Metadata of the chapter that will be visualized online

Series Title	Springer Protocols Handbooks	
Chapter Title	Functional Gene Arrays for Analysis of Microbial Communities on Ocean Platform	
Chapter SubTitle		
Copyright Year	2012	
Copyright Holder	Springer Science + Business Media, LLC	

Corresponding Author	Family Name	Tiquia
	Particle	
	Given Name	**Sonia M.**
	Suffix	
	Division	Department of Natural Sciences
	Organization	University of Michigan
	Address	Dearborn, MI, USA
	Email	smtiquia@umd.umich.edu
Author	Family Name	McKindles
	Particle	
	Given Name	**Katelyn M.**
	Suffix	
	Division	Department of Natural Sciences
	Organization	University of Michigan
	Address	Dearborn, MI, USA
	Email	

Abstract	Recently there has been an increased use of microarray technology as a tool to determine the presence of functional genes in a population of hard to culture communities (e.g., soil, extreme environments). A functional gene array (FGA), or GeoChip, uses probes to screen for specific functional genes vital in biological systems such as nitrogen and carbon cycling, and has even been expanded to include aquatic conditions. However, the time delay from when the sample is taken from the ocean to evaluating the test results back in the lab still posed a problem. The Environmental Sample Processor (ESP) minimizes this time difference by housing a robotic system placed in the ocean for a long period of time that can collect a small sample, concentrate the DNA, run a microarray, and take a picture of the array before sending the data ashore to be evaluated by a researcher. The included protocol and reagents list goes through both lab microarray procedures as well as the procedures list for the ESP, which briefly mentions deployment and data acquisition. The protocols described here should advance applications in microbial oceanography using robotic instrumentation.

Key words: (separated by '-')	Autonomous platform - Functional gene array - Environmental sample processor - Microarray - Hybridization - Genosensor - DNA - GeoChip - Oligonucleotide - Biogeochemical cycling genes

Chapter 9

Functional Gene Arrays for Analysis of Microbial Communities on Ocean Platform

Katelyn M. McKindles and Sonia M. Tiquia

Abstract

Recently there has been an increased use of microarray technology as a tool to determine the presence of functional genes in a population of hard to culture communities (e.g., soil, extreme environments). A functional gene array (FGA), or GeoChip, uses probes to screen for specific functional genes vital in biological systems such as nitrogen and carbon cycling, and has even been expanded to include aquatic conditions. However, the time delay from when the sample is taken from the ocean to evaluating the test results back in the lab still posed a problem. The Environmental Sample Processor (ESP) minimizes this time difference by housing a robotic system placed in the ocean for a long period of time that can collect a small sample, concentrate the DNA, run a microarray, and take a picture of the array before sending the data ashore to be evaluated by a researcher. The included protocol and reagents list goes through both lab microarray procedures as well as the procedures list for the ESP, which briefly mentions deployment and data acquisition. The protocols described here should advance applications in microbial oceanography using robotic instrumentation.

Key words: Autonomous platform, Functional gene array, Environmental sample processor, Microarray, Hybridization, Genosensor, DNA, GeoChip, Oligonucleotide, Biogeochemical cycling genes

1. Introduction

Over the past decade, environmental scientists have been casting a wider net in their attempts to understand complex environmental processes on a molecular scale. Microarray technology is used in gene expression studies of individual microorganisms (1–4), and has more recently been used to understand how genes are important regulators of earth-scale processes as carbon and nitrogen cycling (5, 6). DNA microarrays show great promise as a revolutionary tool for large-scale parallel analysis of microbial community structure and activities (5, 7–9).

[AU1]

Sonia Tiquia-Arashiro (ed.), *Molecular Biological Technologies for Ocean Sensing*, Springer Protocols Handbooks,
DOI 10.1007/978-1-61779-915-0_9, © Springer Science+Business Media, LLC 2012

Different types of microarrays have been developed to monitor microbial community dynamics in environmental studies (Table 1) (10), including functional gene arrays (FGAs) or GeoChip (5–7, 10–15), community genome arrays (CGA) (16, 17), and phylogenetic oligonucleotide arrays (POAs) (10). FGA is a gene microarray chip designed to identify "functional genes" involved in important nutrient cycles. It allows the identification of genes in an environmental sample that regulate carbon fixation, decomposition, and atmospheric nitrogen fixation, to name a few. Understanding what functional genes are available in a system allows scientists to both understand the potential of that system for cycling nutrients and better predict how that system will respond to environmental change. Imagine a glass floor divided into hundreds of identical squares. Each of these squares contains a different fragment of DNA, reconstructed from known DNA sequences. To probe an environmental sample for specific DNA sequences, the samples are hybridized over the floor. Fragments of DNA will stick to their complementary sequence on the floor, causing a square to light up. The array can be

Table 1
Major differences of various types of microarrays for environmental studies (10)

	CGAs	PCR-product-based FGAs	Oligonucleotides-based FGAs	POAs
Probe size	Entire genomic DNAs	Individual functional genes (200–1,000 bp)	Individual functional genes (50–70 bp)	Ribosomal rRNA (18–25 bp)
Types of information provided	Phylogenetic	Functional	Functional	Phylogenetic
Construction of comprehensive arrays	More difficult	More difficult	Easier	Intermediate
Reagent handling and tracking	Intermediate	More difficult	Easier	Easier
Targeted microorganisms	Culturable	Culturable and non-culturable	Culturable and non-culturable	Culturable and non-culturable
Specificity	Species	<80–85 % sequence homology	<86–90 % sequence homology	Single nucleotide difference
Sensitivity (ng of pure genomic DNA)	~0.2	~1	~8	Undetermined
Quantitation	Yes	Yes	Yes	Unknown
Taxonomic resolution	Genus–species	Genus–species	Species–strains	Species–strains

read by identifying fluorescently lit spots where environmental DNA has attached. This information is then used to develop a picture of the functional genes present in that system (9).

In Antarctica, FGA has already been used to answer important ecological questions (18). For example, scientists are finding that genes for nitrogen fixation, the crucial ecosystem process that produces plant-useable nitrogen in the soil, occur in lichen-rich areas. Lichens are believed to be among the earliest land colonizers, and the ability of lichen-dominated systems to add nitrogen to the soil may be an important finding in reconstructing the early colonization of terrestrial systems. Other findings include carbon-fixation genes in plots that lack vegetation, indicating microbial communities that are able to perform some sort of photosynthesis in the absence of plants (18).

In two more recent studies, FGA technology was used in aquatic conditions, specifically in acid mine drainage (AMD) (15) and the hydrothermal vent at the Juan de Fuca Ridge (14). While both of these microbial communities have been thoroughly studied in the past, very little was understood about their functional gene and physiological diversity. The use of GeoChip 2.0 when studying the microbial community of an AMD found that almost all major metabolic processes could be found in this ecosystem, including carbon and nitrogen fixation, carbon degradation, methane metabolism, ammonification, and more (15). Similarly, GeoChip was used to study the community housed in the chemical and thermal gradients of a vent chimney at the Juan de Fuca Ridge (14), through which it was discovered and reaffirmed that high-throughput microarray technology has a great potential in understanding ecosystem dynamics. In the hydrothermal vent study, GeoChip revealed the presence of functional communities involved in CO_2 fixation, methane cycling, nitrogen cycling, and metal resistance (14). Unfortunately, there is one limitation of this method of research, and that is that DNA-based GeoChip analysis can only detect the functional potential of a community, not the specific population (15).

The Environmental Sample Processor (ESP) (Figs. 1a, b and 2) takes FGA technology in the ocean one step further by making it an almost fully automatic system. The ESP is a relatively new technology which enables scientists to analyze the microorganisms of the ocean in an almost real-time situation by remotely collecting samples from the oceans subsurface, administering reagents as needed for a selected few processes such as probe arrays, and collecting data to be sent as images to be processed ashore via radio mooring (19) (Figs. 3 and 4). This instrument has a uniform methodology, which makes it ideal in detecting a variety of targets using one system, and consists of three major sections: (1) the core sample processor (Fig. 5a), (2) the sampling modules, and (3) the analytical modules (20). The system uses a rotating carousel to house the sample chambers or "pucks," (Fig. 5b) which contain

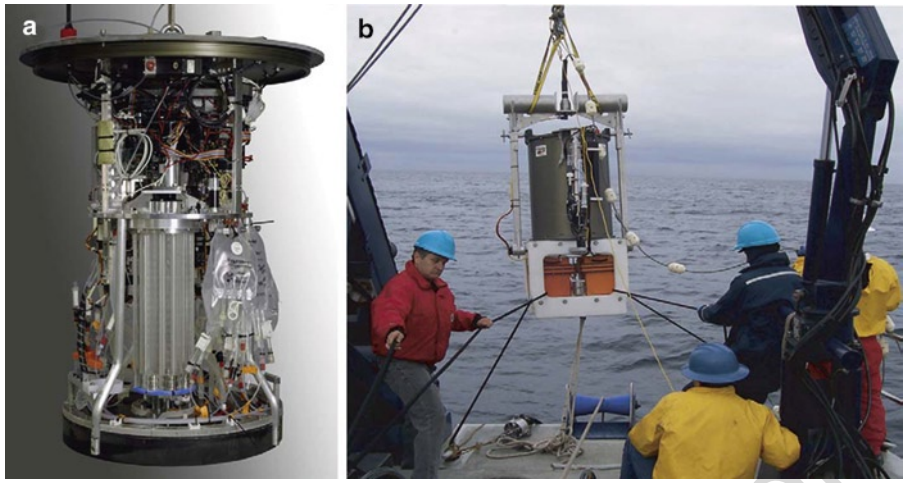


Fig. 1. (**a**) 2 G ESP without pressure housing, showing pucks at bottom of clear plastic carousel tubes (*lower center*) and several reagent bags (*lower right*). (**b**) Field deployment of 2 G ESP contained in pressure housing and mounted with two battery packs (*orange boxes*) (64).

Fig. 2. The second-generation Environmental Sample Processor (2 G ESP) being tested in a seawater tank ahead of deployment in Monterey Bay. The instrument is moored subsurface and an electromechanical cable provides for communications between a remote station and the ESP's surface buoy. An integral conductivity-temperature depth (CTD) package is visible at left. The ESP operates on 12-V rechargeable batteries (at bottom, above the anchor). *Photo credit: Todd Walsh, Monterey Bay Aquarium Research Institute* (19).

