Encyclopedia of Metagenomics
DOI 10.1007/978-1-4614-6418-1_760-4

GeoChip-Based Metagenomic Technologies for Analyzing Microbial Community Functional Structure and Activities

Zhili He[a]*, Joy D. Van Nostrand[a] and Jizhong (Joe) Zhou[a,b,c]
[a]Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA
[b]Department of Environmental Science and Engineering, Tsinghua University, Beijing, China
[c]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Synonyms

Functional gene array; Metagenomic technology

Definition

Functional gene arrays (FGAs) are a special type of microarray 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, antibiotic resistance, energy processing, and stress response. GeoChips are considered to be the most comprehensive FGAs and an important metagenomic tool for microbial community analysis.

Introduction

Microorganisms are the most diverse group of organisms known in terms of phylogeny and functionality. However, they do not live alone but form distinct communities and play integrated and unique roles in ecosystems, such as biogeochemical cycling of carbon (C), nitrogen (N), sulfur (S), phosphorus (P), and metals (e.g., iron, copper, zinc), biodegradation or stabilization of environmental contaminants, and interaction with hosts. Therefore, one of the most important goals of microbial ecology is to understand the diversity, composition, structure, function, dynamics, and evolution of microbial communities and their relationships with environmental factors and ecosystem functioning. Toward this goal, several challenges remain. First, microorganisms are generally too small to see or characterize with most approaches used for plant or animal studies. Second, microbial communities are extremely diverse. It is estimated that 1 g of soil contains 2,000–50,000 microbial species (Torsvik et al. 2002) and even up to millions of species (Gans et al. 2005). Third, a vast majority of microorganisms (>99 %) are uncultured (Whitman et al. 1998), making it difficult to study their functional ability and molecular mechanisms. Finally, 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 necessary.

Indeed, many culture-independent approaches are available including PCR-based cloning analysis, denaturing gradient gel electrophoresis (DGGE), terminal-restriction fragment length

*Email: zhili.he@ou.edu

DOI 10.1007/978-1-4614-6418-1_760-4

polymorphism (T-RFLP), quantitative PCR, and in situ hybridization. However, these methods only provide snapshots of a microbial community but fail to provide a comprehensive view. Therefore, high-throughput metagenomic technologies are necessary for providing a rapid, specific, sensitive, and quantitative analysis of microbial communities and their relationships with environmental factors and ecosystem functioning.

Microarray-based technology can examine thousands of genes at one time, providing a much more comprehensive analysis of microbial communities. This technology, like GeoChip, has been developed and adopted to analyze microbial communities (He et al. 2007, 2010a; Hazen et al. 2010) and has been used to profile the functional diversity, composition, structure, and dynamics of microbial communities from different habitats (He et al. 2011, 2012a, b). A variety of studies demonstrate that microarrays can provide phylogenetic and functional information on a microbial community in a rapid, high-throughput, and parallel manner.

This overview is focused on the analysis of functional diversity, structure, and activity of microbial communities using GeoChip-based metagenomic technologies but also includes a brief introduction of GeoChips, GeoChip development, and GeoChip hybridization and data analysis.

GeoChips as the Most Comprehensive Functional Gene Arrays

Functional gene arrays (FGAs) are special microarrays containing probes for key genes involved in microbial functional processes, such as biogeochemical cycling of carbon (C), nitrogen (N), phosphorus (P), sulfur (S), and metals, antibiotic resistance, biodegradation of environmental contaminants, energy processing, and stress response. Since the exact functions of selected genes on FGAs are known, this type of array is especially useful for examining the functional diversity, composition, and structure of microbial communities across different times and scales. Several FGAs have been reported and evaluated, and they generally target specific functional processes, populations, or environments, including *nodC* and *nifH* arrays, a methanotroph gene (*pmoA*) array, a virulence marker gene (VMG) array, pathogen detection/diagnosis arrays, and a bioleaching array (He et al. 2012b). However, GeoChips are the most comprehensive FGAs to date, especially the later versions (GeoChips 2.0, 3.0, and 4.0), which target a variety of key microbial functional processes, such as C, N, P, and S cycling, contaminant bioremediation, and antibiotic resistance (He et al. 2012a).

