

Microarray Applications in Microbial Ecology Research

T.J. Gentry, G.S. Wickham, C.W. Schadt, Z. He and J. Zhou

(1) Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Received: 23 February 2006 / Accepted: 7 April 2006 / Online publication: 8 August 2006

Abstract

Microarray technology has the unparalleled potential to simultaneously determine the dynamics and/or activities of most, if not all, of the microbial populations in complex environments such as soils and sediments. Researchers have developed several types of arrays that characterize the microbial populations in these samples based on their phylogenetic relatedness or functional genomic content. Several recent studies have used these microarrays to investigate ecological issues; however, most have only analyzed a limited number of samples with relatively few experiments utilizing the full high-throughput potential of microarray analysis. This is due in part to the unique analytical challenges that these samples present with regard to sensitivity, specificity, quantitation, and data analysis. This review discusses specific applications of microarrays to microbial ecology research along with some of the latest studies addressing the difficulties encountered during analysis of complex microbial communities within environmental samples. With continued development, microarray technology may ultimately achieve its potential for comprehensive, high-throughput characterization of microbial populations in near real time.

Introduction

Research over the last two decades has begun to reveal the incredible diversity of microorganisms in natural environ-

ments. Torsvik *et al.* [91, 92, 94], for example, have estimated that as many as 10,000 bacterial species may be present in a gram of surface soil or marine sediment. A more recent study has estimated the number of distinct microbial genomes in soil to be over a million [29]—two orders of magnitude greater than earlier estimates. However, the factors that control these populations, and their often extreme heterogeneity in the environment, are largely unknown [111]. The ultimate goal of microbial ecology research is to elucidate these factors by examining the interactions of microorganisms with each other and the biotic and abiotic characteristics of the environment. The development of technologies such as polymerase chain reaction (PCR) fingerprinting, reverse transcriptase PCR, real-time PCR, reporter genes, and fluorescence *in situ* hybridization (FISH), have made it possible to study the dynamics of simple communities or small groups of dominant organisms in natural settings. However, to fully understand the ecology of complex environments such as surface soils, it is necessary to analyze the dynamics and/or activity of hundreds to thousands of different microbial populations simultaneously.

Microarrays have the unprecedented potential to achieve this objective as specific, sensitive, quantitative, and high-throughput tools for microbial detection, identification, and characterization in natural environments. Due to rapid advances in printing technology, microarrays can now be produced that contain thousands to hundreds of thousands of probes. Microarrays have been primarily developed and used for gene expression profiling of pure cultures of individual organisms, but major advances have recently been made in their application to environmental samples. However, the analysis of environmental samples presents several challenges not encountered during the analysis of pure cultures. Like most other techniques, microarrays currently detect only the dominant populations in many environmental samples [20, 72]. In addition, some environments contain low levels of biomass, making it difficult to obtain enough material for use in

Present address of T.J. Gentry: Department of Soil & Crop Sciences, Texas A&M University, College Station, TX 77843, USA

Present address of Z. He: Institute for Environmental Genomics (IEG), Department of Botany and Microbiology, University of Oklahoma, 101 David L. Boren Blvd., Norman, OK 73019, USA

Correspondence to: J. Zhou at present address: Institute for Environmental Genomics (IEG), Department of Botany and Microbiology, University of Oklahoma, 101 David L. Boren Blvd., Norman, OK 73019, USA; E-mail: jzhou@ou.edu

microarray analysis without first amplifying the nucleic acids. Such techniques, even if applied with the utmost care, may introduce biases into the analyses [71], but perhaps the greatest challenge to the analysis of environmental samples using microarrays is the vast number of unknown DNA sequences in these samples. The importance of an organism, which may be dominant and critical to the ecosystem under study, can be completely overlooked if the organism does not have a corresponding probe on the array. Probes designed to be specific to known sequences can also cross-hybridize to similar, unknown sequences from related or unrelated genes, resulting in either an underestimated signal due to weaker binding of a slightly divergent sequence or a completely misleading signal due to binding of a different gene. Furthermore, it is often a challenge to analyze microarray results from environmental samples due to the massive amounts of data generated and a lack of standardized controls and data analysis procedures.

Despite these challenges, several types of microarrays have been successfully applied to microbial ecology research. These arrays can be divided into at least five categories based on the genes targeted by the array: (1) Phylogenetic oligonucleotide arrays (POAs) are designed based on a conserved marker such as the 16S ribosomal RNA (rRNA) gene, which is used to compare the relatedness of communities in different environments. (2) Functional gene arrays (FGAs) are designed for key functional genes that code for proteins catalyzing various biogeochemical processes, such as the carbon, nitrogen, and sulfur cycles and may also provide information on the

microbial populations controlling these processes. (3) Community genome arrays (CGAs) contain the whole genomic DNA of cultured organisms and can describe a community based on its relationship to these cultivated organisms. (4) Metagenomic arrays (MGA) are a potentially powerful technique because, unlike the other arrays, they contain probes produced directly from environmental DNA itself and can be applied with no prior sequence knowledge of the community. (5) Whole-genome open reading frame (ORF) arrays (WGA) contain probes for all of the ORFs in one or multiple genomes. These arrays have traditionally been used for functional genomic analyses of individual organisms, but they can also be used for comparative genomic analyses or to investigate the interactions of multiple organisms at the transcriptional level (Table 1).

