Molecular Microbial Ecology Manual, Second Edition **1.8.11**: 1–19, 2004.

Oligonucleotide-based functional gene arrays for analysis of microbial communities in the environment

SONIA M. TIQUIA*, SONG C. CHONG, MATTHEW W. FIELDS^ and JIZHONG ZHOU

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
Present address: Department of Natural Sciences, University of Michigan, Dearborn, MI 48128, USA
^*Present address: Department of Microbiology, Miami University, Oxford, OH 45056*

Introduction

The inability to assess microbial diversity and structure rapidly constrains microbial ecologists from the establishment of testable theories as well as any holistic assessment of the ecosystem. In the past several years, DNA microarray technology has played an important role in gene expression studies of individual microorganisms [2, 4, 16], however, microarray technology has not been readily used for high-throughput analysis of microbial communities and/or gene distribution and expression in environmental samples [29, 32]. Environmental samples pose a variety of obstacles, including sequence divergence, the existence of contaminants (i.e., humic materials, organic contaminants, metals), low biomass, and quantification.

DNA microarrays show great promise as a revolutionary tool for large-scale parallel analysis of microbial community structure and activities [7, 29]. Recently, microarray technology has been extended to studies of microbial communities in the environment [7, 22, 25, 29; @@@ see chapter 3.4]. Several types of microarrays have been developed to monitor microbial community dynamics in environmental samples, including functional gene arrays (FGAs), community genome arrays (CGAs), and phylogenetic oligonucleotide arrays (POAs) [31]. FGAs contain genes, or portions thereof, encoding key enzymes involved in various ecological processes such as carbon fixation, nitrification, denitrification, and sulfate reduction. Both DNA fragments and oligonucleotides (oligos) derived from functional genes can be used for FGA construction [29]. To avoid confusion, the former is referred to as DNA-based FGAs, whereas the latter is referred to as the oligonucleotide-based FGAs. FGAs not only detect the existence of particular gene sequences, but can also be useful in studying functional activities of microbial communities in natural environments [29, 31]. CGAs are constructed using genomic DNA isolated from individual microorganisms in monoculture and can be used to describe microbial community dynamics with reference to the community's cultivable component [31]. POAs are constructed with oligonucleotides (approximately 20-mers) based on SSU rRNA genes, and can be used for phylogenetic analyses of microbial community composition and structure in environmental samples [31]. Depending on the objective of the experiment, targets (molecules to be detected in the a given sample) may be PCR products, genomic DNA, total RNA, mRNA, cDNA, plasmid DNA, or oligonucleotides.

Recently, oligonucleotide microarrays containing probes longer than 40 bases have been evaluated and used for whole genome expression studies [10, 15]. Several studies have demonstrated that oligonucleotide-based microarrays can be advantageous over DNA-based microarrays in terms of construction. Due to the better specificity and easier construction, oligonucleotide arrays provide an important alternative, array-based approach for monitoring gene expression. We have recently developed and evaluated a 50-mer-based FGA containing probes for genes involved in nitrification [ammonia monooxygenase (*amoA*)], denitrification [nitrite reductase (*nirS* and *nirK*)], nitrogen fixation [nitrogenase (*nif*H)], methane oxidation [methane monooxygenase (*pmoA*)] and sulfate reduction [sulfite reductase (*dsr*A/B) genes] for environmental applications. Here, we will briefly review the performance of a FGA and provide detailed protocols for microarray construction, application, and data analysis.

Specificity, sensitivity, and quantitation potential of 50-mer FGAs

Substantial technical hurdles related to specificity, sensitivity, and quantification need to be overcome in order to facilitate the efficacy of the prototype FGA with complex environmental samples. We have constructed 50-mer oligonucleotide arrays containing 763 gene probes involved in nitrogen cycling and sulfate reduction. All of the probes on the arrays have less than 85% similarity. Our results have demonstrated that the developed 50-mer FGA is potentially specific, sensitive, and quantitative for environmental applications, and could be useful in monitoring the composition, activities and dynamics of microorganisms in environmental samples.

To understand the taxonomic resolution of the 50-mer based array hybridization, we have compared sequence similarities of *dsr*AB, *nir*S, *nir*K, *nif*H, *amo*A, and *pmo*A genes from pure cultures in terms of taxonomic classification. Our results revealed that at the strain level, the average sequence similarity for *amo*A was 99%, whereas it was lower (91–95%) for the other five functional gene groups. At the species level, the average similarity was between 70 and 82%. The similarity further decreased at the genus level (67–75%) and was 57–66% at the family or higher level. Under the hybridization conditions of 50 °C with 50% formamide, genes having <86–90% sequence identity were differentiated (Fig1A). These results indicated that species-level resolution could be achieved with the particular probes tested with the designed 50-mer FGAs.

