11 Design and Use of Functional Gene Microarrays (FGAs) for the Characterization of Microbial Communities

Christopher W Schadt[1], Jost Liebich[2], Song C Chong[1], Terry J Gentry[1], Zhili He[1], Hongbin Pan[1] and Jizhong Zhou[1]

[1] Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6038, USA; [2] Institute of Chemistry and Dynamics of the Geosphere IV: Agrosphere, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

CONTENTS

INTRODUCTION AND OVERVIEW OF FGAs

The recent development of microarrays as powerful, high-throughput genomic technology has spurred investigators toward their use for the study of various biological processes. Although microarray technology has been used successfully to analyze global gene expression in pure cultures or tissue samples for many different organisms (Lockhart *et al.*, 1996; DeRisi *et al.*, 1997; Schena *et al.*, 1995, 1996; Ye *et al.*, 2000; Thompson *et al.*, 2002; Liu *et al.*, 2003a,b), adapting microarrays for use in environmental studies presents great challenges in terms of design, use and data analysis (Zhou and Thompson, 2002; Zhou, 2003). Recently, various formats of environmental microarrays have been proposed, developed and evaluated for species detection and microbial community analyses in complex environments as reviewed recently by Zhou (2003). These studies have indicated that microarray-based genomic technologies have great potential as specific, sensitive, quantitative, and high-throughput tools for microbial detection, identification and characterization in natural environments. This chapter will focus on and discuss recent work on the development and use of functional gene microarray

METHODS IN MICROBIOLOGY, VOLUME 34
0580-9517 DOI:10.1016/S0580-9517(04)34011-0

(FGA) technology and introduce readers to issues and methodology surrounding their design and use. As you will see in the following discussions this methodological debate is far from over, as this revolutionary technology is still very much in a state of continuing development.

The genes encoding functional enzymes involved in various biogeochemical cycling (e.g. nitrogen, carbon and sulfur) and bioremediation processes are very useful as signatures for monitoring the potential activities and physiological status of microbial populations and communities that drive these processes in the environment. Microarrays containing functional gene sequence information are often referred to as functional gene arrays (FGAs) because they are primarily used for analysis of microbial community activities in the environment (Zhou and Thompson, 2002; Wu *et al.*, 2001). Similar to the microarrays used for monitoring gene expression, both oligonucleotides and DNA fragments derived from functional genes can be used for fabricating FGAs. To construct microarrays containing large DNA fragments as probes, the fragments are generally amplified by polymerase chain reaction (PCR) from environmental clones or from pure culture genomic DNAs (Wu *et al.*, 2001). Obtaining all the diverse environmental clones and bacterial strains from various sources as templates for amplification can be an overwhelming obstacle. As a result, construction of comprehensive FGAs based on PCR gene fragments that adequately encompass diverse environmental sequences is a near impossibility.

To circumvent this problem, FGAs containing synthetic oligonucleotides (oligos) have been developed for use. The main advantage of oligo FGAs is that construction is much easier than DNA-based FGAs because the probes can be directly designed and synthesized based on sequence information from public databases. Therefore, comprehensive arrays representing the extreme diversity of environmental sequences can be constructed. Several studies have applied and evaluated the usefulness of this approach for select groups of microorganisms (Taroncher-Oldenburg *et al.*, 2002; Koizumi *et al.*, 2002; Bodrossy *et al.*, 2003; Denef *et al.*, 2003; Tiquia *et al.*, 2004; Rhee *et al.*, 2004). In one such recent study originating from our laboratory, a 50mer oligo FGA was constructed and evaluated encompassing 1033 genes involved in nitrogen transformations (*nirS*, *nirK*, *nifH* and *amoA*), methane consumption (*pmoA*) and dissimilatory sulfate reduction (*dsrA/B*) from sequences available in public databases and our own environmental sequence collections (Tiquia *et al.*, 2004). Under the hybridization conditions of 50°C and 50% formamide, genes having $<86-90\%$ sequence identity could be clearly differentiated. This level of hybridization specificity is higher than those of PCR fragment-based FGAs (Wu *et al.*, 2001). Based on our comparisons of sequences from pure cultures involved in nitrification, denitrification, nitrogen fixation, methane oxidation and sulfate reduction, the average percent similarity of such functional genes at species level is usually much less than 85%, suggesting that oligo-based FGAs can provide species-level resolution. Also the detection limits of approximately 8–10 ng for pure genomic DNA was 10 times lower than the PCR fragment-based FGAs (Zhou and

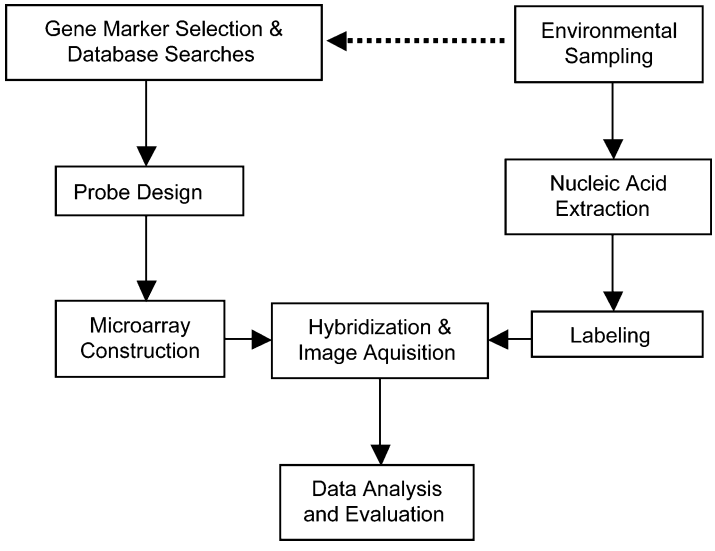

Figure 11.1. Schematic representation of the processes involved in designing and utilizing FGAs for the study of microbial communities.

Thompson, 2002; Wu *et al.*, 2001). In addition, similar to the DNA fragment-based FGAs, a strong linear relationship was observed between signal intensity and target DNA concentrations from 8 to 1000 ng for all six different functional gene groups ($r^2 = 0.96 - 0.98$). Furthermore, 5 μg of bulk community DNA from marine sediments was sufficient to obtain reasonably good hybridizations useful in profiling differences between communities. These results suggest that the developed 50mer FGA has potential as specific, sensitive, and potentially quantitative parallel tools for characterizing the composition, structure, activities and dynamics of microbial communities in natural environments. Based on such results and the methods presented below, a much more comprehensive 50mer FGA of several thousand gene probes is being designed and tested in our laboratory (Figure 11.1) (based on genes in Table 11.1 and others). The probes on the arrays represent very diverse groups of functional genes involved in nitrogen cycling, carbon cycling, sulfate reduction, phosphorus utilization, organic contaminant degradation and metal resistance.

◆◆◆◆◆ FUNCTIONAL GENE DIVERSITY AND DATA ACQUISITION FOR PROBE DESIGN

The first and one of the most important steps in design of oligo FGAs is to identify the optimal target genes that will be useful for tracking the microbial processes of interest (Figure 11.1). In this section, we have provided an overview of the diversity of publicly available functional gene sequence data for several microbially driven processes and suggested suitable genetic markers that could be used in probe design.

Table 11.1. Various microbial functional gene sequences available from public databases of potential use in FGAs. All categories exclude data from vascular plants and metazoans. The availability of environmental sequence data (*) and a representative reference of such is also provided. Categories including multiple genes or protein subunits are indicated as (all)

Gene/category	Sequences #	Example reference
Nitrogen cycling		
Nitrogenase-nifH	1784*	Hurek *et al.* (1997)
Nitrogenase-nifD	180*	Ueda *et al.* (1995)
Nitrogenase-nifK	89	
Ammonium monooxygenase-amoA	1158*	Nold *et al.* (2000)
Hydroxylamine oxidoreductase	15	
Nitrate reductase-napA	148	
Nitrate reductase-narB	50	
Nitrate reductase-narG	544*	Philippot *et al.* (2002)
Nitrate reductase-nasA	120*	Allen *et al.* (2001)
Nitrite reductase-nirK	264*	Liu *et al.* (2003a,b)
Nitrite reductase-nirS	411*	Liu *et al.* (2003a,b)
Nitrous oxide reductase-nosZ	273*	Scala and Kerkhof (1999)
Nitric oxide reductase-norB	68*	Braker and Tiedje (2003)
Urease (all)	1707	
Subtotal	6811	
Methane oxidation & reduction		
Soluble methane monooxygenase (all)	250*	Horz *et al.* (2001)
Particulate methane monooxygenase (all)	503*	Nold *et al.* (2000)
Methyl coenzyme M reductase	670*	Hallam *et al.* (2003)
Subtotal	1423	
Carbon polymer degradation		
Exoglucanase/cellobiohydrolase (all)	120	
Cellulase/endoglucanase (all)	920	
Chitinase (all)	1544*	Metcalfe *et al.* (2002)

(*continued*)

Table 11.1. *Continued*

Gene/category	Sequences #	Example reference
Laccase (all)	194*	Lyons *et al.* (2003)
Lignin peroxidase	15	
Mannanase (all)	400	
Polygalacturonase	156	
Subtotal	3349	
Carbon fixation		
Formyltetrahydrofolate synthetase (FTHFS)	206*	Leaphart and Lovell (2001)
Rubisco small subunit	747	
Rubisco large subunit	3474*	Elsaied and Naganuma (2001)
Rubisco small subunit (all euks)	112	
Rubisco large subunit (all euks)	787*	Elsaied and Naganuma (2001)
Carbon monoxide dehydrogenase (CODH)	14	
Subtotal	4441	
Sulphate reduction		
Sulfite reductase (*dsrA/B*)	924*	Wagner *et al.* (1998)
Adenosine phosphosulfate (*apsA*)	81*	Deplancke *et al.* (2000)
Subtotal	1006	
Phosphorus cycling		
Exopolyphosphatase	207	
Polyphosphate glucokinase	19	
Phytase	27	
Polyphosphate kinase	185*	McMahon *et al.* (2002)
Subtotal	438	
Organic remediation pathways		
Aniline (all)	36	
Atrazine (all)	113*	Martin-Laurent *et al.* (2003)
Benzene (all)	83	
Biphenyl (all)	237	
Dibenzothiophene (all)	101*	Duarte *et al.* (2001)

(*continued*)

Table 11.1. *Continued*

Gene/category	Sequences #	Example reference
2,4-Dichlorophenoxyacetic acid (all)	82*	Shaw and Burns, unpublished
Naphthalene (all)	392*	Baldwin *et al.* (2003)
Naphthalene dioxygenase α subunit	140*	Baldwin *et al.* (2003)
Naphthalene dioxygenase β subunit	21	
Naphthalene dioxygenase ferredoxin	23	
Naphthalene dioxygenase ferredoxin reductase	14	
Dihydrodiol naphthalene dehydrogenase	22	
Dihydroxynaphthalene dioxygenase	18	
Hydroxychromene carboxylate isomerase	42	
Hydroxybenzylidenepyruvate hydratase-aldolase	18	
Salicylaldehyde dehydrogenase	21	
Transcriptional regulator	35*	Park *et al.* (2002)
Nitrotoluene (all)	33	
n-Octane (all)	64	
Pentachlorophenol (all)	55*	Beaulieu *et al.* (2000)
Pyrene (all)	140	
Toluene-aerobic (all)	98*	Cassidy (2002)
Toluene-anaerobic (all)	32	
Trichloroethylene/perchloroethylene (all)	30	
Xylene (all)	77	
Other organic remediation genes	3786	
Subtotal	5359	
Metal resistance and efflux		
Aluminum resistance (all)	36	
Arsenic resistance (all)	407	
Cadmium resistance (all)	224*	Oger *et al.* (2003)
Chromium resistance (all)	125	
Cobalt resistance (all)	44	
Copper resistance (all)	311	
Lead resistance (all)	12	
Mercury resistance (all)	649	
Mercuric reductase-merA	149	
Organomercurial lyase-merB	33	
Mercury binding-merP	82	

(*continued*)

Table 11.1. *Continued*

Gene/category	Sequences #	Example reference
Mercury transport-merT	87	
Mercury transport-merC	45	
Mercury transport-merE	18	
Mercury transport-merF	7	
Phenylmercury resistance-merG	5	
Transcriptional regulator-merR	152	
Additional regulator-merD	50	
Nickel resistance (all)	52	
Selenium resistance (all)	4	
Silver resistance (all)	53	
Tellurium resistance (all)	321	
Vanadium resistance (all)	8	
Zinc resistance (all)	52	
Subtotal	2301	

We utilized a variety of combinatorial search strings within the GenBank ENTREZ interface in order to acquire these sequences, in combination with a locally executable program to extract and download pertinent information (database of origin, species name, gene descriptions, phylogenetic lineage, DNA coding sequence, protein sequence, etc.) starting only from a list of GI (gene identification) numbers in tab delimited format. These databases were then further screened to remove duplicate and non-homologous sequences. We found that such an approach greatly reduces the time and effort involved in sorting through and collecting the vast amounts of sequence data available for probe design. By conducting searches within the protein sequence database, many duplicated entries associated with the nucleotide database are avoided, and direct retrieval of coding DNA sequences is possible that avoids problems associated with intron containing sequences. Additionally, we have recently developed the program further so that once keywords are decided upon it automatically updates the databases (based on modification date) at user defined intervals and will also remove duplicate sequences. To illustrate the selection of genetic markers based on the data available from such searches, we discuss below some examples of this process for selected key functional genes. Table 11.1 additionally summarizes the results of many such searches and amounts of available sequence information.