GeoChips, constructed with 50-mer oligonucleotide probes, have evolved over several generations. The prototype GeoChip contained 89 PCR-amplicon probes for N-cycling genes (*nirS*, *nirK*, *amoA*, and *pmoA*) derived from pure-culture isolates and marine sediment clone libraries (Wu et al. 2001). The first-generation GeoChip (GeoChip 1.0) was constructed with 763 gene variants involved in nitrogen cycling (*nirS*, *nirK*, *nifH*, *amoA*), methane oxidation (*pmoA*), and sulfite reduction (*dsrAB*). Then, an expanded array was developed with 2,402 genes involved in organic contaminant biodegradation and metal resistance to monitor microbial populations and functional genes involved in biodegradation and biotransformation (Rhee et al. 2004). Specificity evaluation with representative pure cultures indicated that the designed probes appeared to be specific to their corresponding target genes. The detection limit was 5–10 ng of genomic DNA in the absence of background DNA and 50–100 ng of pure-culture genomic DNA in the presence of background DNA. Real-time PCR analysis was very consistent with the microarray-based quantification (He et al. 2011).

Although the prototype and GeoChip 1.0 arrays were used to probe specific functional groups or activities, they lacked a truly comprehensive probe set covering key microbial functional processes

DOI 10.1007/978-1-4614-6418-1_760-4

occurring in different environments. Therefore, more comprehensive GeoChips have been developed and evaluated. For example, GeoChip 2.0, containing 24,243 (50-mer) oligonucleotide probes, targeting ~10,000 functional gene variants 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 (He et al. 2007). After 2 years, GeoChip 3.0 was developed, which contained about 28,000 probes and targeted ~57,000 sequences from 292 gene families (He et al. 2010a). 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, 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 (He et al. 2010a). Based on GeoChip 3.0, GeoChip 4.0 was developed, which contains ~84,000 probes and targeting >152,000 genes from 410 functional families. GeoChip 4.0 not only contains all functional categories from GeoChip 3.0 but also includes additional functional categories, such as genes from bacterial phages and those involved in stress response and virulence (Hazen et al. 2010; He et al. 2012a). All evaluation and studies demonstrate that GeoChip is a powerful tool for specific, sensitive, and quantitative analysis of microbial communities from a variety of habitats (He et al. 2011, 2012a, b).

GeoChip Development

GeoChip development involves several major steps, including selection of target genes, sequence retrieval and verification, oligonucleotide probe design, probe validation, and array construction as well as future automatic update, which are generally implemented by a GeoChip development and data analysis pipeline (► http://ieg./ou.edu/) (He et al. 2010a).

Selection of Target Genes and Sequence Retrieval

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, contaminant bioremediation, antibiotic resistance, and stress response. For example, 292 functional gene families were selected for GeoChip 3.0 with 41 for C cycling, 16 for N cycling, 3 for P utilization, 4 for S cycling, 173 for biodegradation of a variety of organic contaminants, 41 for metal reduction and resistance, 11 for antibiotic resistance, and 2 for energy processing. In addition, a phylogenetic marker (*gyrB*) was also chosen (He et al. 2010a). More importantly, when sequences for a known functional gene are available, they can be added in an updated GeoChip. For example, when GeoChip was updated to GeoChip 4.0, functional gene families involved in stress responses, bacterial phages, and virulence were added, resulting in 410 functional gene families on GeoChip 4.0 (Hazen et al. 2010; He et al. 2012a). Generally, genes are chosen for key enzymes or proteins with the corresponding function(s) of interest. If a process involves multiple steps or a protein complex, those genes responsible for catalytic subunits or with the active site(s) will be selected (He et al. 2011).

Sequence retrieval is performed generally with a pipeline with a database integrated for managing all retrieved sequences and subsequently designed probes. For each functional gene, the first step is to submit a query to the GenBank protein database and fetch all candidate amino acid sequences. The key words may include the name of the target gene/enzyme, its abbreviation and enzyme commission number (EC), and affiliated domains of bacteria, archaea, and fungi. Second, retrieved sequences are validated by seed sequences (those sequences that have been experimentally confirmed to produce the protein of interest and that the protein functions as expected) with the

DOI 10.1007/978-1-4614-6418-1_760-4

HMMER program. Finally, all confirmed protein sequences are searched against GenBank again to obtain their corresponding nucleic acid sequences for probe design (He et al. 2010a).

Oligonucleotide Probe Design

A new version of CommOligo (He et al. 2012a) 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) a gene-specific probe must have $\leq$90 % sequence identity, $\leq$20-base continuous stretch, and ≥ -35 kcal/mol free energy; (ii) a group-specific probe has to meet the above requirements for nontarget groups, and it also must have $\geq$96 % sequence identity, $\geq$ 35-base continuous stretch, and ≤ -60 kcal/mol free energy within the group. Computational and experimental evaluation indicates that these designed probes are highly specific to their targets (He et al. 2007, 2010a).

Probe Validation and GeoChip Construction

All designed probes are subsequently verified against the GenBank (NR) nucleic acid database for specificity. Normally, multiple (e.g., 20) probes for each sequence or each group of sequences are designed, but only the best probe set for each sequence or each group of closely related sequences will be chosen for array construction. GeoChip can be constructed in-house, such as GeoChips 2.0 and 3.0 (He et al. 2007, 2010a), or commercially, like GeoChip 4.0 (Hazen et al. 2010; He et al. 2012a).

