Front. Environ. Sci. Engin. China 2011, 5(1): 1–20
DOI 10.1007/s11783-011-0301-y

FEATURE ARTICLE

Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities

Zhili HE[1], Joy D. VAN NOSTRAND[1], Ye DENG[1], Jizhong ZHOU (✉)[1,2,3]

1 Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA
2 Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
3 Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China

Abstract Functional gene arrays (FGAs) are a special type of microarrays containing probes for key genes involved in microbial functional processes, such as biogeochemical cycling of carbon, nitrogen, sulfur, phosphorus, and metals, biodegradation of environmental contaminants, energy processing, and stress responses. GeoChips are considered as the most comprehensive FGAs. Experimentally established probe design criteria and a computational pipeline integrating sequence retrieval, probe design and verification, array construction, data analysis, and automatic update are used to develop the GeoChip technology. GeoChip has been systematically evaluated and demonstrated to be a powerful tool for rapid, specific, sensitive, and quantitative analysis of microbial communities in a high-throughput manner. Several generations of GeoChip have been developed and applied to investigate the functional diversity, composition, structure, function, and dynamics of a variety of microbial communities from different habitats, such as water, soil, marine, bioreactor, human microbiome, and extreme ecosystems. GeoChip is able to address fundamental questions related to global change, bioenergy, bioremediation, agricultural operation, land use, human health, environmental restoration, and ecological theories and to link the microbial community structure to environmental factors and ecosystem functioning.

Keywords functional gene arrays (FGAs), GeoChip, microbial communities, functional diversity/composition/structure, environmental factor, ecosystem functioning

Received October 7, 2010; accepted November 10, 2010

E-mail: jzhou@ou.edu

1 Introduction

Microorganisms are the most diverse groups of organisms known in terms of phylogeny and functionality. They can be found in the most inhospitable environments and play important and distinctive roles in ecosystems, such as biogeochemical cycling of carbon (C), nitrogen (N), sulfur (S), phosphorus (P), and metals (e.g., iron, copper, and zinc) and biodegradation or stabilization of environmental contaminants. Therefore, one of the most important goals of microbial ecology is to understand the diversity, composition, function, dynamics, and evolution of microbial communities and their relationships with environmental factors and ecosystem functioning. Toward such a goal, several challenges remain. First, microorganisms are generally too small to see or characterize with approaches for plant or animal studies. Second, an extremely high diversity has been observed in microbial communities. It has been estimated that one gram of soil contains 2000 to 50000 microbial species [1–5] and even up to millions of species [6]. Characterizing such a vast diversity and then attempting to understand environmental factors shaping it presents numerous obstacles. Third, a vast majority of microorganisms ($>99\%$) are unculturable [7], making it even more difficult to study their functional ability. While studies focused on culturable populations or isolates are still important, they only provide a limited view of the microbial community diversity and function, indicating that culture-independent approaches are necessary in order to study even a fraction of microorganisms in the environment. Finally, although microorganisms control, at least to some degree, various ecosystem processes, in most cases, we do not know what they are largely doing in ecosystems; thus, establishing mechanistic linkages between microbial diversity and ecosystem functioning is

even more difficult. To address these challenges, culture-independent high-throughput technologies for analysis of microbial communities are more important.

Many culture-independent approaches are available, including polymerase chain reaction (PCR)-based cloning analysis, denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism (T-RFLP), quantitative PCR, and *in situ* hybridization. However, both resolution and coverage of these methods are limited to provide a comprehensive view of a microbial community. For example, for functional genes, clone libraries composed of over 2000 clones were still insufficient to cover the *nifH* diversity of the Chesapeake Bay [8,9]. In fact, the clone library sizes commonly used may be too small by a factor of 10 or greater to capture the true diversity of microbial communities [10]. Also, most of these approaches rely on an initial PCR amplification step, which introduces well-known biases [11–13]. In addition, these PCR-based analyses are time-consuming and expensive, especially when many genes or samples are examined or if multiple runs are necessary to increase their coverage. Therefore, high-throughput metagenomic technologies are necessary for providing a rapid, specific, sensitive, quantitative, and comprehensive analysis of microbial communities and their relationships with environmental factors and ecosystem functioning.

Microarray-based technology can examine thousands of genes at one time, overcoming many of these obstacles and providing a much more comprehensive coverage of microbial communities studied. The first microarray reported was designed to monitor gene expression in the plant *Arabidopsis thaliana* [14]. Since then, hundreds to thousands of organism-specific arrays have been developed to examine gene expression under different conditions. The potential application for microarrays was greatly expanded when the utility of microarrays for environmental studies was demonstrated with 16S rRNA gene probes for detection of key genera of nitrifying bacteria [15]. Many studies demonstrate that microarrays can provide information on a microbial community in a rapid, high-throughput, and parallel manner. Also, microarrays are often more cost-effective than other molecular techniques due to the amount of data provided. In addition, since microarrays have a defined set of genes or microorganisms, they are ideal for comparing environmental samples from different sites, conditions, or times. These comparisons are difficult to make if clone libraries or high-throughput sequencing technologies are used since the chance that the same populations of a given microbial community are sampled over multiple sampling events would theoretically be quite small if the process is completely random [10,16]. These features make microarrays excellent tools for comparing microbial community structure, functions, activities, and dynamics as well as linking those properties with ecosystem functioning in natural settings. In this review, we focus on functional gene arrays (FGAs), especially the development and applications of GeoChip for analysis of functional diversity, composition, structure, function, and dynamics of microbial communities from different ecosystems.

2 Functional gene arrays

FGAs are composed of probes for key genes involved in microbial functional processes, such as biogeochemical cycling of carbon, nitrogen, phosphorus, sulfur, and metals, biodegradation of environmental contaminants, and stress responses [17–21]. Since the exact functions of selected genes on FGAs are known, this type of arrays is especially useful for providing insights into our understanding of the functional diversity, composition, and structure of microbial communities as well as linkages between microbial community structure and ecosystem functioning. Unlike PCR-based techniques, which limit the number of genes that can be examined at one time, FGAs allow for the simultaneous examination of many functional genes. Several FGAs have been reported and evaluated and generally target specific functional processes, populations, or environments, which are briefly described as follows.

2.1 FGAs for nitrogen cycling

The first reported FGA contained PCR-amplicon probes targeting 89 N-cycling genes (*nirS*, *nirK*, *amoA*, and *pmoA*) derived from pure culture isolates and clone libraries from marine sediments [18]. In addition, ten 16S rRNA genes were included as positive controls and five yeast genes as negative controls. A probe concentration of 200 $ng \cdot \mu L^{-1}$ in 50% dimethyl sulfoxide (DMSO) was determined to be optimum. An evaluation of different hybridization temperatures showed that different levels of specificities could be obtained by varying hybridization strengths. The sensitivity of this array was approximately 1.0 ng for pure culture DNA and 10 to 25 ng for community DNA. This array was then used to examine marine sediment and soil communities. Genes for *nirS*, *nirK*, and *amoA*/*pmoA* were detected in both samples with higher abundance observed from soil samples than marine sediment [18]. This FGA was served as the prototype of GeoChip in later development.

Taroncher-Oldenburg and colleagues [22] developed two oligonucleotide (70-mer) arrays to examine denitrification, nitrogen fixation, ammonia oxidation, and nitrite reduction with *amoA*, *nifH*, *nirS*, and *nirK* genes. Hybridizations were performed at 65°C, which provided a high specificity. Under these conditions, sequences with at least 84% to 90% similarity were detected with a detection limit of ~10^7 DNA copies. These arrays were

used to examine N-cycling communities from a river system and differences in the denitrifying community were observed. These variations were consistent with differences observed in the overall microbial community [22].

A macroarray was also developed to study nitrogenase gene diversity in the environment using PCR amplicons of *nifH* as probes on a nylon membrane [8]. At 60°C and 40% formamide, the array was able to detect sequences with $>78\%$ sequence similarities, and the detection limit was detected to be 13 pg target mL^{-1}. An expanded *nifH* macroarray was subsequently used to examine diazotroph communities in the Chesapeake Bay [9], indicating a high diversity of N-fixing microorganisms in this environment. In addition, differences in specific phylotypes in the surface water communities at different sampling stations across the Chesapeake Bay were observed.

A specific *nifH* microarray was applied to analyze the diversity (based on DNA) and activity (based on mRNA) of diazotrophs in roots of wild rice samples from Namibia. The results demonstrated that only a small subset of diazotrophs present in the sample actually had nitrogen fixation activity, suggesting that the developed *nifH* microarray was a highly reproducible and semiquantitative method for mapping the variability of diazotrophic diversity and allowed rapid comparisons of the relative abundance and activity of diazotrophic prokaryotes in the environment. A further refined *nifH* microarray with 194 oligonucleotide probes covered more than 90% of sequences in the *nifH* database at that time [23].

An FGA composed of probes for the *nodC* gene, which is a common *nod* gene essential for nodulation in all rhizobial species investigated so far, was developed and used to detect known and new *nodC* sequences [24]. One hundred thirty probes (41- and 50-mer) were designed from two conserved regions of the *nodC* gene. The array was then used to characterize rhizobial isolates. Subsequent sequencing of *nodC*-positive strains indicated that novel *nodC* sequences had been detected.

2.2 FGA for assessing methane metabolism

An FGA composed of oligonucleotide probes for 59 *pmoA/amoA* genes was developed to study methanotrophs [25]. Hybridizations were carried out using RNA extracted from samples taken at a landfill in Seiberdorf, Austria, and amplified using primers containing the T7 promoter. Fifty of the probes were positive for perfect-match or one mismatched sequences when hybridized at 55°C. For environmental samples, the detection limit for this array was approximately 5% of the total population. This array was expanded to include 68 oligonucleotide probes and used to study the effects of plant cover on methanotrophic communities from simulated landfills [26]. Using community DNAs, *Methylocystis* and *Methylocaldum* spp. were found to be the dominant members of the methanotrophic community, and supplemental biogas (CH_4 and CO_2 mix, 3:2) increased the overall abundance of methanotrophs. Using mRNA, *Methylocystis* and *Methylocaldum* spp. were also detected as the dominant methanotrophic genera in the simulated landfills, although *Methylocaldum* spp. activity was quite low and no activity was detected for *Methylocystis* spp. [27]. The results indicated that mRNA-based microarray analyses may provide additional information on the composition and functioning of microbial communities.

2.3 FGAs for pathogen detection and medical diagnosis

A virulence marker gene (VMG) array was developed to detect waterborne pathogens [28]. The VMG array contained two sets of probes, one (n = 791) targeting 35 VMGs from 12 pathogens and the second (n = 2,034) targeting 67 VMGs from 17 pathogens. This array was able to detect pathogens in tap water, effluent from a wastewater treatment plant, and river water at a relative abundance of as little at 0.01%. An antibiotic resistance array was developed as a potential high-throughput tool to identify antimicrobial resistance genes or to screen related strains [29]. This array was composed of PCR amplicons of ~550 bp in length derived from tetracycline genes. Of the 48 isolates tested, 38 showed positive hybridizations to one *tet* gene, one isolate hybridized to two probes, while the remainder did not have positive hybridizations. These results indicated that microarray technology could be used for screening a large number of different antibiotic resistance genes.