Nitrogen Cycling

Microbial nitrogen transformations involve three major microbially driven processes; nitrogen fixation, nitrification and denitrification. These involve many different enzymes including nitrogenase, ammonium monooxygenase, hydroxylamine oxidoreductase, nitrite oxidase, nitrate

reductase, nitrite reductase, nitric oxide reductase, nitrous oxide reductase, assimilatory nitrate reductase and urease (Table 11.1). Thus, it is essential to understand the diversity of functional genes contributing to the nitrogen cycle contained in microbial communities. Over 6000 nitrogen cycle related gene sequences are available from GenBank and other databases (Table 11.1). However, it is beyond the scope of this chapter to discuss all of these processes and enzymes, so we have focused on the genes involved in microbial nitrogen fixation to illustrate the selection of appropriate markers.

Nitrogen is one of the major components of living cells and about 79% of the earth atmosphere is nitrogen in the form of N_2 gas. However, nitrogen is often a limiting factor for growth and biomass production in both aquatic and terrestrial environments (Vitousek and Howarth, 1991). In order to use nitrogen for growth, N_2 must be fixed to ammonium (NH_4) or nitrate (NO_3) which occurs primarily through microbial nitrogen fixation. Nitrogen fixation is exclusively performed by prokaryotes. Some live independently of other organisms (free living nitrogen-fixing bacteria such as *Azotobacter*, *Beijerinckia*, *Desulfovibrio*, purple sulfur bacteria, purple non-sulfur bacteria, and green sulfur bacteria) while others live in symbiotic association with plants (*Rhizobium*, *Frankia*, *Azospirillum*) and some are found in both situations (e.g. *Bradyrhizobium*). Microorganisms perform nitrogen fixation with an evolutionarily conserved nitrogenase protein complex, and all consist of two major proteins found in both archaea and eubacteria (Howard and Rees, 1996). Component I consists of an $\alpha_2\beta_3$ tetramer molybdoferredoxin (MoFe protein) or dinitrogenase (EC 1.18.6.1), the active site for N_2 reduction, and is encoded by the genes *nifD* and *nifK*. Component II consists of a homodimer, azoferredoxin (Fe protein) or dinitrogenase reductase (EC 1.19.6.1), that couples ATP hydrolysis to interprotein electron transfer, and is encoded by *nifH* (Dean *et al.*, 1993). Dinitrogenase genes are further divided into four clusters (Normand and Bousquet, 1989; Normand *et al.*, 1992; Chien and Zinder, 1994). Cluster I contains most of eubacterial MoFe containing dinitrogenase. Cluster II consists of archaea, eubacterial alternative dinitrogenase with non-Mo, non-V. Cluster III, contains Mo dinitrogenase (*nif-1*) genes from the Gram positive eubacteria *Clostridium* and the *nif-2* genes from the archaea *Methanosarcia* and the sulfate reducer *Desulfovibrio*. Cluster IV, containing methanogens, utilizes a distantly related gene similar to bacteriochlorophyll. Numerous other genes are also known to be related to or interact with *nif* proteins such as *nifJ*, *nifF*, *nifM*, *nifB* and are often present in a complex extended *nif* operon.

However, of all these potential targets for a functional marker, the *nifH* gene is the most widely used. Dinitrogenase reductase (*nifH*) is especially evolutionarily conserved and has often been used as a genetic marker for detecting nitrogen fixing microorganisms in natural environments (Kirshtein *et al.*, 1991; Zehr *et al.*, 1995; Widmer *et al.*, 1999) and also as a basis for the phylogenetic analysis of *nifH* containing organisms (Zehr and McReynolds, 1989). Over 1700 *nifH* gene sequences are available from public DNA databases and this is expanding rapidly. However, *nifD* and *nifK* have also been used as additional information, especially for

resolving differences between closely related sequences. For these reasons, the *nifH* gene makes an excellent functional marker, but the importance of *nifD* and *nifK* for resolving finer scale differences should not be discounted.

Carbon Cycling

Microbially driven aspects of the carbon cycle could play important roles in determining the amplitude of anthropogenic effects on climate change and their potential mitigation. Microorganisms play important roles through direct effects on such processes as methane oxidation and production, carbon fixation, and the breakdown and decomposition of organic substrates. While a discussion of all these aspects is not possible, we provide potential target genes for monitoring these processes in Table 11.1. To illustrate some of the issues surrounding selection of suitable genetic markers for these processes, we will proceed with an overview of aerobic methane oxidation. Methanotrophic bacteria oxidize methane for energy production and biosynthesis of organic compounds. These organisms are ubiquitous in environments such as oceanic and inland waters, wetlands, soils, groundwaters and even the deep subsurface and are of great interest to microbial ecologists because of the potentially important role they could play in mitigating global warming (Holmes *et al.*, 1995; Dunfield *et al.*, 1999). They are also of interest for use within certain industrial applications such as bioremediation or steps in the synthesis of certain organic compounds (Sullivan *et al.*, 1998).

The first step in the oxidation of methane is the conversion of methane to methanol by the enzyme methane monooxygenase (EC 1.14.13.25). After this step methanol is converted to formaldehyde via a non-specific alcohol dehydrogenase (e.g. EC 1.1.99.8), where it enters one of two different pathways (RuMP or serine-isocytrate) for the production of C_3 compounds. As a specific genetic marker for methanotrophy we are therefore left with methane monooxygenase. This key enzyme exists in two forms: the cytoplasmic, soluble, methane monooxygenase (sMMO) and the membrane-bound, particulate, methane monooxygenase (pMMO). Of the two forms of MMO, all known methanotrophs carry pMMO whereas only some select methanotrophs (Type II) carry both pMMO and sMMO (Murrell *et al.*, 2000). However, pMMO is also evolutionarily related to the important nitrogen cycling enzyme ammonium monooxygenase (AMO) and they share a high degree of sequence similarity in some regions of these genes (Holmes *et al.*, 1995). Because it is present in all known methanotrophs, pMMO carries distinct advantages as a specific functional marker for methanotrophy (Murrell *et al.*, 1998). Practically speaking, however, because of the high degree of similarity between the two genes (AMO and pMMO) any probe design effort must most likely consider both simultaneously. Both pMMO and AMO are made up of two different polypeptide subunits, the smaller of which (*pmoA* and *amoA*) contain the active sites and have been much

more widely sequenced and studied as functional and phylogenetic markers.

While many organisms can utilize methanol and some other methyl compounds (methylotrophy), currently true methanotrophy is only known to occur within the beta and gamma-proteobacteria. However, studies of environmental libraries of *pmoA* suggest diversity within these bacteria is high, and many lineages contained within them are only known from such environmental libraries. Additionally, several studies have suggested that there may be divergent lineages related to either *pmoA* or *amoA* that could possibly fall outside the proteobacteria (Nold *et al.*, 2000). Currently ~503 *pmoA* genes are listed in GenBank. However, because the closely related *amoA* can often be amplified with the same PCR primers (Nold *et al.*, 2000), care should be taken to ensure that such gene annotations are correctly identified in sequences retrieved from databases.

Sulfur Cycling

The most widespread biochemical reaction of the sulfur cycle is probably the assimilatory sulfate reduction in which inorganic sulfate is reduced to become integrated in amino acids and proteins where they play an essential role in the formation of secondary structures. This process is followed by the release of reduced hydrogen sulfide by the degradation (or desulfurylation) of the organic sulfur compounds. This process is important for almost every life form on Earth and therefore can be found throughout all biological kingdoms. More specific to prokaryotes are sulfide/sulfur oxidation and dissimilatory sulfate reduction (sulfur respiration) for which we discuss probe selection. Sulfate respiration, using oxidized sulfur forms as a terminal electron acceptor, is widespread often leading to the formation of black metal sulfides (e.g. iron sulfide) and the toxic gas hydrogen sulfide (H_2S) (Rabus *et al.*, 2000). Sulfate reducers occur in a wide variety of environments such as marine sediments, deep-sea hydrothermal vents, freshwater systems, anaerobic sludge and as endosymbionts (Jørgensen, 1982; Singleton, 1993; Finegold and Jousimies-Somer, 1997; Manz *et al.*, 1998; Dubilier *et al.*, 2001; Laue *et al.*, 2001; Castro *et al.*, 2002; Liu *et al.*, 2003a,b; Nakagawa *et al.*, 2004). Some sulfate reducing bacteria are also capable of oxidizing organic contaminants of anthropogenic origin (e.g. petroleum hydrocarbons) along with other naturally occurring substances (Lovley, 1997; Kleikemper *et al.*, 2002). Others are able to reduce many different metals beyond sulfur species, for example, some aid in immobilizing potentially hazardous metals like uranium via such reductions (Chang *et al.*, 2001). Both features make this group of interest for bioremediation processes in anoxic environments. In many cases, sulfate reducing bacteria may also compete with methanogens and denitrifiers for electron donors and sulfate reducing bacteria seem to often dominate if sufficient sulfate supply is available (Lovley and Klug, 1983).

Dissimilatory sulfate reduction is generally a three-enzyme transformation involving activation of sulfate by ATP sulfurylase (EC 2.7.7.4), reduction of the product adenosine-phosphosulfate (*aps*) to sulfite by adenylylsulfate reductase (EC 1.8.99.2) and subsequent further reduction to hydrogen sulfide by dissimilatory sulfite reductase (*dsr*, EC 1.8.99.3), the latter enzyme consists of at least three subunits. However, only APS reductase and *dsr*-genes are suitable indicators of dissimilatory sulfate reduction in the environment, since the first step is also involved in assimilatory sulfate reduction that is widely evolutionarily distributed. Furthermore, several organisms are known to use only sulfite as an electron acceptor but not sulfate, and they lack adenylylsulfate reductase genes (Huber *et al.*, 1997; Holliger *et al.*, 1998; Molitor *et al.*, 1998; Laue *et al.*, 2001). Because of this limitation, *aps* genes have only more recently been applied as a marker for this process, primarily to distinguish sulfate/sulfite reducers from those only capable of reducing sulfite (Deplancke *et al.*, 2000; Friedrich, 2002). As a result, in our recent searches, only 81 nucleotide sequences of APS genes from different species or subspecies were found in publicly available databases, whereas 924 sequences can be retrieved for *dsrA* and *dsrB* genes. Because of this better studied diversity, *dsr* genes are a more indicative marker for this overall function. *dsrA* and *dsrB* genes occur in all known sulfate or sulfite reducing bacteria and can be targeted by a single set of conserved primers (Karkhoff-Schweizer *et al.*, 1995; Wagner *et al.*, 1998) allowing probe design for both subunits. A third subunit for dissimilatory sulfite reductase, encoded by *dsrD*, is also known but available sequence information to date is mostly limited to cultured species (Karkhoff-Schweizer *et al.*, 1995). A large portion of the *dsrA* and *dsrB* sequence data were obtained from uncultivated organisms, after the discovery of the conserved nature of dissimilatory sulfite reductases over different phyla made their detection by PCR possible (Karkhoff-Schweizer *et al.*, 1995; Wagner *et al.*, 1998). Many of these sequences are phylogenetically divergent relative to those known sequences from cultured organisms (Liu *et al.*, 2003a,b; Nakagawa *et al.*, 2004).

According to sequence information deposited in GenBank and other publicly accessible sequence databases, 99 species have been identified to belong to this functional group spreading over five bacterial phyla (Chlorobi, Firmicutes, Nitrosospira, Proteobacteria, Thermodesulfobacteria) and two archaeal phyla (Crenarchaeota, Euryarchaeota), with members of the genus *Desulfovibrio* forming the largest group. However, more species existing in nature are able to perform sulfate respiration, some of which are likely to fall into novel to date yet unknown lineages as shown recently by Mori *et al.* (2003). Sequence comparison of these highly conserved genes over distant phylogenetic groups of archaeal and bacterial origin suggests that these genes may have been horizontally transferred in some lineages (Larsen *et al.*, 1999; Klein *et al.*, 2001; Friedrich, 2002). Thus, *dsr*-gene sequences can only partially be used as a phylogenetic marker for sulfate and sulfite reducing microorganisms, but they are very suitable for functional diversity studies.