GeoChip Operation and Data Analysis

Generally, GeoChip operation and data analysis include target preparation, GeoChip hybridization, image and data preprocessing, and data analysis (Fig. 1).

Target Preparation

Target preparation involves a few steps, including nucleic acid extraction and purification, labeling, and hybridization (Fig. 1a). The most important step for successful GeoChip analysis is nucleic acid extraction and purification from environmental samples generally using a well-established method, which is able to produce large fragments of DNA. High-quality DNA should have ratios of $A_{260}/A_{280} \sim 1.8$ and $A_{260}/A_{230} > 1.7$. Low A_{260}/A_{230} ratios indicate impurities in the DNA sample and can negatively influence subsequent labeling and hybridization. Generally, since 1–5 μg of DNA or 5–20 μg of RNA is required for GeoChip hybridization, whole-community genome amplification (WCGA) for DNA and whole-community RNA amplification (WCRA) for RNA are necessary (He et al. 2012b). Non-amplified or amplified nucleic acids are then labeled with fluorescent dye (e.g., Cy3, Cy5) using random priming with the Klenow fragment of DNA polymerase for DNA and SuperScript™ II/III RNase H-reverse transcriptase for RNA. The labeled nucleic acids are then purified and dried for hybridization (Fig. 1a).

Hybridization, Imaging, and Data Preprocessing

Labeled nucleic acid target is suspended in a hybridization buffer containing 40–50 % formamide and hybridized on GeoChip at 42–50 °C (He et al. 2007, 2010a, 2012b). The hybridization

DOI 10.1007/978-1-4614-6418-1_760-4

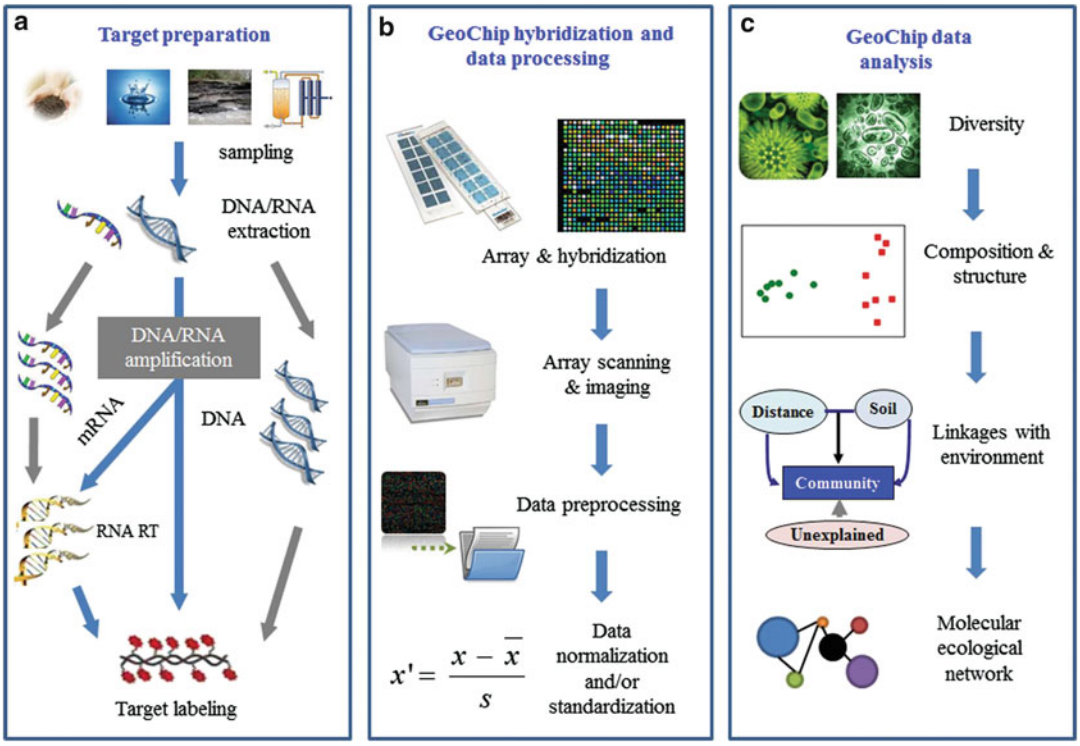

Fig. 1 A schematic presentation of target preparation, GeoChip operation, and data analysis of microbial communities from a variety of habitats. (**a**) Target preparation, (**b**) GeoChip hybridization and data processing, (**c**) GeoChip data analysis (This figure is adapted from Fig. 1 by He et al. (2012b))

stringency can be adjusted by changing the temperature and/or formamide concentration. For every 1 % increase in formamide, the effective temperature increases by 0.6 °C (He et al. 2011).