One potential application of FGAs is for high-throughput medical diagnosis or water and food safety testing. A preliminary medical diagnostic array was tested using oligonucleotide probes (70-mer) targeting *Salmonella*-specific virulence factors (*invA* and *sopB*) and an *Enterobacteriaceae* 16S rRNA gene-specific positive control probe [30]. Whole-genome amplified DNA from *Escherichia coli* DH5α and *Salmonella enterica* serovar Senftenberg labeled with Cy-5 were used for evaluating the array. When hybridizations were carried out at 42°C and 40% formamide, this array was able to detect the target sequences with as little as 1% of the original community.

A prototype array was designed to detect pathogens from the bloodstream [31]. The array was composed of 120 PCR amplicon probes (200–800 bp) for virulence genes, antibiotic resistance genes, and metabolic or structural genes specific to *Staphylococcus aureus*, *Escherichia coli*, and *Pseudomonas aeruginosa*. The array was tested using 45 clinical or reference strains of the target organisms. The hybridizations were specific for each strain tested. The presence of antibiotic resistance genes detected on the array was confirmed for *S. aureus* strains using traditional culture-based methods, which agreed with microarray results.

Another pathogen detection array was designed using 930 genes from 10 different pathogens (*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp., *Proteus* spp., *Klebsiella* spp., *Stenotrophomonas* spp., *Enterobacter* spp., *Acinetobacter* spp., *E. coli*, *Pseudomonas aeruginosa*, and *Candida albicans*), antimicrobial resistance genes, negative control probes from human, mouse, amoeba, and grass genes, and bacterial 16S rRNA gene probes as positive controls [32]. A multiplex PCR amplification was used with primers specific for all genes carried on the array (800 primer pairs). Hybridizations were carried out at 42°C and 25% formamide. With multiplex PCR amplification, the detection limit was 1.0 ng DNA. Swabs taken from an infected wound were used to evaluate the amplification method and the array; positive hybridizations to multiple *Enterococcus faecium* and *Staphylococcus epidermidis* specific probes were observed. These results were later confirmed by culturing.

2.4 FGA for analyzing bioleaching systems

An acid mine drainage (AMD) array was designed using 50-mer oligonucleotide probes for genes (n = 1071), both 16S rRNA and functional genes, relevant to AMD and bioleaching systems [33]. Samples from an AMD site (pH 2.0), a bioleaching system (pH 2.5), and a less acidic site (pH 5.0) were evaluated using this array. Functional genes involved in geochemical cycling of carbon, nitrogen, and sulfur, iron metabolism, DNA replication and repair, and metal resistance were detected. More 16S genes were detected in the less acidic samples, but higher functional diversity was observed in the lower pH samples. This study indicated that this AMD array could provide specific, sensitive, and quantitative detection of AMD microbial communities like bioleaching systems.

2.5 GeoChips

GeoChips constructed with 50-mer oligonucleotide probes have evolved over several generations. The first generation of GeoChip was constructed with 763 genes involved in nitrogen cycling (*nirS*, *nirK*, *nifH*, and *amoA*), methane oxidation (*pmoA*), and sulfite reduction (*dsrAB*). An evaluation showed that the developed array could provide species/strain-level resolution for analyzing microorganisms involved in nitrification, denitrification, nitrogen fixation, methane oxidation, and sulfite reduction [34]. Then, a similar expanded array was developed with 2402 genes involved in organic contaminant biodegradation and metal resistance to monitor microbial populations and functional genes involved in biodegradation and biotransformation [19]. Specificity tests with representative pure cultures indicated that the designed probes on the arrays appeared to be specific to their corresponding target genes. The detection limit was 5 to 10 ng genomic DNA in the absence of background DNA and 50 to 100 ng pure-culture genomic DNA in the presence of background DNA. Real-time PCR analysis was very consistent with the microarray-based quantification. This array was applied to analyze naphthalene-amended enrichment and soil microcosms, and the results indicated that this technology had the potential as a specific, sensitive, and quantitative tool in revealing a comprehensive picture of the compositions of biodegradation genes and the microbial community in contaminated environments [19]. Those arrays were referred to as GeoChip 1.0.

Although all of the FGAs mentioned above can be used to probe specific functional groups or activities, they lack a truly comprehensive probe set covering key microbial functional processes occurring in different environments. Also, many of these probe sets may not provide the specificity, sensitivity, and/or quantitative capability necessary to distinguish gene variants with high homologies. In addition, no standardized methods are available for analyzing such microarray data. To fill these gaps, a more comprehensive second generation of GeoChips (GeoChip 2.0) was developed and evaluated. For example, GeoChip 2.0 containing 24243 (50-mer) oligonucleotide probes, targeting ~10,000 functional genes from 150 gene families involved in the geochemical cycling of C, N, and P, sulfate reduction, metal reduction and resistance, and organic contaminant degradation, was developed as the first comprehensive FGA [21]. Two years ago, the third generation of GeoChip (GeoChip 3.0) was developed, which contains about 28000 probes and targeting ~57000 sequences from 292 gene families [17]. GeoChip 3.0 is more comprehensive and has several other distinct features compared to GeoChip 2.0, such as a common oligo reference standard (CORS) for data normalization and comparison [35], a software package for data management and future updating, the *gyr*B gene for phylogenetic analysis, and additional functional groups including those involved in antibiotic resistance and energy processing [17]. Based on GeoChip 3.0, the latest generation of GeoChip (GeoChip 4.0) in the NimbleGen format has been recently developed, which contains ~84000 probes and targeting > 152000 genes from 410 functional families. GeoChip 4.0 not only contains functional categories from GeoChip 3.0 but also includes additional functional categories, such as genes involved in stress responses, bacterial phages, and virulence. In addition, the current GeoChip also contains ~47000 probes derived from human microbiomes.

GeoChip technology has been recognized as a groundbreaking and cutting-edge technology. GeoChip was cited as an important tool for linking microbial communities to ecosystem functioning by the National Ecology Observatory Networks (NEON) Program in 2006. The Nature Publishing Group issued a press release to highlight this novel technology when the paper describing GeoChip 2.0

was published in the ISME Journal in May 2007. GeoChip was also highlighted in the Metagenomics Report by the National Academy of Sciences in 2007. Based on GeoChip 3.0, OU GeoChip was also awarded a 2009 R&D100 Award by R&D Magazine, which recognizes the 100 most innovative scientific and technical breakthroughs of the year that show great promise for commercialization. Therefore, in the following sections, the development and applications of GeoChip for microbial community analysis is focused on.

3 GeoChip development

GeoChip development involves several major steps, including selection of target genes, sequence retrieval and verification, oligonucleotide probe design, probe validation, array construction, and automatic update (Fig. 1(a)), and such processes are normally realized by a GeoChip development and data analysis pipeline [17].

3.1 Selection of target genes

A variety of functional genes can be used as functional markers targeting different processes, such as biogeochemical cycling of C, N, S, P, and metals, bioremediation of environmental contaminants, and antibiotic resistance. For example, 292 functional gene families were selected for GeoChip 3.0, and for the N cycle, 16 genes/enzymes were selected, including (i) *nifH* (dinitrogenase reductase) for N fixation; (ii) *amoA* (ammonia monooxygenase) and *hao* (hydroxylamine oxidoreductase) for nitrification; (iii) *gdh* (glutamate dehydrogenase) and *ureC* (urease) for ammonification; (iv) *nap*A (nitrate reductase) and *nrf*A (c-type cytochrome nitrite reductase) for dissimilatory N reduction to ammonium (DNRA); (v) *nasA* (nitrate reductase) and *nirA*/*nirB* (nitrite reductase) for assimilatory N reduction to ammonium; (vi) *narG* (nitrate reductase), *nirS*/*nirK* (nitrite reductase), *norB* (nitric oxide reductase), and *nosZ* (nitrous oxide reductase) for denitrification; and (vii) *hzo* (hydrazine oxidoreductase) for anaerobic ammonium oxidation (anammox). Similarly, a total of 48 genes/enzymes were selected for the N, P, and S cycles, 41 for metal resistance, 173 for biodegradation of a variety of organic contaminants, and 13 for antibiotic resistance and energy processing [17]. In addition, phylogenetic markers, such as 16S rRNA and DNA gyrase (GyrB) genes, are commonly used to examine the phylogenetic diversity, composition, and structure of microbial communities [37–41]. More importantly, if sequences for a known functional gene are available, they can be added in an updated GeoChip. For example, functional gene families have increased to 410 on GeoChip 4.0 from 292 on GeoChip 3.0. Most of those added functional gene families are involved in stress responses, bacterial phages, and virulence.

Generally, genes are chosen for key enzymes or proteins with the corresponding function(s) of interest. If a process

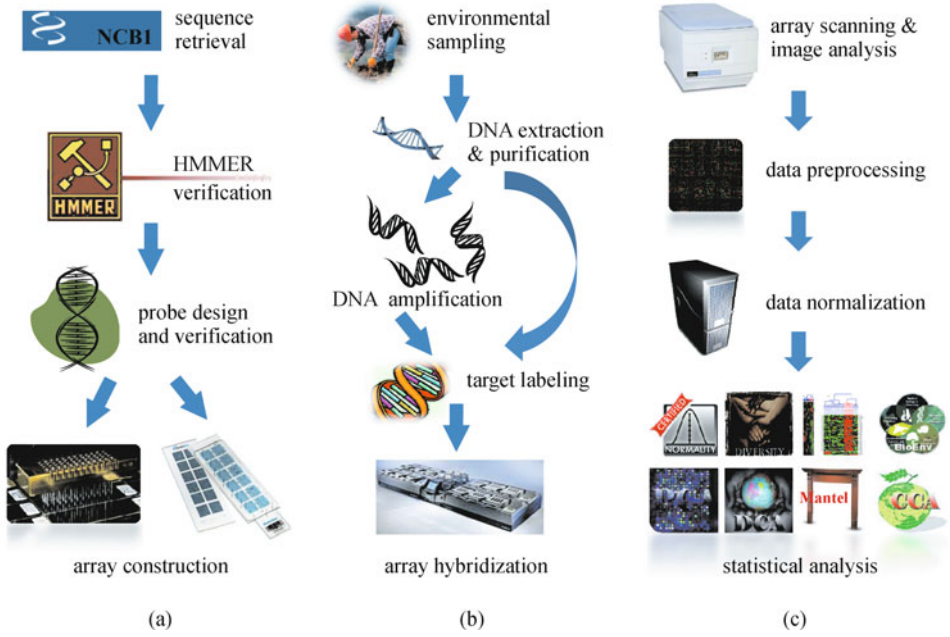

Fig. 1 Schematic presentation of GeoChip development and applications in analysis of microbial communities from a variety of habitats. (a) GeoChip development; (b) Target preparation; (c) GeoChip data analysis. This figure is adapted from Fig. 1 by Zhou et al. [36]

involves multiple steps or a protein complex, those genes responsible for catalytic subunits or with the active site(s) will be selected. For example, the CnrCBA transporter complex is composed of three proteins that confer resistance to Ni and Co via efflux [42–44]. The transporter is composed of an outer membrane factor (OMF) (CnrC) that passes through the outer membrane, a trimer of resistance-nodulation-cell division (RND) protein family members (CnrA) that forms a tube, and a membrane fusion protein (MFP) (CnrB) that maintains the alignment of the OMF and the RND trimer [45–47]. Since the MFP (CnrB) is not actively involved in Ni or Co efflux, this gene would not be included, but the OMF (CnrC) and the RND proteins (CnrA) responsible for Ni and Co transportation and specificity are selected.