In this review, we primarily discuss specific applications of these five major types of arrays to microbial ecology research along with the challenges of applying this technology to environmental samples and some of the latest research addressing these issues. Earlier reviews that may provide additional information of interest have also been published [6, 17, 79, 80, 112, 113].

Phylogenetic Oligonucleotide Arrays

Small-Subunit Ribosomal RNA as a Phylogenetic Marker. Although a number of biomolecules have potential as phylogenetic markers, it was the studies of Carl Woese and colleagues, beginning in the 1970s, that identified rRNA as uniquely suited for molecular

Table 1. Characteristics of microarrays for microbial ecology research^a

Property	Type of array				
	POA	FGA	CGA	MGA	WGA
Probe template	Ribosomal rRNA genes	Functional genes	Whole genome	Environmental DNA	Open reading frames in whole genome
Probe length	~18–25 nt	~50–70 nt oligos or ~200–1000 nt PCR products	Whole genome	~1,000+ nt	~25 nt oligos or ~200–3000 nt PCR products
Targeted microorganisms	Cultured and uncultured	Cultured and uncultured	Cultured	Cultured and uncultured	Cultured
Information provided	Phylogenetic	Functional	Phylogenetic	Functional	Phylogenetic and functional
Specificity	Species level or single nucleotide difference	<80–90% sequence homology	Species–strain	≥Strain	Single nucleotide (oligos)/strain (PCR-based probes)
Sensitivity (ng of pure genomic DNA)	~500 ^b	~1–8	~0.2	Undetermined	~0.1 (PCR-based probes)
Quantitative	Depends on array design ^b	Yes	Yes	Undetermined	Depends on array design

POA = phylogenetic oligonucleotide array; FGA = functional gene array; CGA = community genome array; WGA = whole-genome ORF array; MGA = metagenomic array.

^aAdapted from Zhou [112].

^bUndetermined for POAs based on perfectly matched and mismatched probe pairs.

phylogenetic analyses [98]. Several factors make rRNA in general, and 16S-like, or small-subunit rRNA (ssu rRNA), in particular, ideal for the study of evolutionary relationships [59]: (1) The rRNAs are found in all organisms, enabling a universal phylogeny. (2) Lateral transfer of rRNAs between organisms is extremely rare, ensuring that the evolutionary history of the rRNA reflects the evolutionary history of the organism. (3) The longer rRNAs (16S- and 23S-like) contain regions of highly conserved, moderately variable, and highly variable sequence. The highly conserved regions act as alignment guides to ensure that only homologous nucleotides are compared among organisms. In addition, conserved regions serve as convenient sites to which “universal” primers can be annealed for sequencing or for amplification by the PCR. Conversely, the variable regions serve as targets for group- and organism-specific hybridization probes and provide a phylogenetic signal for determining the relationships among organisms.

Most POAs to date have contained short oligonucleotide probes complementary to specific regions of the ssu rRNA gene. The use of rRNAs for the assessment of microbial diversity in naturally occurring microbial communities was first conceptualized, developed, and applied by Pace and co-workers in the early 1980s [64] and has since revolutionized our view of microbial diversity on earth, providing a framework for comparing and relating microorganisms to one another based on the evolutionary information contained in this conserved and universally distributed molecule. Inherent in this approach is the abrogation of the need to culture microorganisms in the laboratory before they can be characterized. This “characterization without cultivation” has freed microbiologists from the onerous, often seemingly impossible burden of cultivating microorganisms in the laboratory.

Lest there be any doubt about the impact of rRNA-based methodology on the field of microbial ecology, today, more than 20 years after the publication of the first article that used rRNA to characterize uncultivated naturally occurring microbial communities [84], the latest issues¹ of two of the leading environmental microbiology journals, *Environmental Microbiology* and *Microbial Ecology*, devoted a full 60% and 67%, respectively, to studies that employed rRNA-based approaches, whereas the microbial ecology section of *Applied and Environmental Microbiology* committed 67% of its research articles to the subject.