Hybridized arrays are imaged with a microarray scanner having a resolution of at least 10 μm for homemade arrays and 2 μm for commercially manufactured 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, and positive spots are called generally based on signal-to-noise ratio [SNR; SNR = (signal mean – background mean)/background standard deviation] or signal-to-both-standard-deviations ratio [SSDR; SSDR = (signal mean – background mean)/(signal standard deviation – background standard deviation)] (He et al. 2012b).

Raw GeoChip data are further evaluated via the GeoChip data analysis pipeline (He et al. 2010a). The quality of individual spots, evenness of control spot hybridization signals across the slide surface, and background levels are assessed to determine 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 (He et al. 2011). The signal intensities are then normalized for further statistical analysis (Fig. 1b).

GeoChip Data Analysis

Data analysis is the most challenging part in the use of GeoChip for microbial community analysis, and a variety of methods have been used to address fundamental microbial ecology questions (Fig. 1c). First, various diversity indices (e.g., richness, evenness, diversity) based on the number of functional genes detected and their abundances are used to examine the functional diversity of microbial communities. 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

DOI 10.1007/978-1-4614-6418-1_760-4

percentage of genes shared by different samples can also be calculated to compare microbial communities examined. Second, for statistical analysis of the overall microbial community composition and structure with FGA data, ordination techniques can be used such as principal component analysis (PCA), detrended correspondence analysis (DCA), cluster analysis (CA), and nonmetric multidimensional scaling (NMDS). PCA and DCA are multivariate statistical methods, which reduce the number of variables needed to explain the data and highlight the variability between samples. CA groups samples based on the overall similarity of gene patterns. NMDS finds both a nonparametric monotonic relationship between the dissimilarities in the item-item matrix and the Euclidean distances between items and the location of each item in the low-dimensional space. Also, the response ratio can be used to determine changes of specific functional genes between the control and the treatment. In addition, analysis of variation (ANOVA), analysis of similarities (ANOISM), nonparametric multivariate analysis of variance (Adonis), and multi-response permutation procedure (MRPP) can be used to discern dissimilarities of microbial communities over time and space (He et al. 2011, 2012b). Third, if environmental data or other metadata are available, GeoChip data can be used to correlate environmental variables with the functional microbial community structure. These include the Pearson's correlation coefficient (PCC), canonical correspondence analysis (CCA), and Mantel test. PCC measures the strength of linear dependence between two variables, such as functional gene abundances detected by GeoChip, and environmental variables. CCA has been used in many cases in GeoChip-based studies to better understand how environmental factors affect the community structure (He et al. 2011, 2012b). Also, based on the results of the CCA, the relative influence of environmental variables on the microbial community structure can be determined using variance partitioning analysis (VPA). In addition, further correlations of GeoChip data with environmental parameters can be performed with the Mantel test (He et al. 2007, 2010a, b, 2011, 2012b). Finally, GeoChip data can be used to infer functional molecular ecological networks for revealing interactions of functional genes and their associated populations. A recent study indicated that elevated CO_2 substantially altered the network interaction of soil microbial communities and the shift in network structures is significantly correlated with soil properties (He et al. 2012b; Zhou et al. 2010) (Fig. 1c).

GeoChip Applications

Different versions of GeoChip have been used to analyze microbial communities from different habitats, such as aquatic systems, soils, extreme environments, human microbiomes, and bioreactors for addressing fundamental scientific questions related to global change, bioenergy, bioremediation, agricultural management, land use, and human health and disease as well as ecological theories (He et al. 2011, 2012b). Several recent studies are highlighted, especially with a focus on soil and water microbial communities. A list of representative studies with different GeoChip versions is shown in Table 1.

Soils

Soil may harbor the most complex microbial communities among known habitats, and recently GeoChips have been used to investigate soil microbial communities to address fundamental ecological questions related to global change (e.g., elevated CO_2, elevated O_3, warming), bioremediation of oil-contaminated fields, land use, agricultural management, and livestock grazing.