3.2 Sequence retrieval and verification

Sequence retrieval and verification are performed generally by a pipeline. For example, such a pipeline was implemented in Perl scripts integrated with the Common Gateway Interface (CGI) Web standard protocol and some open-source modules, such as Bioperl and DBI [17]. A database is integrated with the pipeline to store and manage all sequence and probe data, such as databases for all initially retrieved candidate sequences, verified sequences, seed sequences, keywords, output probes, and the best probes for GeoChip. For each functional gene, the first step is to submit a query to GenBank Protein Database through the National Center for Biotechnology Information (NCBI) Entrez Programming Utilities (eUtils) and fetch all candidate amino acid sequences. Each query submitted usually includes various keywords, such as the name of the target gene/enzyme, its abbreviation and enzyme commission number (EC), and affiliated domains of bacteria, archaea, and fungi. The sequences retrieved by the query should be as broad as possible to avoid missing any real sequences but also exclude non-target sequences. Thus, all these candidate sequences have to be verified by a sequence-region-consensus finding program, HMMER [48]. Generally, more than five full-length sequences with an experimentally verified function are selected as seed sequences for each functional gene; then, ClustalW [49] and hmmbuild [48] are used to construct a profile hidden Markov model (HMM) based on the alignment of all seed sequences. This model is then used to search against all candidate sequences through both HMMER local and global algorithms, and e-values were obtained for the hits. Only those hits with e-values of the global alignment less than a cutoff (e.g., 0.1) are considered as highly confident targets in this design process. Other hits with local e-values between 0.1 and 1.0 may be listed and manually determined to be targets. Finally, all confirmed protein sequences are searched against GenBank again to obtain their corresponding nucleic acid sequences for probe design.

3.3 Oligonucleotide probe design

Although many software tools have been used for oligonucleotide probe design, a new version of CommOligo [50] with group-specific probe design features can be used to design both gene- and group-specific oligonucleotide probes with different degrees of specificity based on the following criteria: (i) gene-specific probes: $\leqslant 90\%$ sequence identity, $\leqslant 20$-base continuous stretch, and -35 or more kcal/mol free energy [51], and (ii) group-specific probes: a group-specific probe has to meet the above requirements for non-target groups and must have $\geqslant 96\%$ sequence identity, $\geqslant 35$-base continuous stretch, and -60 or less kcal/mol free energy within the group [52]. Computational and experimental evaluation indicates that those designed probes are highly specific to their targets [17,21].

3.4 Probe validation and automatic update

All designed probes are subsequently verified against the GenBank (NR) nucleic acid database for specificity and the criteria for nonspecific hits were $> 90\%$ sequence identities, > 20 bp continuous stretches, or less than -35 kcal/mol free energy by MegaBLAST [53]. Normally, multiple (e.g., 20) probes for each sequence or each group of sequences are designed so that only the best probe set for each sequence or each group of closely related sequences is chosen for array construction. To reflect the current status of functional genes, a pipeline should have a feature to update gene sequences and their probe design information automatically.

3.5 GeoChip construction

GeoChip can be constructed in-house or commercially by NimbleGen (Roche NimbleGen, Madison, WI). For in-house array construction, all designed probes are synthesized commercially (e.g., Invitrogen, Carlsbad, CA), and the concentration of all oligonucleotides is adjusted to a final concentration (e.g., $100\ \mathrm{pmol \cdot \mu L^{-1}}$). In addition, positive control probes (e.g., 16S rRNA gene) and negative control probes (e.g., human genes) are synthesized and spotted on each sub-grid at least two times. All oligonucleotide probes and controls were arrayed onto Corning UltraGAPS (Corning, New York, NY) slides using a array spotter (e.g., Microgrid II Arrayer, Genomic Solutions, Ann Arbor, MI) as described previously [17,21]. NimbleGen uses micromirror array synthesis, which synthesizes the probes directly onto the glass array surface in different formats with up to 2.1 million probes per slide.

4 GeoChip analysis

Generally, GeoChip analysis includes target preparation, GeoChip hybridization, image and data preprocessing, and

data analysis (Figs. 1(b) and (c)).

4.1 Target preparation

The most important step for successful hybridizations is nucleic acid extraction and purification. For experiments using GeoChip, environmental DNA is generally extracted using a well-established method that used freeze-grinding combined with detergent lysis [54,55]. This method provides very large fragments of DNA since it is gentler than the popular bead beating protocols. These large fragments are needed for the subsequent DNA amplification step. DNA can be purified using various methods, including various column purification methods. Our laboratory has successfully used an agarose gel purification followed by a phenol-chloroform-butanol extraction [56], especially for soil or sediment samples with large amounts of humic acids or other impurities. The purified DNA should have $A_{260}:A_{280} > 1.8$ and $A_{260}:A_{230} > 1.7$. The $A_{260}:A_{230}$ ratio has the most influence on hybridization success [57]. Impurities in the DNA can inhibit subsequent amplification, labeling, and hybridization steps. As such, extra time and attention should be taken for DNA extraction and purification steps to assure the success of subsequent steps.

Since so much DNA (1–5 μg) or RNA (10–20 μg of total RNA) is required for GeoChip hybridization, an amplification step is often necessary. For DNA amplification, the whole-community genome amplification (WCGA) method has been developed [58]. This method uses the Templiphi 500 amplification kit (phi 29 DNA polymerase, GE Healthcare, Piscataway, NJ) with a modified buffer containing single-stranded binding protein and spermidine. The buffer modification increases amplification yield and provides representative and even amplification with minimal bias. Using this method as little as 10 fg DNA can be amplified, the equivalent of one to two bacterial cells although generally 1 to 100 ng DNA is used for amplification. It has also been shown to be highly representative. When pure cultures were tested, less than 0.5% of the genomic DNA showed a greater than 2-fold change when compared to unamplified controls and no genes showed a greater than 3-fold change. Amplified DNA was quantitative, showing a linear relationship (r^2 = 0.91) with starting DNA concentrations of 0.1 to 1,000 ng. When environmental DNA was tested, minimal bias was observed when 1.0 ng or greater DNA was used for amplification [58].

Use of RNA for GeoChip hybridization is still challenging mainly because environmental mRNA is not stable with a rapid turnover, usually contains contaminants, and may be in a low abundance for some mRNA species. Several methods are available for extraction of microbial community RNA from environmental samples. Hurt et al. [55] developed a method for the simultaneous extraction of DNA and RNA. RNA can also be extracted with bead-beating [59]. Another method can separate mRNA from total RNA using gel electrophoresis [60]. To overcome the low quantity of mRNA obtained, a whole-community RNA amplification (WCRA) was developed [61]. A fusion primer consists of a short (6–9) series of random nucleotides and a T7 promoter. Representative amplification was obtained using 50 to 100 ng total RNA. The original relationship of mRNA quantity was obtained when 50 ng or more of RNA is used [61].

The amplified nucleic acids are then labeled with fluorescent dye (e.g., Cy3 and Cy5) using random priming with the Klenow fragment of DNA polymerase [58]. RNA is labeled using Superscript™ II/III RNase H-reverse transcriptase [62]. The labeled nucleic acids are then purified and dried for hybridization.

4.2 Hybridization

Labeled nucleic acid target is suspended in hybridization buffer containing 50% formamide. The nucleic acid targets are heated to 99°C to melt the nucleic acids in preparation for hybridization. GeoChips can be hybridized at 42°C to 50°C and 50% formamide [17,21,56,63–65]. The hybridization stringency can be adjusted by changing the temperature and formamide concentration. For every 1% increase in formamide, the effective temperature increases by 0.6°C.

Several different methods for hybridization are available for glass-based arrays. Manual hybridizations are performed using a hybridization oven or in manual hybridization chambers using a water bath. Semi-automated systems are also available, which control temperature and provide mixing via specially designed mixers, acoustic agitation, or with rotation (e.g., Mai Tai® from SciGene, SlideBooster from Advalytix, and Maui from BioMicro Systems). Post-hybridization washes can be done manually or with an automated wash station (e.g., Maui Wash Station, BioMicro Systems). Completely automated systems are also available, which perform all steps from pre-hybridization slide preparation to post-hybridization washes (e.g., Tecan HS4800Pro, TECAN US).

4.3 Imaging and data processing

Hybridized arrays are imaged with a microarray scanner having a resolution of at least 10 μm for homemade arrays and 2 μm for NimbleGen arrays. Microarray analysis software is then used to quantify the signal intensity (pixel density) of each spot. Spot quality is also evaluated at this point using predetermined criteria. Poor- or low-quality spots, which do not meet the criteria, are flagged for later removal. “Good” spots are then evaluated to determine positive spots. Generally, this is determined based on signal-to-noise ratio (SNR; SNR = (signal mean – background mean) / background standard deviation). Other

calculations can also be used. A signal-to-both standard deviations ratio (SSDR; SSDR = (signal mean – background mean) / (signal standard deviation – background standard deviation)) has been reported, which resulted in fewer false-positives and false-negatives than the SNR calculation [66].

Raw GeoChip data are further evaluated via the GeoChip data analysis pipeline [17]. The quality of individual spots, evenness of control spot hybridization signals across the slide surface, and background levels are assessed to determine the overall array quality. Spots flagged as poor or low quality are removed along with outliers (positive spots with (signal – mean signal intensity of all replicate spots) greater than three times the replicate spots signal standard deviation [66,67]). The signal intensities are then normalized and stored in an experiment database for further statistical analysis.

4.4 GeoChip data analysis

Data analysis is the most challenging aspect of FGA because of the large amounts of data generated. Several methods have been frequently used in GeoChip studies. These include various diversity indices (e.g., richness, evenness, and diversity) based on the number of functional genes detected. The relative abundance of specific genes or gene groups can be determined based on the total signal intensity of the relevant genes or the number of genes detected. The percent of genes shared by different samples can also be calculated to compare communities. The response ratio can be used to determine changes in gene abundance based on different treatments or conditions. For statistical analysis of FGA data, several methods are commonly used. These include ordination techniques, such as principal component analysis (PCA) or detrended correspondence analysis (DCA), cluster analysis (CA), and neural network analysis (NNA) [67]. PCA and DCA are multivariate statistical methods that reduce the number of variables needed to explain the data and highlight the variability between samples. CA group samples based on overall similarity of gene patterns. NNA is used to visualize relationships between genes or gene groups. Response ratios compare the signal intensities between two samples, generally control versus treatment [68]. This method was used to compare community responses to oil contamination [56] and elevated CO_2 [69]. If environmental data are available, several statistical methods are available to correlate environmental variables with the functional community structure. These include canonical correspondence analysis (CCA) [70], variance partitioning analysis (VPA) [71,72], or other correlation analyses (e.g., Mantel test). CCA has been used in many cases in GeoChip-based studies to better understand how environmental factors are affecting the community structure [16,63,64,73,74]. Based on the results of the CCA, the relative influence of environmental variables on the microbial community can be determined using VPA. Further correlations of GeoChip data with environmental parameters can be performed with the Mantel test [17,21,63,64,69,74].

5 Key issues for GeoChip analysis

Although GeoChip technology has been demonstrated to be specific, sensitive, and quantitative for analyzing microbial communities from different habitats [16–19,21,34,58,63,64,66,69,74–77], some key issues and challenges remain, including nucleic acid quality, probe coverage, specificity, sensitivity, quantitative capability, and the detection of active microbial communities.