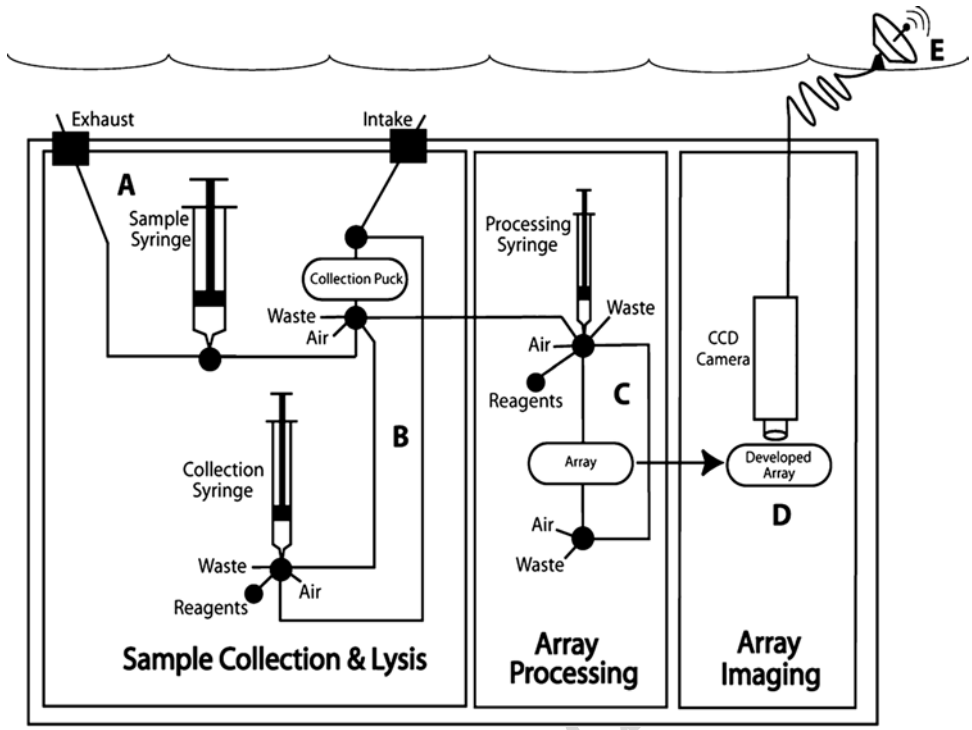


Fig. 3. Schematic diagram of the fluid path within the ESP instrument for sample collection and array processing. Seawater is brought into the instrument and filtered through a puck until the specified volume is reached or until the filter clogged. (**a**) The collection syringe presents various reagents to the particulates collected on the filter in the puck, resulting in cell lysis followed by dilution of the lysate. (**b**) The diluted lysate is passed to the processing syringe. The processing syringe delivers the lysate to a puck containing an array; then after incubation cleared to waste. The process is repeated for subsequent reagents. (**c**) The array is positioned under CCD camera and photographed. (**d**) The resulting image file is sent ashore via surface radio mooring. (**e**) Black circles represent valves that make connections between the syringes and puck, reagents, air or waste (22).

filter media and probe arrays, but utilize robotic mechanisms to load the pucks into their proper processing positions after the ESP has collected a sample (21). The samples are run through the sandwich hybridization array (SHA) format, which detects 16S rRNAs indicative of phylogenetically distinct groups of marine bacterioplankton (22) (Fig. 6), a variety of invertebrates, and harmful algal species (23). In addition to collecting the samples, the ESP houses chemical and physical sensors which enable the samples to be evaluated in respect to the environmental conditions (23).

The technology originated as a means to study the emergence of harmful algal blooms (HABs) (24, 25), but has since included the detection of marine bacterioplankton (22). The HABs produce a toxin that severely disrupts both the ecosystem and the surrounding human population, a problem which demands further research. The use of the ESP in the ocean allowed for periodic sampling of the water during near real-time intervals which enabled researchers to find trends in phytoplankton abundances (24). Later deployments

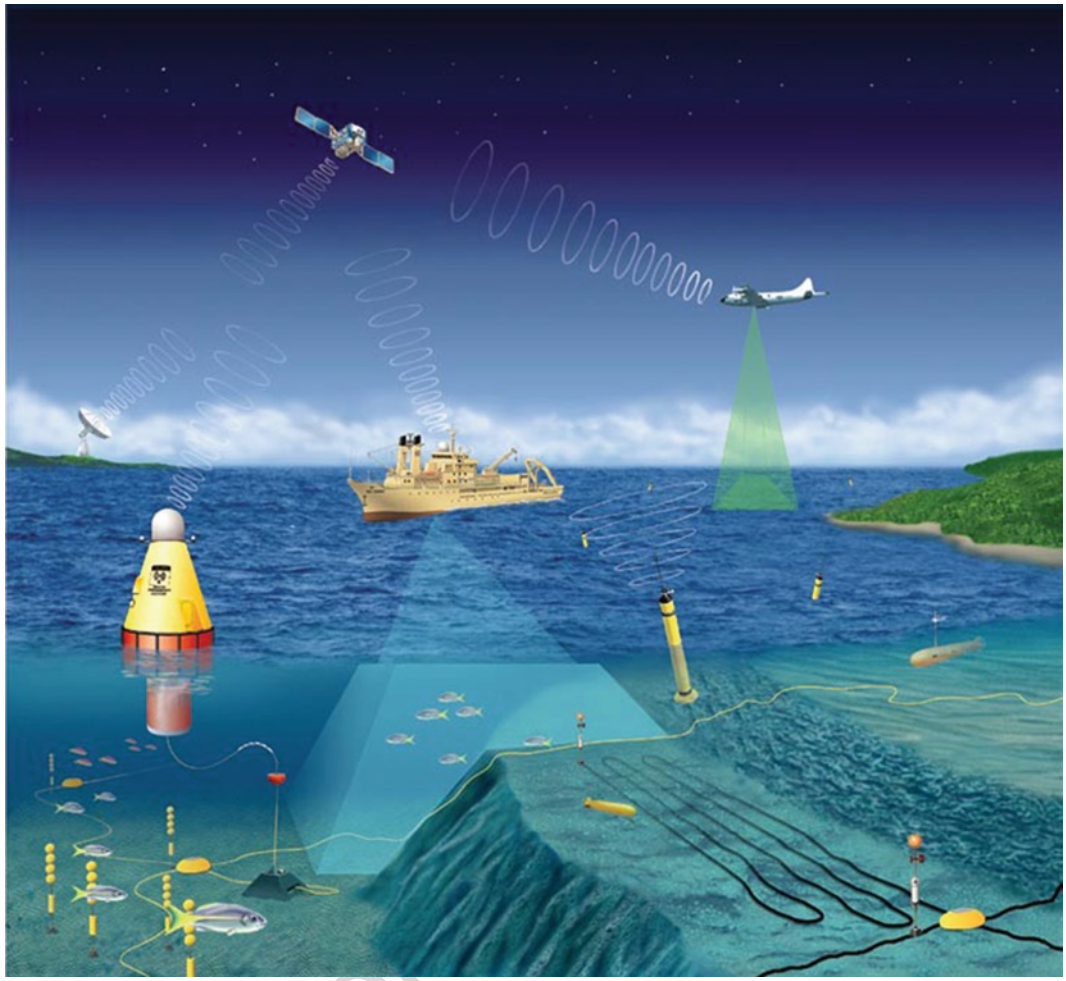
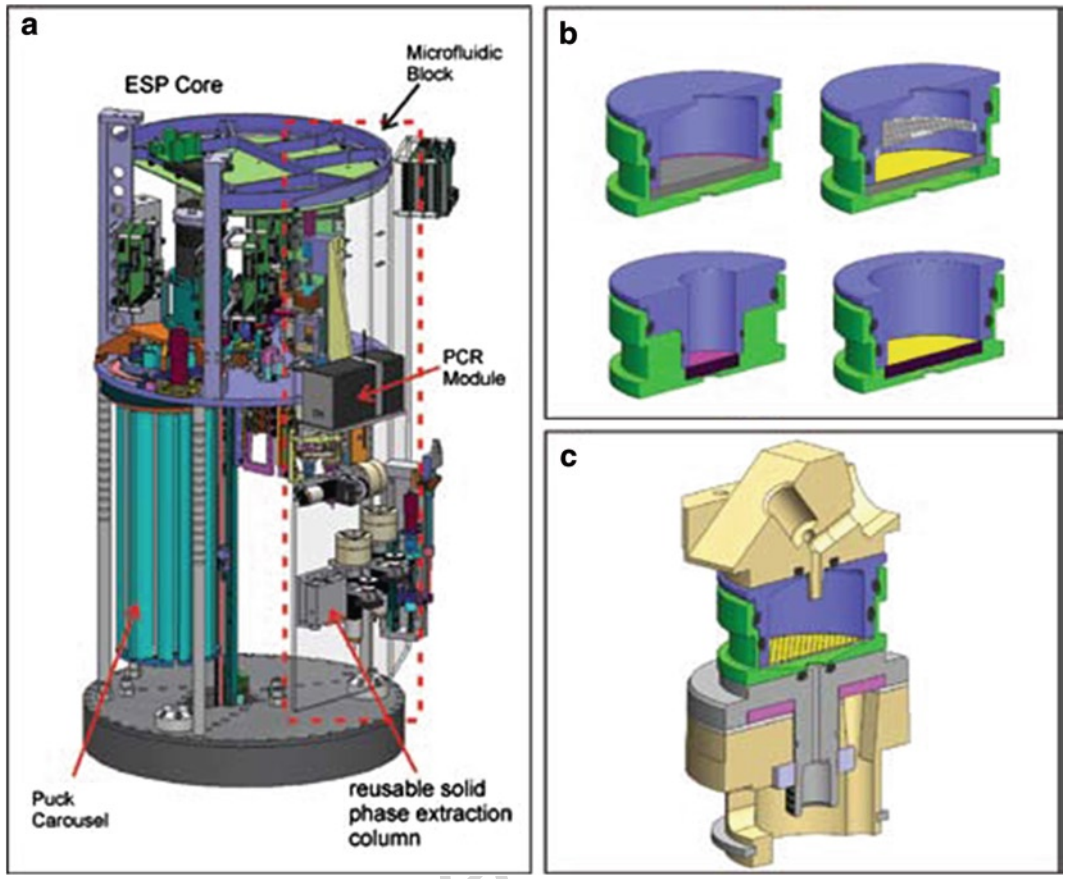