With the 50-mer oligonucleotide arrays, *dsr*B, *nir*S, *nir*K, *nif*H, *amo*A, and *pmo*A genes could be detected with 8 ng of pure genomic DNA using our optimized

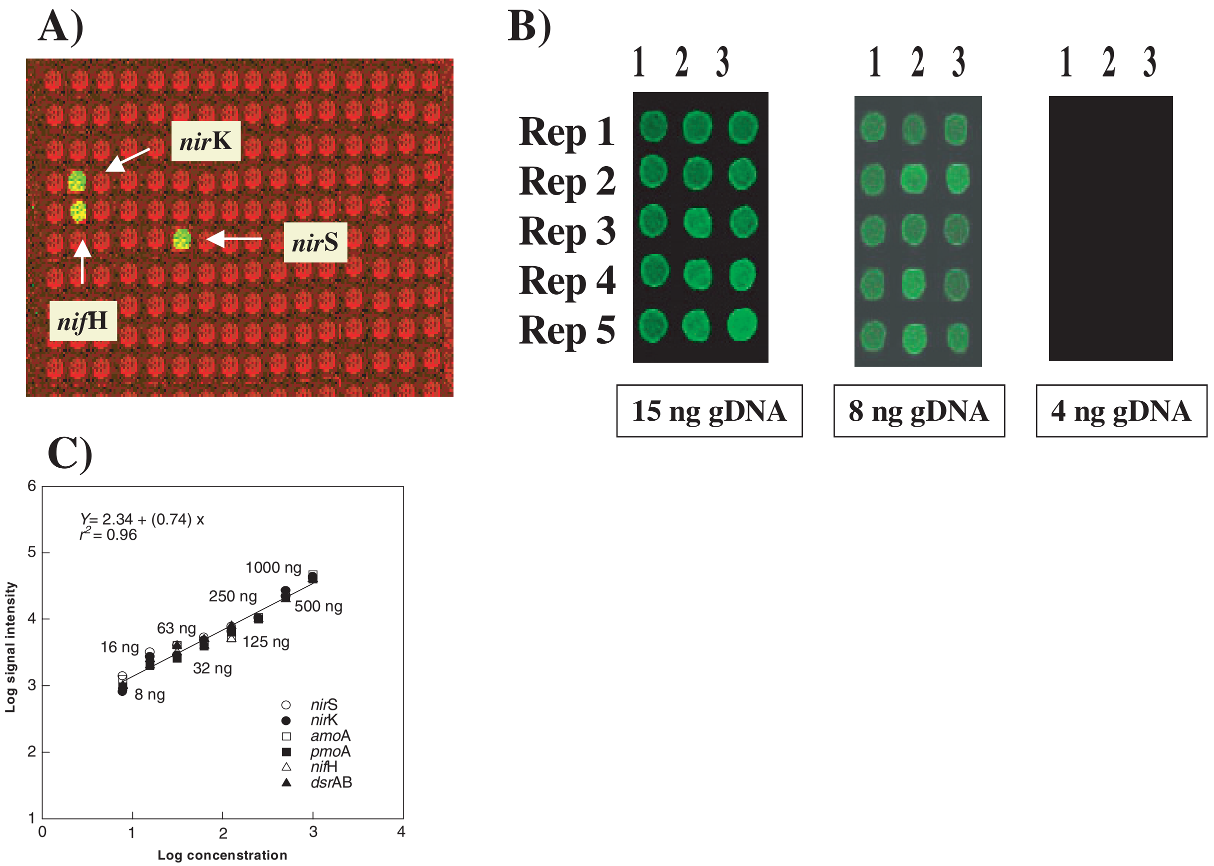


Figure 1. Specificity, sensitivity, and quantitation potential of 50-mer oligonucleotide microarrays. (A) Image demonstrating specific target signals for specific *nir* H, *nir* K and *nir* S genes. (B) Detection sensitivity of the oligonucleotide microarrays. A small array was constructed with 5 replicates of each probe. The probes were as follows: Lane 1: *Desulfovibrio vulgaris dsr* B ; Lane 2: *Pseudomonas stutzeri nir* S ; Lane 3: Pseudomonas sp G-179 *nir* K (C) Relationship between hybridization signal intensities and different DNA target concentrations.

protocol (Fig 1B). Our results also showed that the hybridization signal and the amount of genomic DNA correlated well (Fig. 1C). This result is consistent with those found by Relogio et al. [20] using oligonucleotide probes and by Wu et al. [29] using a DNA-based FGA. Fairly good hybridizations were obtained with 50-mer FGAs when using 5 μg of bulk community DNA from marine sediments. Because the probes in the developed 50-mer arrays were derived from sequences from a variety of environments ranging from marine sediments, soils, salt marshes, and contaminated and non-contaminated ground waters, the developed arrays should represent diverse genes involved in these biogeochemical processes. In addition, the arrays should be useful in monitoring the composition, dynamics and activities of microbial populations involved in these functional processes across different natural environments.

The following sets of protocols are intended to serve as a basic introduction to microarray construction and microarray experimental design. The four fundamental steps required in oligonucleotide-based FGA construction and experimentation are: (1) Oligonucleotide microarray construction, (2) Labeling and quantitation of labeled DNA, (3) Hybridization, and (4) Image processing and data analysis. The schematic diagram for these steps is illustrated in Fig. 2. It is our hope that the methods presented here will serve as an initial and useful tool to study the functional gene profiles of microbes in diverse environmental samples.