However, the power and utility of POAs lay not only in the unique properties of the rRNA and the high-

throughput capacity of microarray technology, but also in the vast amount of rRNA sequence data available via the Ribosomal Database Project (RDP) [16, 43, 65]. Now in its ninth release, the RDP II contains 136,355 aligned bacterial rRNA sequences: 57,325 from cultured microorganisms and 79,030 from environmental clones [16]. The sequences span the breadth of known phylogenetic diversity within the bacterial domain as well as archaeal and eukaryal domains.

In addition, a major difference between POAs and FGAs is that the former are well suited for the design of broad-range group and universal probes, whereas, due to the degeneracy of the genetic code, FGA probes are more typically designed to be only organism specific.

Secondary Structure and Probe Design. One of the major challenges for rRNA-based analysis is the inherent secondary structural properties of these molecules. There are numerous software programs that can identify self-complementarity in oligonucleotide probes, but it is more difficult to predict the effects of target secondary structure on hybridization efficiency. In a study not restricted to rRNAs, computer models suggested that many random secondary structures may be formed by genome-wide RNA transcripts and single-stranded DNA targets due to intramolecule base pairing [69]. This could result in decreased binding to the microarray-bound oligonucleotide probes and corresponding signal reduction or elimination. These models suggested that greater than 50% of Mfold-predicted RNA conformers could assume stable secondary structures [116]. Likewise, close to 30% of the nucleotides in single-stranded DNA molecules were potentially involved in stable secondary structure conformations at 67°C [69]. The effect was even more pronounced in RNAs, where greater than 60% of the nucleotides were involved in stable secondary structures at that temperature [69].

However, unlike the situation in mRNA and single-stranded DNA where medium to long stretches of secondary structure occur primarily as a consequence of random base pairing, the secondary structure of rRNA has evolved to be specific and ordered, allowing the problem of target accessibility to be overcome by judicious use of the extensive rRNA sequence and secondary structure databases as aids in probe design. The secondary structures of the ssu [104] and large subunit² (lsu) [60] rRNAs were originally established more than two decades ago by a combination of comparative sequence analysis, T₁ oligonucleotide catalogs, and chemical and enzymatic modification experiments [61, 105]. The explosion in the number of rRNA sequences from both

¹ The latest issues at the time of this writing were the March 2006, January 2006, and February 2006 issues of *Environmental Microbiology*, *Microbial Ecology*, and *Applied and Environmental Microbiology*, respectively.

² For the remainder of the review, when we refer to the rRNA, we will be referring specifically to the ssu rRNA.

cultured and uncultured microorganisms in that time has enabled additional comparative sequence analysis, resulting in the refinement of the original secondary structure model to the point that the single- and double-stranded regions of the molecule can be reliably predicted [32]. As such, there are at least three ways to overcome the problem of target secondary structure: designing probes to target regions of rRNA not involved in secondary structure, shearing or fragmenting the target before hybridization, and disrupting intrastand pairing in the target rRNA. The first solution is obvious given the ready availability of reliable secondary structure models [10], but this severely restricts the regions of rRNA that can serve as a target. Thus, we will limit our discussion to the latter two approaches.

Several investigators have studied the effects of rRNA target secondary structure on oligonucleotide microarray-based detection [12, 13, 66, 83]. Generally, these studies suggest that false positives (nonspecific binding of oligonucleotide probes to rRNA targets) can be eliminated by judicious choice of hybridization conditions: high hybridization temperature and/or the inclusion of formamide in the hybridization buffer [66]. Elimination of false negatives (instances where the probe is unable to bind as a consequence of secondary structure in the target molecule) is another concern, however. Under hybridization conditions where false positives were virtually eliminated, Peplies *et al.* [66] found that 17 of 41 array-bound, rRNA-specific oligonucleotide probes were unable to bind to their specific rRNA targets. Although significant, this problem may be overcome by fragmenting or shearing the target before hybridization [12, 13, 69] or by the addition of oligonucleotide “helper probes” to the hybridization mixture [66]; the helper probe is designed to specifically bind to the target adjacent to the probe binding site and disrupt the local secondary structure. It should be pointed out, however, that stabilization by base stacking may result in nonspecific binding of the rRNA target if the helper probe is designed to bind too closely to the 5′ or 3′ end of the capture probe binding site [12]. In addition, disruption of secondary structure in one region may result in the formation of suboptimal, yet stable, secondary structures in other regions of the target, affecting the binding sites of other probes [66]. Since the helper probes also need to have a level of specificity similar to the capture probes, it will be much more difficult to design suitable helper probes and use this approach with higher density arrays.