Organic Contaminant Degradation

There are thousands of different gene sequences available in public databases encoding various enzymes that transform one or more of hundreds of complex organic chemicals. A review of all available contaminant degradation genes is beyond the scope of this chapter, so for the purposes of illustrating the diversity of available degradative genes and how they may be used in microarrays, we have focused on the naphthalene degradation pathway. Naphthalene is the most widely studied member of the polycyclic aromatic hydrocarbons (PAHs) which are composed of multiple, fused aromatic rings. In addition to naturally occurring sources, environmental PAH-contamination often results from anthropogenic deposition of various fossil fuel-derived chemicals such as the wood preservative creosote (Sun *et al.*, 2003).

In Table 11.1, we have listed the enzymes which sequentially convert naphthalene to salicylate (Bosch *et al.*, 1999; Takizawa *et al.*, 1999). The initial reaction in the pathway occurs via a multi-component naphthalene dioxygenase (NDO) (EC 1.14.12.12) (Kauppi *et al.*, 1998). The α subunit of the NDO iron sulfur protein is believed to confer the specificity of the enzyme and, as indicated in Table 11.1, is the most studied of the naphthalene genes (Parales *et al.*, 2000; Wackett, 2002). The other three NDO subunits probably have limited impact on enzymatic specificity (Parales *et al.*, 1998; Romine *et al.*, 1999). For these reasons, the α subunit is potentially the best candidate for use as a microarray marker gene. It has been identified in numerous eubacteria including *Burkholderia*, *Comamonas*, *Cycloclasticus*, *Marinobacter*, *Neptunomonas*, *Polaromonas*, *Pseudoalteromonas*, *Pseudomonas*, *Ralstonia* and *Rhodococcus* spp. (Kurkela *et al.*, 1988; Denome *et al.*, 1993; Takizawa *et al.*, 1994; Fuenmayor *et al.*, 1998; Hedlund *et al.*, 1999, 2001; Larkin *et al.*, 1999; Melcher *et al.*, 2002; Jeon *et al.*, 2003; Kasai *et al.*, 2003). The α subunit gene is referred to by various names, including *doxB*, *nagAc*, *nahAc*, *narAa*, *ndoB*, *pahAc* and *phnAc*, that were initially chosen in part based on the substrate on which the host bacterium was isolated (Habe and Omori, 2003). Most of the α subunit sequence information from isolated organisms is derived from Gram negative bacteria, primarily *Pseudomonas* spp. containing *nah*-like NDO genes (Habe and Omori, 2003). But, it has recently been demonstrated that other NDO genes such as the *phn* genes, of which much less is known, may be prevalent in environmental samples (Lloyd-Jones *et al.*, 1999; Wilson *et al.*, 2003). Information on naphthalene degradation genes in Gram positive bacteria is limited and even less information is available for the fungal genes (Andreoni *et al.*, 2000; Larkin *et al.*, 1999). Several different primers have been designed for the α subunit genes (Hamann *et al.*, 1999; Lloyd-Jones *et al.*, 1999; Wilson *et al.*, 1999; Baldwin *et al.*, 2003), but only a handful of studies have directly amplified and sequenced the genes from soil, sediment, or groundwater samples (Wilson *et al.*, 1999; Stach and Burns, 2002; Jeon *et al.*, 2003). To our knowledge, the diversity of the α subunit gene has not been thus assessed in marine environments even though the gene has been identified in

several marine bacteria (Geiselbrecht *et al.*, 1998; Hedlund *et al.*, 1999, 2001).

Another potential application of environmental microarrays is in the determination of horizontal gene transfer events. Gene transfer has contributed to the evolution of metabolic pathways and has also played a role in the spread of NDO genes (Herrick *et al.*, 1997; McGowan *et al.*, 1998; Habe and Omori, 2003; Wilson *et al.*, 2003). Bacteria have been identified that contain multiple copies of the *nah* genes, and it also appears that some isolates have mosaics of *nah* operons from different organisms (Bosch *et al.*, 1999; Ferrero *et al.*, 2002). Microarray technology could be used to quickly assess if an isolate contains multiple copies of a given gene and if some of the genes in a pathway are similar to those in one organism while other genes are similar to another organism. This knowledge could help to determine the factors involved in microbial adaptation following environmental contamination with xenobiotics (Rensing *et al.*, 2002; Top and Springael, 2003). However, for successful application of microarray technology to the study of pathway evolution, probes would be required for all of the genes in a given pathway, which in the case of naphthalene would necessitate more sequence information from many organisms for the other genes besides the NDO α subunit.

Metals Resistance

A recent search of public databases revealed 2303 sequences either identified as, or similar to, genes encoding microbial resistance to 15 different metals and metalloids. Sequences for the various mercury resistance genes were the most numerous; therefore, we have focused on these genes for illustration. Mercury compounds are widely distributed around the Earth, and their presence in the environment can occur through natural or anthropogenic processes with the latter estimated to account for approximately two-thirds of the worldwide Hg input (Mason *et al.*, 1994). Not surprisingly, mercury resistance is among the most common phenotypes observed in bacteria (Barkay *et al.*, 2003). It has been proposed that the emergence of the basic mercury resistance genes (*mer*) predates the divergence of Gram negative and positive bacteria (Osborn *et al.*, 1997). Of the *mer* functional genes, *merA*, which encodes the mercuric ion reductase enzyme (EC 1.16.1.1) that converts Hg^{2+} to the volatile species Hg^{0}, is the most studied and has the most available sequence data (Table 11.1) (Barkay *et al.*, 2003; Nascimento and Chartone-Souza, 2003). The *merA* gene is therefore a good candidate for microarray fabrication and monitoring bacterial mercury resistance, although other genes may be more appropriate for specific samples such as *merB* encoding organomercurial lyase (EC 4.99.1.2) for samples contaminated with organomercury compounds.

The *merA* gene is widely spread among both Gram negative and Gram positive bacteria including *Alcaligenes*, *Bacillus*, *Delftia*, *Exiguobacterium*, *Pantoea*, *Pseudomonas*, *Shigella*, *Staphylococcus* and *Xanthomonas* spp.

(Laddaga *et al.*, 1987; Yurieva *et al.*, 1997; Bogdanova *et al.*, 1998; Reniero *et al.*, 1998; Kholodii *et al.*, 2000; Venkatesan *et al.*, 2001; Sota *et al.*, 2003). Whole genome sequences of archaea have also indicated the presence of *mer*-like sequences, and the first report of functional, archaeal *merA* and *merR* genes has recently been published for *Sulfolobus solfataricus* (Schelert *et al.*, 2004). There is, however, very limited sequence data in the literature and databases regarding the diversity of *merA* genes from uncultured microorganisms in environmental samples. In fact, the only culture-independent *merA* sequences we found were from plasmids isolated from sewage sludge and soil via an exogenous plasmid isolation method (Schluter *et al.*, 2003; Schneiker *et al.*, 2001). Researchers have amplified *merA* from environmental samples; however, the PCR products were not sequenced but analyzed with other procedures (Felske *et al.*, 2003; Hart *et al.*, 1998). Further information on the diversity of uncultured *merA* sequences would be useful prior to microarray construction since the sequences from isolated organisms may not comprehensively represent the genetic diversity in the environment (Bruce *et al.*, 1995; Marchesi and Weightman, 2003). Additionally, only limited information is available for mercury resistance genes in fungi, although several genes involved in mercury (metal) resistance in yeast have recently been identified which may help to expand this knowledge base in the future (Furuchi *et al.*, 2002; Nguyên-nhu and Knoops, 2002; Westwater *et al.*, 2002; Gueldry *et al.*, 2003).

◆◆◆◆◆◆ DESIGN OF SPECIFIC OLIGONUCLEOTIDE PROBES FOR FGAs

Oligonucleotide-based microarrays are becoming more popular because they offer a number of advantages over cDNA microarrays. First, as stated previously, only sequence information is required and PCR amplification can therefore be avoided. Secondly, more flexibility to control specificity of hybridization can be achieved in probe design by the ability to strictly delimit parameters such as melting temperature (T_m), overall similarity (% homology) and other factors. Thirdly, oligonucleotide synthesis costs have dropped considerably in the last few years (Relogio *et al.*, 2002). In addition, oligonucleotide arrays provide potential solutions to some of the more complicated problems involved in environmental studies. For example, short oligonucleotides may be used to avoid highly conserved regions of orthologous genes that would not be possible with PCR amplification using conserved primers. The challenge for probe design is how to identify the optimum probes for each gene or each group of genes.

There are a number of pre-existing programs available for automated selection of oligonucleotide probes for DNA microarrays (Table 11.2). OligoArraySelector (Zhu *et al.*, 2003) runs on Linux/Unix systems and uses a BLAST approach to search for sequence similarity and compute the thermodynamic properties for only the most probable non-specific

Table 11.2. Oligonucleotide probe design programs

Name	OS	Reference
ArrayOligoSelector	Linux	Zhu *et al.* (2003), http://sourceforge.net/projects/arrayoligosel/
OligoArray	Windows and Unix/Linux	Rouillard *et al.* (2002), http://berry.engin.umich.edu/oligoarray
OligoArray 2.0	Unix/Linux	Rouillard *et al.* (2003), http://berry.engin.umich.edu/oligoarray2
OligoPicker	Linux	Wang and Seed (2003), http://pga.mgh.harvard.edu/oligopicker/index.html
OligoWiz (Web-based)	Unix	Nielsen *et al.* (2003), http://cbs.dtu.dk/services/oligowiz/
PRIMEGENS	Unix/Linux	Xu *et al.* (2002), http://compbio.ornl.gov/structure/primegens/
PROBEmer (Web-based)	Linux	Emrich *et al.* (2003), http://probemer.cs.loyola.edu
ProbeSelect	Unix/Linux	Li and Stormo (2001)
ROSO (Web-based)	Windows and Unix	Reymond *et al.* (2004), http://pbil.univ-lyon1.fr/roso
ArrayDesigner (Commercial)	N/A	TeleChem International Inc., http://arrayit.com
Sarani Goldminer (Commercial)	N/A	Strand Genomics, http://mail.strandgenomics.com/index.html

hybridization. Oligopicker (Wang and Seed, 2003) runs on Linux platforms and relies on BLAST search and 15-base stretch filtering. Those two programs select 70mer oligonucleotides for whole genomes. OligoArray (Rouillard *et al.*, 2002) can run on Windows, Unix or Linux systems and the oligonucleotide specificity is checked using BLAST (Altschul *et al.*, 1997) and possible secondary structures are predicted by the Mfold server (Zuker *et al.*, 1999). Its sister program, OligoArray 2.0 (Rouillard *et al.*, 2003), runs on Linux or Unix systems and the probe specificity is based on a comparison of sequence similarity between the specific target and putative non-specific targets. PRIMEGENS (Xu *et al.*, 2002) uses BLAST search and sequence alignment to select gene-specific fragments and then feeds those fragments to the Primer3 program (Rozen and Skaletsky, 2000) to design PCR primer pairs or probes on a genome scale. The program runs on Linux or Unix platforms and can also be used from a web interface. ProbeSelect (Li and Storomo, 2001) runs on Linux or Unix and uses a suffix tree to search for sequence similarity and the Myersgrep program (Myers, 1998) to search for matching sequences with few mismatches. This program can choose short (20–25 bases) or long (50 or 70 bases) oligonucleotides. Recently, some web-based probe design programs have been developed. OligoWiz (Nielsen *et al.*, 2003) is implemented as a client-server application. The server is responsible for the calculation of scores and utilizes the BLAST program for homology search. The client is used to submit jobs to the server, to visualize the scores and to fine-tune the placement of oligonucleotides. PROBEmer (Emrich *et al.*, 2003) uses suffix tree-based algorithms to identify common substrings. The program can design oligonucleotide probes for a single sequence or a defined group of sequences (16S rRNA gene) or PCR primers. ROSO (Reymond *et al.*, 2004) separates the time-consuming BLAST search from the fast step of thermodynamic analysis. The program can be used to select oligonucleotide probes or PCR primers. Probe design parameters, such as oligonucleotide length, number of probes for each gene and target T_m can be changed by users for some of the programs described.