Fig. 4. Vision of the components of an ocean-observing system, including cabled observatories, autonomous underwater vehicles, gliders, buoys, moorings, satellites, and a traditional observing platform (research vessel) (19).

of the ESP confirmed the hypothesis that environmental conditions, such as the strength of upwelling (a wind driven motion which greatly enhances nutrient supply to the surface of the ocean) and associated patterns in stratification effected the HAB species composition. *Alexandrium catenella*, a motile dinoflagellate, is favored by relatively strong stratification while *Pseudonitzschia* is favored by a strong upwelling pulse (25). The main goal of both of these studies was to be able to determine either the cause or any trends in HABs in order to effectively and efficiently monitor and predict their occurrence (25). The study done by Preston et al. (22) had a slightly different goal in mind to report the first in situ DNA probe-based detection of marine bacterioplankton. This study discusses the ability of the ESP to become a diverse detection tool, as well as the limitations of the system, such as its inability to detect low copy number targets. No matter what the target microbe the ESP has



[AU2] Fig. 5. SolidWorks® models showing: (**a**) the core ESP with a microfluidic block (MFB) and polymerase chain reaction module attached on the right-hand side, (**b**) different types of pucks used for processing samples, and (**c**) the sample-collection station clamp in the closed position holding a sample puck. Puck assemblies are specially designed for different operations. In (**b**), the pucks shown are designed for collecting and homogenizing large-volume samples (*top left*), archiving material for microscopy (*top right*), and developing probe arrays printed on 12-mm or 25-mm membranes (*bottom left* and *right*, respectively). All pucks conform to the same overall size and shape so that robotic systems used to move and utilize the pucks can be standardized against a constant form factor. ESP with MFB is ~0.5-m diameter and ~1-m tall. A puck is ~30-mm diameter × 17-mm tall (20).

been programmed to detect, this tool has such potential in real-world and autonomous applications.

The following sets of protocols are intended to serve as a basic introduction to microarray construction and the steps required in microarray experimental design. There are four fundamental steps required in oligonucleotide-based FGA construction and experimentation: (1) FGA 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. 7. 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 the ocean.

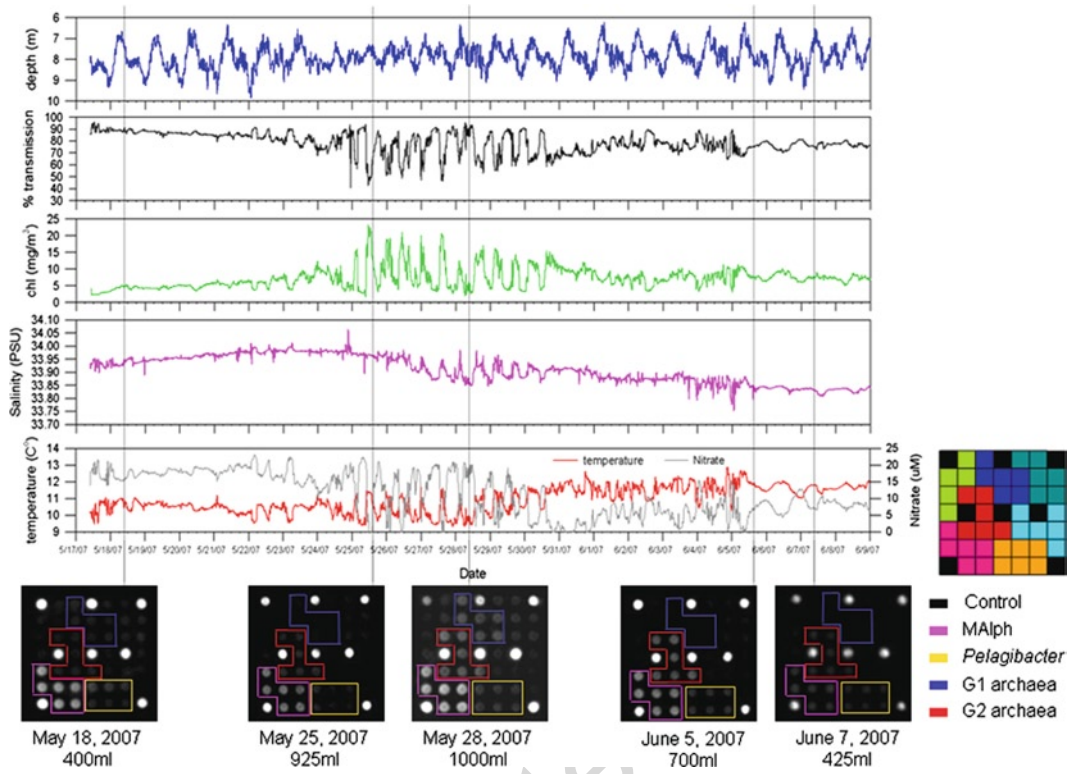


Fig. 6. In situ detection of marine bacterioplankton using ESP DNA probe arrays during a spring field deployment. The ESP was moored in Monterey Bay, California, 17 May to 11 June 2007. Top four graphs show physical and chemical data collected by contextual sensors on the ESP mooring during the deployment. The bottom images shows DNA probe arrays targeting rRNA indicative of various groups of marine bacterioplankton. Sample volume is shown underneath the array. The arrays shown are 15 mm × 15 mm (20).

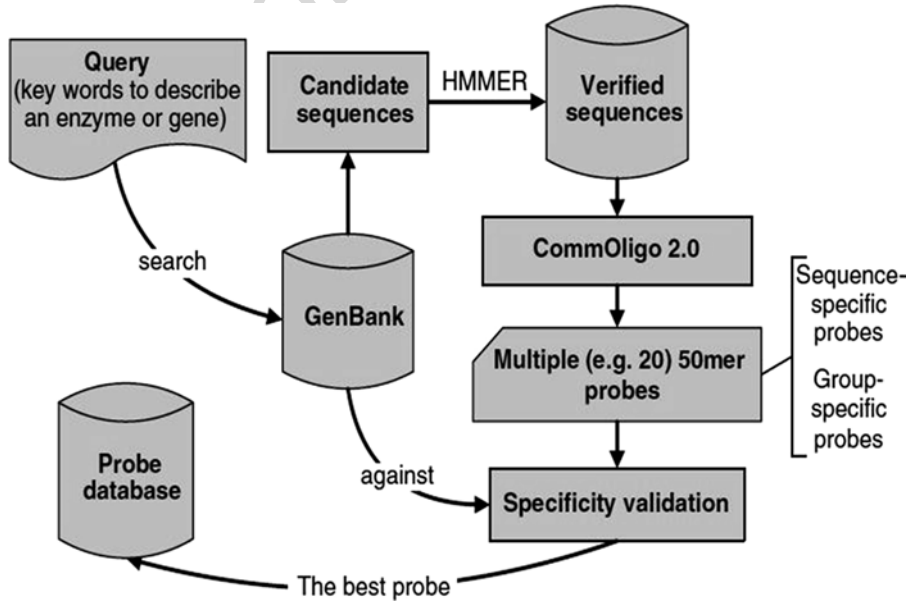


Fig. 7. The design pipeline for FGA construction (11).

2. Reagents and Equipment

2.1. Oligo Microarray Fabrication

50 % D MSO printing buffer

0.1 % SDS buffer

2.1.1. Reagents

2.1.2. Equipment and Materials

Aminosilane-coated glass slides (SuperAmine) (Telechem # SMA)

384-Well Printing Plate and lid

Orbital Shaker

Centrifuge with rotor for microtitre plates

Array Printer (PixSys 5500) (Cartesian Technologies, Irvine, CA)

200–500 μm spaced split pins

Slide box

GS Gene Linker® UV ChamberOligonucleotide probes

2.1.3. Software and Web-Based Resources

Genbank, EMBL, or Swiss Prot

UniGene (UniGene Laboratories, Boonton, NJ)

PRIMEGENS (Digital Biology Laboratory, University of Missouri-Columbia) Primer 3 (Whitehead Institute for Biomedical Research, Cambridge, MA), or Web Primer (Stanford University, Stanford, CA)

OligoArray (University of Michigan, Ann Arbor), Array Designer (PREMIER Biosoft, Palo Alto, CA), or Sarani (Strand Genomics, Burlingame, CA)

2.2. Target Preparation

Temliphi amplification kit

PicoGreen (Quant-iT™ PicoGreen® dsDNA kit; Invitrogen, Carlsbad, CA)

2.3. Labeling and Quantitation of Target DNA

2.3.1. Reagents

1 mM Cy3 or Cy5 dCTP (Amersham Pharmacia Biotech #PA55021)

RNase- and DNase-free water (Ambion, Inc. # 9934)

750 ng μL^{-1} random octamer primers (Invitrogen # Y01393)

5 mM dATP, dTTP, dGTP, and 2.5 mM dCTP (Biopioneer Inc., San Diego, CA)

40 U μL^{-1} Klenow fragment (Invitrogen # Y01396)

20 pmol PCR primers

25 mM dATP, dCTP, dGTP, and 15 mM dTTP (New England Biolabs)

10 mM aminoallyl-dUTP (Sigma, St. Louis, MO)

Taq DNA polymerase

0.1 M carbonate buffer (pH 9.0)

N-Hydroxy Succinimide esters Cy3 or Cy5 (NHS-Cy3 or Cy5; Amersham Pharmacia Biotech, Piscataway, NJ)

100 mM NaOAC

2.3.2. Equipment and Materials

QIAquick columns (Qiagen, Valencia, CA)

Speed-vac (e.g., SPD 1010 SpeedVac system; Thermo Savant Waltham, MA)

Spectrophotomer (NanoDrop™ ND-1000 spectrophotomer; Nanodrop Technologies, Wilmington, DE)

2.4. Hybridization

2.4.1. Reagents

Hybridization solution: RNase-free water, formamide, 20× saline sodium citrate (SSC) solution, 5 % SDS buffer, Herring sperm DNA

SSC solution

0.2 % SDS buffer

2.4.2. Equipment and Materials

Hybridization chamber (product number 2551; Corning, Lowell, MA)

Hybridization oven (e.g., PersonalHyb Hybridization oven; Strategene, La Jolla, CA)

Array coverslip (Structure Probe, West Chester, PA)

Centrifuge with rotor for microtitre plates

2.5. Image Processing and Data Analysis

Fluorescence reader (e.g., FLUOstar OPTIMA; BMG Labtech, Durham, BC)

2.5.1. Equipment

2.5.2. Software

1. ScanArray 5000 System (GSI Lumonics, Watertown, MA).
2. ArrayStat™ (Imaging Research, Inc., Ontario, Canada).
3. Imagene (BioDiscovery, El Segundo, CA), GenPix Pro (Axon Instruments, Union City, CA), Array Pro (Media Cybernetic, Carlsbad, CA), Quant Array (Packard Biosciences, Boston, MA), or TIGR Spot Finder (The Institute of Genomic Research TIGR, Rockville, MD).