It is possible that the limitations imposed on probe design by the secondary structure of the rRNA target may be overcome altogether. In two recent studies, Yilmaz *et al.* [109, 110] have shown that all regions of the ssu rRNA from *Escherichia coli* are accessible by oligonucleotide probes as long as the thermodynamic affinity of probe for target is sufficient (average $\Delta G^{\circ}_{\text{overall}} = -13.5$

kcal/mol) and the incubation period of the hybridization is extended to optimize the kinetics of target unfolding and probe binding. If the results of these experiments, which were performed on whole cells in solution, are extensible to ssu rRNA hybridizations where the probes are bound to a solid support, and the target percent GC is significantly higher than that of the ssu rRNA from *E. coli*, then it should considerably relax the current theoretical restrictions in probe design.

Due to the conserved nature of rRNA genes, it is often necessary to use short oligonucleotides (~20-mer) for POAs in order for the probes to be specific to individual organisms. One of the more common formats consists of arraying multiple probes that perfectly match a given target along with corresponding probes containing a single mismatch, usually at the central position [26, 66, 98, 103]. Greater signal intensity for the perfectly matched probes compared to the mismatched probes indicates detection of the target sequence. This approach enables very specific detection of target sequences but does have some potential disadvantages. These drawbacks are discussed in a later section on specificity along with potential approaches to improve specificity including the use of dissociation profiles for probe–target duplexes.

Applications. The POAs are among the most commonly used microarrays to date due to the widespread application of rRNA as a tool for characterizing naturally occurring microbial communities. There are several noteworthy studies that have employed POAs in environmental investigation of both total microbial populations using the rRNA gene as the target [11, 53, 55, 76] and active microbial populations using rRNA as the target (Table 2) [1, 26, 42, 83]. These studies have examined a diverse set of environments including lake water and sediments [11, 76], estuary sediments [26] and enrichments [42], soil extracts [55, 83], activated sludge [1], and hypersaline cyanobacterial mats [53]. However, the targeted organisms have been more restricted in scope, focusing primarily on ammonia oxidizers [1], cyanobacteria [11, 76], and metal- and sulfate-reducing prokaryotes [26, 42, 53, 55, 83].

In one of the most illustrative ecological applications of POAs to date, Loy *et al.* [55] used a POA containing 132 probes (18-mer) to characterize sulfate-reducing bacteria at four depths (ranging from 0 to 30 cm) in two acidic, low-sulfate fens (wetland soils) in Germany. The POA consisted of probes specific to the rRNAs of individual and groups of organisms, spanning and inclusive of all known lineages and individuals of sulfate-reducing bacteria. The fens differed in iron content, vegetation, acidity, and to some degree, seasonal water saturation. The POA results indicated that stable sulfate-reducing populations varied little with depth within each of the two

Table 2. Highlights from selected studies that applied microarray analyses to microbial ecology research]

Type of array	No. probes	Probe length (nt)	Sample	DNA or RNA	Process investigated/major results	Ref.
POA	132	18	Fen soils	rDNA	Detection of sulfate-reducing prokaryotes	[55]
POA	31,179 ^a	20	Air	rDNA	Characterization of airborne bacteria and fungi. Largest POA published	[103]
POA	10	16–22	Enrichment, estuary sediment	rRNA	Detection of sulfate-reducing bacteria in a toluene-degrading and an ethylbenzene-degrading consortium	[42]
POA	8	18–20	Activated sludge	rRNA	Combined with isotope techniques for parallel detection of ammonia-oxidizing bacteria and CO ₂ fixation in nitrifying activated sludge	[1]
POA	12	17–22	Soil extract	rRNA	Detected metal- and sulfate-reducing bacteria. First report of direct rRNA detection (without PCR) in nonaqueous environmental samples	[83]
POA	22	16–22	Sediment cores	rRNA	Direct detection of rRNA from estuary sediments without PCR amplification	[26]
FGA	763	50	Marine sediment	DNA	Detection of nitrifying, denitrifying, and sulfate-reducing bacteria	[90]
FGA	104	~300–800	Soil, marine sediment	DNA	Indicated differences in nitrifying and denitrifying bacterial communities in soil and marine sediment	[106]
FGA	68	17–27	Soil	DNA	Linked methanotrophic community structure in landfill-simulating lysimeters with vegetation and level of methane oxidation	[86]
FGA	22	45 & 75	River sediment	DNA	Detection of polychlorinated biphenyl-degrading bacteria. Modified labeling method for increased sensitivity	[20]
FGA	64	70	River sediment	DNA	Denitrifier communities differed between samples implying an environmental gradient effect on microbial populations	[88]
FGA	64	~271–1300	Mixed culture, reactor	mRNA	Increased expression of functional genes during degradation of the herbicide 2,4-D and resin acid	[21]
FGA	1662	50	Enrichment, soil	DNA, mRNA	Microbial composition and activity during degradation of naphthalene. Largest FGA published to date	[72]
CGA	67	Whole genome	River and marine sediment, soil	DNA	Demonstrated differences in microbial communities from different sources. Results consistent with biogeochemical data. Allowed differentiation of pure-culture bacteria to the strain level	[107]
WGA	~285,000 ^a	25	Legume nodule	mRNA	Simultaneous gene expression in legume and bacterial symbiont. Results indicated most bacterial, but not legume, genes induced in nodule were linked to N fixation	[2]
MGA	~672	~1000	Riverine biofilm enrichment	DNA	Composition of biofilm enrichment from aquifer. Approach allowed detection and isolation of genes from uncultured organisms	[81]

Abbreviations as in Table 1.