Some of the programs mentioned above were used to design probes for the whole genome of *Methanococcus maripaludis* and a separate group of sequences of *nirS* and *nirK* (nitrite reductase) based on those publicly available and our own sequence collections. The results are summarized in Table 11.3. Most programs worked well for the whole genome data. However, serious problems occurred when they were used to design 50mer oligonucleotides for the *nirSK* group of sequences. First, too few specific probes and too many non-specific probes were designed. For the purposes of this test we categorized probes as non-specific if they had $>85\%$ similarity, identical stretches >15 bp, or mismatch free energy <-30 kcal/mol. Secondly, a large majority of sequences did not have probes if only unique oligonucleotides were selected. A similar situation would be expected in probe selection for many other genes that might be used for FGAs because the nature of sequence data for these arrays is quite different from whole genome data. For example, sequences for FGAs are often highly homologous, and many sequences originating

Table 11.3. Number and specificity of designed probes (50mer) by different programs

Programs used	Whole genome sequences of *M. maripaludis* (1766 ORFs)					Group sequences of *nirSK* (842 gene sequences)				
	Total ORFs	ORFs rejected	Probes designed	Specific probes	Non-specific	Total ORFs	ORFs rejected	Probes designed	Specific probes	Non-specific
ArrayOligoSelector	1766	7	1759	1415	344	842	0	842	117	725
OligoArray	1766	68	1698	1654	44	842	35	807	70	737
OligoArray 2.0	1766	68	1698	1464	234	842	51	791	35	756
OligoPicker	1766	18	1748	1745	3	842	657	185	141	44
CommOligo	1766	21	1745	1745	0	842	695	147	147	0

from phylogenetic and environmental studies are incomplete. Additionally, as little is known about the diversity that might be encountered in any given environmental sample, probe design should focus on the conserved, well-known, regions of a sequence. For example, when full-length sequences and shorter environmental sequences are used together, most algorithms will design what are thought to be unique probes for the full-length sequences outside of the conserved region used for environmental data.

For these reasons, a new probe design software tool called CommOligo is currently being developed and tested in our laboratory that will select optimal oligonucleotide probes for whole genomes, meta-genomes or groups of orthologous sequences such as those involved in FGAs. A multiple sequence alignment (MSA) approach is used to pre-process sequence data, and then users can choose regions for designing probes by masking based on the MSA results. The program then uses a new global alignment algorithm to design single or multiple unique probes for each gene and designate allowable parameters such as maximal similarity (default = 85%), maximal number of continuous match stretches (15 bases) and free energy (-30 kcal/mol) all of which can be controlled independently and simultaneously. The program is also able to design single or multiple group-specific probes for related groups of genes if it is not possible to select unique probes for a sequence. This new algorithm selects probes which have maximal similarities within a group and minimal similarities outside groups. Using defaults, group-specific probes should have a minimal similarity of 96% within a group and the same parameters as unique probes outside a group. Other filters, such as self-binding, mismatch position and GC content can also be used. While the above parameters are used as defaults, users may adjust all these parameters and values to meet their own needs. The program was evaluated using both whole genome and orthologous gene sequence data and compared with other software. For example, 147 specific (unique) probes were designed for the *nirS* and *nirK* sequences, and most importantly, this program did not choose any non-specific probes (Table 11.3). A group-specific probe algorithm has been evaluated using small to medium size data sets with documented phylogenetic relationships. For example, for *nirS* and *nirK* sequences, the program automatically formed 59 groups based on the default settings and each group had 2–30 sequences. Single or multiple probes could be designed for each group. Those group-specific probes covered an additional 180 sequences for which no unique probes were possible for single sequences because of their close similarities. For this particular data set, only 40% of sequences had unique or group-specific probes under default conditions. Relaxing design parameters will produce more probes but it also could potentially jeopardize the probe quality.

In summary, the application of oligonucleotide arrays for environmental studies presents many problems for probe design. To achieve the optimal specificity of oligonucleotide arrays, probe design criteria need to be further investigated and better algorithms are needed to facilitate these intensive computations. We are currently still in the pre-release stage of

testing solutions for this problem presented in the discussion of CommOligo. In the meantime, great care must be taken when applying tools originally designed for whole genome data to the design of FGA probes.

◆◆◆◆◆◆ MICROARRAY CONSTRUCTION, LABELING, HYBRIDIZATION AND IMAGE ACQUISITION

Reproducibility is one of the most critical requirements for microarray fabrication. For reliable and reproducible data, the uniformity of spots across the entire array is crucial to simplify image analysis and enhance the accuracy of signal detection. While array construction (oligo synthesis and printing) of FGAs does not significantly differ from that of other types of arrays, several recommendations can be made. Various factors will affect the uniformity of spots including array substrate, slide quality, printing 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. Additionally, the printing buffer is critical for obtaining homogeneous spots. Using saline sodium citrate (SSC) buffer, the spot homogeneity as well as binding efficiency is often poor, largely because of high evaporation rates. We and others have found more uniform spots can be obtained with the printing buffer containing 50% DMSO (dimethyl sulfoxide) and between 50 and 100 pmol/ul probe concentration (Hegde *et al.*, 2000; Diehl *et al.*, 2001; Wu *et al.*, 2001; Tiquia *et al.*, 2004). We use this in combination with aminopropyl silane coated glass slides (e.g. UltraGAPS, Corning, Corning, NY) and UV cross-linking at 200 mJ. In general, more cross-linking time or energy may bind oligos more strongly to slides, but may also interfere with proper hybridization. However, the individual slide manufacturer's recommendations for printing, cross-linking and pre-hybridization should always be consulted, as slight variations in slide chemistry and preparation procedure greatly affect these processes.

Protocol 1: Microarray Printing

1. Prepare printing oligo probe solution in a 384-well, v-bottom, printing plate. Final concentration will be 50–100 pmol/μl probe and 50% DMSO (generally 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 microtiter plates at 500 rpm for 5 min.
4. Setup the array printer and software (we use a PixSys 5500 printer; Cartesian technologies, Inc. Irvine, CA). 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 at least 2 h before proceeding to UV cross-linking and post-processing (according to slide manufacturer's protocol).

Total genomic DNAs are generally used as targets for functional gene studies. Thus, effective and repeatable DNA extraction from the environment is therefore a key step for FGA studies. We suggest several criteria for evaluating extraction methods following Hurt *et al.* (2001): (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 labeling and hybridization. (3) The extraction and purification protocol should be robust and repeatable. The DNA extraction and purification protocol described by Zhou *et al.* (1996) and modified for simultaneous DNA and mRNA extraction by Hurt *et al.* (2001), fulfill the above criteria for soils and sediments. However, other methods may be suitable or superior depending upon the sample type of interest.

Direct labeling procedures and PCR labeling amplifications with Cy3 or Cy5 fluorescent dye modified deoxynucleotides (dNTPs) based on Schena *et al.* (1995) are the most common labeling methods for whole genome array studies and have also been used successfully in environmental samples (Wu *et al.*, 2001; Rhee *et al.*, 2004). For DNA samples, direct labeling with random primers and Klenow fragment DNA polymerase I is widely used. Given the current sensitivity limits for detection (Cho and Tiedje, 2002) and the diverse nature of microbial communities, the likelihood of detecting genes present in lower numbers will increase with the amount of DNA template used for hybridization. We are routinely able to efficiently label 2–5 μg of target DNA using the methods outlined below. Targets can also be labeled via PCR using gene-specific primers, however, this method introduces biases inherent in such procedures and is used most often for detection of specific targets that might be of low number in an environmental sample or for validation of probe specificity.

Protocol 2: Direct Community DNA Labeling Procedure

1. In a 0.2 ml PCR tube combine:
   (a) 2–5 μg[1] purified community DNA (in 10 μl nuclease-free water).
   (b) 20 μl (750 ng/μl) random octamer primers (Invitrogen # Y01393).
2. Mix well and denature at 99.9°C for 5 min.
3. Place immediately on ice.

[1] The DNA template amount will vary, but the higher the amount used, the higher the likelihood of detecting genes present in low numbers. Additional positive control templates may be added as well to the same labeling reactions.

4. 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) Klenow fragment (Invitrogen # Y01396).
   (d) 1.25 μl DTT (Invitrogen # Y00147).
   (e) 13.75 μl DNase- and RNase-free water.
5. Add this mixture to the 0.2 ml PCR from step 1 (total volume = 50 μl).
6. Mix well and centrifuge the mixture briefly at maximum speed.
7. Incubate at 37°C for 6 h or overnight.
8. Purify labeled target DNA using QIAquick PCR purification columns according to the manufacturer's instructions (Qiagen, Valencia, CA).
9. Quantify labeling efficiency as below.

Protocol 3: Quantifying Labeling Efficiency of Cy-Labeled DNA Targets

1. Use a spectrophotometer to quantify the OD at 550 for Cy3 and OD 650 for Cy5. Also, measure OD at 230, 260 and 280 to assess purity. This can be done by using only 1 μl of the labeled DNA and a NanoDrop™ ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Montchanin, DE) or equivalent.
2. Calculate the amount of DNA as well as the specific activity of the labeled DNA. The specific activity is calculated as follows:

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

3. Dry in vacuum centrifuge (45°C) for 1 h. Do not use higher heat levels or heat lamps to accelerate evaporation as the fluorescent dyes could be degraded.

Temperature, concentration of formamide and the volume of the hybridization mixture are critically important parameters for all microarray hybridizations. Temperature and formamide concentration together control the specificity of the resulting hybridization. While this is true of all microarray hybridizations, this can be especially important in FGAs as many orthologous and highly similar genes may be present in any given sample. Uneven hybridizations resulting from fluctuations in volume across the array can also result in spurious signal strengths and strong backgrounds. However, in FGA studies where detection limits are always an issue, it is highly desirable to minimize the volume of hybridization solution. We utilize a 22 × 22 mm glass LifterSlip cover slip (Erie Scientific, Portsmouth, NH) that allows even hybridizations with as little as 15 μl of hybridization solution. These methods and materials are not the only ones available; other procedures have been used successfully by different researchers. We present several protocols here for illustration of the steps involved based on those currently employed in our laboratory.

Protocol 4: Hybridization and Washing

Buffer	Volume (μl)	Final concentration
Nuclease-free water	3.3	
Formamide	7.5	50%
20× SSC	2.5	3.33 ×
10% SDS	0.5	0.33%
Herring sperm DNA (10 mg/ml)	1.2	10.2 μg

1. Preheat microarray slide in hybridization chamber (Corning #2551) for 20 min at 50°C.
2. Resuspend sample in hybridization solution, spin down and heat at 95°C for at least 5 min in a thermocycler.
3. Dispense 15 μl of 3× SSC solution into the chamber hydration wells.
4. Deposit the hybridization (15 μl) solution directly onto the immobilized DNA probes and place a cover slip over the array, avoid bubble formation.
5. Close the hybridization chamber and ensure a proper seal is formed.
6. Incubate the chamber in a 50°C water bath for 12–15 h (overnight).

Post-hybridization wash

1. Place the slides, with the coverslips still affixed, in a pre-warmed washing buffer I (2× SSC and 0.1% SDS) and allow the coverslips to fall from the slide.
2. Place the slides in a pre-warmed washing buffer I (2× SSC and 0.1% SDS) and wash for 5 min with gentle shaking. Repeat this wash once.
3. Place the slides into fresh buffer II (0.1× SSC and 0.1% SDS) at ambient temperature for 5 min. Repeat this wash once.
4. Place the slides in buffer III (0.1× SSC) at ambient temperature for 1 min. Repeat wash four times.
5. 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 microtiter plates. As evaporation can be quite rapid, it is suggested that the slide be placed in the centrifuge immediately upon removal from the jar to avoid residual salt deposition.
6. Slides should be stored in the dark until ready for scanning.

Notes. We found that non-specific hybridization could be significant when the microarray slides were not warmed or the hybridization mixture remained at the room temperature for several minutes after hybridization and prior to washing. To minimize potential non-specific hybridization, the slides should be pre-warmed and the hybridization mixture should be kept above the hybridization temperature through all hybridization steps prior to washing. The above post-washing procedure is based on that of Corning and may vary by slide manufacturer, but we have found it is critical to proceed immediately to the first wash step when slides are removed from the chamber.

Microarray image processing is critical to control signal variation due to high background and weak signals, and to remove false positive signals (Schuchhardt *et al.*, 2000). One of the critical steps in the analysis of microarray image data is filtering noise versus true signals. Problematic false signals introduced by impurities in the arrays can be identified and removed by visual examination of each spot. This is particularly a critical procedure for reducing misinterpretations in the final results. With the current sensitivity limits of environmental FGAs, it is most often required to use very high laser power and photomultiplier tube (PMT) settings for detection of microarray signals. Compared to whole genome microarrays, that are usually scanned at much lower settings, one may thus find very high levels of background for FGAs. Because of this, it is important to prevent impurities from being introduced to the process and to examine all spots critically to distinguish real signals and false signals (Figure 11.1). This can most easily be done after the image is imported to software such as ImaGene™ for spot identification and quantification, and before proceeding to data analysis. The image in Figure 11.2 shows typical hybridization results for an FGA of nitrogen and sulfur cycle genes (Laser power 100%, PMT gain 95%).