Procedures

Protocol 1. Oligo microarray fabrication

a) Oligo design

The 50-mer FGAs can be constructed with probes based upon sequences recovered from a wide variety of environments, designed to represent the known microbial population diversity involved in the biogeochemical processes of interest. Also, sequences can be retrieved from public databases such as GenBank, EMBL and SwissProt. For the design of such 50-mer oligonucleotide probes, we use a modified version of PRIMEGENS (*http://compbio.ornl.gov/structure/primegens/*), which was originally developed for designing gene-specific primers for whole genome cDNA microarrays [30]. The software initially compares each gene sequence against the entire sequence database using BLAST, and produces an alignment with the other sequences that have more than the desired threshold sequence similarity (e.g. 85%) using dynamic programming. Based on the global optimal alignments, segments of 50 bp oligonucleotides with less than the threshold identity

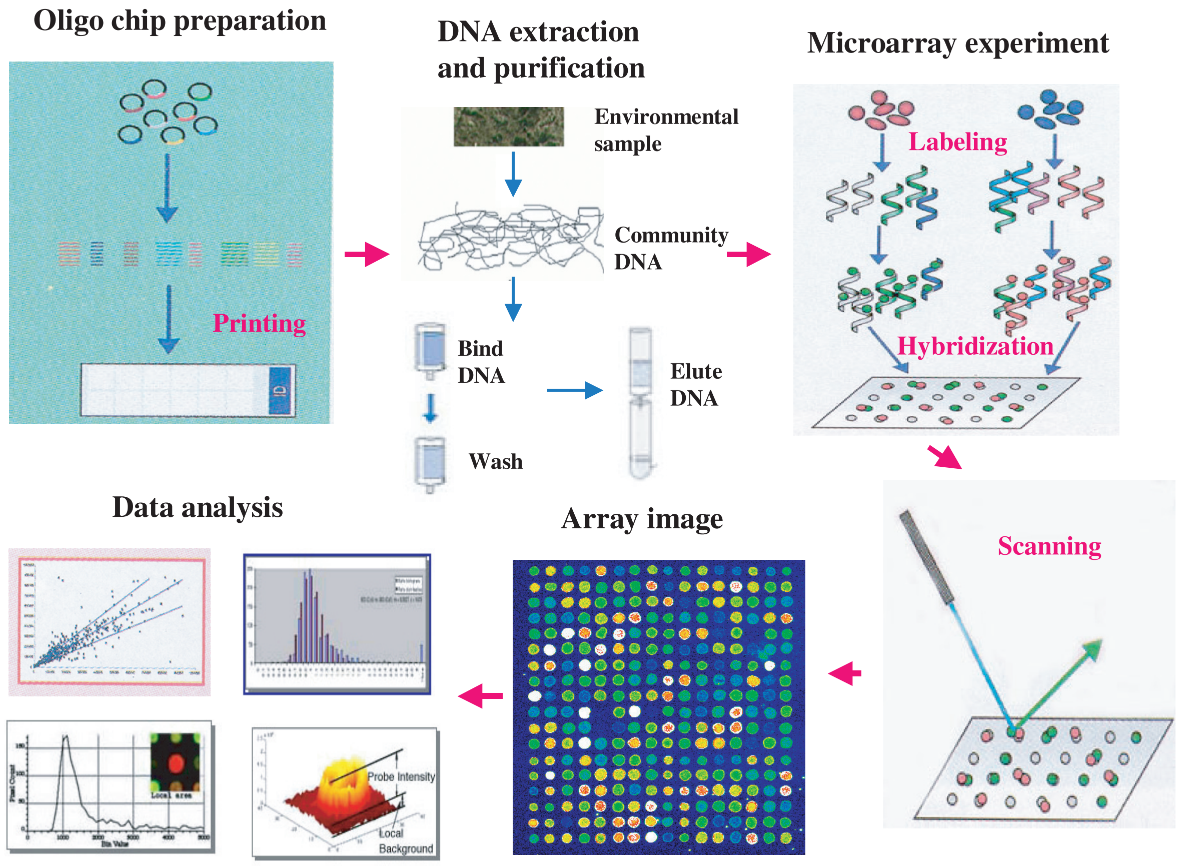


Figure 2. Schematic diagram of microarray construction and experiments.

to the corresponding aligned regions of any BLAST hit sequences are selected as potential probes. Among these identified potential probes, a final probe is selected by considering the GC content, melting temperature, and self-complementarity. Outputs of the designed probes are imported into Excel and a pivot table is constructed containing the sequence information of each probe. There are several other free and commercial software packages for designing oligonucleotides. OligoArray [21] is a free software that designs gene-specific oligonucleotides for genome-scale microarray construction. Array Designer (Biosoft International, Palo Alto, CA) and Sarani (Strand Genomics, Burlingame, CA) are commercial softwares for automatic large-scale design of optimal oligonucleotide probes for microarray experiments.

Oligos are synthesized at the desired scale at a final concentration of 100 pmol ul^{-1} without any modification, and diluted to 30–40 pmol μl^{-1} with 50% DMSO. Thereafter, oligonucleotides are printed onto aminosilane-coated glass slides such as SuperAmine (Telechem International, Inc., Sunnyvale, CA, Corning Incorporated, Corning, NY). SuperAmine slides contain covalent amine groups that allow stable attachment of nucleic acids.

b) Oligo array printing

1. Prepare printing oligo solutions to a final concentration of 50 pmol μl^{-1}using 50% DMSO in a 384-well printing plate (5 μl probe and 5 μl DMSO)
2. Cover the plate with the plastic lid and mix in an orbital shaker at 700 rpm for 3 min.
3. Spin the printing plate using a centrifuge equipped with a rotor for microtitre plates at 500 rpm for 5 min.
4. Setup the array printer (PixSys 5500 printer; Cartesian technologies, Inc. Irvine, CA) and print slides according to the manufacturer's protocol. The ideal relative humidity should be between 40 and 60% at room temperature (20–25 °C). The spot size should be approximately 100 to 150 μm, with 200 to 500 μm spacing distance using split pins from Telechem.
5. Allow the slides to dry for 2 h prior to UV cross-linking.