5.1 Nucleic acid quality

A critical factor for the success of FGA hybridizations is the quality of nucleic acids used. There are many nucleic acid extraction methods available. However, methods that result in large pieces of DNA or non-degraded RNA are preferred as fragmentation of nucleic acids decreases the amplification efficiency. Fresh DNA or RNA is best since even short-term storage of nucleic acids can result in degradation or fragmentation if any contaminants or impurities remain in the nucleic acid solution. The presence of inhibitors and contaminants is also problematic since this can inhibit amplification and labeling steps. Gel purification has worked very well in our laboratory. However, fresh, large-sized DNA in relatively large amounts is necessary since some DNA is lost in this process, and some samples cannot be purified to the required level. Thus, newer methods of purification that result in high quality but little loss of DNA are needed.

An even greater challenge is the use of environmental mRNA. Because of the low abundance, rapid turnover, and instability of environmental mRNA, extraction is difficult. Only a limited number of studies used environmental mRNA for GeoChip analysis [23,55,61,78]. The use of environmental mRNA often requires tedious and time-consuming amplification steps. More efficient amplification methods are needed.

5.2 Probe coverage

Earlier versions of FGAs were composed of a limited number of probes and only covered one or a few functional genes. The GeoChip arrays greatly expanded the number of probes and gene groups covered on a single array. GeoChip 2.0 is the first comprehensive FGA targeting more than 10000 gene sequences in 150 functional gene families [21], while GeoChip 3.0 targets about 57000 gene sequences in 292 functional gene families [17]. The latest

version GeoChip 4.0 is more comprehensive than GeoChip 3.0 with 152000 gene sequences targeted in 410 functional gene families. However, no matter how comprehensive a GeoChip is now, the continuous addition of new sequences to the public databases will result in exponential increases in available functional sequences. Therefore, frequent updates are necessary to keep GeoChip to the current status, and an automatic update feature has been implemented in GeoChip design pipeline [17].

5.3 Specificity

Another critical issue with microarray detection is specificity. This is especially important when environmental samples are used since there are most likely many gene variants within each sample. Array specificity is controlled by probe design and hybridization conditions. A challenge in designing specific probes is the number of unknown sequences in the environment and many gene variants with high homologies. The use of carefully selected design criteria can produce highly specific oligonucleotide probes. Probe design criteria-based sequence identity, continuous sequence stretches, and free energy have been systematically evaluated and experimentally established in order to increase specificity [51,52]. Also, probe specificity as well as coverage can be improved with probe design software. A novel microarray probe design software tool *CommOligo* [50] and its improved versions was used to design probes for GeoChip 2.0, GeoChip 3.0, and GeoChip 4.0. Experimental evaluations of GeoChip 2.0 and GeoChip 3.0 indicated that a minimal number of false positives (0.002–0.025%) were observed [17,21].

Optimization of hybridization conditions can also be used to increase specificity. Specificity is generally controlled by the temperature and formamide concentration. With hybridizations at 50°C and 50% formamide, GeoChip could differentiate sequences with similarities less than 90% to 92% [19,51,79], which provides a resolution at the species-strain level [34,51,79]. With PCR amplicon probes, sequences with similarities as low as 70% to 75% were distinguishable at 45°C, while sequences with similarities $> 87\%$ were distinguishable at 65°C without formamide added [18].

5.4 Sensitivity

Sensitivity is a major concern with any detection technology, especially with the samples from complex environments where many gene variants are expected to be in low abundance. The current level of sensitivity for oligonucleotide arrays using environmental samples is approximately 50 to 100 ng or 10^7 cells [18,19,25,34,80] or approximately 5% of the microbial community [25], providing coverage of only the most dominant community members. This is an important issue especially for most environmental samples from soils, groundwater, and marine water columns since the abundance in these samples is generally lower than the detection limit. As such, the use of microarrays for many environmental studies is very limited.

Several strategies have been utilized to increase sensitivity. The most important is the use of target DNA or RNA amplification to increase sensitivity [32,58,61]. WCGA can amplify all of the community DNAs, including sequences in a low abundance [58]. WCGA has been shown to provide a representative, sensitive, and quantitative amplification of microbial community DNA using starting amounts of 1.0 to 250 ng. Use of amplified DNA for hybridization increased the sensitivity of GeoChip hybridization to 10 fg (equivalent to 1–2 cells) [58].

Several array modifications can increase sensitivity, although generally there is a trade-off with a decreased specificity, another important concern with microarrays. Increasing the length of oligonucleotide probes increases sensitivity [62,81], although this does decrease specificity [82]. Another option is to increase the amount of probe per spot [82–84]. This strategy is based on the fact that membrane-based arrays are more sensitive than glass-based arrays probably because of the higher probe concentration ($> 1\ \mu g \cdot spot^{-1}$ for membranes; $< 20\ pg \cdot spot^{-1}$ for glass slides) [83]. However, again, this decreases the array density and may actually decrease sensitivity by reducing overall signal intensity [81].

Decreasing the volume of hybridization solution can also increase sensitivity. Wu et al. (2004) demonstrated that decreasing the hybridization buffer to 3 μL from the typical 15 μL increased the sensitivity from 5 ng in the presence of multiple DNA to 0.2 ng for pure culture DNA. Mixing during hybridization can also increase sensitivity with SNR values three times that of static hybridizations [85]. Hybridization results were consistent over a range of RNA concentrations (5–20 μg), indicating that mixing during hybridization allows for detection of lower amount of nucleic acid [85].

Dyes or other labeling techniques can also increase the detection limit of GeoChip. Cyanine dye-doped nanoparticles [86] or tyramide signal amplification labeling [81] has been shown to increase sensitivity up to 10-fold. Development of signal detection systems with higher sensitivities would also increase the detection sensitivity of GeoChips [83,84]. In addition, improving air quality where hybridization and array scanning occurs may improve sensitivity since ozone has been implicated in loss of cy-dye signal [87]. Ozone control systems are available for equipment, work spaces, or entire rooms. Because positive signals must be distinguished from background noise, increased sensitivity can be accomplished by reducing background signal on arrays. The use of unmodified glass slides for array construction [88,89] may decrease

background levels since unmodified glass has 4- to 12-fold less background intensity [88]. Modifications can also be made to the calculations used to distinguish true signals from background [66].

5.5 Quantitative capability

An important goal in microarray analysis is to provide quantitative information. GeoChip has been shown to have a linear relationship between target DNA or RNA concentrations and hybridization signal intensities. PCR-based amplicon probes demonstrated a linear relationship ($r = 0.94$) between signal intensity and DNA quantity over a concentration range of 0.5 to 100 ng [18]. A linear relationship (r = 0.98–0.99) over a concentration range of 8 to 1,000 ng was observed for 50-mer oligonucleotide probes as well [34]. This relationship also existed for hybridizations using RNA as the target as well (r = 0.895–0.942; 50–100 ng) [61]. Significant correlations between the gene copy number measured by quantitative real-time PCR and the signal intensity were observed for GeoChip 2.0 (r = 0.530, $n = 85$, and $p < 0.0001$) and GeoChip 3.0 (r = 0.724, $n = 91$, and $p < 0.0001$) with soil samples from the BioCON experimental site, indicating that GeoChip data are quantitative [69]. This linear relationship was observed for pure culture, mixed culture, and environmental samples with or without amplification by WCGA [18,19,34,58].

While signal intensity can be used to provide quantitative information, it can be affected by sequence divergence (i.e., the more divergent the sequence, the lower the signal intensity). As such, strategies and techniques that can distinguish true signal from noise signal can improve the quantitative ability of microarrays as well as the specificity. Mismatch probes have been proposed as a method to discern true signals from noises [79]. While this strategy has been used successfully for 16S rRNA probes [41,90], this strategy may be more difficult to implement for environmental samples since longer oligonucleotide probes are used for GeoChips. In addition, so many unknown sequences exist in the environment. Another strategy to overcome potential cross-hybridization is to use relative comparisons across samples rather than absolute comparisons. By utilizing ratios of gene abundance in control versus experimental systems, any cross-hybridization can be cancelled out assuming the communities are similar [21].

5.6 Activity

While DNA detection provides information on the presence of functional genes in the environment, it does not provide unconditional evidence for microbial activity. Population changes can be used to infer microbial activity, but this may not be accurate [27]. To monitor microbial activity mRNA or other markers of activity must be used [27,61,78,91]. mRNA is used frequently in pure culture studies; however, since environmental mRNA is usually in low abundance, has a rapid turnover, and is only a small proportion of the total RNA, a large quantity of RNA (10–20 μg) is required for hybridization. There are some extraction protocols for environmental RNA [55,59,60], but these can be time-consuming and laborious. Improved extraction techniques are needed. In addition, because of the low quantity of environmental mRNA, amplification techniques are needed as well. WCRA was developed to overcome some of the limitations of environmental mRNA use [61]. This method was able to provide representative detection of environmental RNA with 50 to 100 ng starting material. Positive signals were detected with as little as 10 ng RNA, which provides sufficient sensitivity for many natural settings. However, the current amplification protocol is time-consuming and laborious. Therefore, improved RNA extraction methods are necessary for environmental samples. Another option for determining microbial activity is the use of stable isotope probing (SIP) of active community members [91].

6 GeoChip applications

Different versions of GeoChips have been used to analyze microbial communities from different habitats, such as water and marine systems, soils, extreme environments, human microbiomes, and bioreactors for addressing fundamental scientific questions related to global change, bioenergy, bioremediation, agricultural management, land use, and ecological theories as well as human health and disease. In this following, we will provide brief highlights for various representative results.

6.1 Groundwater and aquatic ecosystems

GeoChip has been used to analyze microbial communities (water and sediment samples) from aquatic ecosystems, especially U.S. DOE contamination sites like the Field Research Center, Oak Ridge, TN (OR-FRC), which originates from the former S-3 waste collection ponds that received nitric acid, metals, and radionuclide (U and Tc) contaminated waste until 1983 [92]. Those ponds have been drained and treated, but the subsurface near the source zone remains heavily contaminated with acidic pH (3.4–3.6) and high concentrations of numerous chemicals, such as uranium (50 $mg \cdot L^{-1}$) and nitrate (8–12 $g \cdot L^{-1}$).

A series of contaminated wells and an uncontaminated background well were examined using GeoChip 1.0 [58]. Higher functional gene numbers and diversity were observed in the uncontaminated well and in low contamination wells compared to the high-contamination wells, and *dsr* genes that may be involved in U(VI) reduction were detected along with genes involved in

denitrification, organic contaminant degradation, and metal resistance. These wells and an additional well were further analyzed using GeoChip 1.0 [64]. The results showed that the microbial community structure was closely correlated with the geochemistry of each well and that the environmental parameters with the most influence on community structure were pH, sulfate, and the combination of U and Tc, with pH and the combination of U and Tc each showing the highest level of variance, approximately 21% for pH and 40% for the combination of U and Tc [64].

A pilot-scale field test system was established near the S-3 ponds to examine the feasibility of *in situ* bioremediation and immobilization of U(VI) [93], resulting in a successful reduction of U(VI) through microbial bioreduction and immobilization [94]. GeoChip 2.0 was used to successfully track the dynamics of metal-reducing bacteria (MRB) and associated communities for an in situ bioremediation and showed that the abundance of cytochrome c and *dsr* genes was significantly correlated with the U(VI) concentration [21]. To further monitor the long-term effects of uranium bioremediation in this site, GeoChip 2.0 has been used to examine several experimental periods of this field test system. During the active U(VI) reduction phase, functional genes associated with denitrification, sulfate reduction, and Fe(III) reduction increased as U(VI) concentrations decreased, indicating that these populations were important in U(VI) reduction, and these populations leveled off after the initial U(VI) reduction and remained relatively stable during the maintenance phase.[1)] Ethanol, nitrate, pH, sulfate, and U(VI) were found to be the most important environmental parameters in shaping the functional structure of the microbial community. In a later period of this system, the stability of the bioreduced U(IV) was examined by halting ethanol injections and introducing dissolved O_2 (DO) into the system and then allowed the system to recover by reintroducing ethanol and decreasing DO [94]. Consistently, a shift in the microbial community structure was observed following reintroduction of ethanol to the system, and COD, temperature, and sulfate and U(VI) were the most important environmental parameters in controlling the community structure and were able to explain almost 65% of the variation observed [63].