2.6. Environmental Sample Processor

Signal probe cocktail in 2 M guanidinium thiocyanate (GuSCN) signal buffer

2.6.1. Reagents

Anti-Dig HRP (Pierce; Rockford, IL)

Stabilized diluent blocker (1 mL) (Pierce)

Substrate (Pierce SuperSignal West Femto Maximum Sensitivity Substrate: Stable Peroxidase Buffer and Luminol/Enhancer Solution mixed 1:1 before delivery to the array)

2.6.2. Equipment and Materials

First- or Second-Generation ESP.

5-μm pore size hydrophilic Durapore backing filter (Millipore).

Charge-coupled device (CCD) camera (1G ESP: Santa Barbara Instruments model ST-8EI [Santa Barbara, CA, USA] with a Fujinon [Japan] HF35A-2M1 lens; 2G ESP: Starlight Xpress model SXV-H9 [England] with a Fujinon model HF16HA-1B lens).

V++ Precision Digital Imaging System, v. 4.0 (Digital Optics, Auckland, NZ).

2.7. General Laboratory Supplies and Equipment Required

Pipettes (with appropriate tips)

0.1 ml PCR tubes

1.8 mL microcentrifuge tubes

Laminar flow hood or PCR workstation hood (e.g., AirClean 600 PCR Workstation, AirClean Systems, Raleigh, NC)

Thermocycler

Water bath

Heating block

Gel electrophoresis unit (with appropriate agarose, buffers, and staining supplies)

Ice bucket

Nuclease-free or PCR quality water

MilliQ water

Coplin jars or similar container

Diamond pen

Slide rack

3. Protocol

3.1. Oligo Microarray Fabrication

3.1.1. Oligo Design

The 50-mer FGAs can be constructed with the probes from sequences recovered from a variety of environments to represent the known microbial population diversity involved in the biogeochemical processes of interest. Sequences can be retrieved from public databases such as GenBank, EMBL, and SwissProt. However, these databases contain redundant sequences and it can be difficult to retrieve all sequences of interest. It might be better to retrieve sequences from the UniGene database (http://www.ncbi.nlm.nih.gov/UniGene/query_tips.html). This database is a collection of unique GenBank sequences grouped by organism and gene, and gives all sequences in one entry with links to the GenBank entries.

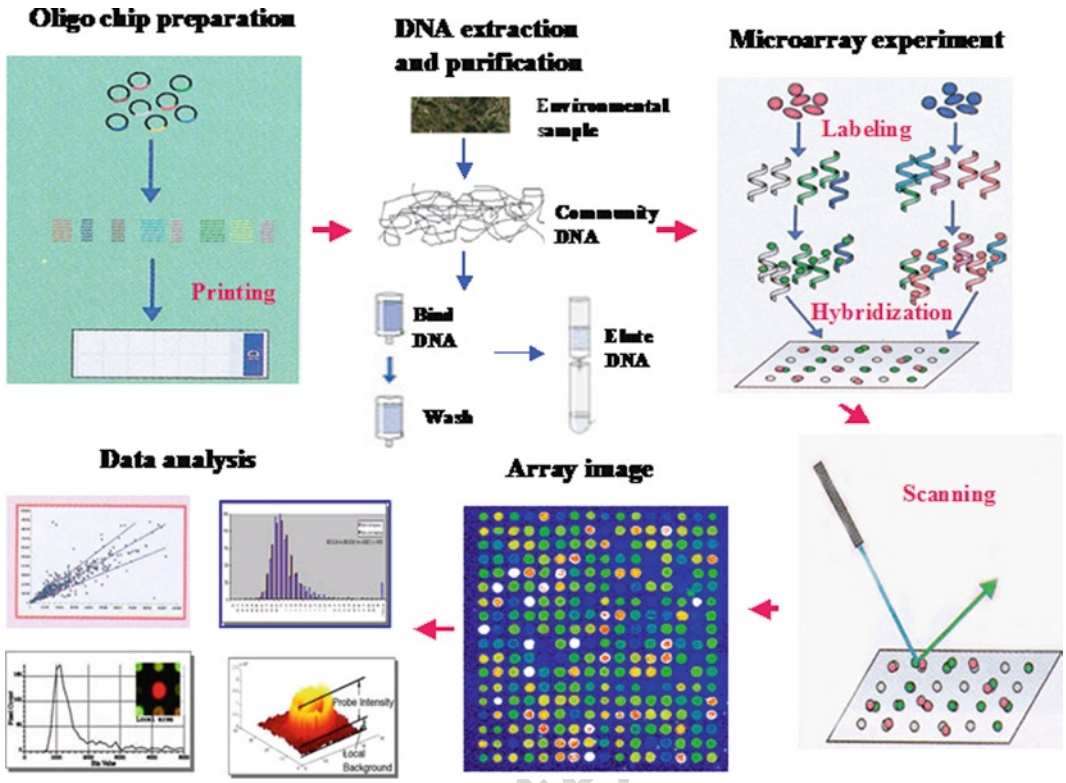


Fig. 8. Schematic diagram of microarray construction and experiments.

Alternatively, sequence retrieval can be carried out using the GeoChip design pipeline (11). The whole pipeline runs on a Web-based Common Gateway Interface (CGI) server and the scripts are written in Perl (Fig. 8). For each functional gene, a query of words is first submitted to GenBank Protein Database to fetch all candidate amino acid sequences. All candidate sequences for each functional gene are retrieved by key words and confirmed by HMMER 2.3.2. (Ashburn, VA, USA) (26), with seed sequences. In addition, all confirmed protein sequences are used to obtain the nucleic acid sequences from GenBank for probe design. A new version of CommOligo 2.0 with group-specific probe design features is used to design 50-mer oligonucleotide probes (both gene-specific and group-specific) using the same criteria as described for GeoChip 2.0 (13). After the specificity of all designed probes is computationally checked with currently available databases (GenBank), the best probe for each sequence or each group of sequences is selected to synthesize for GeoChip 3.0 construction. Because all seed sequences and key words are stored in databases, automatic updates can be performed in the future (see Note 1).

To design 50-mer oligonucleotide probes, oligo design software such as PRIMEGENS (http://compbio.ornl.gov/structure/primegens/), Primer 3 (http://www.bioinformatics.nl/cgi-bin/

primer3plus/primer3plus.cgi), and Web Primer (http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer) can be used. These online programs have been used primarily for primer design but they can also be used to design oligo probes. The PRIMEGENS program has been used to design gene-specific primers for whole genome cDNA microarrays (27), and oligo probes for FGAs. 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 to the corresponding aligned regions of any of 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 free and commercial software packages for designing oligonucleotides. OligoArray (28) 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. Thousands of gene sequences can be analyzed together, and the best available oligonucleotide probes with uniform thermodynamic properties and minimal similarity to nonspecific genes can be selected using these software.

Oligos are synthesized at the desired scale at the 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 # SMA). SuperAmine slides contain covalent amine groups that allow stable attachment of nucleic acids.

3.1.2. Oligo Array Printing

1. Prepare printing oligo solution 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 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 100–150 µm, with 200–500 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 after each printing.

Environmental conditions have significant effects on spot uniformity and size (29). 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. As a result, the quality of the deposited DNA will be changed. Also, reducing evaporation can help the spotted volume of DNA to have more time to bind at equal rates across the entire spot. As a result, DNA spots of increased homogeneity will be obtained (30). The printing buffer is also critical for obtaining homogeneous spots. With the widely used SSC buffer, the spot homogeneity as well as binding efficiency is often poor. Using the printing buffer containing 1.5 M betaine improves spot homogeneity as well as binding efficiencies (30). This is because betaine increases the viscosity of a solution and reduces the evaporation rate. More uniform spots can also be obtained with the printing buffer containing 50 % DMSO (dimethyl sulfoxide) (9, 29).

3.1.3. 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 in a GS Gene Linker® UV Chamber for 30 s.
2. Wash slides at room temperature first with 0.1 % SDS and then with water:

		Washing time
Wash 1	0.1 % SDS	4 min
Wash 2	Water	2 min

3. Transfer slides to a ten-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 before hybridization.

3.2. Target Preparation

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 (9). Successful application of microarray for microbial community analysis relies on the effective recovery of nucleic acids from the environment. Hurt et al. (31) and Zhou et al. (32) pointed out some criteria for ideal recovery of DNA or RNA from environmental samples: (1) 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; (2) the DNA should be of sufficient purity for reliable hybridization; (3) the extraction and purification protocol should be robust and reliable. The DNA extraction and purification protocol described by Hurt et al. (31) fulfills the above criteria (see Note 2). Of course it should be possible to substitute other protocols that meet these criteria.

The FGA requires 2–5 μg of genomic DNA for hybridization. Depending on the amount of DNA available, whole genome amplification (WGA) may be required. WGA can be performed using either phage Φ29 (16, 33) or *Bacillus stearothermophilus* DNA polymerases (34, 35). Amplification using Φ29 (Templiphi, GE Healthcare; Piscataway, NJ) for WGA of microbial community DNA has been systematically evaluated and shown to provide sensitive (10 fg detection limit) and representative amplification (<0.5 % of amplified genes showed more than twofold different from unamplified) (16). The following protocol uses the Templiphi amplification kit (GE Healthcare) and is based on a previously published protocol (16). All steps should be carried out in a laminar flow hood or PCR workstation hood.