Reproducibility is one of the most critical requirements for microarray fabrication. For reliable and reproducible data, the uniformity of individual spots across the entire array is crucial for simplifying image analysis and enhancing the accuracy of signal detection. Several factors will affect the uniformity of spots, including array substrate, pins, printing buffer, and environmental controls. For instance, significant variations could be caused by pin characteristics due to the mechanical difference in pin geometry, pin age and sample solutions. Movement of the pin across the surface in the XY direction may cause the tip to bend. Tapping the pins on the surface may result in deformation of the pin tips. Also, dragging the pin tip across the surface may cause clogging of the pin sample channel. Therefore, great care is needed in handling pins. Pins should be cleaned with an ultrasonic bath for 5 minutes after each printing.

Environmental conditions have significant effects on spot uniformity and size [8]. Humidity control is crucial to prevent sample evaporation from source plates and the pin channel during the printing process. Sample evaporation can cause changes in DNA concentration and viscosity, affecting the quality of the deposited DNA. Also, reducing evaporation can help the spotted volume of DNA to have more time to bind at equal rates across the entire spot, resulting in more homogeneous DNA spots. As a result, DNA spots of increased homogeneity will be obtained [5]. The printing buffer is also critical for obtaining homogeneous spots. With the widely used saline sodium citrate (SSC) buffer, the spot homogeneity as well as binding efficiency is often poor. Using a printing buffer containing 1.5 M betaine improves spot homogeneity as well as binding efficiencies [5]. This is because betaine increases the viscosity of a solution and reduces the evaporation rate. More uniform spots can also be obtained with a printing buffer containing 50% DMSO (dimethyl sulfoxide) [8, 29].

c) UV cross-linking and slide processing.

At the end of the print, remove slides from the printer, label each slide with an identifier and the slide number by writing on the edge of the slide with a diamond pen and place slides in a dust-free slide box. It is useful to etch a line, which outlines the printed area of the slide, onto the first slide. This serves as a guide to locate the area after the slides have been processed.

1. Expose the slides, printed face up, to a 80 mJ dose of ultraviolet irradiation for 30 s.
2. Wash slides at room temperature first with 0.1% SDS and then with water:

		Washing time
Washing 1	0.1% SDS	4 min
Washing 2	water	2 min

3. Transfer slides to a 10-slide glass rack and place the rack into a glass tank.
4. Remove the slides and spin using a centrifuge equipped with a rotor for microtitre plates at 500 rpm for 5 min to dry.
5. Transfer the slides to a clean, dust-free slide box and let it stand overnight prior to hybridization.

Protocol 2. Labeling and quantitation of target DNA

Once the oligo microarrays are printed, targets are prepared for hybridization. For community analysis of environmental samples, genomic DNAs from pure cultures or environmental clones are normally used as target, and human genes as controls [29]. Successful application of microarrays for microbial community analysis relies on the effective recovery of nucleic acids from the environment. Hurt et al. [11] and Zhou et al. [33] pointed out some criteria for ideal recovery of DNA or RNA from environmental samples: (i) The nucleic acid recovery efficiency should be high and not biased so that the final nucleic acids are representative of the total nucleic acids within the naturally occurring microbial community; (ii) The DNA should be of sufficient purity for reliable hybridization; (iii) The extraction and purification protocol should be robust and reliable. The DNA extraction and purification protocol described by Hurt et al. [11] fulfills the above criteria. Of course it should be possible to substitute other protocols that meet these criteria.

a) Labeling

Random primer and PCR amplification labeling with Cy3 or Cy5 fluorescent dyes are the most common means used for target detection in environmental samples [29]. Random primer labeling with Klenow fragment of DNA polymerase I is particularly useful for labeling genomic DNA fragments. Targets can also be labeled by PCR using gene-specific primers. PCR labeling targets using gene-specific primers is particularly important for increasing detection sensitivity.

Random priming labeling method:

1. In a 0.2 ml PCR tube combine:
   a. 2–5 μg purified community DNA (in 10 μl RNase-free water).
   b. 20 μl (750 ng μl^{-1}) random octamer primers (Invitrogen # Y01393)
2. Mix them well and denature at 99.9 °C for 5 min.
3. Place immediately on ice for 5 minutes.
4. Centrifuge the mixture for 3 minutes at maximum speed.
5. In a 1.5 ml microcentrifuge tube, combine
   a. 2.5 μl dNTP's (5 mM dATP, dTTP, dGTP and 2.5 mM dCTP)
   b. 1 μl (1 mM) Cy3 or Cy5 dCTP
   c. 1.5 μl (40 U μl^{-1}) Klenow fragment (Invitrogen # Y01396)
   d. 1.25 μl DTT (Invitrogen #Y00147)
   e. d. 13.75 μl DNase- and RNase-free water
6. Add this mixture to the 0.2 ml PCR tube that contains DNA (total volume of the mixture = 50 μl).
7. Mix well and incubate at 37 °C for 6 h or overnight.
8. After incubation, boil the mixture at 100 °C for 5 min and chill on ice.
9. Purify labeled target DNA using QIAquick columns according to the manufacturer's instructions (Qiagen, Valencia, CA).