^aAlso contained the same number of mismatch probes.

sites but were different between the sites. Members of the *Syntrophobacteraceae* were detected in the upper 30 cm of both sites, but *Desulfomonile* spp. were only found in one soil, which also contained a more diverse sulfate-reducing community. These results were confirmed by direct PCR amplification with the appropriate group-specific rRNA primers and by the detection of the corresponding *dsrAB* genes from the samples. Development of this particular "PhyloChip" was made possible by the availability of a large sulfate-reducing bacteria-specific rRNA probe database, dubbed "ProbeBase"³ [54]. This is particularly noteworthy because it illustrates the importance and power of a comprehensive probe database in the POA-based analysis of naturally occurring microbial communities. More recently, the same research group developed a POA targeting all of the cultured and uncultured members of the *Rhodocyclales* [56]. The array detected *Rhodocyclales* populations representing less than 1% of the total community, following *Rhodocyclales*-selective PCR amplification. The POA indicated the presence of several uncultured *Zoogloea*-, *Ferribacterium/Dechloromonas*-, and *Sterolibacterium*-like organisms in activated sludge from an industrial wastewater treatment plant, which was corroborated by the results from a 16S rRNA gene clone library. The results also demonstrated that the *Rhodocyclales* community in the reactor, thought to represent the major denitrifiers in the system, had dramatically changed, possibly as a result of alterations in treatment plant operations.

The most comprehensive POA developed so far contained 31,179 perfectly matched hierarchical 20-mer probes (with a corresponding number of single mismatch probes as negative, mismatch controls) targeting 1945 prokaryotic and 431 eukaryotic sequences from the RDP [103]. The array was developed with the Affymetrix system in which oligonucleotides were directly synthesized on the array, thus enabling higher printing densities than possible using other methods. The POA correctly identified 15 of 17 tested pure bacterial cultures. The array was then used to investigate microorganisms collected from a 1.4-million-liter air sample. The POA results generally agreed with those from an rRNA gene clone library, but could only resolve differences to the third level of phylogenetic rank, as defined by RDP, and could not identify individual species. Eight of 10 phylogenetic clusters detected by the array were represented in the rRNA gene clone libraries, and the organisms not detected had relatively low signals on the array. Approximately 7% of the clones were not detected by the POA, but these were from novel organisms not represented in the RDP or on the array. In contrast, there was not a good correlation

between the relative numbers of clones in each group and the signal intensity of that group detected by the array indicating a potential limitation with respect to microbial quantitation with this system.

Despite the excitement generated by the combination of the rRNA-based phylogenetic analysis with the high-throughput potential of microarrays, there have been relatively few studies that have used POAs for comprehensive environmental studies, and there are still technological limitations to POA analysis as the previously discussed study illustrates [103]. As these technological issues continue to be resolved, POAs will undoubtedly find wide application in microbial ecology research.

Functional Gene Arrays

Selection of Probe Targets. Unlike POAs, which are designed primarily for the detection of specific microorganisms and phylogenetic differentiation between samples, FGAs measure genes involved in some process of interest, and thus not only provide a degree of phylogenetic classification but also give information on genetic capacity for, or activity of, a given process in the environment under study. Genes encoding key enzymes in metabolic processes are often good targets, and several categories of these genes have been used for FGAs including those involved in biogeochemical cycles [5, 15, 38, 72, 85, 86, 88, 90, 106] and contaminant remediation [15, 20, 21, 72]. A key point to consider when selecting functional genes for inclusion on a FGA is the vast differences in available sequence data for various genes, even within a given pathway. For example, for the microbial nitrogen fixation genes *nifD*, *nifH*, and *nifK*, there are 1784 *nifH* genes in public databases, but only 89 *nifK* and 180 *nifD* gene sequences are available [79]. An ideal candidate gene for an FGA (1) encodes a critical enzyme or protein in the process of interest, (2) is evolutionarily conserved but has enough sequence divergence in different microorganisms to allow probe design for individual species, and (3) has substantial sequence data from isolates and environmental samples available in public databases. If only limited sequence data are available, it may be beneficial to initially do clone libraries for the gene and environment of interest in order to obtain the necessary sequence information for FGA probe design.