Protocol 5: Image Acquisition and Processing

1. Scan the slide initially at a low resolution of 50 μm to obtain a quick display image and then finally 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 observed in the low resolution scan so that the signals for most of the spots are not saturated.
2. Save the scanned display as a 16-bit TIFF and BMP file and quantify the intensity of each spot using ImaGene™ (BioDiscovery, Los Angeles, CA) or equivalent.
3. Assess spot quality and reliability, and perform background subtraction of the microarray data.

Notes. Besides ImaGene software, there are other software packages available for image processing, spot identification, quantitation and normalization. These imaging 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 TECHNIQUES

Data analysis techniques for FGAs should allow for the detection of genes that are significantly different between samples. However, FGAs pose

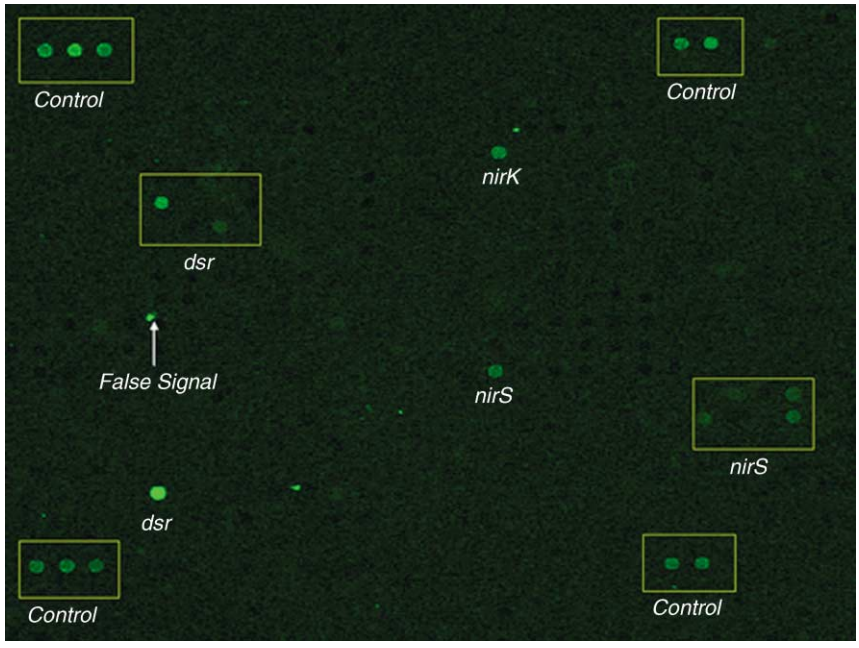

Figure 11.2. 50mer oligonucleotide microarray hybridization with a marine sediment genomic DNA sample. 3 ug of extracted marine sediment DNA was labeled Cy5 using methods outlined in the protocols outlined in this chapter. A small portion of an array image is shown to illustrate typical FGA images and true versus false signals.

a special challenge for environmental analysis since these samples often contain genomic DNA from a highly diverse range of organisms for which most of the genomic information is not known or knowable. Besides uncertain quantification of whole genome microarrays due to the inherently high variation associated with array fabrication, probe labeling, hybridization and image processing (Beißbarth *et al.*, 2000; Zhou and Thompson, 2002), the targeted genes in FGAs are diluted in a complex pool of genomic DNA from both target and non-target organisms. Thus, potential target genes make up only an extremely small portion of the overall DNA in any given sample and thus sensitivity is a major issue in this kind of microarray analysis.

There are several software programs available, which deal with the task of data preparation and subsequently can be used for statistical analysis and evaluation (e.g. *ArrayStat*, Imaging Research, Inc., Ontario, Canada; *Cluster* 3.0, Human Genome Center, University of Tokyo, Japan; *GenePix Pro*, Axon, Foster City, CA; *GeneSpring*, Silicon Genetics, Redwood City, CA). In our lab, we routinely use ImaGene 5.5.4 (BioDiscovery, El Segundo, CA). Unlike the analysis of expression data using whole genome microarrays, FGAs focusing on environmental data are hampered by a low amount of positive spots for the reasons stated above. Therefore, a good allocation of positive and negative spots might be one of the first challenges for this type of microarray. If too few positive

spots are present, placing the grid correctly can be very time consuming. We suggest users consider adding several positive control spots to each grid sector of a printed array. This allows for better spot locating ability as well as an easy way to implement localized quality control measures.

Because all raw microarray data, and especially FGAs, are affected by unspecific background binding of labeled DNA to the slides, correction for background signal intensity of the raw data is necessary. To perform this task, usually the mean or median value of the local background intensity of a spot is subtracted from the image intensities of the spot. Generally speaking, the application of median values is preferable because extreme outliers are not taken into calculations which could otherwise falsify the real value (Beißbarth *et al.*, 2000). However, both approaches have been used to process microarray data. The subtraction of local background data is especially imperative, whenever uneven background staining makes the usage of a global background values impossible. Other methods of background correction are also possible: Fluorescence intensity may also be corrected from empty spots (array positions that do not contain any DNA) and negative control spots (e.g. probes targeting human genes in soil or groundwater samples). To distinguish between background and real hybridizations, signal-to-noise ratios (SNR) are calculated and only those spots above a certain threshold (usually SNR ≥ 2 or 3) are considered as positive hybridizations. The SNR can be calculated as follows:

$$\text{SNR} = \frac{\text{signal mean} - \text{background mean}}{\text{background standard deviation}}.$$

However, the method of calculation sometimes varies between authors and thus the method must always be included when data are presented.

Meaningful interpretation and comparisons of microarray data require standardization of the measured raw signal intensities. Variability due to pre-hybridization handling and uneven labeling efficiency can be accounted for by the analysis of replicate microarray slides for each sample. However, if this does not lead to adequate information, raw data may be corrected by dividing their intensity values by negative control spots (Dennis *et al.*, 2003). If two or more samples are compared in two-color experiments, standardization of all data by calculating the ratio of cy5 labeled samples and a cy3 labeled control is always necessary. As mentioned above, in contrast to whole genome arrays, where overall expression levels between two differently labeled samples are compared, only a small fraction of genes can often be detected when environmental samples are hybridized to FGAs, thus quantitative comparisons (like generation of ratios and relative abundances of certain genes) in this type of microarray can only be generated for genes detectable in both samples (Talaat *et al.*, 2002). As a consequence, many standard normalization procedures are not applicable for this type of microarray analysis and the user should use great care in selecting such automated procedures.

In most cases, two or more replicates of each probe are printed on a microarray. While many software programs offer automatic calculation

of the accordant means, data concerning the overall variability might be lost and thus should rather be handled separately (Beißbarth *et al.*, 2000). Microarray data is not in most cases normally distributed. Logarithmic transformation ($\log_{10}$ or $\log_2$) is needed and the transformed values will thus largely reflect the degree of abundance or expression increase of a treated sample versus the control (Dennis *et al.*, 2003). More information concerning normalization and data transformations that goes beyond the scope of this chapter may be found at Zar (1999). However, new approaches for normalization and interpretation of microarray data are constantly being considered (Piétu *et al.*, 1996; Chen *et al.*, 1996; Richmond *et al.*, 1999; Beißbarth *et al.*, 2000; Dudley *et al.*, 2002; Talaat *et al.*, 2002), but the user must be aware that most have not been fully evaluated for application to FGAs and their inherently different problems.

After processing the raw data, several statistical tests can be employed to discover prominent genes within specific samples or relationships between different samples. This can be done by describing similarities and distances of data sets by methods such as cluster analysis, principal component analysis (PCA) or with the aid of self-organizing maps (SOM). Similarity comparisons are the most obvious methods to compare microarray data. They can be calculated either from Euclidean distances, which calculate the absolute distances between two data points in space, or as Pearson correlations, which are insensitive to the amplitude of the signal intensity and as a result are most often used. In PCA, a multivariant table with P columns in a P-dimensional (Euclidean) space is reduced to the two or three most representative dimensions. The first dimension explains as many of the differences in the data sets as possible and the second dimension as many as possible of what cannot be explained by the first one, and so on (Gilbert *et al.*, 2000). This approach has been widely employed to analyze microarray data for gene expression analysis (Hilsenbeck *et al.*, 1999; Thomas *et al.*, 2001; Spanakis and Brouty-Boyé, 1997).

Clustering analyses apply one of the above-mentioned similarity measurements to groups of genes with similar expression profiles, in order to detect clusters of genes that are presumably involved in a common process or that respond to a given treatment (Eisen *et al.*, 1998). A data set is divided in several subsets on the basis of their similarities. In hierarchical clustering, first the most closely related data are combined to form a cluster. Subsequently, the next cluster will be formed by a subset of two other data possibly including the already formed clusters. As a result, the data are combined to form a phylogenetic tree, the branch lengths of which represent the degree of similarity between the sets. However, several clustering methods exist, which might lead to slightly different interpretations (Gilbert *et al.*, 2000). Applied to FGAs, this type of analysis helps to identify similarities and differences in the genes present in the microbial communities at different sites as well as specific changes that might be due to a specific experimental treatment, etc. Such differences in clustering may indicate presumed differences in the function of interest, for example, nitrate reduction. SOM are a kind of artificial neural networks which can be considered as a type of mathematical cluster analysis. Data

are iteratively relocated leading to adjoining clusters with high similarities and more distant clusters further apart. Similar patterns will occur as neighbors in SOM (Tamayo *et al.*, 1999).

◆◆◆◆◆◆ EVALUATION AND VALIDATION OF FGA RESULTS

Probes should be highly specific for the target gene in order to prevent cross-hybridization from similar environmental sequences which could lead to inaccurate results (Rhee *et al.*, 2004; Wu *et al.*, 2001). Specificity can be assessed by test hybridizations using pure culture genomic DNAs and PCR amplified genes, or by synthetic oligonucleotides. While templates may not always be available for testing with genomic DNA or PCR product, synthetic oligonucleotides can be synthesized to test the specificity of any probe. Additionally, more comprehensive testing can be achieved with oligos, as mismatches can be designed in any number and at any position along the probe template hybrid. Using such methods, 50mer FGAs tested in our laboratory have been shown to be specific when % homology is not $>85–88\%$ (Tiquia *et al.*, 2004; Rhee *et al.*, 2004). However, the free energy of potential probe target hybrids is possibly a better predictor of specificity, and with the techniques outlined here we have found that probe-hybrid combinations with ΔG values of > -30 kcal/mol were very specific (Rhee *et al.*, 2004; Liebich, unpublished data). However, as in sensitivity measurements, these values are dependent upon the specifics of the protocols in use, and have to be empirically determined for each study.

Evaluation of the sensitivity (e.g. lower detection limit) and the quantitative relationship between hybridization strength and DNA amount are critical for understanding the results of microarray-based approaches for detecting genes in environmental samples. Dilutions of pure culture genomic DNA hybridized against corresponding probes provide a rather straightforward approach for measuring sensitivity and evaluating the quantitative nature of FGAs. However, unlike whole genome arrays, detection limits for FGA and other environmental microarrays must account for the effects of heterogeneous non-target DNA sequences characteristic of environmental samples (Cho and Tiedje, 2002; Rhee *et al.*, 2004). Cho and Tiedje (2002) proposed that the detection limit of a PCR fragment of the denitrification gene *nirS* in an environmental sample was $\sim$10 pg. This means only genes of organisms with a total DNA amount $\sim$50 ng are detectable (assuming a 4 Mbp average genome size). In their experiments, 1 μg of total environmental DNA was analyzed, suggesting that a particular organism containing *nirS* must contribute at least 1/20 of the applied DNA amount to be detectable. However, increasing the amount of environmental DNA applied and other methods can improve these detection limitations. Using the same protocols suggested in this chapter, Rhee *et al.* (2004) estimated that several genes involved in biodegradation of naphthalene could be

detected with 5 ng genomic DNA in the absence of background DNA. However, detection limits were 50–100 ng of pure culture genomic DNA when diluted in a heterogeneous background of *Shewanella oneidensis*. In this case, however, hybridizations were carried out with 5 μg of total sample DNA, suggesting the FGA could detect cells present at a level of about 1/50–1/100 of the total. After logarithmic transformation, the relationship between signal intensity and applied DNA amount is most often linear to at least 1 μg of target (Wu *et al.*, 2001; Cho and Tiedje, 2002; Tiquia *et al.*, 2004; Rhee *et al.*, 2004). However, the quantitative nature of the relationship is dependent upon the specifics of the protocols in use, and has to be empirically determined for each study. Ideally, quantitative controls should be introduced to every slide using a series of control spots and corresponding control DNA that is co-labeled with each sample (Chen *et al.*, 1996; Dudley *et al.*, 2002).