1. Add 10 μl of sample buffer (supplied with kit) to a PCR tube or micro-well plate.
2. Transfer 10–100 ng DNA to the sample buffer. The total volume of DNA added should be no more than 5 μl and the sample volume should be the same for all samples. Use nuclease-free water to bring the volume up, if necessary. See Note 3.
3. Mix the DNA and buffer thoroughly and incubate 10 min at room temperature.
4. While DNA and buffer are incubating, prepare the Templiphi premix [for each reaction: 10 μl reaction buffer, 0.6 μl enzyme

mixture (both supplied in the kit), and single-stranded binding protein (USB; Cleveland, OH) and spermidine to a final concentration of 260 ng μl^{-1} and 0.1 μM, respectively].

5. Transfer 12.85 μl of the Templiphi premix to the DNA/buffer mixture (or the equivalent volume for one sample).
6. Incubate the reaction at 30 °C for 3 h and then heat-inactivate the enzyme at 65 °C for 10 min.
7. To evaluate the amplification quality, run approximately 2 μl of amplified product on a gel. The product should produce a smear rather than a single band.
8. Quantify the amplified DNA using a dye-binding assay, such as PicoGreen (Quant-iT™ PicoGreen® dsDNA kit; Invitrogen, Carlsbad, CA). The amplified product cannot be measured using 260/280 ratios due to primers and dNTPs remaining in the sample. There should be at least 2 μg of amplified DNA. If there is less than this, the amplification should be repeated.

3.3. Labeling and Quantitation of Target DNA

DNA for hybridization is generally labeled using fluorescent dyes, primarily Cy3 or Cy 5. The DNA can be labeled directly (dyes are directly integrated into the target DNA) or indirectly (targets are labeled after hybridization) (see Note 4).

3.3.1. 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 (9). 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) 1–2 μg purified community DNA (in 10 μl RNase-free water)
   (b) 20 μl (750 ng μl^{-1}) random octamer primers (Invitrogen # Y01393)
   (c) 5 μl DNase- and RNase-free water
2. Mix them well and denature at 100 °C for 5 min.
3. Place immediately on ice for at least 30 s.
4. In a 1.5-ml microcentrifuge tube, combine:
   (a) 0.2 μl dNTP's (5 mM dATP, dTTP, dGTP, and 2.5 mM dCTP)
   (b) 0.4 μl (1 mM) Cy3 or Cy5 dCTP
   (c) 1 μl (40 U μl^{-1}) Klenow fragment (Invitrogen # Y01396)
   (d) 13.4 μl DNase- and RNase-free water

5. Add this mixture to the 0.2 ml PCR tube that contains DNA (volume = 35 µl).
6. Mix well and incubate at 37 °C for 3 h or overnight.
7. After incubation, boil the mixture at 100 °C for 5 min and chill on ice.
8. 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, 15 mM 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).

3.3.2. 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 results from 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 (36). 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 and 550 nm for Cy3 or 650 for Cy5. A suitable labeling reaction should have 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 Cye 3 and OD 650 for Cy5. Also, measure OD at 230, 260 and 280 to assess purity.
2. Take 1 µl of the labeled DNA OD using 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 (see Note 5). The specific activity is calculated as follows:

$$\text{Specific activity} = \frac{\text{amount of target DNA} \times 1{,}000}{\text{pmol of dye incorporated} \times 324.5}.$$

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.

3.4. Hybridization

Hybridizations using glass arrays can be carried out manually or using automated or semi-automated hybridization stations. Manual hybridizations are performed using a water bath or hybridization oven and specially designed hybridization chambers that maintain humidity levels within the chamber. Several hybridization stations provide incubation at controlled temperatures and mixing (e.g., Mail Tai from SciGene, SlideBooster from Advalytix, Maui from BioMicro Systems). Washing after hybridization can be accomplished manually or using an automated wash station (e.g., Maui Wash Station, BioMicro Systems). Other systems are completely automated from pre-hybridization through post-hybridization wastes (e.g., Tecan HS4800Pro, TECAN, USA).

An array covered by a 22×22 mm coverslip will require ~15 µl of hybridization solution. For a 15 µl hybridization solution, combine the following components (see Note 6):

Hybridization buffer	Volume (µl)	Final concentration
a. RNase-free water	2.5	
b. Formamide	7.5	50 %
c. 20× SSC	2.5	3.25×
d. 5 % SDS	2.0	0.31 %
e. Herring sperm DNA (Promega)	1.2	0.775 µg

3.4.1. Hybridization Protocol

1. Heat the hybridization solution at 95 °C for 2 min in a thermocycler, cool quickly to 25 °C, and spin down at 14,000 × *g* for 5 min *(*see Note 7*)*.
2. 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 (see Notes 8 and 9).
3. Put the slide in the hybridization chamber.
4. Dispense 20 μl of 3× SSC solution into the hydration wells on both sides.
5. Close the hybridization chamber. Make sure the seal is formed along the O-ring.
6. Incubate the chamber in a 50 °C water bath for 12–15 h (see Notes 10 and 11).

3.4.2. Post-hybridization Wash

1. Place slides, with the coverslips still affixed, in a jar filled with 1× SSC and 0.2 % SDS buffer and wash for 5 min. Allow the coverslips to fall from the slide and then remove the coverslips from the jar with forceps.
2. Transfer the slides to a fresh jar filled with 0.1× SSC and 0.2 % SDS wash buffer. Wash the slides for 5 min.
3. Wash slides with 0.1× SSC for 30 s.
4. Transfer the slides to a slide rack and immediately spin the slides dry at 600 rpm for 5 min in a centrifuge with a horizontal rotor for microtitre plates. As the rate of drying can be quite rapid, it is suggested that the slide be placed in the centrifuge immediately upon removal from the jar.
5. Slides are ready for scanning (see Note 12).

3.5. Image Processing and Data Analysis

The objective of microarray image processing is to measure and quantify the relative abundance of the signal intensity of the arrayed spots. It is therefore important that the spots on the array image be correctly identified. Microarray images are comprised of arrays of spots arranged in grids. An ideal microarray image for easy spot detection should have the following properties: (1) the location of spots should be centered on the intersections between the row and column lines, (2) the spot size and shape should be circular and homogeneous, (3) the location of the grids on the images should be fixed, (4) the slides should have no dust or other contaminants, and (5) the background intensity should be low and uniform across the entire image.

3.5.1. 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 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) (see Note 13).

2. Save the scanned display as a 16-bit TIFF image file and quantify the intensity of each spot. Many methods are available for resolving the spot location errors, spot size, shape irregularities, and contamination problems (37) to accurately estimate spot intensities. 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 (38). Be sure to subtract local background for each spot and then flag and remove poor quality spots from the data set for further analysis.
4. Compute signal-to-noise ratio (SNR) for each spot to discriminate true signals (SNR≥3.0) from noise (SNR<3.0) (39). 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™ (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.

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).

3.5.2. Data Analysis

Data analysis is the most challenging aspect of FGA because of the large amount of data generated. Several methods have been frequently used in FGA studies (Table 2). These include various

Table 2 t4.1
[AU3] **Microarray data analysis methods** t4.2

Analysis methods	Information provided	
Descriptive statistics		t4.4
Richness, evenness, diversity	Commonly used descriptive methods. For microarray data, functional genes (or probes) would be considered "species" and signal intensity would be used for abundance	t4.5 t4.6 t4.7
Relative abundance	Percent of all genes detected that belong to a certain functional group of gene. Signal intensity of gene number can be used for this calculation	t4.8 t4.9
Shared/unique genes	Percent of all genes detected that were found in two or more samples. Unique genes are those that are only detected in one sample	t4.10 t4.11
Response ratios	Determine changes in gene abundance based on different treatments or conditions by comparing the signal intensities between two samples, generally control vs. treatment (46)	t4.12 t4.13 t4.14
Ordination techniques		t4.15
Principal component analysis (PCA)	Ordination method that reduces the number of variables needed to explain the data and highlight the variability between samples. In the ordination plot, the distance between sample points indicates how similar or dissimilar samples are	t4.16 t4.17 t4.18 t4.19
Detrended correspondence analysis (DCA)	Similar to the PCA but uses detrending to remove artefacts (i.e., the arch effect) typically found in correspondence analysis (65)	t4.20 t4.21
Parallel coordinate analysis	Plots microarray data such that data points that showed similar signal intensities are clustered together	t4.22 t4.23
Methods for environmental data		t4.24
Canonical correspondence analysis (CCA)	Provides information on how abiotic and biotic factors impact and drive the community structure. Ordination plots show similarity between samples based on distance and how much influence environmental variables (shown as arrows) have on a given sample	t4.25 t4.26 t4.27 t4.28
Variation partitioning analysis (VPA)	Uses data obtained in the CCA to determine the relative influence of environmental variables on the microbial community. Data is shown as a percentage of variation	t4.29 t4.30 t4.31
Similarity comparisons of microarray data		t4.32
Euclidean distance	Uses the square root of the summation of the squares of the differences between all pair-wise comparisons	t4.33 t4.34
Pearson correlation coefficient	Correlates between two variables X and Υ to identify profiles with similar shapes	t4.35 t4.36
Other multivariate statistical analyses		t4.37
Cluster analysis	Group samples based on overall similarity of gene patterns or profiles	t4.38
Neural network analysis	Based on the random matrix theory (66) and creates microbial ecological networks to visualize relationships between genes or gene groups	t4.39 t4.40
Self-organizing maps (SOMs)	Data points are mapped onto a grid and clustered in such a way that those points closest to each other are the most closely related	t4.41 t4.42
Mantel test	A multivariate correlation analysis used to compare environmental factors and functional genes	t4.43 t4.44

diversity indices (e.g., richness, evenness, diversity) based on the number of functional genes detected. The relative abundance of specific 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), neutral network analysis (NNA), and parallel coordinate analysis (5, 40). PCA and DCA are multivariate statistical methods that reduce the number of variables needed to explain the data and highlight the variability between samples. They provide 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 (41, 42). It also can be used to visualize clusters of genes that behave similarly across different experiments. CA groups samples based on the overall similarity of gene patterns. It has been used to identify groups of genes, or clusters that have similar expression profiles (43). 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 (44). NNA is used to visualize relationships between genes or gene groups. Parallel coordinates technique is a multivariate visualization technique that allow for efficient analysis and understanding of complex data (45). In this study, parallel coordinates technique was used to cluster microarray data and determine the most dominant genes at different depths. This technique uses hierarchical clustering algorithm that aims at grouping items so that items in a cluster are similar as possible and as different from data items in the other clusters as possible. In order to determine the most dominant genes in a sample, microarray data with similar signal intensities were clustered.