With the rapid advances in printing technologies, the development of comprehensive FGAs is limited only by the availability of requisite cultures or sequence data and the capital necessary for array construction. The largest FGA published to date contained 1662 probes for genes involved in the carbon, nitrogen, and sulfur cycles, organic contaminant degradation, and metals resistance and reduction [72], but this FGA has recently been expanded to over 24,000 probes [Schadt *et al.*, unpublished].

³ ProbeBase also contains complete probe lists for a number of other published POAs.

Either PCR products from amplification of various functional genes [15, 106] or shorter, synthesized oligonucleotide probes designed from these genes can be used for FGAs [20, 72]. A major advantage of PCR-derived probes is that they can be amplified from various isolates without prior sequence knowledge by using primers designed from conserved regions of the gene in other organisms. However, it can be virtually impossible to acquire all of the necessary isolates and environmental clones from their various sources to produce a comprehensive PCR fragment-based FGA. The major advantage of synthesized, oligonucleotide probes is that they can be designed directly from available sequence data. Use of oligonucleotide probes also allows the researcher more control and flexibility, such as the avoidance of highly conserved regions, in the probe design. The choice of PCR fragment-based or oligonucleotide probes also has implications for specificity and sensitivity as discussed later in the review.

Applications. Several recent studies have used FGAs to investigate microbial involvement in environmental processes including nitrogen fixation, nitrification, denitrification, and sulfate reduction in freshwater and marine systems [38, 85, 88, 90, 106]; degradation of organic contaminants including polychlorinated biphenyls (PCBs) [20] and polycyclic aromatic hydrocarbons (PAHs) [72] in soils and sediments; and methane-oxidizing capacity and diversity in landfill-simulating soil [5, 86] (Table 2). However, many of these applications were conducted primarily as proofs of concept and did not analyze enough samples or treatments to enable biologically meaningful conclusions to be formed.

For example, Tiquia *et al.* [90] used an FGA containing 50-mer oligonucleotide probes for 763 genes involved in nitrogen cycling and sulfate reduction to characterize microbial populations in a marine sediment from the Gulf of Mexico. Tests with pure cultures indicated that the array could achieve species-level resolution of microorganisms. The array detected several genes encoding nitrogenases (*nifH*), ammonia monooxygenase (*amoA*), nitrite reductase (*nirS/K*), methane monooxygenase (*pmoA*), and dissimilatory sulfite reductase (*dsrAB*) indicating its potential for comprehensive analysis of environmental samples.

In the largest-scale FGA application to date, Stalis-Pavese *et al.* [86] used an array containing 68 different 17- to 27-mer probes, primarily targeting the particulate methane monooxygenase (*pmoA*) genes of several methanotrophs, to investigate the impact of five different plant covers on methanotrophic activity in lysimeters under landfill-simulating conditions. The lysimeters contained sewage sludge compost, and half received a constant feed of artificial biogas (CH_4/CO_2 , 3:2) to simulate landfill emissions. The authors linked the

methanotrophic community structure in the vegetated lysimeters, in which type II methanotrophs had a competitive advantage over type Ia methanotrophs, with increased methane oxidation relative to the nonvegetated lysimeters. Not surprisingly, the relative abundances of methanotrophs were lower in the lysimeters that did not receive biogas. These results have obvious implications for the management of methane emissions at landfill sites and illustrate how FGA technology can be used to address real-world issues. The authors did note that their approach was useful only for comparing very similar samples due to the potential biases introduced when the relative abundances of genes were PCR-amplified before FGA analysis.

A major potential benefit of FGAs is that they cannot only be used to determine the presence of important genes in an environment by measuring DNA, but they can also be used to determine the expression of these genes by measuring mRNA. However, only a handful of studies have used FGAs for mRNA analysis [2, 21, 72]. Dennis *et al.* [21] constructed a PCR product FGA (271- to 1300-bp fragments) containing probes for 64 genes including several from the 2,4-dichlorophenoxyacetic acid (2,4-D)-degradation pathway of *Ralstonia eutropha* JMP134 and related organisms. Mixed cultures were created consisting of four isolates from a batch reactor treating pulp mill effluent and varying concentrations of *R. eutropha* JMP134. The *R. eutropha* JMP134 concentrations were 3.7, 0.37, 0.037, and 0.0037% of a total population of 10^8 cells/mL. The cultures were amended with 2 mM of 2,4-D and incubated 6 h before mRNA extraction. Significant induction of 2,4-D degradation genes was detected from populations as low as 0.0037% (3.7×10^3 cells in 10^8 total community) to 3.7%, depending on the specific genes detected and sequence similarity of the probes that were used. The authors also detected significant increases in resin acid degradation genes in a pulp mill effluent-treating bioreactor after it was spiked with a resin compound.