PCR amplification specific labeling method

1. In a PCR tube, combine the following and make up to 30 μl volume using RNAse-free water:
   10 pg of plasmid containing the desired target gene
   20 pmol PCR primers (specific primers for gene of interest)
   25 mM of dATP, dCTP, dGTP, 15mM dTTP (New England Biolabs),
   10 mM aminoallyl-dUTP (Sigma, St. Louis, MO)
   0.5 U *Taq* DNA polymerase
2. Place PCR mixture in a thermocyler using the following amplification conditions: 1 cycle at 80 °C for 30 s, 94 °C for 2 min followed by 25 cycles of 94 °C for 30 s, 57 °C for 1 min, and 72 °C for 1 min, with a final extension step at 72 °C for 7 min. Note that the annealing temperature may vary depending on primers used.
3. Purify PCR product using QIAquick columns (Qiagen, Valencia, CA).
4. Dry PCR product in speed-vac for 30 min and resuspend in 4.5 μl

0.1 M carbonate buffer (pH 9.0).

5. Mix the solution with (4.5 μl) N-hydroxy succinimide esters Cy3 or Cy5 (NHS-Cy3 or Cy5; Amersham Pharmacia Biotech, Piscataway, NJ) and incubate in the dark for 1 h.
6. After incubation, add 35 μl of 100 mM NaOAC (pH 5.2).
7. Purify labeled target PCR products using QIAquick columns (Qiagen, Valencia, CA).

b) Quantifying the amount and specific activity of Cy-labeled DNA targets

Labeling is a critical step for obtaining high-quality microarray data. The experimental problem most often encountered is that microarray hybridization signal varies greatly from time to time. In many cases, poor hybridization signal is a result of poor dye incorporation. Decreased dye incorporation (<1 dye per 100 nucleotides) gives unacceptably low hybridization signals. However, studies have shown that very high dye incorporation (e.g., >1 dye molecules per 20 nucleotides) is also not desirable, because high-dye incorporation significantly destabilizes the hybridization duplex [28][1]. Thus, it is important to measure dye incorporation efficiency prior to hybridization. The specific activity of dye incorporation can be determined by measuring the absorbance at wavelengths of 260 nm and 550 nm for Cy3 and 260 nm and 650 for Cy5. A suitable labeling reaction should have a 8–15 A_{260}/A_{550} ratio for Cy3 and 10–20 A_{260}/A_{650} for Cy5.

1. Use a spectrophotometer to quantify the OD at 550 for Cy 3 and OD 650 for Cy5. Also, measure OD at 230, 260 and 280 to assess purity.
2. Determine the OD of 1 μl of labeled DNA OD using a NanoDrop™ ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Montchanin, DE).
3. Calculate the amount of DNA and as well as the specific activity of the labeled DNA. The specific activity is calculated as follows:

$$\text{Specific activity} = \frac{\text{amount of target DNA} \times 1000}{\text{pmole of dye incorporated} \times 324.5}$$

4. Dry in speed-vac (no heat) for 1–2 h. Do not use high heat or heat lamps to accelerate evaporation. The fluorescent dyes could be degraded[1].

Note

1. Corning (*http://www.corning.com/cmt*) recommends not using labeled target if the specific activity is more than 75. Check specific activity of labeled target before use. Specific activity should be 1 dye molecule per 25 to 50 nucleotides for good hybridization.

Protocol 3. Hybridization

Since microarray hybridization is generally performed in the absence of mixing, the hybridization solution should be mixed well so that the labeled targets are evenly distributed on the array surface to obtain optimal target-probe interactions across the entire microarray. Otherwise, the availability of the labeled target molecules to the arrayed spots could be significantly different across the microarray surface. Labeled target molecules may be depleted in some areas, yet abundant in others. As a result, significant differences in signal intensity could be observed. Non-uniform hybridization is a common problem associated with microarray experiments. Thus, it is essential to have replicate spots well separated on a slide. It is also imperative to determine the volume of hybridization solution required. An array covered by a 22 × 22 mm glass LifterSlip (Erie Scientific company, Portsmouth, NH) coverslip will require ∼15 μl of hybridization solution. The volume of the hybridization solution is critical. When too little solution is used, it is difficult to place the coverslip without introducing air bubbles over some portion of the arrayed oligos. If the coverslip is bowed toward the slide in the center, there will be less labeled DNA in that area and the hybridization will be non-uniform. When too much volume is applied, the coverslip will move easily during handling, which may lead to misplacement relative to the arrayed oligos, and non-hybridization in some areas of the array may occur.

For 15 μl of hybridization solution, combine the following components:

Hybridization buffer:	Volume	Final concentration
a. Labeled DNA dissolved with RNase-free water	3.3 μl	
b. Formamide	7.5 μl	50%
c. 20 × SSC	2.5 μl	3.33 ×
d. 10% SDS	0.5 μl	0.33%
e. Herring sperm DNA (10mg/ml)	1.2 μl	10.2 μg

1. Heat the hybridization solution at 95 °C for 5 min in a thermocycler, cool quickly to 25 °C, and spin down at 14,000 × g for 5 min.
2. Preheat microarray slide for 20 minutes at 50 °C.
3. Deposit the hybridization (15 μl) solution directly onto the immobilized DNA prior to placing a cover slip (6.25 mm × 8 mm) over the array, avoiding bubble formation. It is helpful to practice this operation with buffer and plain slides before attempting actual samples.
4. Put the slide in the hybridization chamber.
5. Dispense 20 μl of 3 × SSC solution into the hydration wells on both sides.
6. Close the hybridization chamber. Make sure the seal is formed along the O-ring.
7. Incubate the chamber in a 50 °C water bath for 12–15 h or overnight.