Three recent studies focused on the response of soil microbial communities to global change, including elevated CO_2, temperature, and O_3. First, GeoChip 3.0 was used to analyze soil microbial

DOI 10.1007/978-1-4614-6418-1_760-4

Table 1 Summary of representative GeoChip applications. If no references are cited, those studies are described in a previous review (He et al. 2012b)

Habitat or ecosystem	Ecosystem/sample type	GeoChip	Objectives of study/biological questions
Aquatic systems	Marine sediment	GeoChip 1.0	Functional microbial community structure of marine sediments in the Gulf of Mexico
	Ebro and Elbe river sediment	GeoChip 2.0	Pesticide impacts on European rivers
	Coral-associated marine water	GeoChip 2.0	Microbial communities in healthy and yellow-band diseased coral (*Montastraea faveolata*)
Soils	Antarctic latitudinal transect soil	GeoChip 2.0	Microbial C and N cycling across an Antarctic latitudinal transect
	Deciduous forest soil	GeoChip 2.0	Gene-area relation in microorganisms
	Native grassland soil	GeoChip 2.0	Afforestation impacts soil microbial communities and their functional potential
	Strawberry farmland soil	GeoChip 2.0	Microbial responses to farm management
	Grassland soil	GeoChip 2.0	Microbial responses to plant invasion
	Agricultural soil	GeoChip 2.0	Agricultural practices/land use (Xue et al. 2013)
	Grassland soil	GeoChip 3.0	Global change (elevated CO_2) (He et al. 2010b)
	Grassland soil	GeoChip 3.0	Global change (warming) (Zhou et al. 2012)
	Wheat rhizosphere soil	GeoChip 3.0	Global change (elevated O_3) (Li et al. 2013)
	Citrus rhizosphere soil	GeoChip 3.0	Rhizosphere microbial community responses to *Candidatus Liberibacter asiaticus*-infected citrus trees
	Grassland soil	GeoChip 4.0	The effect of grazing on microbial communities (Yang et al. 2013)
Contaminated sites	U-contaminated underground water (Oak Ridge, TN)	GeoChip 1.0	Bioremediation of U-contaminated groundwater
		GeoChip 2.0	Bioremediation of U-contaminated groundwater (Van Nostrand et al. 2011)
	U-contaminated sediment (Oak Ridge, TN)	GeoChip 2.0	Bioremediation of U-contaminated sediments
	U-contaminated underground water (Rifle, CO)	GeoChip 2.0	Bioremediation of U-contaminated groundwater (Liang et al. 2012)
	PCB-contaminated soil	GeoChip 2.0	Microbial bioremediation of PCB-contaminated soil
	Oil-contaminated soil	GeoChip 2.0	Bioremediation of oil-contaminated soil
	Arsenic-contaminated soil	GeoChip 3.0	Rhizosphere microbial community responses to arsenic contamination and phytoremediation
	Landfill groundwater	GeoChip 3.0	Microbial responses to landfill-derived contaminants in groundwater (Lu et al. 2012)
	Oil-spill seawater	GeoChip 4.0	Microbial bioremediation of oil-spill sites (Hazen et al. 2010)

(*continued*)

DOI 10.1007/978-1-4614-6418-1_760-4

Table 1 (continued)

Habitat or ecosystem	Ecosystem/sample type	GeoChip	Objectives of study/biological questions
Extreme environments	Deep-sea hydrothermal vent (chimney)	GeoChip 2.0	Functional gene diversity of deep-sea hydrothermal vent microbial communities
	Deep-sea basalt samples	GeoChip 2.0	Functional gene diversity and structure of deep-sea basalt microbial communities
	GSL hypersaline water	GeoChip 2.0	Functional gene diversity and structure of hypersaline water microbial communities
	Acid mine drainage (water)	GeoChip 2.0	Functional gene diversity of microbial communities in acid mine drainage (AMD) systems
Bioreactors	Fluidized bed reactor for bioremediation	GeoChip 2.0	Microbial bioremediation of hydrocarbon-contaminated water
	Microbial electrolysis cell for hydrogen production	GeoChip 3.0	Microbial hydrogen production using wastewater