In another study, GeoChip was used to examine the functional composition and structure of microbial communities of sediment samples collected from both inner and outer loop wells in this site [95]. Distinct microbial communities were established wells with a higher microbial functional gene number, diversity, and abundance wells in the inner loop than in the outer loop wells. Also, MRB, such as *Desulfovibrio*, *Geobacter*, *Anaeromyxobacter* and *Shewanella*, and other bacteria (e.g., *Rhodopseudomonas*, *Pseudomonas*), were highly abundant in the inner loop wells. In addition, the richness and abundance of microbial functional genes were highly correlated with the mean travel time of groundwater from the inner loop injection well, pH, and sulfate concentration in groundwater. These results suggest that the indigenous microbial communities can be successfully stimulated for U bioremediation in the groundwater ecosystem, and their structure and performance can be manipulated or optimized by adjusting geochemical and hydrological conditions [95].

The above GeoChip studies with groundwater and sediment samples suggest that key microbial functional processes are occurring at this contaminated site and that those indigenous microbial communities, including MRB, denitrifiers, and sulfate-reducing bacteria (SRB) as well as organic contaminant degraders, may play important roles in bioremediation of such a contaminated site.

Another study was to examine microbial communities from marine sediments in the Gulf of Mexico using GeoChip 1.0 [74]. Samples were collected from two stations (200 and 800 m) at different depths (0.75–25.5 cm for the 200 m station and 0.25–16.0 cm for the 800 m station). The results showed that the microbial communities became more distinct with depth and that genes for all major functional categories were detected by this array, including carbon degradation, nitrification, denitrification, nitrogen fixation, sulfur reduction, phosphorus utilization, contaminant degradation, and metal resistance. Statistical analyses suggest that sediment depth, porosity, and ammonium, Mn(II), phosphate, and silicic acid concentrations might play major roles in shaping the microbial community structure in the marine sediments [74].

Microbial communities from two pesticide-impacted European rivers, the Ebro in Spain and the Elbe in Germany, were examined using GeoChip 2.0 [96]. Samples were collected from river sediments and rice fields within the river delta. GeoChip analysis showed that spatial and temporal fluctuations were reflected not only in the abundance of *Dehalococcoides* spp. but also in the composition of the populations and reductive dehalogenase (Rdh) gene diversity. The results indicated that the strongest drivers determining the functional structure within those systems were the C/N ratio, depth, total N, and location. This study provides new insights into the natural occurrence and dynamics of active *Dehalococcoides* spp. in HCB-contaminated river basins [96].

GeoChip 2.0 was also used to examine the functional structure of microbial communities associated with healthy

1) Van Nostrand J D, Wu L, Wu W, Gentry T J, Huang Z, Deng Y, Carley J, Carroll S, He Z, Gu B, Luo J, Criddle C S, Watson D B, Jardine P M, Marsh T L, Tiedje J M, Hazen T C, Zhou J. Dynamics of microbial community composition and function correlate with changes in site geochemistry during the in situ bioremediation of a uranium-contaminated aquifer. Applied and Environmental Microbiology (in review)

and yellow band diseased (YBD) coral, *Montastraea faveolata* [97]. The results suggest a more generalized shift in functional potential involving various aspects of nutrient cycling, metal transformations, and contaminant degradation and provide insights into biogeochemical cycling capacity in healthy and diseased coral-associated microbial communities.

6.2 Soil

GeoChip has been used to investigate soil microbial communities to address fundamental ecological questions, such as global climate change, bioremediation/phytoremediation, land use, and ecology theories.

GeoChip 1.0 was used to study the effects of different land use strategies (a primitive fir forest, two spruce plantations, and cropland) on the microbial community [98]. The results suggest that the number of functional genes and the gene diversity index are correlated with increasing amounts of soil organic carbon, except in the primitive *Abies faxoniana* forest site, and that primitive forest soil was clustered more closely to soil from the spruce plantation established about 40 years ago [98].

GeoChip 2.0 in combination with SIP was used to monitor active microbial populations from a polychlorinated biphenyl (PCB)-contaminated site [91]. Microcosms were set up from soil samples, fed ^{13}C-labeled or unlabeled biphenyl, and sampled at 1, 4, and 14 days. Genes involved in the degradation of biphenyl (BP) and benzoate as well as genes for a variety of aromatic ring hydroxylating dioxygenase (ARHD) subunits and the ketoadipate pathway were detected by GeoChip 2.0. This study improves our understanding of BP degradation and carbon flow in soil [91].

Microbial communities associated with Antarctic sediments were characterized using GeoChip 2.0 to better understand N and C cycling in this environment [73]. GeoChip analysis showed that the detected N- and C-cycle genes were significantly different across different sampling locations and vegetation types. Statistical analyses showed that pH and the number of freeze–thaw cycles were significant factors for all gene groups detected and that the C:N ratio had a significant effect on the C- and N-cycling functional genes. In addition, soil temperature was found to be positively correlated with cellulose degradation and denitrification genes. This study provides insight into our understanding of the forces driving important processes of terrestrial Antarctic nutrient cycling [73].

GeoChip 2.0 was used to examine whether microbial species richness increases with area in forest soil, as has been observed in macroorganisms [99], and to better understand spatial scaling in microbial communities [16]. Samples were collected from forest soils in a nested design with samples taken over a scale of less than 1 m to 1 km. The gene-area relationship was determined using the functional genes detected with GeoChip 2.0, and the results indicated that the forest soil microbial community had a relatively flat gene-area relationship with less turnover than observed for plants and animals [16].

Recently, GeoChip 3.0 has been used to analyze soil microbial communities associated with plant diversity and global changes (e.g., elevated CO_2 and warming). Two studies were conducted at a multifactor grassland experiment site, BioCON (Biodiversity, CO_2, and Nitrogen deposition) in the Cedar Creek Ecosystem Science Reserve, MN [100]. One was to investigate how plant diversity affects the functional diversity, composition, and structure of soil microbial communities, and DCA of all detected 4,012 genes showed that the microbial community structure was significantly different between monospecies (1 species) and multispecies (4, 9, and 16 species), suggesting that mono-cultures of plants may lead to a decrease in the soil microbial community diversity [17]. The other was to examine how elevated CO_2 affects the functional composition and structure of soil microbial communities [69]. The results showed that the microbial community structure was markedly different between ambient CO_2 and elevated CO_2 as indicated by DCA of GeoChip 3.0 data as well as 16S rRNA gene-based pyrosequencing data. Also, those involved in labile C degradation and C and N fixation were significantly increased under elevated CO_2, although the abundance of genes involved in decomposing recalcitrant C remained unchanged. In addition, changes in microbial communities were significantly correlated with soil C and N contents and plant productivity [69]. Another study was conducted at a long-term experimental warming site in a tall grass prairie ecosystem in the U.S. Great Plains of Central Oklahoma to explore microbial-mediated carbon cycle feedback in response to 8-year warming by approximately 2°C, and the results indicate that microorganisms critically regulate ecosystem carbon cycling in response to climate warming[1)].

GeoChip 3.0 was also used to understand how microbial communities respond to arsenic contamination in the rhizosphere of *Pteris vittata*, which can tolerate and accumulate arsenic from arsenic-contaminated soils [76]. Arsenic resistance, sulfur reduction, phosphorus utilization, and denitrification genes were remarkably distinct between *P. vittata* rhizosphere and non-rhizosphere soils, which provides evidence for a strong linkage among the level of arsenic contamination, the rhizosphere, and the functional gene distribution. CCA revealed that arsenic is the main driver in reducing the soil functional gene

1) Zhou J, Xie J, Wu L, Xue K, Fei S, Deng S, He Z, Deng Y, Van Nostrand J D, Luo Y. Microbial mediation of carbon cycle feedbacks to climate warming. Science (submitted)

diversity, suggesting that rhizobacteria may play an important role during soil arsenic uptake and hyperaccumulation processes by *P. vittata* [76].

6.3 Extreme environments

GeoChip has been used to investigate the functional diversity, composition, structure, and metabolic potential of microbial communities from extreme environments, including deep-sea hydrothermal vents, deep-sea basalts, hypersaline lake, and AMD.

Deep-sea hydrothermal vent microbial communities were investigated using GeoChip 2.0 [75]. Samples were taken from a mature chimney and the inner and outer portions of a 5-day-old chimney. The inner chimney communities were less diverse than those from the outer portion of the 5-day-old chimney or the mature chimney, and genes involved in methanogenesis, aerobic and anaerobic methane oxidation, nitrification, denitrification, sulfate reduction, degradation of complex carbon substrates, and metal resistance were also detected, suggesting that the hydrothermal microbial communities are metabolically and physiologically highly diverse and that the communities appear to be undergoing rapid dynamic succession and adaptation in response to the steep temperature and chemical gradients across the chimney [75].

Microbial communities from deep-sea basalt were examined with GeoChip 2.0 to understand key biogeochemically functional processes associated with this environment [65]. GeoChip analysis showed that genes coding for previously unreported processes, such as carbon fixation, methane oxidation, methanogenesis, and nitrogen fixation, were present, suggesting that these processes may be important in this environment [65].

Horizontal gene transfer (HGT) among microorganisms of microbial communities from a hypersaline lake (the Great Salt Lake, Utah, USA) where NaCl concentration ranges from near seawater to saturation, with exceptionally high concentrations of sulfate and heavy metals, was assessed using GeoChip 2.0 and Phylochip [101]. High levels of chromium and sulfate across the lake resulted in a high diversity of *dsr* and chromium resistance gene variants in all samples, suggesting that the high selection pressure from these elements resulted in an increased gene transfer [101]. This study suggests that active HGT could be assessed at the population level in microbial communities and that these biogeographic patterns may serve as a model to study bacteria adaptation and speciation.

AMD is an extreme environment, which is generally thought to have a simple microbial community structure [102]. GeoChip 2.0 was used to analyze the functional diversity, structure, and metabolic potential of AMD microbial communities with five samples obtained from three copper mines in China. Almost all key functional genes were detected in the AMD microbial communities, including carbon fixation, carbon degradation, methane generation, nitrogen fixation, nitrification, denitrification, ammonification, nitrogen reduction, sulfur metabolism, metal resistance, and organic contaminant degradation, suggesting that the functional gene diversity may be higher than previously thought. This study presents an overview of functional diversity and structure of AMD microbial communities and also provides insights into our understanding of metabolic potential in AMD ecosystems [102]. All those studies of microbial communities from extreme environments suggest that the functional diversity is higher than it was previously thought.

6.4 Bioreactor systems

GeoChip has been also used for analysis of microbial communities from controlled environments like bioreactor systems. GeoChip 2.0 was used to examine microbial communities associated with a bioremediation system in Vega Baja, Puerto Rico [103]. This system was composed of a fluidized bed reactor supplied with oxygenated diesel-contaminated groundwater and sulfate or nitrate compounds as electron acceptors. Genes related to the degradation of diesel fuel and other organic contaminants (phthalate, methyl tert-butyl ether, phthalate, protocatechuate, biphenyl, aniline, cyclohexanol, toluene, acetylene, benzoate, and naphthalene) were detected. An increase in the signal intensity of genes involved in anaerobic benzoate degradation increased over time, suggesting that the system shifted to an anaerobic process, and subsequent increases in *nir*S, *nir*K, *nos*Z, and *dsr* genes were also consistent with a shift to anaerobiosis [103].