While FGAs have the potential to rapidly quantify thousands of different DNA/RNA sequences in environmental samples simultaneously (Cho and Tiedje, 2002; Dennis *et al.*, 2003; Rhee *et al.*, 2004; Tiquia *et al.*, 2004; Wu *et al.*, 2001), it may be useful to validate selected results using other techniques such as quantitative PCR (qPCR). For example, Rhee *et al.* (2004) used a real-time PCR-based qPCR approach to verify FGA data obtained from a PAH contaminated soil. The FGA analysis had indicated the presence of numerous organic contaminant degradation genes in the soil including several from the naphthalene catabolic pathway. The researchers designed primers for six of these naphthalene genes, four of which generated single PCR products from the soil DNA and were subsequently used for qPCR. The qPCR results corroborated the FGA data for each of the four genes by demonstrating significant correlations between the gene copy number and the FGA hybridization signals ($r^2 = 0.74$ for all genes and 0.96 for genes with SNR > 3). This approach for validation, however, is most useful when probes target relatively unique genes. When data originating from numerous orthologous gene sequences are used for probe design, the specific primers necessary for qPCR may not be possible to design.

Acknowledgements

The authors would like to thank Sonia Tiquia, Sung K. Rhee, and Liyou Wu for their pioneering work in the initial development of many of the protocols and ideas contained in this paper. The authors' efforts in preparing this chapter were supported by the US DOE Office of Science as part of its Biological and Environmental Research Programs in Natural and Accelerated Bioremediation Research, Genomes To Life, Biotechnology Investigations-Ocean Margins, and Carbon Sequestration (as part of the consortium on research to enhance Carbon Sequestration in Terrestrial Ecosystems-CSiTE). Jost Liebich's work was also supported by the postdoctoral program of the German Academic Exchange Service [DAAD]. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725.

References

Allen, A., Booth, M. G., Frischer, M. E., Verity, P. G., Zehr, J. P. and Zani, S. (2001). Diversity and detection of nitrate assimilation genes in marine bacteria. *Appl. Environ. Microbiol.* **67**, 5343–5348.

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Miller, W. and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Andreoni, V., Bernasconi, S., Colombo, M., van Beilen, J. B. and Cavalca, L. (2000). Detection of genes for alkane and naphthalene catabolism in *Rhodococcus* sp. strain 1BN. *Environ. Microbiol.* **2**, 572–577.

Baldwin, B. R., Nakatsu, C. H. and Nies, L. (2003). Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR. *Appl. Environ. Microbiol.* **69**, 3350–3358.

Barkay, T., Miller, S. M. and Summers, A. O. (2003). Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiol. Rev.* **27**, 355–384.

Beaulieu, M., Becaert, V., Deschenes, L. and Villemur, R. (2000). Evolution of bacterial diversity during enrichment of PCP-degrading activated soils. *Microb. Ecol.* **40**, 345–356.

Beißbarth, T., Fellenberg, K., Brors, B., Arribas-Prat, R., Boer, J. M., Hauser, N. C., Scheideler, M., Hoheisel, J. D., Schütz, G., Poustka, A. and Virgon, M. (2000). Processing and quality control of DNA array hybridization data. *Bioinformatics* **16**, 1014–1022.

Bodrossy, L., Stralis-Pavese, N., Jurrell, J. C., Radejewski, S., Weilharter, A. and Sessitsch, A. (2003). Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ. Microbiol.* **5**, 566–582.

Bogdanova, E. S., Bass, I. A., Minakhin, L. S., Petrova, M. A., Mindlin, S. Z., Volodin, A. A., Kalyaeva, E. S., Tiedje, J. M., Hobman, J. L., Brown, N. L. and Nikiforov, V. G. (1998). Horizontal spread of *mer* operons among Gram-positive bacteria in natural environments. *Microbiology* **144**, 609–620.

Bosch, R., Garcia-Valdes, E. and Moore, E. R. B. (1999). Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene* **236**, 149–157.

Braker, G. and Tiedje, J. M. (2003). Nitric oxide reductase (norB) genes from pure cultures and environmental samples. *Appl. Environ. Microbiol.* **69**, 3476–3483.

Bruce, K. D., Osborn, A. M., Pearson, A. J., Strike, P. and Ritchie, D. A. (1995). Genetic diversity within *mer* genes directly amplified from communities of noncultivated soil and sediment bacteria. *Mol. Ecol.* **4**, 605–612.

Cassidy, S. L. (2002). Microbial activity and biodiversity as indicators of hydrocarbon bioremediation. Thesis. University of London, United Kingdom.

Castro, H., Reddy, K. R. and Ogram, A. (2002). Composition and function of sulfate-reducing prokaryotes in eutrophic and pristine areas of the Florida Everglades. *Appl. Environ. Microbiol.* **68**, 6129–6137.

Chang, Y. J., Peacock, A. D., Long, P. E., Stephen, J. R., McKinley, J. P., Macnaughton, S. J., Hussain, A. K., Saxton, A. M. and White, D. C. (2001). Diversity and characterization of sulfate-reducing bacteria in groundwater at a uranium mill tailings site. *Appl. Environ. Microbiol.* **67**, 3149–3160.

Chen, Y., Dougherty, E. R. and Bittner, M. (1996). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.* **2**, 364–374.

Chien, Y. T. and Zinder, S. H. (1994). Cloning, DNA sequencing, and characterization of a nifD-homologous gene from the archaeon *Methanosarcina*

barkeri 227 which resembles nifD from the eubacterium *Clostridium pasteurianum*. *J. Bacteriol.* **176**, 6590–6598.
Cho, J. C. and Tiedje, J. M. (2002). Quantitative detection of microbial genes by using DNA microarrays. *Appl. Environ. Microbiol.* **68**, 1425–1430.
Dean, D. R., Bolin, J. T. and Zheng, L. M. (1993). Nitrogenase metalloclusters: structures, organization, and synthesis. *J. Bacteriol.* **175**, 6737–6744.
Denef, V. J., Park, J., Rodrigues, J. L. M., Tsoi, T. V., Hashsham, S. A. and Tiedje, J. M. (2003). Validation of a more sensitive method for using spotted oligonucleotide DNA microarrays for functional genomics studies on bacterial communities. *Environ. Microbiol.* **5**, 933–943.
Dennis, P., Edwards, E. A., Liss, S. N. and Fulthorpe, R. (2003). Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Appl. Environ. Microbiol.* **69**, 769–778.
Denome, S. A., Stanley, D. C., Olson, E. S. and Young, K. D. (1993). Metabolism of dibenzothiophene and naphthalene in *Pseudomonas* strains – complete DNA-sequence of an upper naphthalene catabolic pathway. *J. Bacteriol.* **175**, 6890–6901.
Deplancke, B., Hristova, K. R., Oakley, H. A., McCracken, V. J., Aminov, R., Mackie, R. I. and Gaskins, H. R. (2000). Molecular ecological analysis of the succession and diversity of sulfate-reducing bacteria in the mouse gastrointestinal tract. *Appl. Environ. Microbiol.* **66**, 2166–2174.
DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
Diehl, F., Grahlmann, S., Beier, M. and Hoheisel, J. D. (2001). Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res.* **29**, E38.
Duarte, G. F., Rosado, A. S., Seldin, L., de Araujo, W. and van Elsas, J. D. (2001). Analysis of bacterial community structure in sulfurous-oil-containing soils and detection of species carrying dibenzothiophene desulfurization (*dsz*) genes. *Appl. Environ. Microbiol.* **67**, 1052–1062.
Dubilier, N., Mulders, C., Ferdelman, T., de Beer, D., Pernthaler, A., Klein, M., Wagner, M., Erseus, C., Thiermann, F., Krieger, J., Giere, O. and Amann, R. (2001). Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**, 298–302.
Dudley, A. M., Aach, J., Steffen, M. A. and Church, G. M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc. Natl Acad. Sci.* **99**, 7554–7559.
Dunfield, P. F., Liesack, W., Henckel, T., Knowles, R. and Conrad, R. (1999). High-affinity methane oxidation by a soil enrichment culture containing a type II methanotroph. *Appl. Environ. Microbiol.* **65**, 1009–1014.
Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci.* **95**, 14863–14868.
Elsaied, H. and Naganuma, T. (2001). Phylogenetic diversity of ribulose-1, 5-bisphosphate carboxylase/oxygenase large-subunit genes from deep-sea microorganisms. *Appl. Environ. Microbiol.* **67**, 1751–1765.
Emrich, S. J., Lowe, M. and Delcher, A. (2003). PROBEmer: a web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res.* **31**, 3746–3750.
Felske, A. D. M., Fehr, W., Pauling, B. V., von Canstein, H. and Wagner-Dobler, I. (2003). Functional profiling of mercuric reductase (*mer* A) genes in biofilm communities of a technical scale biocatalyzer. *BMC Microbiol.* **3**, 22, (published online- http://www.biomedcentral.com/1471-2180/3/22).

Ferrero, M., Llobet-Brossa, E., Lalucat, J., Garcia-Valdes, E., Rossello-Mora, R. and Bosch, R. (2002). Coexistence of two distinct copies of naphthalene degradation genes in *Pseudomonas* strains isolated from the western Mediterranean region. *Appl. Environ. Microbiol.* **68**, 957–962.

Finegold, S. M. and Jousimies-Somer, H. (1997). Recently described clinically important anaerobic bacteria: medical aspects. *Clin. Infect. Dis.* **25**(Suppl. 2), S88–S93.

Friedrich, M. W. (2002). Phylogenetic analysis reveals multiple lateral transfers of adenosine-5′-phosphosulfate reductase genes among sulfate-reducing microorganisms. *J. Bacteriol.* **184**, 278–289.

Fuenmayor, S. L., Wild, M., Boyes, A. L. and Williams, P. A. (1998). A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. *J. Bacteriol.* **180**, 2522–2530.

Furuchi, T., Hwang, G. W. and Naganuma, A. (2002). Overexpression of the ubiquitin-conjugating enzyme Cdc34 confers resistance to methylmercury in *Saccharomyces cerevisiae*. *Mol. Pharmacol.* **61**, 738–741.

Geiselbrecht, A. D., Hedlund, B. P., Tichi, M. A. and Staley, J. T. (1998). Isolation of marine polycyclic aromatic hydrocarbon (PAH)-degrading *Cycloclasticus* strains from the Gulf of Mexico and comparison of their PAH degradation ability with that of Puget Sound *Cycloclasticus* strains. *Appl. Environ. Microbiol.* **64**, 4703–4710.

Gilbert, D. R., Schroeder, M. and van Helden, J. (2000). Interactive visualization and exploration of relationships between biological objects. *Trends Biotechnol.* **18**, 487–494.

Gueldry, O., Lazard, M., Delort, F., Dauplais, M., Grigoras, I., Blanquet, S. and Plateau, P. (2003). Ycf1p-dependent Hg(II) detoxification in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **270**, 2486–2496.

Habe, H. and Omori, T. (2003). Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Biosci. Biotechnol. Biochem.* **67**, 225–243.

Hallam, S. J., Girguis, P. R., Preston, C. M., Richardson, P. M. and DeLong, E. F. (2003). Identification of methyl coenzyme M reductase A (mcrA) genes associated with methane-oxidizing archaea. *Appl. Environ. Microbiol.* **69**, 5483–5491.

Hamann, C., Hegemann, J. and Hildebrandt, A. (1999). Detection of polycyclic aromatic hydrocarbon degradation genes in different soil bacteria by polymerase chain reaction and DNA hybridization. *FEMS Microbiol. Lett.* **173**, 255–263.

Hart, M. C., Elliott, G. N., Osborn, A. M., Ritchie, D. A. and Strike, P. (1998). Diversity amongst *Bacillus mer A* genes amplified from mercury resistant isolates and directly from mercury polluted soil. *FEMS Microbiol. Ecol.* **27**, 73–84.

Hedlund, B. P., Geiselbrecht, A. D., Bair, T. J. and Staley, J. T. (1999). Polycyclic aromatic hydrocarbon degradation by a new marine bacterium, *Neptunomonas naphthovorans* gen. nov., sp. nov. *Appl. Environ. Microbiol.* **65**, 251–259.