Response ratios compare the signal intensities between two samples, generally control versus treatment (46). If environmental data is available, several statistical methods are available to correlate environmental variables with functional community structure. These include canonical correspondence analysis (CCA) (47), variable partitioning analysis (VPA) (48, 49), self-organizing maps (SOMs) or other correlation analyses (e.g., Mantel test). CCA has been used many times in FGA studies to better understand how environmental factors are affecting the community structure (6, 50, 51). Based on the results of the CCA, the relative environmental variables on the microbial community can be determined using VPA. SOMs are a more robust and accurate method for grouping large data sets (52). 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. Further correlations can be made with Mantel test (6, 13, 50, 51). 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 (53). The other approach is the Pearson correlation coefficient, which is ideal for identifying profiles with similar shape (54, 55).

There are 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, PCA, and SOMs. Free-computer programs (i.e., CLUSTER and TREEVIEW) to ascertain hierarchical relationships of different spots are also available (43).

3.6. Application of Microarray Technology in Autonomous Platform

The following protocol was taken from Greenfield et al. (56), except where noted, which briefly outlines the use of an ESP, including the deployment (Fig. 9), array processing, and sample archiving.

1. When the ESP is deployed in the past, it has been fielded different platforms, including moorings, piers, remotely operated vehicles (ROVs) and benthic "elevators" for a period of around 30 days (57). So far, all of the ESP deployments have taken place in Monterey Bay, California, and at a depth range of the surface to 1,000 m for the ESP and continuing down to 4,000 m for the D-ESP (23). For further references on ESP deployment, see Preston et al. (22), Scholin et al. (23), and Jones et al. (58).
2. Printed arrays prepared above are loaded into a clean array puck on top of a 5-μm pore size hydrophilic Durapore backing filter (Millipore), leaving a 0.009-in. gap above the array, and the puck is then placed into the instrument's rotating carousel.
3. An aliquot of sample is drawn into the ESP's processing syringe. The ESP then loads the puck to the SHA (Sandwich Hybridization Assay) processing position and heats the puck to 25–30 °C. The 1G and 2G ESPs add sample and reagents in the same sequence, but the 1G uses 2 ml of each whereas the 2G uses 1 ml each. Lysate and processing solutions (see Goffredi et al. (59), except where noted) are applied automatically as follows: lysate, 20 min; 1× wash, 2 min; signal probe cocktail in 2 M GuSCN signal buffer, 10 min; 2× wash, 2 min each; anti-Dig HRP (Pierce; Rockford, IL) diluted 1:1,500 in stabilized diluent blocker (1 mL) (Pierce), 5 min; 4× wash,

[AU4]

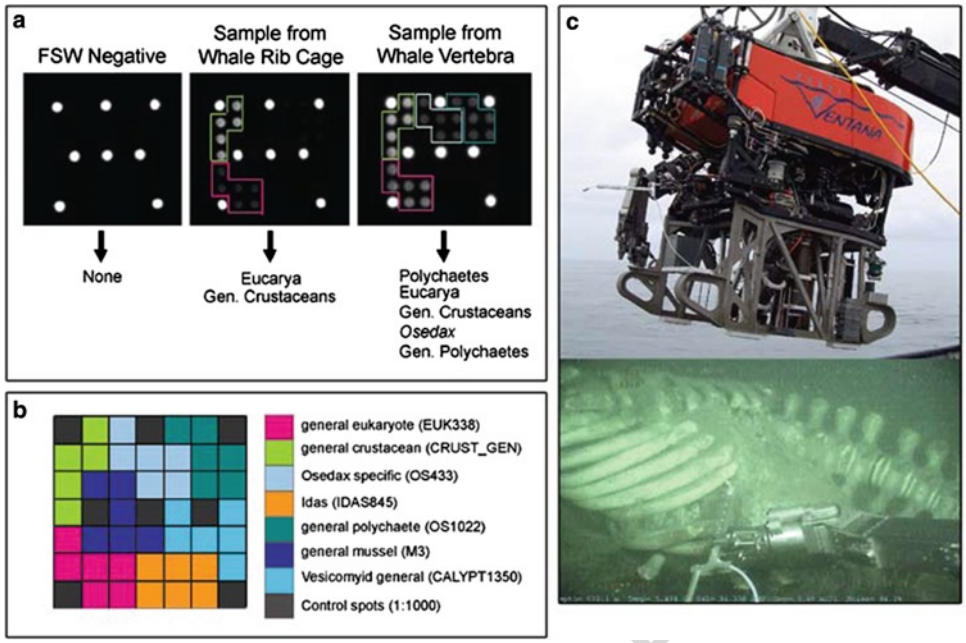


Fig. 9. Application of ESP for detection of invertebrates associated with a whale fall at 633-m depth in Monterey Bay, California, in August 2007. (**a**) From *left* to *right* are arrays from pre-deployment negative control (filtered seawater [FSW]) and two arrays from material collected from different portions of the carcass. Different sets of probes reacted positively depending on sample source. Actual size of the arrays is ~15 mm × 15 mm. (**b**) Array key showing locations of probes for different invertebrate rRNA sequences, including universal probe for Eucarya. *Colored boxes* surrounding probe spots on arrays (**a**) correspond to invertebrate species detected. (**c**) The *top picture* shows deployment of ROV Ventana with the D-ESP mounted below; the sampling wand is held in a robotic arm. The *bottom picture* is a video frame grab showing the sampling wand extended during sampling of the rib cage (23).

2 min each; HRP substrate (Pierce SuperSignal West Femto Maximum Sensitivity Substrate: Stable Peroxidase Buffer and Luminol/Enhancer Solution mixed 1:1 before delivery to the array), 10 s.

4. Afterward, the puck is immediately positioned beneath the camera and the image is captured using a CCD camera (1G ESP: Santa Barbara Instruments model ST-8EI [Santa Barbara, CA, USA] with a Fujinon [Japan] HF35A-2M1 lens; 2G ESP: Starlight Xpress model SXV-H9 [England] with a Fujinon model HF16HA-1B lens).
5. The final phase of ESP array processing, image analysis, is done using V++ Precision Digital Imaging System, v. 4.0 (Digital Optics, Auckland, NZ). The grand mean (±standard error [SE]) spot intensity per DNA probe is determined for each image by measuring a 10-by-10 pixel area per spot then recording average (±standard deviation [SD]) pixel intensity. An array spot with a grand mean intensity significantly higher than background (array region where no probe spotting occurred)

indicates a positive reaction for that probe. Background intensity is determined as above by taking the grand mean of three randomly selected unspotted regions of the imaged array.

6. The resulting image taken by the CCD camera is sent ashore via surface radio mooring (22). An electromechanical cable provides for communications between a remote station and the ESP's surface buoy (19) (Fig. 4).

4. Typical Protocol Results

The hybridization image indicated that the 50-mer oligonucleotide arrays hybridized well with the DNAs from marine sediment (Fig. 10). The DNA content of the sediments ranged from 1.69 to 18.10 μg, with highest yield from sediment samples collected at 0–0.5 cm. Within the mixed zone (0–25.5 cm depth), the DNA content ranged between 5.23 and 18.10 μg, whereas it was between 1.69 and 3.56 μg in the unmixed zone (50–84.5 cm depth). The hybridization image indicated that the microarrays hybridized reasonably with DNA's from Puget Sound marine sediments (Fig. 10). Strong signals were obtained with some nitrogenases (*nif*H) dissimilatory sulfate reductase (*dsr*AB), ammonia monooxygenase (*amo*A), methane monooxygenase (*pmo*A), and nitrite reductase (*nirK/*S). Microarray analysis indicated difference in the number of positive hybridization signals and signal intensity between sediments from shallow (bioturbation zone) and deeper

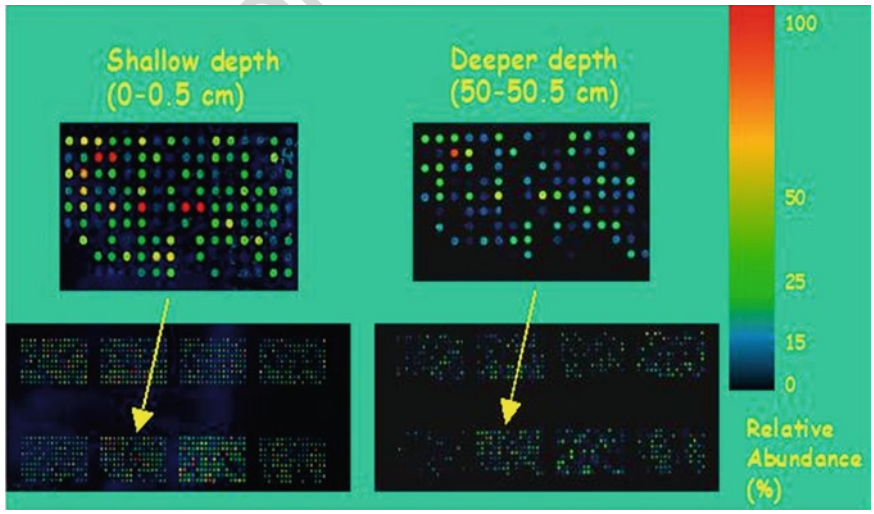


Fig. 10. Hybridization images hybridization images showing the profiles of different *dsr*AB, *nir*S, *nir*K, *nif*H, *amo*A, and *pmo*A genes. Community DNA (2 μg) from marine sediment was labeled with Cy5 using random primer labeling method, and hybridized at 50 °C for 15 h to the oligonucleotide arrays printed in replicate.