communities under elevated CO_2 at a multifactor grassland experiment site, BioCON (biodiversity, CO_2, and nitrogen deposition), in the Cedar Creek Ecosystem Science Reserve, MN (He et al. 2010b). The results showed that the functional microbial community structure was markedly different between ambient CO_2 and elevated CO_2 as indicated by DCA of GeoChip 3.0 data and 16S rRNA gene-based pyrosequencing data. Also, genes involved in labile C degradation and C and N fixation were significantly increased under elevated CO_2 although the abundance of recalcitrant C degradation genes remained unchanged. In addition, changes in the microbial community structure were significantly correlated with soil C and N contents and plant productivity (He et al. 2010b). Second, GeoChip 3.0 was used to understand the effect of increased temperature on soil microbial communities and their roles in regulating soil carbon dynamics at a tallgrass prairie ecosystem in the US Great Plains of Central Oklahoma. The results suggest soil microorganisms may regulate soil carbon dynamics through three primary feedback mechanisms: (i) shifting microbial community composition, leading to the reduced temperature sensitivity of heterotrophic soil respiration; (ii) differentially stimulating labile C but not recalcitrant C degradation genes to maintain long-term soil carbon stability and storage; and (iii) enhancing nutrient-cycling processes to promote plant growth (Zhou et al. 2012). Third, GeoChip 3.0 was used to investigate the functional composition, and structure of rhizosphere microbial communities from O_3-sensitive and O_3-relatively-sensitive wheat (*Triticum aestivum* L.) cultivars under elevated O_3 (eO_3). Based on GeoChip hybridization signal intensities, although the overall functional structure of rhizosphere microbial communities did not significantly change by eO_3 or cultivars, the results showed that the abundance of specific functional genes involved in C fixation and degradation, N fixation, and sulfite reduction did significantly alter in response to eO_3 and/or wheat cultivars. Also, the O_3-sensitive cultivar appeared to harbor microbial functional communities in the rhizosphere more sensitive in response to eO_3 than the O_3-relatively sensitive cultivar. In addition, CCA suggested that the functional structure of microbial communities involved in C cycling was largely shaped by soil and plant properties including pH, dissolved organic carbon (DOC), microbial biomass C, C/N ratio, and grain weight (Li et al. 2013). Those studies indicate that global change significantly impacts soil microbial communities, which may in turn regulate ecosystem functioning through different feedback mechanisms.

Various agriculture management practices may have significant influences on soil microbial communities and their ecological functions. GeoChip 2.0 was used to evaluate the potential

DOI 10.1007/978-1-4614-6418-1_760-4

functions of soil microbial communities under conventional (CT), low-input (LI), and organic (ORG) management systems at an agricultural research site in Michigan. Compared to CT, a high diversity of functional genes was observed in LI. The functional gene diversity in ORG did not differ significantly from that of either CT or LI. The abundance of genes encoding enzymes involved in C, N, P, and S cycling was generally lower in CT than in LI or ORG, but functional genes involved in lignin degradation, methane generation/oxidation, and assimilatory N reduction remained unchanged. Also, significant correlations were observed between NO_3^- concentration and denitrification gene abundance, NH_4^+ concentration and ammonification gene abundance, and N_2O flux and denitrification gene abundance, indicating a close linkage between soil N availability or utilization and associated functional potential of soil microbial communities (Xue et al. 2013).

Livestock grazing is a type of global land-use activity. However, the effect of free livestock grazing on soil microbial communities at the functional gene level remains unclear. GeoChip 4.0 was used to examine the effects of free livestock grazing on the microbial community at an experimental site in Tibet, a region known to be very sensitive to anthropogenic perturbation and global warming. The results showed that grazing changed the microbial community functional structure, in addition to aboveground vegetation and soil geochemical properties. Further statistical analysis showed that microbial community functional structures were closely correlated with environmental variables and variations in microbial community functional structures were mainly controlled by aboveground vegetation, soil C/N ratio, and NH_4^+-N. Therefore, these results indicated that soil microbial community functional structure was very sensitive to livestock grazing and revealed the role of soil microbial communities in the regulation of soil N and C cycling, supporting the necessity to include microbial components in evaluating the consequence of land use and/or climate change (Yang et al. 2013).

Groundwater and Aquatic Ecosystems

Due to human activities, groundwater and aquatic ecosystems are often contaminated from various sources (e.g., mining, oil spill, landfill) and with a variety of toxic compounds (e.g., heavy metals, herbicides, antibiotics, pesticides) and conditions (e.g., low pH, high salinity). To understand how such contamination impacts groundwater and aquatic ecosystems, GeoChips were used to investigate those microbial communities to explore the potential of in situ bioremediation of contaminated sites by indigenous microbial communities.

A pilot-scale system was established to examine the feasibility of in situ U(VI) immobilization at a highly contaminated aquifer in Oak Ridge, TN. Ethanol was injected intermittently as an electron donor to stimulate microbial U(VI) reduction, leading to a decrease of U(VI) concentrations below the Environmental Protection Agency drinking water standard. GeoChip 2.0 was used to monitor microbial communities in three wells during active U(VI) reduction and maintenance phases. The results showed that the overall microbial community structure exhibited a considerable shift over the remediation phases examined and functional populations of Fe(III)-reducing bacteria (FeRB), nitrate-reducing bacteria (NRB), and sulfate-reducing bacteria (SRB) reached their highest levels during the active U(VI) reduction phase (days 137–370), in which denitrification, Fe(III) reduction, and sulfate reduction occurred sequentially, suggesting that these functional populations could play an important role in both active U(VI) reduction and maintenance stability of reduced U(IV) (Van Nostrand et al. 2011).