Hedlund, B. P., Geiselbrecht, A. D. and Staley, J. T. (2001). *Marinobacter* strain NCE312 has a *Pseudomonas*-like naphthalene dioxygenase. *FEMS Microbiol. Lett.* **201**, 47–51.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. and Quackenbush, J. (2000). A concise guide to cDNA microarray analysis. *BioTechniques* **29**, 548–560.

Herrick, J. B., StuartKeil, K. G., Ghiorse, W. C. and Madsen, E. L. (1997). Natural horizontal transfer of a naphthalene dioxygenase gene between bacteria native to a coal tar-contaminated field site. *Appl. Environ. Microbiol.* **63**, 2330–2337.

Hilsenbeck, S. G., Friedrichs, W. E., Schiff, R., O'Connell, P., Hansen, R. K., Osborne, C. K. and Fuqua, S. A. (1999). Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl Cancer Inst.* **91**, 53–459.

Holliger, C., Hahn, D., Harmsen, H., Ludwig, W., Schumacher, W., Tindall, B., Vazquez, F., Weiss, N. and Zehnder, A. J. (1998). *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetra- and trichloroethene in an anaerobic respiration. *Arch. Microbiol.* **169**, 313–321.

Holmes, A. J., Costello, A., Lidstrom, M. E. and Murrell, J. C. (1995). Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* **132**, 203–208.

Horz, H. P., Yimga, M. T. and Liesack, W. (2001). Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of *pmoA*, *mmoX*, *mxaF*, and 16S rRNA and ribosomal DNA including *pmoA*-based terminal restriction fragment length polymorphism profiling. *Appl. Environ. Microbiol.* **67**, 4177–4185.

Howard, J. B. and Rees, D. C. (1996). Structural basis of biological nitrogen fixation. *Chem. Rev.* **96**, 2965–2982.

Huber, H., Jannasch, H., Rachel, R., Fuchs, T. and Stetter, K. O. (1997). *Archaeoglobus veneficus* sp. nov., a novel facultative chemolithoautotrophic hyperthermophilic sulfite reducer, isolated from abyssal black smokers. *Syst. Appl. Microbiol.* **20**, 374–380.

Hurek, T., Egener, T. and Reinhold-Hurek, B. (1997). Divergence in nitrogenases of *Azoarcus* spp., Proteobacteria of the beta subclass. *J. Bacteriol.* **179**, 4172–4178.

Hurt, R. A., Qui, X., Wu, L., Roh, Y., Palumbo, A. V., Tiedje, J. M. and Zhou, J. (2001). Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* **67**, 4495–4503.

Jeon, C. O., Park, W., Padmanabhan, P., DeRito, C., Snape, J. R. and Madsen, E. L. (2003). Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proc. Natl Acad. Sci. USA* **100**, 13591–13596.

Jørgensen, B. B. (1982). Mineralization of organic-matter in the sea bed – the role of sulfate reduction. *Nature* **296**, 643–645.

Karkhoff-Schweizer, R. R., Huber, D. P. and Voordouw, G. (1995). Conservation of the genes for dissimilatory sulfite reductase from *Desulfovibrio vulgaris* and *Archaeoglobus fulgidus* allows their detection by PCR. *Appl. Environ. Microbiol.* **61**, 290–296.

Kasai, Y., Shindo, K., Harayama, S. and Misawa, N. (2003). Molecular characterization and substrate preference of a polycyclic aromatic hydrocarbon dioxygenase from *Cycloclasticus* sp. strain A5. *Appl. Environ. Microbiol.* **69**, 6688–6697.

Kauppi, B., Lee, K., Carredano, E., Parales, R. E., Gibson, D. T., Eklund, H. and Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Struct. Fold. Des.* **6**, 571–586.

Kholodii, G., Yurieva, O., Mindlin, S., Gorlenko, Z., Rybochkin, V. and Nikiforov, V. (2000). Tn*5044*, a novel Tn3 family transposon coding for temperature-sensitive mercury resistance. *Res. Microbiol.* **151**, 291–302.

Kirshtein, J. D., Paerl, H. W. and Zehy, J. (1991). Amplification, cloning, and sequencing of a nifH segment from aquatic microorganisms and natural communities. *Appl. Environ. Microbiol.* **57**, 2645–2650.

Kleikemper, J., Schroth, M. H., Sigler, W. V., Schmucki, M., Bernasconi, S. M. and Zeyer, J. (2002). Activity and diversity of sulfate-reducing bacteria

in a petroleum hydrocarbon-contaminated aquifer. *Appl. Environ. Microbiol.* **68**, 1516–1523.

Klein, M., Friedrich, M., Roger, A. J., Hugenholtz, P., Fishbain, S., Abicht, H., Blackall, L. L., Stahl, D. A. and Wagner, M. (2001). Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J. Bacteriol.* **83**, 6028–6035.

Koizumi, Y., Kelly, J. J., Nakagawa, T., Urakawa, H., El-Fantroussi, S., Al-Muzaini, S., Fukui, M., Urushigawa, Y. and Stahl, D. A. (2002). Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA–DNA membrane hybridization, and DNA microarray technology. *Appl. Environ. Microbiol.* **68**, 3215–3225.

Kurkela, S., Lehvaslaiho, H., Palva, E. T. and Teeri, T. H. (1988). Cloning, nucleotide-sequence and characterization of genes encoding naphthalene dioxygenase of *Pseudomonas putida* strain NCIB9816. *Gene* **73**, 355–362.

Laddaga, R. A., Chu, L. E., Misra, T. K. and Silver, S. (1987). Nucleotide sequence and expression of the mercurial-resistance operon from *Staphylococcus aureus* plasmid pI258. *Proc. Natl Acad. Sci. USA* **84**, 5106–5110.

Larkin, M. J., Allen, C. C. R., Kulakov, L. A. and Lipscomb, D. A. (1999). Purification and characterization of a novel naphthalene dioxygenase from *Rhodococcus* sp. strain NCIMB12038. *J. Bacteriol.* **181**, 6200–6204.

Larsen, O., Lien, T. and Birkeland, N. K. (1999). Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: phylogenetic and structural implications from gene sequences. *Extremophiles* **3**, 63–70.

Laue, H., Friedrich, M., Ruff, J. and Cook, A. M. (2001). Dissimilatory sulfite reductase (desulfoviridin) of the taurine-degrading, non-sulfate-reducing bacterium *Bilophila wadsworthia* RZATAU contains a fused DsrB–DsrD subunit. *J. Bacteriol.* **183**, 1727–1733.

Leaphart, A. B. and Lovell, C. R. (2001). Recovery and analysis of formyltetrahydrofolate synthetase gene sequences from natural populations of acetogenic bacteria. *Appl. Environ. Microbiol.* **67**, 1392–1395.

Li, F. and Stormo, G. (2001). Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**, 1067–1076.

Liu, X., Tiquia, S. M., Holguin, G., Wu, L., Nold, S. C., Devol, A. H., Luo, K., Palumbo, A., Tiedje, J. M. and Zhou, J. (2003a). Molecular diversity of denitrifying genes in continental margin sediments within the oxygen-deficient zone off the pacific coast of Mexico. *Appl. Environ. Microbiol.* **69**, 3549–3560.

Liu, Y., Zhou, J., Omelchenko, M., Beliaev, A., Venkateswaran, A., Stair, J., Wu, L., Thompson, D. K., Xu, D., Rogozin, I. B., Gaidamakova, E. K., Zhai, M., Makarova, K. S., Koonin, E. V. and Daly, M. J. (2003b). Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc. Natl Acad. Sci. USA* **100**, 4191–4196.

Lloyd-Jones, G., Laurie, A. D., Hunter, D. W. F. and Fraser, R. (1999). Analysis of catabolic genes for naphthalene and phenanthrene degradation in contaminated New Zealand soils. *FEMS Microbiol. Ecol.* **29**, 69–79.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680.

Lovley, D. R. (1997). Potential for anaerobic bioremediation of BTEX in petroleum-contaminated aquifers. *J. Ind. Microbiol. Biotechnol.* **18**, 75–81.

Lovley, D. R. and Klug, M. J. (1983). Sulfate reducers can out-compete methanogens at fresh-water sulfate concentrations. *Appl. Environ. Microbiol.* **45**, 187–192.

Lyons, J. I., Newell, S. Y., Buchan, A. and Moran, M. A. (2003). Diversity of ascomycete laccase gene sequences in a southeastern U.S. salt marsh. *Microb. Ecol.* **45**, 270–281.

Manz, W., Eisenbrecher, M., Neu, T. R. and Szewzyk, U. (1998). Abundance and spatial organization of Gram-negative sulfate-reducing bacteria in activated sludge investigated by in situ probing with specific 16S rRNA targeted oligonucleotides. *FEMS Microbiol. Ecol.* **25**, 43–61.

Marchesi, J. R. and Weightman, A. J. (2003). Comparing the dehalogenase gene pool in cultivated α-halocarboxylic acid-degrading bacteria with the environmental metagene pool. *Appl. Environ. Microbiol.* **69**, 4375–4382.

Martin-Laurent, F., Piutti, S., Hallet, S., Wagschal, I., Philippot, L., Catroux, G. and Soulas, G. (2003). Monitoring of atrazine treatment on soil bacterial, fungal and atrazine-degrading communities by quantitative competitive PCR. *Pest. Manag. Sci.* **59**, 259–268.

Mason, R. P., Fitzgerald, W. F. and Morel, F. M. M. (1994). The biogeochemical cycling of elemental mercury – anthropogenic influences. *Geochim. Cosmochim. Acta* **58**, 3191–3198.

McGowan, C., Fulthorpe, R., Wright, A. and Tiedje, J. M. (1998). Evidence for interspecies gene transfer in the evolution of 2,4-dichlorophenoxyacetic acid degraders. *Appl. Environ. Microbiol.* **64**, 4089–4092.

McMahon, K. D., Dojka, M. A., Pace, N. R., Jenkins, D. and Keasling, J. D. (2002). Polyphosphate kinase from activated sludge performing enhanced biological phosphorus removal. *Appl. Environ. Microbiol.* **68**, 4971–4978.

Melcher, R. J., Apitz, S. E. and Hemmingsen, B. B. (2002). Impact of irradiation and polycyclic aromatic hydrocarbon spiking on microbial populations in marine sediment for future aging and biodegradability studies. *Appl. Environ. Microbiol.* **68**, 2858–2868.

Metcalfe, A. C., Krsek, M., Gooday, G. W., Prosser, J. I. and Wellington, E. M. H. (2002). Molecular analysis of a bacterial chitinolytic community in an upland pasture. *Appl. Environ. Microbiol.* **68**, 5042–5050.

Molitor, M., Dahl, C., Molitor, I., Schafer, U., Speich, N., Huber, R., Deutzmann, R. and Truper, H. G. (1998). A dissimilatory sirohaem-sulfite-reductase-type protein from the hyperthermophilic archaeon *Pyrobaculum islandicum*. *Microbiology* **144**, 529–541.

Mori, K., Kim, H., Kakegawa, T. and Hanada, S. (2003). A novel lineage of sulfate-reducing microorganisms: Thermodesulfobiaceae fam. nov., *Thermodesulfobium narugense*, gen. nov., sp. nov., a new thermophilic isolate from a hot spring. *Extremophiles* **7**, 283–290.

Murrell, J. C., Gilbert, B. and McDonald, I. R. (2000). Molecular biology and regulation of methane monooxygenase. *Arch. Microbiol.* **173**, 325–332.

Murrell, J. C., McDonald, I. R. and Bourne, D. G. (1998). Molecular methods for the study of methanotroph ecology. *FEMS Microbiol. Ecol.* **27**, 103–114.

Myers, E. W. (1998). A fast bit-vector algorithm for approximate string matching based on dynamic programming. *Ninth Combinatorial Pattern Matching Conference*. Springer, Piscataway, NJ, pp. 1–13.

Nakagawa, T., Nakagawa, S., Inagaki, F., Takai, K. and Horikoshi, K. (2004). Phylogenetic diversity of sulfate-reducing prokaryotes in active deep-sea hydrothermal vent chimney structures. *FEMS Microbiol. Lett.* **232**, 145–152.

Nascimento, A. M. and Chartone-Souza, E. (2003). Operon mer: bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genet. Mol. Res.* **2**(92), 101.

Nguyên-nhu, N. T. and Knoops, B. (2002). Alkyl hydroperoxide reductase 1 protects *Saccharomyces cerevisiae* against metal ion toxicity and glutathione depletion. *Toxicol. Lett.* **135**, 219–228.

Nielsen, H. B., Wernersson, R. and Knudsen, S. (2003). Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays. *Nucleic Acids Res.* **31**, 3491–3496.