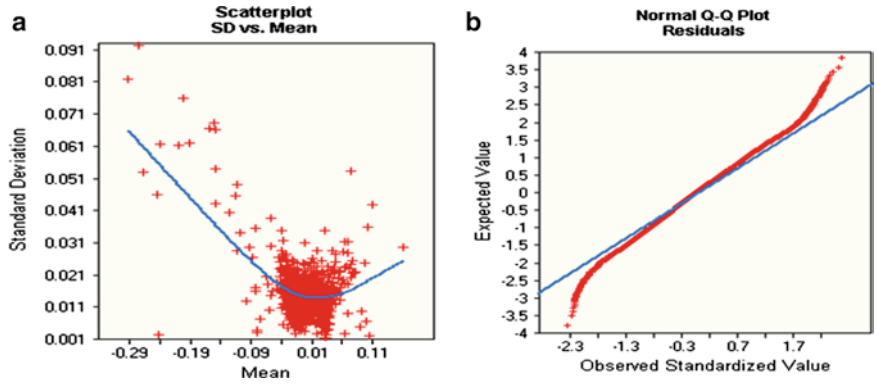


Fig. 11. Quality of microarray data. (**a**) Proportional model showing the relationship between standard deviation and the mean. (**b**) *Q–Q* plot displays.

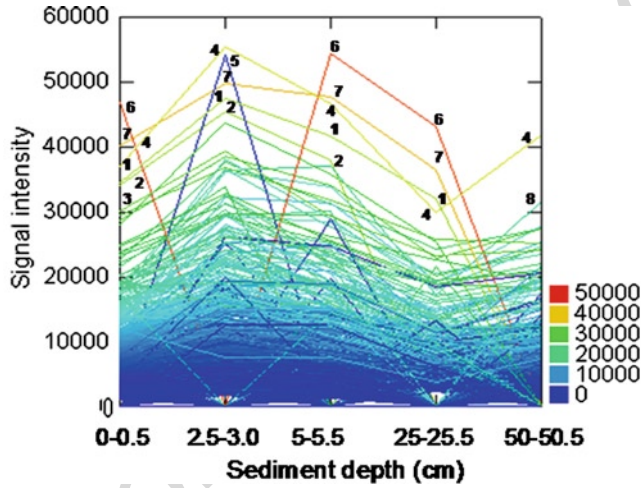


Fig. 12. Ordinate plots from principal component analysis based on the amount of *amo*A/pmoa, *dsr*AB, *nir*S, *nir*K, and *nif*H genes at different sediment depths. Values in *parentheses* indicate percent of total variances of PCA derived from the amount of individual functional gene group data and the combine data.

(below bioturbation zone) depths. The standard variation of the mean is relatively small for most of the gene probes (Fig. 11a). The *Q–Q* (quantile–quantile) plots of the observed standardized residual versus the expected values showed that the majority of the expected values fell closely along the observed values in a wide dynamic range, and only 6.4 % of the spots were outliers (Fig. 11b). Therefore the designed 50-mer FGAs should be useful in monitoring the composition, structure, activities, and dynamics of microbial populations involved in these functional processes across different natural environments. Using the hierarchical rule induction method, the test generated five layers of hierarchy corresponding to five signal intensity groups (1×10^4, 2×10^4, 3×10^4, 4×10^4, and 5×10^4) (Fig. 12). Eight most abundant functional genes (signal

intensities between 3×10^4 and 5×10^4), similar to those found in groundwater, terminate gut, sediments, and known cultivable bacteria. Some of these genes were dominant in all samples collected at different depths, while others are abundant only within the bioturbation zone or below the bioturbation. For example, the nitrite reductase gene similar to *nir*K_NKTH17 was found to be dominant across all samples; the nitrogenase gene similar to that isolated from termite gut (*nif*H_3157499), the nitrogenase-containing bacterium *Pseudanabaena* sp. (*nif*H_1698867), and the ammonia monooxidase gene similar to amoA_E03A16280 were only detected within the bioturbation zone; and nitrite reductase gene similar to *nir*K_WA20 was dominant only below the bioturbation zone.

5. Notes

5.1. Notes for DNA Microarray Fabrication

1. Regardless of how comprehensive the FGA is, sequences are constantly being added to public databases, leading to an exponential increase in the number of functional genes as well as the number of sequences for each particular functional gene. As such, continual updates of the FGA are necessary. Even with advances in probe design software, this process is still time consuming due to the large number of sequences and probes that must be designed and tested.

5.2. Notes for Target Preparation

2. The extracted DNA should be purified as soon as possible after extraction to prevent degradation. We have observed degradation of raw DNA extracts after as little as 1 month, even at −20 °C. Ideally, DNA should have a 260:280 ratio ≥1.8 and 260:230 ratio ≥1.7. Gel purification of soil and sediment DNA has worked very well for our lab. A column purification kit can also be used, especially for samples with low DNA yields. While these steps can be tedious, it is critical that DNA used for hybridizations be as pure as possible; therefore, it is beneficial (both in terms of time management and cost) to spend the time necessary on DNA preparation before proceeding with subsequent steps. Any impurities remaining in the DNA can interfere with amplification, labeling, and hybridization.
3. The best amplification results will be obtained using freshly extracted, high molecular weight DNA of the highest quality obtainable. However, DNA from samples with very low biomass, limited sample size, or that would be impossible to replace may not be of optimum quality yet are important to analyze. These samples can still be amplified, but may require some additional steps. If the DNA sample is very dilute, the DNA can be concentrated so that more DNA can be added to the reaction. If no or poor amplification occurs, try decreasing the amount of sample volume used to dilute out any inhibitors that may be

present. Serial dilution (2–3 dilution steps) of the sample can be used to "wash" the DNA. This approach has been used successfully in our lab for a variety of samples. Serial dilution can also be used if the DNA quality is above the recommended thresholds, but the amplification results are poor. Other options that can be tried to include re-precipitating the DNA (using an ethanol or isopropanol protocol) to try and remove any inhibitors, increasing amplification time or performing multiple amplifications of low product samples and combining the products to increase the total amount of amplified DNA.

5.3. Notes on Labeling and Quantitation of Target DNA

4. High quality, fresh reagents should be used to insure the best possible results (9). The higher the labeling efficiency, the better the hybridization results will be. Cy dyes are light sensitive and should be protected from the light as much as possible. We have stored dried, labeled products for months with no apparent loss of efficacy.
5. 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 one dye molecule per 25–50 nucleotides for good hybridization.

5.4. Notes on Hybridization

6. 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.
7. One common problem in microarray hybridization is the quality of fluorescent dyes. The labeling efficiency and hybridization vary significantly sometimes from batch to batch, especially Cy5. It is very important to use fresh reagents to achieve highly sensitive detection (9).
8. Labeled target molecules may be depleted in some areas, yet abundant in others. As a result, significant differences in signal intensity could be observed. Nonuniform 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.
9. 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 nonuniform. 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.

10. Labeled DNA or RNA is suspended in hybridization buffer. FGA can be hybridized at 42–50 °C and 50 % formamide (5–7, 13, 51, 60, 61). The hybridization temperature and formamide concentration can be adjusted to increase or decrease stringency in order to detect more or less diverse sequences. The effective hybridization temperature can be increased be the use of formamide (0.6 for every 1 %).
11. To reduce handling time of the slides at room temperature after hybridization, take out only one hybridization chamber at a time from the water bath.
12. Completed arrays should be protected from the light until imaged. A black or foil wrapped slide box works well. Imaging should be done within a few hours of removing the arrays from the hybridization chamber to minimize loss of signal.

5.5. Notes on Image Processing and Data Analysis

13. While increasing the PMT increases signal intensity, this also increases background. The hybridization signal can be photobleached, so the number of scans should be limited. The hybridization quality should be evaluated both during scanning (e.g., presence of positive control spots, even hybridization signals across the array, minimal background intensity) and after image analysis (e.g., presence of weak or poor spots). Several sources are available that provide more information regarding image preprocessing and analysis (13, 62, 63).

6. Future Directions

This paper focused on applying the DNA microarray technology in detecting functional genes and attempted to describe the potential ESP technology to extend the methodology in autonomous platform. The availability and the reliability of commercially available instruments, methods, and supplies have made it possible to dig into the molecular underpinnings of just about everything that is "environmental." These achievements, coupled with advances in ocean observatory technology and plans to extend those networks, have fueled the idea of applying molecular sensors in remote settings. In that regard, there is considerable potential to explore the ESP technology for DNA microarray applications as it is a highly configurable platform. Sampling and processing protocols can be tailored to user specifications. For example, assay chemistry can be modified, new probes can be developed, and sampling/processing

procedures (sample volume, archiving, filtration, lysis, etc.) can be amended (56). Since the first-generation ESP trials, the system has evolved to take into account deep-sea deployments on ROVs and benthic observatories. The deep-sea ESP (D-ESP) represents the next step in autonomous platform sampling, but still more work is being done to extend deployment duration, geographic coverage, depth rating, and analytical capacity (23). In the long term, it is our hope that this project will stimulate and inspire ocean scientists and engineers, who will in turn, contribute the major breakthroughs needed to make an impact in this field.

Acknowledgments

This research was part of S.M. Tiquia's work at Oak Ridge National Laboratory. We gratefully acknowledge financial support from the Office of the Vice President (OVPR) Grants, The University of Michigan-Ann Arbor and the United States Department of Energy (US DOE).