Rhee *et al.* [72] used a 50-mer oligonucleotide FGA (containing 1662 probes including organic contaminant degradation genes) to determine both the presence and the expression of naphthalene-degradation genes in soil enrichments. Soil from a site contaminated with PAHs was enriched with naphthalene or pyruvate (as a control). At midgrowth phase, DNA was extracted from aliquots, and the remaining naphthalene enrichment was split into two separate flasks and amended with either pyruvate (control) or additional naphthalene. After 3 h, mRNA was harvested and analyzed. Four different naphthalene-degradation genes, three of which were from *Rhodococcus* spp., were detected at higher levels in the naphthalene-amended enrichment based on DNA analysis. Likewise, the mRNA results indicated that three different *Rhodococcus* sp. genes involved in naphthalene

degradation were up-regulated (40- to 100-fold) in the naphthalene-amended enrichment, including two of the genes detected by DNA analysis. This corroborated the DNA hybridization results and indicated that this strain was actively degrading naphthalene in the enrichment. The results also revealed that other potential naphthalene-degrading organisms, whose genes were detected in the enrichments by the DNA hybridizations, were not responsible for naphthalene degradation under the tested conditions.

All of these studies that analyzed mRNA used relatively simple systems, mixed cultures, or enrichments. There are still several limitations for application of FGAs to mRNA analysis in complex environmental samples, e.g., surface soils and sediments, including difficulties in extracting sufficient quantities of high-quality mRNA from these samples within a reasonable time frame and the lack of sequence knowledge for environmental samples [78]. Advances in RNA extraction techniques [9, 37, 82] and newly available commercial kits (e.g., TruRNA from Atom Sciences, Inc. and FastRNA from Qbiogene) are helping with this process, but it still requires considerable effort to remove impurities and DNA from many samples—often substantially decreasing mRNA recovery. This further exasperates investigation of many low-biomass environments that already do not contain sufficient mRNA for FGA analysis. Unfortunately, unlike eukaryotic mRNAs, which can be amplified via their poly(A) tail with commercially available kits, it is difficult to amplify prokaryotic mRNA. However, new methods, such as that of Botero *et al.* [7] where a poly(A) tail is added to prokaryotic RNA for subsequent amplification, may make this feasible in the near future.

Researchers should ultimately exercise caution when attempting to link FGA results from environmental samples with the capacity for a specific biogeochemical function. The presence, or even activity, of a given pathway gene in a sample does not indicate the presence of all the genes necessary to carry out the complete transformation. For instance, although *dsrAB* is a key gene in sulfate reduction, it is also found in some non-sulfate-reducing bacteria [117]. In addition, a process may proceed via related genes and/or pathways not represented on the FGA.

Community Genome Arrays

Wu *et al.* [107] developed a novel prototype array that contained the entire genomic DNA of 67 different bacteria including α -, β -, and γ -*Proteobacteria* and Gram-positive bacteria with most of the organisms being *Azoarcus*, *Pseudomonas*, or *Shewanella* spp. The array was termed a community genome array (CGA) because it contained whole genomic DNA (one species' genome per spot) and was initially designed as a tool to detect specific micro-

organisms within a natural microbial community. The CGA could achieve species- to strain-level differentiation depending on the hybridization temperature. The CGA was used to compare the microbial populations in four marine sediments, three river sediments, and three soils. Principal components analysis of the CGA results grouped the three types of samples into three distinct groups, indicating that the microbial populations from a given type of sample (e.g., soil) were more similar to one another than those in the other types of samples (e.g., marine and river sediments). The CGA results also correlated well with the differences in biogeochemical and physical properties between the sites. These results demonstrate the potential of CGAs as a comparative tool for determining the relatedness of microbial communities in different samples. The CGAs can also be used to determine the genomic relatedness of isolated bacteria to each other and also the organisms represented on the array.

The CGA is conceptually analogous to membrane-based reverse sample genome probing (RSGP) [30], but uses a nonporous hybridization surface and fluorescence-based detection that enable high throughput analyses but decrease sensitivity [107]. Like RSGP, the main potential disadvantage of CGAs is that only the cultured components of a community are included on the array. However, with recent advances in the generation of large insert-sized metagenomic libraries, it is also possible to use DNA from uncultured organisms for microarray generation as discussed in the next section.

Metagenomic Arrays

Due to the growing evidence that most environmental microorganisms cannot be isolated using current techniques, the field of metagenomics, or the direct extraction and cloning of nucleic acids from environmental samples, has developed [33]. These techniques have even been used to sequence entire communities in an acid mine drainage site [97] and a portion of the community in a sample from the Sargasso Sea [100], but it is not yet possible to assemble even the dominant genomes from more diverse sites such as surface soils and sediments [93, 95, 100]. However, the combination of microarray and metagenomic technologies has the potential to reveal detailed information on these yet to be cultured organisms.