A microbial electrolysis cell (MEC) is a bioelectrochemical system that can produce hydrogen from acetate at high hydrogen recoveries. GeoChip was used to examine the microbial community functional structure in MECs initially operated under different conditions [77]. Hydrogen yields were generally higher for reactors that were always operated as MECs than those initially operated as microbial fuel cells (MFCs) and independent of startup conditions. Also, MECs with the largest hydrogen yields had the highest microbial diversity. In addition, well-known exoelectrogenic bacteria (e.g., *Geobacter*, *Shewanella*, *Desulfovibrio*, and *Anaeromyxobacter*) were found in these systems. Multivariate analyses indicate that reactor operations altered the microbial community composition, and community shifts based on a Mantel test were significantly related to CEs and COD removals in these reactors, suggesting that there were significant changes in microbial community composition as a result of conditions that affected MEC performance [77].

6.5 Oil-contaminated waters and soils

GeoChip has been used to analyze microbial communities from oil-contaminated waters and soils to explore the

potential of bioremediation of oil components by indigenous microbial communities.

Ozonation with a subsequent biodegradation treatment was performed to remove recalcitrant organic compounds from long-term weathered crude oil contaminated soils [104]. GeoChip 2.0 was used to examine microbial community dynamics. The total microbial functional gene numbers and overall genetic diversity decreased after ozonation. Also, most of the key functional genes pertaining to carbon, nitrogen, and sulfur cycling and organic contaminant degradation decreased. However, in the subsequent biodegradation treatments, with and without bioaugmentation, the abundance of key genes in most functional groups recovered. This study provided insight into changes in crude oil composition and microbial responses to ozonation and bioremediation treatments [104].

GeoChip 2.0 was used to examine the response of soil microbial communities to oil contamination [56]. The overall abundance of soil bacteria, archaea, and fungi in the oil-contaminated samples decreased to 10%, 40%, and 80% of those in the pristine soil. Genes involved in carbon and nitrogen fixation and carbon degradation were significantly decreased with oil contamination, especially at high contaminations of oil contamination. However, a few genes encoding key enzymes for catechol, protocatechuate, and biphenyl degradation showed a significant increase at a medium level of oil contamination. The results indicate the potential of in situ bioremediation of oil-contaminated soils by indigenous microbial communities and provide an insight into microbial functional diversity in oil-contaminated soils [56].

GeoChip 2.0 was also used to analyze surface soil samples from five oil-contaminated fields located in different geographic regions of China [105]. The results indicated that samples from the same field clustered together with less effect of oil contamination levels based on all functional gene patterns. However, genes involved in organic contaminant degradation were significantly influenced by oil contaminants. Meanwhile, some genes involved in carbon and nitrogen cycling showed negative relationships with oil contamination. This study provides important information for our comprehensive understanding of the in situ microbial functions in oil-contaminated fields [105].

The latest GeoChip 4.0 and Phylochip have been used to oil-contaminated water microbial communities from the oil spill in the Gulf of Mexico [106]. The results showed that the dispersed hydrocarbon plume stimulated deep-sea indigenous microorganisms closely related to known petroleum-degraders. Also, the microbial community functional composition and structure was dramatically altered ~40 days after the spill. In addition, a variety of metabolic genes/populations involved in both aerobic and anaerobic hydrocarbon degradation were highly enriched in the plume samples, indicating a great potential for *intrinsic* bioremediation of the oil plume in the deep-sea[1]. These results suggest that the metabolically highly diverse marine microbial communities appeared to be undergoing rapid dynamic adaptation and succession in response to the vast oil spill and associated geochemical changes, and they could play significant role in controlling the ultimate fates and consequences of deep-sea oil plumes.

6.6 Human microbiome

GeoChip 4.0 also contains 36062 probes targeting 139 key genes/enzymes involved in 19 pathways associated with the human microbiome and covers 47979 gene sequences. This array was used to examine the functional structure of 10 fecal samples collected from two remote areas with five each from a jungle forest and an Andean grassland of Peru, South America, which are geographically distinct. Significantly more genes were detected in the jungle samples than in the Andean samples, and DCA of all detected genes showed a clear geographic separation. The results indicate that such an array could be used to study the functional diversity, composition, and structure of human microbiomes[2].

7 Challenges and future directions

Great advances in technology development, applications, and data analysis have been made in the last decade since the first FGA was developed. This technology has revolutionized the study of molecular biology and microbial ecology. Numerous studies demonstrate that GeoChip is a powerful tool to provide sensitive, specific, and potentially quantitative analysis of microbial community functional diversity, composition, structure, function, and dynamics for different ecosystems, such as water, soil, sediment, extreme environment, bioreactor, and human microbiome. GeoChips have shown a great promise for addressing fundamental questions relevant to global change, bioremediation, bioenergy, agricultural operation, land use, ecosystem restoration, and human health and linking microbial structure to geochemical processes and ecosystem functioning.

However, in order for FGAs to fully realize their potential, challenges in technology, experiment, and data

1) Lu Z, Deng Y, Van Nostrand J D, He Z, Voordeckers J, Zhou A, Lee Y, Hazen T C, Zhou J. Rapid Succession and adaptation of microbial community functional structure in response to deepwater oil plume. Proceedings of the National Academy of Sciences of the United States of America (in review)

2) Lewis C M, Yu H, Tito R Y, Deng Y, Van Nostrand J D, Wu L, He Z, Zhou J. Human microbiomes from remote native South American communities. Nature Genetics (submitted)

analysis still remain with FGA development and applications. This is especially true for the study of environmental communities because of the vast diversity and many low abundance species. Therefore, future directions may focus on the following key issues: (i) improving experimental design to overcome and minimize inherent variations among array hybridizations, (ii) developing new strategies to improve FGA-based quantitative accuracy, (iii) developing approaches to increase hybridization sensitivity to detect extremely low biomass, (iv) developing novel bioinformatic tools for data analysis and interpretation, and (v) integrating microarray technology into environmental studies to address overarching ecological and environmental questions and hypotheses. It is not only FGA technology should be improved but also data analysis, interpretation, and modeling should be enhanced. Novel strategies for experimental controls are needed in order to fully compare microarray data across samples, experiments, and laboratories. An integrated pipeline with improved bioinformatics tools is needed to facilitate sequence retrieval, probe design and verification, data analysis, and future updates. Higher-order analysis techniques and mathematical modeling are needed to expand FGA data analysis into network analysis, environmental and cellular modeling, simulation, and prediction.

Recently, next-generation, high-throughput sequencing technologies are available for analysis of microbial communities. Barcode-based 454 pyrosequencing allows us to obtain hundreds of thousands of sequences from many environmental samples in a single run [107–109]. Pyrosequencing of aromatic dioxygenase genes was performed to understand the relationship between gene diversity and function for important environmental processes, and the greater functional gene diversity revealed by this gene-targeted approach provides deeper insights into genes potentially important in environmental processes to better understand their ecology, functional differences, and evolutionary origins [110]. However, a target gene needs to be PCR amplified with conserved primers designed. This is fine with some phylogenetic gene markers, such as 16S rRNA gene and *gyrB*, but conserved primers have been designed for only a few functional genes with a reasonable coverage of environmental gene variants, which greatly reduce the number of functional genes of interest to be examined. Also, even if conserved primers are designed, it is sometimes difficult to amplify gene variants from environmental samples. In addition, as mentioned above, it is known that the initial PCR step usually introduces biases for quantitative analysis [11–13]. Large-scale (576.7 gigabases of sequences) Illumina-based shot-gun metagenomic sequencing was conducted with fecal samples from 124 European individuals, and assembly and characterization of those sequences showed that 3.3 million non-redundant microbial genes were predicted, which is 150 times larger than the human gene complement [111]. Although those sequencing-based technologies can discover novel sequences, they are too expensive if shot-gun sequencing of a community is performed with a reasonable coverage or suffer without appropriate conserved primers if a target gene is pyrosequenced. Sequencing-based technologies may also have a disadvantage for under-sampling, making it difficult to compare different samples. However, since microarrays have a defined set of sequences, comparisons are easier to perform among different samples. Therefore, due to the unique features and advantages and disadvantages provided by both microarray- and sequencing-based technologies, it is preferred that they are complementarily used for microbial community analysis in order to address fundamental questions in microbial ecology and environmental biology.

Acknowledgements This work has been partially supported through contracts DE-SC0004601 and DE-AC02-05CH11231 (as part of ENIGMA, a Scientific Focus Area) by the U.S. Department of Energy, Office of Science, Office of Biologic and Environmental Research, Genomics: GTL Foundational Science and Environmental Remediation Science Program (ERSP) Programs, and Oklahoma Applied Research Support (OARS), Oklahoma Center for the Advancement of Science and Technology (OCAST), the Oklahoma Bioenergy Center (OBC), and the State of Oklahoma through the Project AR062-034.

References

1. Torsvik V, Øvreås L, Thingstad T F. Prokaryotic diversity—magnitude, dynamics, and controlling factors. Science, 2002, 296(5570): 1064–1066
2. Roesch L F, Fulthorpe R R, Riva A, Casella G, Hadwin A K, Kent A D, Daroub S H, Camargo F A, Farmerie W G, Triplett E W. Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME Journal, 2007, 1(4): 283–290
3. Schloss P D, Handelsman J. Toward a census of bacteria in soil. PLoS Computational Biology, 2006, 2(7): e92
4. Hong S H, Bunge J, Jeon S O, Epstein S S. Predicting microbial species richness. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(1): 117–122
5. Curtis T P, Sloan W T, Scannell J W. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(16): 10494–10499
6. Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science, 2005, 309(5739): 1387–1390
7. Whitman W B, Coleman D C, Wiebe W J. Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences of the United States of America, 1998, 95(12): 6578–6583
8. Dunbar J, Barns S M, Ticknor L O, Kuske C R. Empirical and theoretical bacterial diversity in four Arizona soils. Applied and Environmental Microbiology, 2002, 68(6): 3035–3045
9. Steward G F, Jenkins B D, Ward B B, Zehr J P. Development and

testing of a DNA macroarray to assess nitrogenase (*nifH*) gene diversity. Applied and Environmental Microbiology, 2004, 70(3): 1455–1465