To better understand the microbial functional diversity changes with subsurface redox conditions during in situ U(VI) bioremediation, GeoChip 2.0 was applied to examine groundwater microbial communities at a uranium mill tailings remedial action (UMTRA) site (Rifle, CO). The results indicated that functional microbial communities altered with a shift in the dominant metabolic

DOI 10.1007/978-1-4614-6418-1_760-4

process and the abundance of *dsrAB* and *mcr* genes increased when redox conditions shifted from Fe-reducing to sulfate-reducing conditions, while cytochrome genes were primarily detected from *Geobacter* species and decreased with lower subsurface redox conditions. Statistical analysis of environmental parameters and functional genes indicated that acetate, U(VI), and redox potential were the most significant geochemical variables linked to the microbial functional gene structures. This study indicates that microbial functional genes could be very useful for tracking microbial community structure and dynamics during bioremediation (Liang et al. 2012).

In another study, GeoChip 3.0 was used to study the functional gene diversity and structure of groundwater microbial communities in a shallow landfill leachate-contaminated aquifer in Norman, OK. Samples were taken from eight wells at the same aquifer depth immediately below a municipal landfill or along the predominant downgradient groundwater flowpath. The results showed that functional gene richness and diversity immediately below the landfill and the closest well were considerably lower than those in downgradient wells and that landfill leachate impacted the diversity, composition, structure, and functional potential of groundwater microbial communities as a function of groundwater pH and concentrations of sulfate, ammonia, and dissolved organic carbon (Lu et al. 2012).

In 2010, the Deepwater Horizon oil spill occurred in the Gulf of Mexico. GeoChip 4.0 was used to examine the functional composition and structure of water microbial communities from the oil plume and control sites. The results indicated that the water microbial community composition and structure were dramatically altered in deep-sea oil plume samples. A variety of functional genes involved in both aerobic and anaerobic hydrocarbon degradation were highly enriched in the plume compared with outside the plume, indicating a great potential for intrinsic bioremediation or natural attenuation in the deep sea. Various other microbial functional genes that are relevant to C, N, P, S, and iron cycling, metal resistance, and bacteriophage replication were also enriched in the plume. Overall, this study suggests that indigenous microbial communities could have a significant role in biodegradation of oil spills in deep-sea environments (Hazen et al. 2010).

Other Environments

GeoChips were also used to analyze microbial communities from other habitats/ecosystems, including various contaminated sites (e.g., chromate-contaminated water, U-contaminated sediments, polychlorinated biphenyl- and arsenic-contaminated soils), extreme environments (e.g., acid mine drainage, hypersaline lakes, deep-sea basalts, deep-sea hydrothermal vents), bioleaching systems, and bioreactors as well as the human microbiome (He et al. 2011, 2012b).

Summary

Although GeoChip technology has been demonstrated to be specific, sensitive, and quantitative and applied to analyze microbial communities from different habitats, some key issues and challenges still remain, including probe coverage, specificity, sensitivity, quantitative capability, nucleic acid quality, the detection of microbial community activity, and challenges by high-throughput sequencing technologies. It should be noted that probe coverage on GeoChip is relatively low compared to the availability of functional gene sequences in databases, especially for earlier versions of FGAs. One of the reasons is that some sequences do not have specific probes based on the availability of sequence databases and software. Also, GeoChip probe sets need continuous updates to reflect the current status of functional gene sequence information.

DOI 10.1007/978-1-4614-6418-1_760-4

Critical issues with GeoChip design and detection are specificity, sensitivity, and quantitative capability, which are especially important since many gene variants within each environmental sample are unknown. Array specificity is controlled by probe design and hybridization conditions. A novel microarray probe design software tool, *CommOligo* (He et al. 2012a), and its improved versions were 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 low percentages of false positives (0.002–0.025 %) were observed (He et al. 2007; He et al. 2010a). GeoChip hybridizations are generally performed at 42–50 °C with 50 % formamide. Sensitivity is another major concern since many gene variants are expected to be low abundant in environmental samples. The current level of sensitivity for oligonucleotide arrays using environmental samples is approximately 50–100 ng or 10^7 cells, or approximately 5 % of the microbial community, providing a coverage of only the most dominant community members. Several strategies have been utilized to increase sensitivity. For example, with WCGA and WCRA approaches, the sensitivity of GeoChip hybridization could increase to 10 fg. Also, array surface modifications, a decrease of hybridization solution, and the use of new labeling techniques could increase GeoChip detection sensitivity (He et al. 2011, 2012a). 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. However, this relationship can be affected by sequence divergence (i.e., the more divergent the sequence, the lower the signal intensity). Therefore, two strategies are used to improve quantitative ability: mismatch probes and using relative comparisons across samples rather than absolute comparisons (He et al. 2012a).