References

1. 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
2. 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 U S A 100:4191–4196
3. Haveman S, DiDonato R, Villanueva L, Shelobolina E, Postier B, Xu B, Liu A, Lovley D (2008) Genome-wide gene expression patterns and growth requirements suggest that Pelobacter carbinolicus reduces Fe(III) Indirectly via sulfide production. Appl Environ Microbiol 74:4277–4284
4. Borneman A, Bartowskey E, McCarthy J, Chambers P (2010) Genotypic diversity in Oenococcus oeni by high-density microarray comparative genome hybridization and whole genome sequencing. Appl Environ Microbiol 86:681–691
5. Tiquia SM, Gurczynski S, Zholi A, Devol A (2006) Diversity of biogeochemical cycling genes from Puget Sound sediments using DNA microarrays. Environ Technol 27:1377–1389
6. Van Nostrand JD, Wu WM, Wu L, Deng Y, Carley J, Carroll S et al (2009) GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium contaminated aquifer. Environ Microbiol 11:2611–2626
7. Tiquia SM, Wu L, Chong SC, Passovets S, Xu D, Xu Y et al (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. Biotechniques 36:664–675
8. 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
9. 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
10. Zhou J (2003) Microarrays for bacterial detection and microbial community analysis. Curr Opin Microbiol 6:288–294
11. He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL, Li X, Wu L, Gentry TJ, Yin Y, Liebich J, Hazen TC, Zhou J (2010) GeoChip 3.0 as a high-throughput tool for analyzing

microbial community composition, structure and functional activity. ISME J 4:1167–1179

12. Van Nostrand JD, Khijniak TV, Gentry TJ, Novak MT, Sowder AG, Zhou JZ, Bertsch PM, Morris PJ (2007) Isolation and characterization of four Gram-positive nickel-tolerant microorganisms from contaminated sediments. Microb Ecol 53:670–682
13. He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P, Criddle C, Zhou J (2007) Geochip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J 1:67–77
14. Wang FP, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand JD, Deng Y, He Z, Wu J, Zhou J, Xiao X (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. Proc Natl Acad Sci U S A 106:4840–4845
15. Xie J, He Z, Liu X, Liu X, Van Nostrand JD, Deng Y, Wu L, Zhou J, Qiu G (2011) GeoChip-baded analysis of the functional gene diversity and metabolic potential of microbial communities in Acid Mine Drainage. Appl Environ Microbiol 77:991–999
16. Wu L, Liu X, Schadt CW, Zhou J (2006) Microarray-based analysis of submicrogram quantities of microbial community DNAs by using whole-community genome amplification. Appl Environ Microbiol 72:4931–4941
17. Krause DO, Smith WJM, McSweeney CS (2004) Use of community genome arrays (CGAs) to assess the effects of Acacia angustinssima on rumen ecology. Microbiology 150:2899–2909
18. Yergeau E, Kang S, He Z, Zhou J, Kowalchuk GA (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. ISME J 1: 163–179
19. Paul JH, Scholin C, van Den Engh G, Perry MJ (2007) *In situ* instrumentation. Oceanography 20:58–66
20. Scholin C, Jensen S, Roman B, Massion E, Marin R III, Preston C, Greenfield D, Jones W, Wheeler K (2006) The Environmental Sample Processor (ESP): an autonomous robotic device for detecting microorganisms remotely using molecular probe technology. Paper presented at OCEANS 2006 MTS/IEEE conference, Boston, MA,18–21 September 2006. Marine Technology Society, Columbia, MD
21. Scholin C (2009) What are "ecogenomic sensors?" A review and thoughts for the future. Ocean Sci Discuss 6:191–213
22. Preston CM, Marin R III, Jensen S, Feldman J, Birch J, Massion E, DeLong E, Suzuki M, Wheeler K, Scholin C (2009) Near real-time, autonomous detection of marine bacterioplankton on a coastal morring in Monterey Bay, California, using rRNA-targeted DNA probes. Environ Microbiol 11:1168–1180
23. Scholin C, Doucette G, Jensen S, Roman B, Pargett D, Marin R III, Preston C, Jones W, Feldman J, Everlove C et al (2009) Remote detection of marine microbes, small invertebrates, harmful algae and biotoxins using the Environmental Sample Processor (ESP). Oceanography 22:158–167
24. Greenfield D, Marin R III, Doucette GJ, Mikulski G, Jensen S, Roman B, Alvarado N, Scholin CA (2008) Field applications of the second-generation Environmental Sample Processor (ESP) for remote detection of harmful algae: 2006–2007. Limnol Oceanogr Methods 6:667–679
25. Ryan J, Greenfield D, Marin R III, Preston C, Roman B, Jensen S, Pargett D, Birch J, Mikulski C, Doucette G, Scholin C (2011) Harmful phytoplankton ecology studies using an autonomous molecular analytical and ocean observing network. Limnol Oceanogr 56: 1255–1272
26. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763
27. 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
28. Rouillard JM, Herbert CJ, Zuker M (2002) OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18:486–487
29. 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
30. 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
31. 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
32. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl Environ Microbiol 62:461–468
33. Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and

phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11:1095–1099

34. Lage JM, Leamon JH, Pejovic T, Hamann S, Lacey M, Dillon D, Segraves R, Vossbrinck B, Gonzalez A, Pinkel D, Albertson DG, Costa J, Lizardi PM (2003) Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. Genome Res 13:294–307
35. Aviel-Ronen S, Zhu CQ, Coe BP, Liu N, Watson SK, Lam WL, Tsao MS (2006) Large fragment Bst DNA polymerase for whole genome amplification of DNA from formalin fixed paraffin embedded tissues. BMC Genomics 7:312–321
36. 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. Eaton Publishing, Natick, MA, pp 65–85
37. Zhou YX, Kalocsai P, Chen JY, Shams S (2000) Information processing issues and solutions associated with microarray technology. In: Schena M (ed) Microarray biochip technology. Eaton Publishing, Natick, MA, pp 167–200
38. 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
39. 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. DNA Press, LLC, Eagleville, PA, pp 83–98
40. He Z, Van Nostrand JD, Wu L, Zhou J (2008) Development and application of functional gene array for microbial community analysis. Trans Nonferrous Met Soc Chin 18: 1319–1327
41. 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
42. Raychaudhuri S, Stuart JM, Altman RB (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. Pacific symposium on biocomputing, pp 455–466
43. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–14868
44. 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 U S A 95:334–339
45. Inselburg A (1998) Visual data mining with parallel coordinates. Comput Stat 13:47–63
46. Luo Y, Hui D, Zhang D (2006) Elevated CO_2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. Ecology 87:53–63
47. ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179
48. Okland RH, Eilertsen O (1994) Canonical correspondence analysis with variation partitioning: some comments and applications. J Veg Sci 5:117–126
49. Ramette A, Tiedje JM (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proc Natl Acad Sci U S A 104:2761–2766
50. Wu L, Kellogg L, Devol AH, Tiedje JM, Zhou J (2008) Microarray-based characterization of microbial community functional structure and heterogeneity in marine sediments from the Gulf of Mexico. Appl Environ Microbiol 74:4516–4529
51. Waldron PJ, Van Nostrand JD, Watson DB, He Z, Wu L, Jardine PM, Hazen TC, Zhou JZ (2009) Functional gene array-based analysis of microbial community structure in groundwaters with a gradient of contaminant levels. Environ Sci Technol 43:3529–3534
52. Kohonen T (1997) Self-organizing maps, 2nd edn. Springer, Berlin
53. Knudsen S (2002) A biologist's guide to analysis of DNA microarray data. Wiley, New York
54. Chen Y, Bittner ML, Dougherty ER (1999) Issues associated with microarray data analysis and integration. Nat Genet 22:213–215
55. Jagota A (2001) Microarray data analysis and visualization. Bioinformatics by the Bay Press, Santa Cruz, CA
56. Greenfield DI, Marin R III, Jensen S, Massion E, Roman B, Feldman J, Scholin C (2006) Application of Environmental Sample Processor (ESP) methodology for quantifying Pseudonitzschia australis using ribosomal RNA-targeted probes in sandwich and fluorescent in situ hybridization formats. Limnol Oceanogr Methods 4:426–435
57. Birch J, Preston C, Pargett D, Jensen S, Roman B, Everloce C, Marin III R, Orphan V, Girguis P, Scholin C (2010) Searching for microbes in deep-sea seep and hydrothermal vents using the Environmental Sample Processor. Presented at the Astrobiology Science Conference 2010

58. Jones WJ, Preston C, Marin R III, Scholin C, Vrijenhoek R (2008) A robotic molecular method for in situ detection of marine invertebrate larvae. Mol Ecol Resour 8:540–550

59. Goffredi SK, Jones W, Scholin CA, Marin R III, Vrijenhoek RC (2005) Molecular detection of marine larvae. Mar Biotechnol 8:1–12

60. Mason OU, Di Meo-Savoie CA, Van Nostrand JD, Zhou J, Fisk MR, Giovannoni SJ (2009) Prokaryotic diversity, distribution, and insights into their role in biogeochemical cycling in marine basalts. ISME J 3:231–242

61. Liang T, Li G, Van Norstrand JD, He Z, Wu L, Deng Y, Zhang X, Zhou J (2009) Microarray-based analysis and microbial diversity along an oil contaminated gradient in oil field. FEMS Microbiol Ecol 70:324–333

62. Zhou J, Thompson DK (2002) Microarrays: application in environmental microbiology. In: Bitton G (ed) Encyclopedia of environmental microbiology, Vol 4. Wiley, New York, pp 1968–1979

63. Gentry TJ, Schadt CW, He Z, Zhou J (2007) Functional gene arrays for microbial community analysis. In: Hurst CJ, Crawford RL, Garland JL, Lipson DA, Mills AL, Ststzenbach LD (eds) Manual of environmental microbiology, 3rd edn. ASM, Washington, DC, pp 1052–1062

64. Doucette G, Mikulski C, Jones K, King K, Greenfield D, Marin R III, Jensen S, Roman B, Elliott C, Scholin C (2009) Remote, subsurface detection of the algal toxin domoic acid onboard the Environmental Sample Processor: assay development and field trials. Harmful Algae 8:880–888

65. Hill MO, Gauch HG Jr (1980) Detrended correspondence analysis, an improved ordination technique. Vegetatio 42:47–58

66. Sawa T, Ohno-Machado L (2003) A neutral network-based similarity index for clustering DNA microarray data. Comput Biol Med 33:1–15

Author Queries

Chapter No.: 9 0001520379

Queries	Details Required	Author's Response
AU1	References have been renumbered to maintain sequential order in the text. Please check for correctness.	
AU2	Kindly provide The Better Quality of figures for Figs. 5, 6, and 8 to 12.	
AU3	Please check the layout of this table for correctness.	
AU4	SHA is defined as both "Sandwich Hybridization Assay" and "Sandwich Hybridization Array". Please check.	