10. Jenkins B D, Steward G F, Short S M, Ward B B, Zehr J P. Fingerprinting diazotroph communities in the Chesapeake Bay by using a DNA macroarray. Applied and Environmental Microbiology, 2004, 70(3): 1767–1776
11. Warnecke P M, Stirzaker C, Melki J R, Millar D S, Paul C L, Clark S J. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Research, 1997, 25(21): 4422–4426
12. Lueders T, Friedrich M W. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Applied and Environmental Microbiology, 2003, 69(1): 320–326
13. Suzuki M T, Giovannoni S J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology, 1996, 62(2): 625–630
14. Schena M, Shalon D, Davis R W, Brown P O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, 270(5235): 467–470
15. Guschin D Y, Mobarry B K, Proudnikov D, Stahl D A, Rittmann B E, Mirzabekov A D. Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. Applied and Environmental Microbiology, 1997, 63(6): 2397–2402
16. Zhou J, Kang S, Schadt C W, Garten C T Jr. Spatial scaling of functional gene diversity across various microbial taxa. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(22): 7768–7773
17. He Z, Deng Y, Van Nostrand J D, Tu Q, Xu M, Hemme C L, Li X, Wu L, Gentry T J, Yin Y, Liebich J, Hazen T C, Zhou J. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. The ISME Journal, 2010, 4(9): 1167–1179
18. Wu L, Thompson D K, Li G, Hurt R A, Tiedje J M, Zhou J. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Applied and Environmental Microbiology, 2001, 67(12): 5780–5790
19. Rhee S K, Liu X, Wu L, Chong S C, Wan X, Zhou J. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Applied and Environmental Microbiology, 2004, 70(7): 4303–4317
20. Gentry T J, Wickham G S, Schadt C W, He Z, Zhou J. Microarray applications in microbial ecology research. Microbial Ecology, 2006, 52(2): 159–175
21. He Z, Gentry T J, Schadt C W, Wu L, Liebich J, Chong S C, Huang Z, Wu W, Gu B, Jardine P, Criddle C, Zhou J. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. The ISME Journal, 2007, 1(1): 67–77
22. Taroncher-Oldenburg G, Griner E M, Francis C A, Ward B B. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. Applied and Environmental Microbiology, 2003, 69(2): 1159–1171
23. Zhang L, Hurek T, Reinhold-Hurek B. A nifH-based oligonucleotide microarray for functional diagnostics of nitrogen-fixing microorganisms. Microbial Ecology, 2007, 53(3): 456–470
24. Bontemps C, Golfier G, Gris-Liebe C, Carrere S, Talini L, Boivin-Masson C. Microarray-based detection and typing of the Rhizobium nodulation gene nodC: potential of DNA arrays to diagnose biological functions of interest. Applied and Environmental Microbiology, 2005, 71(12): 8042–8048
25. Bodrossy L, Stralis-Pavese N, Murrell J C, Radajewski S, Weilharter A, Sessitsch A. Development and validation of a diagnostic microbial microarray for methanotrophs. Environmental Microbiology, 2003, 5(7): 566–582
26. Stralis-Pavese N, Sessitsch A, Weilharter A, Reichenauer T, Riesing J, Csontos J, Murrell J C, Bodrossy L. Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. Environmental Microbiology, 2004, 6(4): 347–363
27. Bodrossy L, Stralis-Pavese N, Konrad-Köszler M, Weilharter A, Reichenauer T G, Schöfer D, Sessitsch A. mRNA-based parallel detection of active methanotroph populations by use of a diagnostic microarray. Applied and Environmental Microbiology, 2006, 72(2): 1672–1676
28. Miller S M, Tourlousse D M, Stedtfeld R D, Baushke S W, Herzog A B, Wick L M, Rouillard J M, Gulari E, Tiedje J M, Hashsham S A. In situ-synthesized virulence and marker gene biochip for detection of bacterial pathogens in water. Applied and Environmental Microbiology, 2008, 74(7): 2200–2209
29. Call D R, Bakko M K, Krug M J, Roberts M C. Identifying antimicrobial resistance genes with DNA microarrays. Antimicrobial Agents and Chemotherapy, 2003, 47(10): 3290–3295
30. Kostić T, Weilharter A, Sessitsch A, Bodrossy L. High-sensitivity, polymerase chain reaction-free detection of microorganisms and their functional genes using 70-mer oligonucleotide diagnostic microarray. Analytical Biochemistry, 2005, 346(2): 333–335
31. Cleven B E E, Palka-Santini M, Gielen J, Meembor S, Krönke M, Krut O. Identification and characterization of bacterial pathogens causing bloodstream infections by DNA microarray. Journal of Clinical Microbiology, 2006, 44(7): 2389–2397
32. Palka-Santini M, Cleven B E, Eichinger L, Krönke M, Krut O. Large scale multiplex PCR improves pathogen detection by DNA microarrays. BMC Microbiology, 2009, 9(1): 1
33. Yin H, Cao L, Qiu G, Wang D, Kellogg L, Zhou J, Dai Z, Liu X. Development and evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in Acid Mine Drainages and bioleaching systems. Journal of Microbiological Methods, 2007, 70(1): 165–178
34. Tiquia S M, Wu L, Chong S C, Passovets S, Xu D, Xu Y, Zhou J. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. BioTechniques, 2004, 36(4): 664–670, 672, 674–675
35. Liang Y, He Z, Wu L, Deng Y, Li G, Zhou J. Development of a common oligonucleotide reference standard for microarray data normalization and comparison across different microbial communities. Applied and Environmental Microbiology, 2010, 76(4):

1088–1094
36. Zhou J, He Z, Van Nostrand J D, Wu L, Deng Y. Applying GeoChip analysis to disparate microbial communities. Microbe, 2010, 5: 60–65
37. Wilson K H, Blitchington R B, Greene R C. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. Journal of Clinical Microbiology, 1990, 28(9): 1942–1946
38. Amann R I, Ludwig W, Schleifer K H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological Reviews, 1995, 59(1): 143–169
39. Yamamoto S, Harayama S. PCR amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains. Applied and Environmental Microbiology, 1995, 61(3): 1104–1109
40. Hugenholtz P, Goebel B M, Pace N R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. Journal of Bacteriology, 1998, 180(18): 4765–4774
41. Brodie E L, Desantis T Z, Joyner D C, Baek S M, Larsen J T, Andersen G L, Hazen T C, Richardson P M, Herman D J, Tokunaga T K, Wan J M, Firestone M K. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Applied and Environmental Microbiology, 2006, 72(9): 6288–6298
42. Liesegang H, Lemke K, Siddiqui R A, Schlegel H G. Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. Journal of Bacteriology, 1993, 175(3): 767–778
43. Grass G, Grosse C, Nies D H. Regulation of the cnr cobalt and nickel resistance determinant from *Ralstonia* sp. strain CH34. Journal of Bacteriology, 2000, 182(5): 1390–1398
44. Tibazarwa C, Wuertz S, Mergeay M, Wyns L, van Der Lelie D. Regulation of the *cnr* cobalt and nickel resistance determinant of *Ralstonia eutropha* (*Alcaligenes eutrophus*) CH34. Journal of Bacteriology, 2000, 182(5): 1399–1409
45. Nies D H. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiology Reviews, 2003, 27(2–3): 313–339
46. Saier M H Jr, Tam R, Reizer A, Reizer J. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Molecular Microbiology, 1994, 11(5): 841–847
47. Paulsen I T, Park J H, Choi P S, Saier M H Jr. A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria. FEMS Microbiology Letters, 1997, 156 (1): 1–8
48. Eddy S R. Profile hidden Markov models. Bioinformatics (Oxford, England), 1998, 14(9): 755–763
49. Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 1994, 22(22): 4673–4680
50. Li X, He Z, Zhou J. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. Nucleic Acids Research, 2005, 33(19): 6114–6123
51. Liebich J, Schadt C W, Chong S C, He Z, Rhee S K, Zhou J. Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. Applied and Environmental Microbiology, 2006, 72(2): 1688–1691
52. He Z, Wu L, Li X, Fields M W, Zhou J. Empirical establishment of oligonucleotide probe design criteria. Applied and Environmental Microbiology, 2005, 71(7): 3753–3760
53. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. Journal of Computational Biology, 2000, 7(1–2): 203–214
54. Zhou J, Bruns M A, Tiedje J M. DNA recovery from soils of diverse composition. Applied and Environmental Microbiology, 1996, 62(2): 316–322
55. Hurt R A, Qiu X, Wu L, Roh Y, Palumbo A V, Tiedje J M, Zhou J. Simultaneous recovery of RNA and DNA from soils and sediments. Applied and Environmental Microbiology, 2001, 67 (10): 4495–4503
56. Liang Y, Li G, Van Nostrand J D, He Z, Wu L, Deng Y, Zhang X, Zhou J. Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field. FEMS Microbiology Ecology, 2009, 70(2): 324–333
57. Ning J, Liebich J, Kästner M, Zhou J, Schäffer A, Burauel P. Different influences of DNA purity indices and quantity on PCR-based DGGE and functional gene microarray in soil microbial community study. Applied Microbiology and Biotechnology, 2009, 82(5): 983–993
58. Wu L, Liu X, Schadt C W, Zhou J. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. Applied and Environmental Microbiology, 2006, 72(7): 4931–4941
59. Bürgmann H, Widmer F, Sigler W V, Zeyer J. mRNA extraction and reverse transcription-PCR protocol for detection of *nifH* gene expression by *Azotobacter vinelandii* in soil. Applied and Environmental Microbiology, 2003, 69(4): 1928–1935
60. McGrath K C, Thomas-Hall S R, Cheng C T, Leo L, Alexa A, Schmidt S, Schenk P M. Isolation and analysis of mRNA from environmental microbial communities. Journal of Microbiological Methods, 2008, 75(2): 172–176
61. Gao H, Yang Z K, Gentry T J, Wu L, Schadt C W, Zhou J. Microarray-based analysis of microbial community RNAs by whole-community RNA amplification. Applied and Environmental Microbiology, 2007, 73(2): 563–571
62. He Z, Wu L, Fields M W, Zhou J. Use of microarrays with different probe sizes for monitoring gene expression. Applied and Environmental Microbiology, 2005, 71(9): 5154–5162
63. Van Nostrand J D, Wu W M, Wu L, Deng Y, Carley J, Carroll S, He Z, Gu B, Luo J, Criddle C S, Watson D B, Jardine P M, Marsh T L, Tiedje J M, Hazen T C, Zhou J. GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium-contaminated aquifer. Environmental Microbiology, 2009, 11(10): 2611–2626