The quality and quantification of environmental nucleic acids are one of the most important for successful GeoChip hybridization and reliable data generation. DNA with large fragments and minimal amounts of contaminants are especially important when samples need to be amplified using WCGA. Accurate measurement of DNA yields is also important, so quantification should be based on double-strand DNA (dsDNA) measurement (e.g., PicoGreen) rather than via absorbance. 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. To monitor microbial activity, mRNA should be used. However, since mRNA is easily degraded with rapid turnover, usually has a low abundance, and has a small proportion of the total RNA, improved RNA extraction methods are necessary to use environmental RNA for GeoChip analysis. Alternatively, other techniques, such as stable isotope probing (SIP), enzyme activity, metaproteomic analysis, and metabolite assays, may be used to study the functional activity and ecosystem functions of microbial communities.

High-throughput sequencing technologies (e.g., 454, Illumina) are available for microbial community analysis, which challenge GeoChip technologies. However, although these sequencing-based technologies can discover novel sequences, it can be expensive to do in-depth shotgun sequencing of a community. In addition, it suffers from lack of appropriate conserved primers for many target genes. Also, sequencing-based technologies have a disadvantage of random sampling, and/or under-sampling, making it difficult to compare different samples, while microarray-based technologies have a defined probe set, which is good for community comparisons (He et al. 2012b). Therefore, due to the unique features and advantages and disadvantages of both microarray-based and sequencing-based technologies, it is preferable that they be used complementarily for microbial community analysis in order to address fundamental questions in microbial ecology and environmental biology.

DOI 10.1007/978-1-4614-6418-1_760-4

Acknowledgments

This work conducted by ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) (▶ http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the US Department of Energy under Contract No. DE-AC02-05CH11231 and by the Oklahoma Applied Research Support (OARS), Oklahoma Center for the Advancement of Science and Technology (OCAST), State of Oklahoma, through AR11-035 and AR062-034.

References

Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science. 2005;309:1387–90.

Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, et al. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science. 2010;330:204–8.

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, et al. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 2007;1:67–77.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL, et al. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME J. 2010a;4:1167–79.

He Z, Xu M, Deng Y, Kang S, Kellogg L, Wu L, et al. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO_2. Ecol Lett. 2010b;13:564–75.

He Z, Van Nostrand JD, Deng Y, Zhou J. Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities. Front Environ Sci Engin China. 2011;5:1–20.

He Z, Deng Y, Zhou J. Development of functional gene microarrays for microbial community analysis. Curr Opin Biotechnol. 2012a;23:49–55.

He Z, Van Nostrand JD, Zhou J. Applications of functional gene microarrays for profiling microbial communities. Curr Opin Biotechnol. 2012b;23:460–6.

Li X, Deng Y, Li Q, Lu C, Wang J, Zhang H, et al. Shifts of functional gene representation in wheat rhizosphere microbial communities under elevated ozone. ISME J. 2013;7(3):660–71.

Liang Y, Van Nostrand JD, N'Guessan LA, Peacock AD, Deng Y, Long PE, et al. Microbial functional gene diversity with a shift of subsurface redox conditions during in situ uranium reduction. Appl Environ Microbiol. 2012;78:2966–72.

Lu Z, He Z, Parisi VA, Kang S, Deng Y, Van Nostrand JD, et al. GeoChip-based analysis of microbial functional gene diversity in a landfill leachate-contaminated aquifer. Environ Sci Technol. 2012;46:5824–33.

Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl Environ Microbiol. 2004;70:4303–17.

Torsvik V, Ovreas L, Thingstad TF. Prokaryotic diversity – magnitude, dynamics, and controlling factors. Science. 2002;296:1064–6.

DOI 10.1007/978-1-4614-6418-1_760-4

Van Nostrand JD, Wu L, Wu W-M, Huang Z, Gentry TJ, Deng Y, et al. Dynamics of microbial community composition and function during in situ bioremediation of a uranium-contaminated aquifer. Appl Environ Microbiol. 2011;77:3860–9.

Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci USA. 1998;95:6578–83.

Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol. 2001;67:5780–90.

Xue K, Wu L, Deng Y, He Z, Van Nostrand J, Robertson PG, et al. Functional gene differences in soil microbial communities from conventional, low-input, and organic farmlands. Appl Environ Microbiol. 2013;79:1284–92.

Yang Y, Wu L, Lin Q, Yuan M, Xu D, Yu H, et al. Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland. Glob Chang Biol. 2013;19:637–48.

Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X. Functional molecular ecological networks. mBio. 2010;1(4):e00169.

Zhou J, Xue K, Xie J, Deng Y, Wu L, Cheng X, et al. Microbial mediation of carbon-cycle feedbacks to climate warming. Nat Clim Chang. 2012;2:106–10.