One common problem in microarray hybridization is the quality of fluorescent dyes. The labeling efficiency and hybridization can vary significantly from batch to batch, especially for Cy5. It is very important to use fresh reagents to achieve highly sensitive detection [29].

Post-hybridization wash

1. Place the slides, with the coverslips still affixed, in a prewarmed washing buffer(2 × SSC and 0.1% SDS) and allow the coverslips to fall from the slide.
2. Place the slides in a prewarmed washing buffer (2 × SSC and 0.1% SDS) and wash for 5 min with gentle shaking. Repeat this wash once.

Place the slides to a fresh jar filled with 0.1 ×

Image acquisition and processing

1. Scan the slide initially at a low resolution of 50 μm to obtain a quick display image and then at 5–10 μm using for instance the ScanArray 5000 System (GSI Lumonics, Watertown, MA). The emitted fluorescent signal is detected by a photomultiplier tube (PMT) at 570 nm (Cy3) or 670 nm (Cy5). The percentages of laser power and PMT used should be appropriately selected based on hybridization signal intensity so that the signals for most of the spots are not

saturated. The signal should also be balanced during scanning by using a higher PMT setting for the dye with weaker signal to allow detection of more spots with low signal intensity.

2. Save the scanned display as a 16-bit TIFF and BMP file and quantify the intensity of each spot. Many methods are available for resolving irregularities in spot location, size and shape, as well as contamination problems [34] to accurately estimate spot intensities. A variety of commercial and free software, such as ImaGene$^{\text{TM}}$ from BioDiscovery (Los Angeles, CA), QuantArray$^{\text{TM}}$ from GSI Lumonics, and the software on Axon GenePix$^{\text{TM}}$ systems [1] can be used for microarray image processing. Typically, a user-defined gridding pattern is overlaid on the image and the areas defined by patterns of circles are used for spot intensity quantification.
3. Assess spot quality and reliability, and perform background subtraction of the microarray data. Because of the inherently high variation associated with array fabrication, hybridization, and image processing, the intensity data for some spots may not be reliable. Thus, the first step in data processing is to assess the quality of spots and to remove unreliable, poor spots prior to data analysis. Also, in many cases, because of slide quality, background and contamination, the quality of data can vary significantly among different slides [24]. Be sure to subtract local background for each spot and then flag and remove poor quality spots from the data set prior to further analysis.
4. Compute signal-to-noise ratio (SNR) for each spot to discriminate true signals from noise [26]. Generally, a SNR larger than 3 is considered as positive signal. The SNR ratio is calculated as follows:

$$\text{SNR} = \frac{\text{Signal mean} - \text{Background mean}}{\text{Background standard deviation}}$$

Remove outlying spots (outliers) prior to data analysis using ArrayStat$^{\text{TM}}$ (Imaging Research, Inc., Ontario, Canada). Outliers are extreme values in a distribution of replicates. Outlying spots could be caused by uncorrected image artifacts such as dust or by the factors undetectable by image analysis such as cross-hybridization. Thus, removal of outlying spots is an important step for pre-data analysis. However, distinguishing outliers is very challenging, because there is no general definition for outliers.

Note

Besides Imagene software, there are other software packages available for image processing, spot identification, quantitation, and normalization. These imaging softwares include GenPix Pro (Axon Instruments, Union City, CA), Array Pro (Media Cybernetic, Carlsbad, CA), Quant Array (Packard Biosciences, Boston, MA), and TIGR Spot Finder (The Institute of Genomic Research TIGR, Rockville, MD).

Data analysis

Microarray experiments generate large data sets, and a major challenge in microarray experiments is to extract meaningful information out of the data. One of the key goals for microarray analysis is to identify genes that give statistically significant differences in signal intensity across treatments. Many different statistical methods have been used for analyzing microarray data, such as similarity measurements [3, 2–13], principal components analysis [9, 19] cluster analysis [6] and self organizing maps (SOM) [23]. For similarity comparisons of microarray data, two approaches are generally used for quantifying the relationships among different genes. One approach is to use Euclidean distance, which is defined as the square root of the summation of the squares of the differences between all pair-wise comparisons [13]. The other approach is the Pearson correlation coefficient, which is ideal for identifying profiles with similar shape [3, 12]. PCA provides an easy way of identifying outliers in the data such as genes that behave differently than most of the genes across a set of experiments [9, 19]. It also can be used to visualize clusters of genes that behave similarly across different experiments. Cluster analysis has been used to identify groups of genes, often called clusters, that have similar expression profiles [6] (note that we do not address expression in the protocols described in this chapter). Subsequently, the clusters, and genes within them, can be examined for commonalities in functions, as well as sequences, for better understanding of how and why they behave similarly. Cluster analysis can also help establish functionally related groups of genes and can predict the biochemical and physiological roles of functionally unknown genes [27]. SOMs are a more robust and accurate method for grouping large data sets [14]. In this analysis, the data points are mapped onto a grid, and the positions of the representative points are iteratively relocated in a way that each center has one representative point. Clusters close to each other in the grid are more similar to each other than those further apart.

There are several software packages available to facilitate statistical analyses of array data. For instance, ArrayStat (Imaging Research, Inc. Ontario, Canada) allows analysis of statistical significance, p-values, and standard deviation of microarray data. GeneSpring (Silicon Genetics, CA) permits the analysis of array data for scatter plot, cluster analysis, principal component analysis (PCA) and self-organizing maps (SOM). Free-computer programs (i.e. CLUSTER and TREEVIEW) that can ascertain hierarchical relationships of different spots are also available [6].