64. Waldron P J, Wu L, Van Nostrand J D, Schadt C W, He Z, Watson D B, Jardine P M, Palumbo A V, Hazen T C, Zhou J. Functional gene array-based analysis of microbial community structure in groundwaters with a gradient of contaminant levels. Environmental Science & Technology, 2009, 43(10): 3529–3534
65. Mason O U, Di Meo-Savoie C A, Van Nostrand J D, Zhou J, Fisk M R, Giovannoni S J. Prokaryotic diversity, distribution, and insights into their role in biogeochemical cycling in marine basalts. The ISME Journal, 2009, 3(2): 231–242
66. He Z, Zhou J. Empirical evaluation of a new method for calculating signal-to-noise ratio for microarray data analysis. Applied and Environmental Microbiology, 2008, 74(10): 2957–2966
67. He Z, Van Nostrand J D, Wu L, Zhou J. Development and application of functional gene arrays for microbial community analysis. Transactions of Nonferrous Metals Society of China, 2008, 18(6): 1319–1327
68. Luo Y, Hui D, Zhang D. Elevated CO2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. Ecology, 2006, 87(1): 53–63
69. He Z, Xu M, Deng Y, Kang S, Kellogg L, Wu L, Van Nostrand J D, Hobbie S E, Reich P B, Zhou J. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO_2. Ecology Letters, 2010, 13(5): 564–575
70. ter Braak C J F. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology, 1986, 67(5): 1167–1179
71. Økland R H, Eilertsen O. Canonical correspondence analysis with variation partitioning: some comments and an application. Journal of Vegetation Science, 1994, 5(1): 117–126
72. Ramette A, Tiedje J M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(8): 2761–2766
73. Yergeau E, Kang S, He Z, Zhou J, Kowalchuk G A. Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. The ISME Journal, 2007, 1(2): 163–179
74. Wu L, Kellogg L, Devol A H, Tiedje J M, Zhou J. Microarray-based characterization of microbial community functional structure and heterogeneity in marine sediments from the Gulf of Mexico. Applied and Environmental Microbiology, 2008, 74(14): 4516–4529
75. Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand J D, Deng Y, He Z, Wu L, Zhou J, Xiao X. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(12): 4840–4845
76. Xiong J, Wu L, Tu S, Van Nostrand J D, He Z, Zhou J, Wang G. Microbial communities and functional genes associated with soil arsenic contamination and rhizosphere of the arsenic hyperaccumulating plant *Pteris vittata* L. Applied and Environmental Microbiology, 2010 (in press, AEM.00500–10)
77. Liu W, Wang A, Cheng S, Logan B E, Yu H, Deng Y, Nostrand J D V, Wu L, He Z, Zhou J. Geochip-based functional gene analysis of anodophilic communities in microbial electrolysis cells under different operational modes. Environmental Science & Technology, 2010, 44(19): 7729–7735
78. Dennis P, Edwards E A, Liss S N, Fulthorpe R. Monitoring gene expression in mixed microbial communities by using DNA microarrays. Applied and Environmental Microbiology, 2003, 69 (2): 769–778
79. Deng Y, He Z, Van Nostrand J D, Zhou J. Design and analysis of mismatch probes for long oligonucleotide microarrays. BMC Genomics, 2008, 9(1): 491
80. Chen Q, Yin H, Luo H, Xie M, Qiu G, Liu X. Micro-array based whole-genome hybridization for detection of microorganisms in acid mine drainage and bioleaching systems. Hydrometallurgy, 2009, 95(1–2): 96–103
81. Denef V J, Park J, Rodrigues J L M, Tsoi T V, Hashsham S A, Tiedje J M. Validation of a more sensitive method for using spotted oligonucleotide DNA microarrays for functional genomics studies on bacterial communities. Environmental Microbiology, 2003, 5 (10): 933–943
82. Relógio A, Schwager C, Richter A, Ansorge W, Valcárcel J. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Research, 2002, 30(11): e51
83. Cho J C, Tiedje J M. Quantitative detection of microbial genes by using DNA microarrays. Applied and Environmental Microbiology, 2002, 68(3): 1425–1430
84. Zhou J, Thompson D K. Challenges in applying microarrays to environmental studies. Current Opinion in Biotechnology, 2002, 13(3): 204–207
85. Schaupp C J, Jiang G, Myers T G, Wilson M A. Active mixing during hybridization improves the accuracy and reproducibility of microarray results. BioTechniques, 2005, 38(1): 117–119
86. Zhou X C, Wu L Y, Zhou J Z. Fabrication of DNA microarrays on nanoengineered polymeric ultrathin film prepared by self-assembly of polyelectrolyte multilayers. Langmuir, 2004, 20(20): 8877–8885
87. Branham W S, Melvin C D, Han T, Desai V G, Moland C L, Scully A T, Fuscoe J C. Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements. BMC Biotechnology, 2007, 7(1): 8
88. Kumar A, Larsson O, Parodi D, Liang Z. Silanized nucleic acids: a general platform for DNA immobilization. Nucleic Acids Research, 2000, 28(14): e71
89. Gudnason H, Dufva M, Duong Bang D, Wolff A. An inexpensive and simple method for thermally stable immobilization of DNA on an unmodified glass surface: UV linking of poly(T)10-poly(C)10-tagged DNA probes. BioTechniques, 2008, 45(3): 261–271
90. DeSantis T Z, Brodie E L, Moberg J P, Zubieta I X, Piceno Y M, Andersen G L. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microbial Ecology, 2007, 53(3): 371–383
91. Leigh M B, Pellizari V H, Uhlík O, Sutka R, Rodrigues J, Ostrom N E, Zhou J, Tiedje J M. Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). The ISME Journal, 2007, 1

(2): 134–148

92. Gu B, Brooks S C, Roh Y, Jardine P M. Geochemical reactions and dynamics during titration of a contaminated groundwater with high uranium, aluminum, and calcium. Geochimica et Cosmochimica Acta, 2003, 67(15): 2749–2761
93. Wu W M, Carley J, Gentry T, Ginder-Vogel M A, Fienen M, Mehlhorn T, Yan H, Caroll S, Pace M N, Nyman J, Luo J, Gentile M E, Fields M W, Hickey R F, Gu B, Watson D, Cirpka O A, Zhou J, Fendorf S, Kitanidis P K, Jardine P M, Criddle C S. Pilot-scale in situ bioremedation of uranium in a highly contaminated aquifer. 2. Reduction of u(VI) and geochemical control of u(VI) bioavailability. Environmental Science & Technology, 2006, 40(12): 3986–3995
94. Wu W M, Carley J, Luo J, Ginder-Vogel M A, Cardenas E, Leigh M B, Hwang C, Kelly S D, Ruan C, Wu L, Van Nostrand J D, Gentry T, Lowe K, Mehlhorn T, Carroll S, Luo W, Fields M W, Gu B, Watson D, Kemner K M, Marsh T, Tiedje J, Zhou J, Fendorf S, Kitanidis P K, Jardine P M, Criddle C S. *In situ* bioreduction of uranium (VI) to submicromolar levels and reoxidation by dissolved oxygen. Environmental Science & Technology, 2007, 41(16): 5716–5723
95. Xu M, Wu W M, Wu L, He Z, Van Nostrand J D, Deng Y, Luo J, Carley J, Ginder-Vogel M, Gentry T J, Gu B, Watson D, Jardine P M, Marsh T L, Tiedje J M, Hazen T, Criddle C S, Zhou J. Responses of microbial community functional structures to pilot-scale uranium in situ bioremediation. The ISME Journal, 2010, 4 (8): 1060–1070
96. Tas N, van Eekert M H, Schraa G, Zhou J, de Vos W M, Smidt H. Tracking functional guilds: “Dehalococcoides” spp. in European river basins contaminated with hexachlorobenzene. Applied and Environmental Microbiology, 2009, 75(14): 4696–4704
97. Kimes N E, Van Nostrand J D, Weil E, Zhou J, Morris P J. Microbial functional structure of Montastraea faveolata, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. Environmental Microbiology, 2010, 12(2): 541–556
98. Zhang Y, Zhang X, Liu X, Xiao Y, Qu L, Wu L, Zhou J. Microarray-based analysis of changes in diversity of microbial genes involved in organic carbon decomposition following land use/cover changes. FEMS Microbiology Letters, 2007, 266(2): 144–151
99. Lawton J H. Are there general laws in ecology? Oikos, 1999, 84 (2): 177–192
100. Reich P B, Knops J, Tilman D, Craine J, Ellsworth D, Tjoelker M, Lee T, Wedin D, Naeem S, Bahauddin D, Hendrey G, Jose S, Wrage K, Goth J, Bengston W. Plant diversity enhances ecosystem responses to elevated CO_2 and nitrogen deposition. Nature, 2001, 410(6830): 809–812
101. Parnell J J, Rompato G, Latta L C I V 4th, Pfrender M E, Van Nostrand J D, He Z, Zhou J, Andersen G, Champine P, Ganesan B, Weimer B C, Aziz R K. Functional biogeography as evidence of gene transfer in hypersaline microbial communities. PLoS ONE, 2010, 5(9): e12919
102. Xie J, He Z, Liu X, Liu X, Van Nostrand J D, Deng Y, Wu L, Qiu G, Zhou J.Geochip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. Applied and Environmental Microbiology, 2010,
103. Rodríguez-Martínez E M, Pérez E X, Schadt C W, Zhou J, Massol-Deyá A A. Microbial diversity and bioremediation of a hydrocarbon-contaminated aquifer (Vega Baja, Puerto Rico). International Journal of Environmental Research and Public Health, 2006, 3(3): 292–300
104. Liang Y, Nostrand J D V, Wang J, Zhang X, Zhou J, Li G. Microarray-based functional gene analysis of soil microbial communities during ozonation and biodegradation of crude oil. Chemosphere, 2009, 75(2): 193–199
105. Liang Y, Van Nostrand J D, Deng Y, He Z, Wu L, Zhang X, Li G, Zhou J. Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. The ISME Journal, 2010, (in press)
106. Hazen T C, Dubinsky E A, DeSantis T Z, Andersen G L, Piceno Y M, Singh N, Jansson J K, Probst A, Borglin S E, Fortney J L, Stringfellow W T, Bill M, Conrad M E, Tom L M, Chavarria K L, Alusi T R, Lamendella R, Joyner D C, Spier C, Baelum J, Auer M, Zemla M L, Chakraborty R, Sonnenthal E L, D’haeseleer P, Holman H Y N, Osman S, Lu Z, Van Nostrand J D, Deng Y, Zhou J, Mason O U. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science, 2010, 330(6001): 204–208
107. Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Bemben L A, Berka J, Braverman M S, Chen Y J, Chen Z T, Dewell S B, Du L, Fierro J M, Gomes X V, Godwin B C, He W, Helgesen S, Ho C H, Irzyk G P, Jando S C, Alenquer M L, Jarvie T P, Jirage K B, Kim J B, Knight J R, Lanza J R, Leamon J H, Lefkowitz S M, Lei M, Li J, Lohman K L, Lu H, Makhijani V B, McDade K E, McKenna M P, Myers E W, Nickerson E, Nobile J R, Plant R, Puc B P, Ronan M T, Roth G T, Sarkis G J, Simons J F, Simpson J W, Srinivasan M, Tartaro K R, Tomasz A, Vogt K A, Volkmer G A, Wang S H, Wang Y, Weiner M P, Yu P, Begley R F, Rothberg J M. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 2005, 437(7057): 376–380
108. Binladen J, Gilbert M T P, Bollback J P, Panitz F, Bendixen C, Nielsen R, Willerslev E, Hahn M. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE, 2007, 2 (2): e197
109. Hamady M, Walker J J, Harris J K, Gold N J, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nature Methods, 2008, 5(3): 235–237
110. Iwai S, Chai B, Sul W J, Cole J R, Hashsham S A, Tiedje J M. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. The ISME Journal, 2010, 4(2): 279–285
111. Qin J, Li R, Raes J, Arumugam M, Burgdorf K S, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende D R, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J M, Hansen T, Le Paslier D, Linneberg A, Nielsen H B, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich S D, Wang J. A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 2010, 464(7285): 59–65

Jizhong Zhou is a Presidential Professor in the Department of Botany and Microbiology and Director for the Institute for Environmental Genomics, University of Oklahoma, Norman, OK, Adjunct Senior Scientist at Lawrence Berkeley National Laboratory, and Adjunct Professor at Tsinghua University, Beijing, China. He received a B.S. in Plant Pathology and Entomology in 1981, an M.S. in Insect Mathematical Ecology in 1984 from Hunan Agricultural University, China, and a Ph.D. in Molecular Genetics and Cell Biology in 1993 from Washington State University. He studied theoretical ecology and ecosystem modeling for three years at the Eco-Environmental Research Center, Chinese Academy of Sciences, Beijing, China. He worked at the Center for Microbial Ecology, Michigan State University, from 1993 to 1995 as a Postdoctoral Research Associate with James Tiedje. Before moving to OU in 2005, he worked at Oak Ridge National Laboratory as a Staff Scientist, Senior Staff Scientist, and then Distinguished Staff Scientist for 10 years. His expertise is in molecular biology, microbial genomics, microbial ecology, molecular evolution, theoretical ecology and genomic technologies. His laboratory has pioneered the development and use of genomic technologies for environmental studies in which GeoChip was awarded a 2009 R&D100 Award. He received a Presidential Early Career Award for Scientists and Engineers in 2001.