeoChip analysis correlated community changes with changes in soil carbon and nitrogen as well as plant productivity. For example, genes involved in fixing nitrogen and carbon significantly increased when carbon dioxide levels rose, while genes involved in degrading recalcitrant soil carbon remained unchanged. These results have important implications for the feedback responses of ecosystems to atmospheric carbon dioxide and to global climate change.
- One question for microbial communities in forest soils is whether they fall into spatial patterns resembling those of plant and animal communities, in which there is a positive-power law relationship among species (or taxa), richness, and area. Such taxa-area relationships are fundamental to understanding the distribution of global biodiversity, but they prove elusive for microbial communities, especially in soil habitats. GeoChip analysis suggests that forest soil microbial communities exhibit relatively flat gene-area relationships, while varying across different functional and phylogenetic groups. Moreover, the spatial turnover of microorganisms appears to be lower than that of plants and animals.
- GeoChip analysis suggests that the diversity of microbial communities in and around

deep-sea hydrothermal vents is lower in the inner than the outer part of the chimney. Also, these microbial communities appear capable of fixing carbon, oxidizing and generating methane, and fixing nitrogen. Overall, the hydrothermal microbial communities are metabolically and physiologically highly diverse, and undergo rapid dynamic succession and adaptation in response to the steep temperature and chemical gradients across the chimney.

- When GeoChip was used to analyze basalts of varying ages from the East Pacific Rise and from neighboring seamounts, the arrays detected genes encoding carbon fixation, methane oxidation, methanogenesis, and nitrogen fixation, suggesting that basalts also harbor extensive metabolic diversity.
- GeoChip has been used to monitor how efficiently microbial communities degrade and remediate contaminants. In its favor, GeoChip can quickly identify key functional genes and microbial populations that may be responding to changes in environmental conditions, such as shifts in heavy metals, oxygen, and pH. For example, sulfate-, nitrate/nitrite-, and metal-reducing microorganisms were responsible for rapid uranium (VI) reduction. Ethanol, which served as an electron donor and carbon source, was an important driver in determining community structure, which remained relatively constant with changes in oxygen or nitrate.
- GeoChip analysis revealed significant shifts in microbial community structure that could account for changes that lead them to corrode pipes and other industrial components.
- Analysis of microbial communities associated with healthy and yellow band-diseased coral, *Montastraea faveolata*, suggests that genes encoding cellulose degradation and nitrification pathways may provide competitive advantages to coral pathogens.

Future Prospects

GeoChip is a powerful, high-throughput metagenomic tool for analyzing microbial community functional structure, metabolic potential, and diversity, and for linking microbial community structure to ecosystem functioning. Significant advances in microarray-based detection are

making it possible to use functional gene arrays to analyze environmental samples with sufficient specificity, sensitivity, and quantitative accuracy. One bottleneck is in processing the enormous amount of data generated by GeoChip hybridizations. Synthesizing and interpreting the hybridization data within a biological context is even more difficult. Thus, novel mathe-

matical and bioinformatic tools are urgently needed to speed data processing, data mining, and visualization.

GeoChip is only a tool, and as such it should be integrated with studies focused on clear and addressable ecological and environmental questions and hypotheses. A new era of quantitative predictive microbial ecology is coming.

ACKNOWLEDGMENTS

Preparation of this manuscript was supported by The United States Department of Energy under the Genomics: GTL program through the Virtual Institute of Microbial Stress and Survival (VIMSS; <http://vimss.lbl.gov>), and Environmental Remediation Science Program, as well as by the Oklahoma Center for the Advancement of Science and Technology under Oklahoma Applied Research Support Program.

SUGGESTED READING

- Deng, Y., Z. He, J. D. Van Nostrand, and J.-Z. Zhou. 2008. Design and analysis of mismatch probes for long oligonucleotide microarrays. *BMC Genomics* 9:491.
- Gao, H., Z. K. Yang, T. J. Gentry, L. Wu, C. W. Schadt, and J.-Z. Zhou. 2007. Microarray-based analysis of microbial community RNAs by Whole community RNA amplification (WCRA). *Appl. Environ. Microbiol.* 73: 563–571.
- He, Z., T. J. Gentry, C. W. Schadt, L. Wu, J. Liebich, S. C. Chong, W.M. Wu, B. Gu, P. Jardine, C. Criddle, and J.-Z. Zhou. 2007. GeoChip: A comprehensive microarray for investigating biogeochemical, ecological, and environmental processes. *ISME J.* 1:67–77.
- He, Z., L. Wu, X. Li, M. W. Fields, and J.-Z. Zhou. 2005. Empirical establishment of oligonucleotide probe design criteria using perfect match and mismatch probes and artificial targets. *Appl. Environ. Microbiol.* 71:3753–3760.
- He, Z., M. Xu, Y. Deng, S. Kang, L. Kellogg, L. Wu, J. D. Van Nostrand, S. E. Hobbie, P. B. Reich, and J.-Z. Zhou. 2010. Metagenomic analysis reveals a marked divergence in the functional structure of below ground microbial communities at elevated CO₂. *Ecol. Lett.*, in press.
- Kimes, N. E., J. D. Van Nostrand, E. Weil, J.-Z. Zhou, and P.J. Morris. 2009. Microbial functional structure of *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. *Environ. Microbiol.* Online
- Li, X., Z. He, and J.-Z. Zhou. 2005. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res.* 33:6114–6123.
- Mason, O. U., C. A DiMeo-Savoie, J. D. Van Nostrand, J.-Z. Zhou, M. R. Fisk, and S. J. Giovannoni. 2009. Prokaryotic diversity, distribution, and preliminary insights into their role in biogeochemical cycling in marine basalts. *ISME J.* 3: 231–242.
- Wang, F., H. Zhou, J. Meng, X.T. Peng, L. Jiang, P. Sun, C. Zhang, J. D. Van Nostrand, Y. Deng, Z. He, L. Wu, J.-Z. Zhou, and X. Xiao. 2009. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proc Nat. Acad. Sci.* 106:4840–4845.
- Wu, L., X. Liu, C. W. Schadt, and J.-Z. Zhou. 2006. Microarray-based analysis of sub-nanogram quantities of microbial community DNAs using whole community genome amplification (WCGA). *Appl. Environ. Microbiol.* 72:4931–4941.
- Zhou, J.-Z., S. Kang, C. W. Schadt, and C.T. Garten, Jr. 2008. Spatial scaling of functional gene diversity across various microbial taxa. *Proc Nat. Acad. Sci.* 105:7768–7773.