Application of the method

To evaluate the potential applicability of the 50-mer FGAs for microbial community analysis, 5 μg of bulk community DNA derived from 5 g of Gulf of Mexico (TX) marine sediment (top 1 cm) was labeled with Cy5 using the random primer labeling method and hybridized with an oligo array containing genes involved in nitrogen cycling, sulfate reduction and carbon cycling. The hybridization image indicated that the 50-mer oligonucleotide arrays hybridized reasonably well with the DNAs from marine sediment (Fig. 3). The abundant genes included those encoding nitrogenage (nifH), dissimilatory sulfite reductase (dsr A/B), ammonia monooxygenase (amoA), methane monooxygenase, and nitrite reducatse (nir S/K) (Fig 4). These results indicated that oligonucleotide microarray technology is potentially useful in monitoring the composition, structure, activities and dynamics of microbial populations involved in these functional processes. However, the application of the 50-mer oligo arrays for environmental samples is still being improved. More rigorous tests within the context of environmental application and validation of the microarray results with other independent methods are needed. The usefulness of the 50-mer oligo arrays should also be evaluated with diverse samples from a variety of environments, thereby addressing its usefulness across a range of ecological questions. We have demonstrated the feasibility of the approaches described in this chapter and current research involves the further validation of these methodologies and their application to a variety of environmental samples to address research questions re-

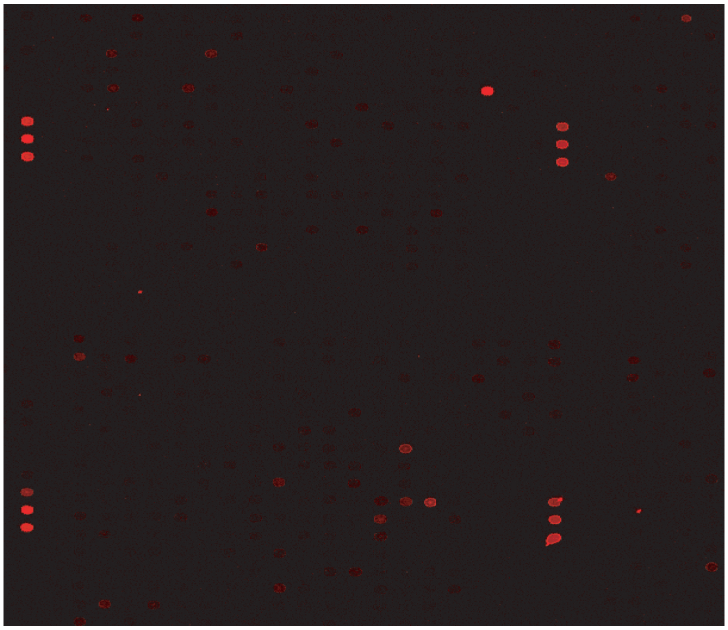
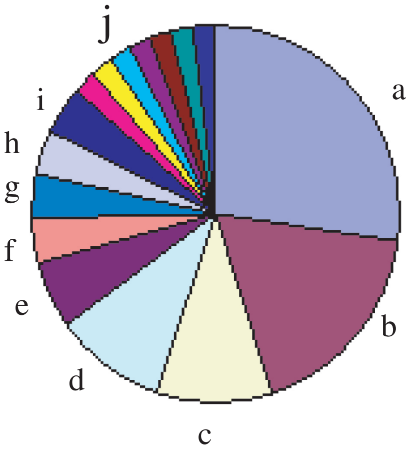


Figure 3. Microarray hybridization with a marine sediment sample. 5 ug of total DNA from marine sediment was labeled with Cy5 using random primer labeling method, and hybridized at 50 °C for overnight to a 50-mer oligonucleotide array. Only a portion of the hybridization image is shown.



a. Nitrogenase (nifH, 27%)
b. Dissimilatory sulfite reductase (dsr A/B, 18 %)
c. Ammonia monooxygenase (10%)
d. Methane monooxygenase (10%)
e. Nitrite reductase (nirS/K, 6%)
f. Endoglucanase (4%)
g. Nitric oxide reductase (4%)
h. Polyphosphate kinase (4%)
i. Xylanase (4%)
j. RuBisCO, chitinase, formyltetrahydrofolate synthetase, nitrous oxide reductase(13%)

Figure 4. Abundance of target genes within a marine sediment microbial community. The hybridization signals are treated with signal to noise ratio >3.

lated to bioremediation and nitrogen and carbon dynamics in both marine and terrestrial habitats.

Acknowledgements

This research was supported by the US DOE Office of Science as part of its Biological and Environmental Research Programs in Biotechnology Investigations-Ocean Margins program, Natural and Accelerated Bioremediation Research, and Carbon Sequestration (as part of the consortium on research to enhance Carbon Sequestration in Terrestrial Ecosystems -CSiTE). Oak Ridge National Laboratory is managed by the University of Tennessee-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