Nold, S. C., Zhou, J. Z., Devol, A. H. and Tiedje, J. M. (2000). Pacific northwest marine sediments contain ammonia-oxidizing bacteria in the beta subdivision of the Proteobacteria. *Appl. Environ. Microbiol.* **66**, 4532–4535.

Normand, P. and Bousquet, J. (1989). Phylogeny of nitrogenase sequences in Frankia and other nitrogen-fixing microorganisms. *J. Mol. Evol.* **29**, 436–447.

Normand, P., Gouy, M., Cournoyer, B. and Simonet, P. (1992). Nucleotide sequence of nifD from Frankia alni strain ARI3: phylogenetic inferences. *Mol. Biol. Evol.* **9**, 495–506.

Oger, C., Mahillon, J. and Petit, F. (2003). Distribution and diversity of a cadmium resistance (*cadA*) determinant and occurrence of IS257 insertion sequences in Staphylococcal bacteria isolated from a contaminated estuary (Seine, France). *FEMS Microbol. Ecol.* **43**, 173–183.

Osborn, A. M., Bruce, K. D., Strike, P. and Ritchie, D. A. (1997). Distribution, diversity and evolution of the bacterial mercury resistance (*mer*) operon. *FEMS Microbiol. Rev.* **19**, 239–262.

Parales, R. E., Emig, M. D., Lynch, N. A. and Gibson, D. T. (1998). Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems. *J. Bacteriol.* **180**, 2337–2344.

Parales, R. E., Lee, K., Resnick, S. M., Jiang, H. Y., Lessner, D. J. and Gibson, D. T. (2000). Substrate specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme. *J. Bacteriol.* **182**, 1641–1649.

Park, W., Padmanabhan, P., Padmanabhan, S., Zylstra, G. J. and Madsen, E. L. (2002). *nahR*, encoding a LysR-type transcriptional regulator, is highly conserved among naphthalene-degrading bacteria isolated from a coal tar waste-contaminated site and in extracted community DNA. *Microbiology* **148**, 2319–2329.

Philippot, L., Piutti, S., Martin-Laurent, F., Hallet, S. and Germon, J. C. (2002). Molecular analysis of the nitrate-reducing community from unplanted and maize-planted soils. *Appl. Environ. Microbiol.* **68**, 6121–6128.

Piétu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Samson, R., Houlgatte, R., Soularue, P. and Auffray, C. (1996). Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* **6**, 492–503.

Rabus, R., Hansen, T. and Widdel, F. (2000). The dissimilatory sulfate- and sulfur reducing Prokaryotes. In *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community* (M. Dworkin et al., eds), 3rd edn. Springer, New York, release 3.3, http://link.springer-ny.com/link/service/books/10125/.

Relogio, A., Schwager, C., Richter, A., Ansorge, W. and Valcarcel, A. (2002). Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res.* **30**, e51.

Reniero, D., Mozzon, E., Galli, E. and Barbieri, P. (1998). Two aberrant mercury resistance transposons in the *Pseudomonas stutzeri* plasmid pPB. *Gene* **208**, 37–42.

Rensing, C., Newby, D. T. and Pepper, I. L. (2002). The role of selective pressure and selfish DNA in horizontal gene transfer and soil microbial community adaptation. *Soil Biol. Biochem.* **34**, 285–296.

Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G. and Fayard, J.-M. (2004). ROSO: optimizing oligonucleotide probes for microarrays. *Bioinformatics* **20**, 271–273.

Rhee, S. K., Liu, X., Wu, L., Chong, S. C., Wan, X. and Zhou, J. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl. Environ. Microbiol.* **70**, 4303–4317.

Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. and Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**, 3821–3835.

Romine, M. F., Stillwell, L. C., Wong, K. K., Thurston, S. J., Sisk, E. C., Sensen, C., Gaasterland, T., Fredrickson, J. K. and Saffer, J. D. (1999). Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *J. Bacteriol.* **181**, 1585–1602.

Rouillard, J.-M., Herbert, C. and Zuker, M. (2002). OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics* **18**, 486–487.

Rouillard, J.-M., Zuker, M. and Gulari, E. (2003). OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using thermodynamic approach. *Nucleic Acids Res.* **31**, 3057–3062.

Rozen, S. and Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. *Meth. Mol. Biol.* **13**, 365–386.

Scala, D. and Kerkhof, L. J. (1999). Diversity of nitrous oxide reductase (nosZ) genes in continental shelf sediments. *Appl. Environ. Microbiol.* **65**, 1681–1687.

Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M. and Blum, P. (2004). Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.* **2**, 427–437.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. and Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA* **93**, 10614–10619.

Schluter, A., Heuer, H., Szczepanowski, R., Forney, L. J., Thomas, C. M., Puhler, A. and Top, E. M. (2003). The 64 508 bp IncP-1β antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1β group. *Microbiology* **149**, 3139–3153.

Schneiker, S., Keller, M., Droge, M., Lanka, E., Puhler, A. and Selbitschka, W. (2001). The genetic organization and evolution of the broad host range mercury resistance plasmid pSB102 isolated from a microbial population residing in the rhizosphere of alfalfa. *Nucleic Acids Res.* **29**, 5169–5181.

Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. and Herzel, H. (2000). Normalization strategies for cDNA microarrays. *Nucleic Acids Res.* **28**, E47.

Singleton, R. Jr. (1993). The sulfate-reducing bacteria: an overview. In *The Sulfate-Reducing Bacteria: Contemporary Perspectives* (J. M. Odom and R. Singleton Jr., eds), *Brock/Springer Series in Contemporary Bioscience*, pp. 1–20. Springer, New York.

Sota, M., Kawasaki, H. and Tsuda, M. (2003). Structure of haloacetate-catabolic IncP-1β plasmid pUO1 and genetic mobility of its residing haloacetate-catabolic transposon. *J. Bacteriol.* **185**, 6741–6745.

Spanakis, E. and Brouty-Boyé, D. (1997). Discrimination of fibroblast subtypes by multivariate analysis of gene expression. *Int. J. Cancer* **71**, 402–409.
Stach, J. E. M. and Burns, R. G. (2002). Enrichment versus biofilm culture: a functional and phylogenetic comparison of polycyclic aromatic hydrocarbon-degrading microbial communities. *Environ. Microbiol.* **4**, 169–182.
Sullivan, J. P., Dickinson, D. and Chase, H. A. (1998). Methanotrophs, *Methylosinus trichosporium* OB3b, sMMO, and their application to bioremediation. *Crit. Rev. Microbiol.* **24**, 335–373.
Sun, C. G., Snape, C. E., McRae, C. and Fallick, A. E. (2003). Resolving coal and petroleum-derived polycyclic aromatic hydrocarbons (PAHs) in some contaminated land samples using compound-specific stable carbon isotope ratio measurements in conjunction with molecular fingerprints. *Fuel* **82**, 2017–2023.
Takizawa, N., Iida, T., Sawada, T., Yamauchi, K., Wang, Y. W., Fukuda, M. and Kiyohara, H. (1999). Nucleotide sequences and characterization of genes encoding naphthalene upper pathway of *Pseudomonas aeruginosa* PaK1 and *Pseudomonas putida* OUS82. *J. Biosci. Bioeng.* **87**, 721–731.
Takizawa, N., Kaida, N., Torigoe, S., Moritani, T., Sawada, T., Satoh, S. and Kiyohara, H. (1994). Identification and characterization of genes encoding polycyclic aromatic hydrocarbon dioxygenase and polycyclic aromatic hydrocarbon dihydrodiol dehydrogenase in *Pseudomonas putida* OUS82. *J. Bacteriol.* **176**, 2444–2449.
Talaat, A. M., Howard, S. T., Hale IV, W., Lyons, R., Garner, H. and Jonston, S. A. (2002). Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res.* **30**, e104.
Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. and Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci.* **96**, 2907–2912.
Taroncher-Oldenburg, G., Griner, E. M., Francis, C. A. and Ward, B. B. (2002). Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl. Environ. Microbiol.* **69**, 1159–1171.
Thomas, J. G., Olson, J. M., Tapscott, S. J. and Zhao, L. P. (2001). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res.* **11**, 1227–1236.
Thompson, D. K., Beliaev, A. S., Giometti, C. S., Tollaksen, S. L., Khare, T., Lies, D. P., Nealson, K. H., Lim, H., Yates III, J., Brandt, C. C., Tiedje, J. M. and Zhou, J. (2002). Transcriptional and proteomic analysis of a ferric uptake regulator (Fur) mutant of *Shewanella oneidensis*: possible involvement of Fur in energy metabolism, transcriptional regulation, and oxidative stress. *Appl. Environ. Microbiol.* **68**, 881–892.
Tiquia, S. M., Wu, L., Chong, S. C., Passovets, S., Xu, D., Xu, Y. and Zhou, J. (2004). Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* **36**, 664–675.
Top, E. M. and Springael, D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr. Opin. Biotechnol.* **14**, 262–269.
Ueda, T., Suga, Y., Yahiro, N. and Matsuguchi, T. (1995). Genetic diversity of N2-fixing bacteria associated with rice roots by molecular evolutionary analysis of a nifD library. *Can. J. Microbiol.* **41**, 235–240.
Venkatesan, M. M., Goldberg, M. B., Rose, D. J., Grotbeck, E. J., Burland, V. and Blattner, F. R. (2001). Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect. Immun.* **69**, 3271–3285.
Vitousek, P. M. and Howarth, R. W. (1991). Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13**, 87–115.

Wackett, L. P. (2002). Mechanism and applications of Rieske non-heme iron dioxygenases. *Enzyme Microb. Technol.* **31**, 577–587.
Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A. and Stahl, D. A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* **180**, 2975–2982.
Wang, X. and Seed, B. (2003). Selection of oligonucleotide probes for protein coding sequences. *Bioinformatics* **19**, 796–802.
Westwater, J., McLaren, N. F., Dormer, U. H. and Jamieson, D. J. (2002). The adaptive response of *Saccharomyces cerevisiae* to mercury exposure. *Yeast* **19**, 233–239.
Widmer, F., Shaffer, B. T., Porteous, L. A. and Seidler, R. J. (1999). Analysis of nifH gene pool complexity in soil and litter at a Douglas fir forest site in the Oregon Cascade mountain range. *Appl. Environ. Microbiol.* **65**, 374–380.
Wilson, M. S., Bakermans, C. and Madsen, E. L. (1999). In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. *Appl. Environ. Microbiol.* **65**, 80–87.
Wilson, M. S., Herrick, J. B., Jeon, C. O., Hinman, D. E. and Madsen, E. L. (2003). Horizontal transfer of *phnAc* dioxygenase genes within one of two phenotypically and genotypically distinctive naphthalene-degrading guilds from adjacent soil environments. *Appl. Environ. Microbiol.* **69**, 2172–2181.
Wu, L. Y., Thompson, D. K., Li, G., Hurt, R. A., Tiedje, J. M. and Zhou, J. (2001). Development and evaluation of functional gene arrays for detection of selected gene in the environment. *Appl. Environ. Microbiol.* **67**, 5780–5790.
Xu, D., Li, G., Wu, L., Zhou, J. and Xu, Y. (2002). PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis. *Bioinformatics* **18**(11), 1432–1437.
Ye, R. W., Tao, W., Beddzyk, L., Young, T., Chen, M. and Li, L. (2000). Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J. Bacteriol.* **182**, 4458–4465.
Yurieva, O., Kholodii, G., Minakhin, L., Gorlenko, Z., Kalyaeva, E., Mindlin, S. and Nikiforov, V. (1997). Intercontinental spread of promiscuous mercury-resistance transposons in environmental bacteria. *Mol. Microbiol.* **24**, 321–329.
Zar, J. H. (1999). *Biostatistical Analysis*. Prentice-Hall, Englewood Cliffs, NJ.
Zehr, J. P. and McReynolds, L. A. (1989). Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium* spp. *Appl. Environ. Microbiol.* **55**, 2522–2526.
Zehr, J. P., Mellon, M., Braun, S., Litaker, W., Steppe, T. and Paerl, H. W. (1995). Diversity of heterotrophic nitrogen fixation genes in a marine cyanobacterial mat. *Appl. Environ. Microbiol.* **61**, 2527–2532.
Zhou, J. (2003). Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol.* **6**, 288–294.
Zhou, J. and Thompson, D. K. (2002). Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol.* **13**, 204–207.
Zhou, J., Bruns, M. A. and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 461–468.
Zhu, J., Bozdech, Z. and DeRisi, J. (2003). *ArrayOligoSelector*. http://sourceforge.net/projects/arrayoligosel/.
Zuker, M., Mathews, D. H. and Turner, D. H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology* (J. Barciszewski and B. F. C. Clark, eds), pp. 11–43. Kluwer Academic Publishers, Dordrecht.