References

1. Bassett DJr, Eisen M, Boguski M (1999) Gene expression informatics-it's all in your mine. Nature Gen Suppl 21: 51–55.
2. Beliaev AS, Thompson DK, Fields MW, Wu L, Lies DP, Nealson KH, Zhou J (2002) Microarray transcription profiling of a *Shewanella oneidensis etr*A mutant. J Bacteriol 184: 4612–4616.
3. Chen Y, Bittner ML, Dougherty ER (1999) Issues associated with microarray data analysis and integration. Nature Genetics 22: 213–215.
4. DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278: 680–686.
5. Diehl F, Grahlmann S., Beier M, Hoheisel JD (2001) Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. Nucl Acid Res 29: E38.
6. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863–14868.
7. Guschin DY, Mobarry BK, Proudnikov D, Stahl DA, Rittman BE, Mitzabekov AD (1997) Oligonucleotide microarrays as genosensors for determinative environmental studies in microbiology. Appl Environ Microbiol 63: 2397–2402.
8. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J (2000) A concise guide to cDNA microarray analysis. Biotechniques 29: 548–560.
9. Hilsenbeck SG, Friedrichs WE, Schiff R, O'Connell P, Hansen RK, Osborne CK, Fuqua SAW (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. J Natl Cancer Inst 91: 453–459.
10. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai HY, He YDD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19: 342–347.
11. Hurt RA, Qui X, Wu L, Roh Y, Palumbo AV, Tiedje JM, Zhou J (2001) Simultaneous recovery of RNA and DNA from soils and sediments. Appl Environ Microbiol 67: 4495–4503.
12. Jagota A (2001) Microarray data analysis and visualization. Bioinformatics by the Bay Press, Santa Cruz, California.
13. Knudsen S (2002) A biologist's guide to analysis of DNA microarray data. John Wiley and Sons, Inc. New York.
14. Kohonen T (1997) Self-organizing maps. Second Edition. Berlin, New York.

15. Lessard IAD, Domingo GJ, Borges A, Perham RN (1998) Expression of genes encoding the E2 and E3 components of *the Bacillus stereothermophilus* pyruvate dehydrogenase complex and the stoichiometry of subunit interaction in assembly in vitro. Eur J Biochem 258: 491–501.
16. Liu Y, Zhou J, Omelchenko M, Beliaev A, Venkateswaran A, Stair J, Wu L, Thompson DK, Xu D, Rogozin IB, Gaidamakova EK, Zhai M, Makarova KS, Koonin EV, Daly MJ (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Proc Natl Acad Sci USA 100: 4191-4196.
17. Ohkubo S, Iwasaki H, Hori H, Osawa S (1986) Evolutionary relationships of denitrifying bacteria as deduced from 5S rRNA sequences. J Biochem 100: 1261–1267.
18. Purkhold U, Pommeriening-Roser A, Juretschko S, Schmid MC, Koops HP, Wagner M (2000) Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amo*A sequence analysis: implications for molecular diversity surveys. Appl Environ Microbiol 66: 5368–5382.
19. Raychaudhuri S, Stuart JM, Altman RB (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. In: Altman RB, Dunker AK, Hunter l< Lauderdale K, Klein TE (eds) Proceedings of the Pacific Symposium on Biocomputing. pp. 455–466. World Scientific Publishing Company, Inc. Singapore.
20. Relogio A, Schwager C, Richter A, Ansorge W, Valcarcel J (2002) Optimization of oligonucleotide-based DNA microarrays. Nucl Acid Res 30: e51.
21. Rouillard JM, Herbert CJ, Zuker M (2002) OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18: 486–487.
22. Small J, Call DR, Brockman FJ, Straub TM, Chandler DP (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Appl Environ Microbiol 67: 4708–4716.
23. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96: 2907–2912.
24. Tseng, GC, Oh MK, Rohlin L, Liao JC, Wong WH (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucl Acid Res 29: 2549–2557.
25. Valinsky L, Vedova GD, Scupham AJ, Figueroa A, Yin B, Hartin RJ, Chroback M, Crowley DE, Jiang T, Borneman J (2002) Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. Appl Environ Microbiol 68: 3243–3250.
26. Verdnik D, Handran S, Pickett, S (2002) Key considerations for accurate microarray scanning and image analysis. In: Shah S, Kamberova G (eds) DNA Array Image Analysis - Nuts & Bolts. pp. 83–98. DNA Press, LLC, Eagleville, PA.
27. Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL, Somogyi R (1998) Large-scale temporal gene expression mapping of central nervous system. Proc Natl Acad Sci USA 95: 334–339.
28. Worley J, Bechtol K., Penn S, Roach D, Hanzel D, Trounstine M, Barker D (2000) A Systems Approach to Fabricating and Analyzing DNA Microarrays. In: Schena M (ed) Microarray Biochip Technology. pp. 65-85. Eaton Publishing. Natick, MA.
29. Wu LY, Thompson DK, Li G, Hurt RA, Tiedjie JM, Zhou J (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol 67: 5780–5790.
30. Xu D, Li G, Wu L, Zhou J, Xu Y (2002) PRIMEGENS: A Computer Program for Robust and Efficient Design of Gene-Specific Probes for Microarray Analysis. Bioinformatics 18: 1432–1437.
31. Zhou J (2003) Microarrays for bacterial detection and microbial community analysis. Curr Opin Microbiol 6: 288-294.
32. Zhou J, Thompson DK (2002) Microarrays: application in environmental microbiology. In: G Bitton (ed) Encyclopedia of Environmental Microbiology, Vol 4. pp. 1968–1979. John and Wiley and Sons, New York.
33. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl

Environ Microbiol 62: 461–468.

34. Zhou YX, Kalocsai P, Chen JY, Shams S (2000) Information Processing Issues and Solutions Associated with Microarray Technology. In: M Schena (ed) Microarray Biochip Technology. pp. 167–200. Eaton Publishing, Natick, MA.