Sebat *et al.* [81] generated an MGA using a cosmid library from a groundwater enrichment. Approximately 1-kb inserts were amplified from 672 cosmids and placed on the array along with several rRNA gene control probes. The MGA was used as a high-throughput library screening technique. Groundwater isolates, reference strains, and community DNA were hybridized to the array. Ten bacteria isolated from the enrichment hybridized specifically to 10 individual probes on the array.

Other probes hybridized to multiple related bacteria, indicating that these probes likely contained conserved genes. Some probes hybridized to community genomic DNA from the enrichment but did not hybridize to any of the isolates, indicating that the organisms bearing these DNA fragments were not cultured. The cosmid inserts corresponding to these probes were sequenced and were related to genes involved in several ecologically important processes including denitrification, hydrogen oxidation, and transposition.

The above MGA contained DNA fragments that were only ~1 kb, but larger fragments (>50 kb) from fosmid or bacterial artificial chromosome libraries could be used to provide higher genomic throughput [4, 107]. If sufficient mRNA could be obtained, it may also be possible to generate MGAs from a cDNA library that could then be used to create a site-specific FGA for measuring microbial activity. The MGA technology is still in the early stages of development, but it has tremendous potential for environmental applications since the enormous amount of unknown sequences in these environments is one of the major limitations for microarray analysis.

Whole-Genome Open Reading Frame Arrays

Organisms that are closely related, based on 16S rRNA genes, can exhibit strikingly different phenotypic characteristics and may actually have substantially dissimilar genomes due to processes such as lateral gene exchange [62, 63]. Whole-genome ORF arrays (WGAs), which contain probes for all of the ORFs in a genome can be very useful for comparative genomics of different organisms with specific application to the processes of lateral gene transfer and microevolution [3, 22, 25, 58, 62, 63, 77]. The WGAs can also be used to study genome-wide transcription in response to different environmental stimuli. This is commonly done in functional genomics for pure cultures [52, 89] but can also be applied to small groups of organisms to study their interactions [2].

Dong *et al.* [22] used a WGA containing 96% of the annotated ORFs in *E. coli* K-12 to comparatively interrogate the genome of the closely related (97% based on 16S rRNA gene) *Klebsiella pneumoniae* 342, which is a maize endophyte. Only 70% of *E. coli* K-12 ORFs were found in *K. pneumoniae* 342 ($\geq 55\%$ similarity cutoff), whereas 24% were not present in *K. pneumoniae* 342. The signal was too low to make a determination for a small portion of the genes ($n = 68$). Highly conserved genes including those for energy, amino acid, and fatty acid metabolism along with cofactor synthesis, cell division, DNA replication, transcription, translation, transport, and regulatory proteins were among those found to be shared. The *E. coli* K-12 ORFs not found in *K. pneumo-*

niae included many hypothetical and putative regulatory proteins, chaperones, and enzymes in addition to genes thought to have been acquired from phage, plasmids, or transposons via lateral transfer. The WGA results agreed with the phenotypic characteristics of the bacteria. Murray *et al.* [58] also used a WGA to discover evidence of lateral gene transfer in several *Shewanella* spp.

Barnett *et al.* [2] gave an excellent example of the results a WGA can produce when used to investigate the interactions of multiple organisms. The authors created a symbiosis chip using the Affymetrix system, as discussed previously for POAs, containing probes for all of the ORFs in the genome of *Sinorhizobium meliloti* and ~10,000 expressed genes in its host organism, the legume *Medicago truncatula* [57]. Over 200 *M. truncatula* genes had increased expression in nodules versus uninfected root tissue. Most of these genes had been previously demonstrated to be up-regulated in nodules, thus confirming the reliability of the approach. Furthermore, the expressed *M. truncatula* genes were very similar in nodules inhabited by wild-type *S. meliloti* and those containing a non-nitrogen-fixing mutant *S. meliloti*, indicating that the plant responses were due to the presence of the bacteria rather than their nitrogen fixation. In contrast, there were large differences in the genes expressed by wild-type and mutant *S. meliloti* in the nodules, indicating that most of the increased bacterial gene expression in nodules was due to nitrogen fixation. Although this relatively simple system was composed of only two organisms, it illustrates that the potential to investigate the interactions of organisms at the transcriptional level in complex systems is becoming more feasible as additional sequence data become available and microarray technologies continue to improve. The WGA approach can also be combined with reporter gene technology for noninvasive, real-time analysis [99].

Other Types of Arrays

Researchers have also developed other types of microarrays that have potential applications in microbial ecology research. Arrays containing probes generated from random genomic fragments have been used in situations where the genome sequences of the target organisms were unknown [14, 40]. Kim *et al.* [40] digested genomic DNA from *Gordonia amarae*, *Zooglear ramigera*, and *Mycobacterium peregrinum* with restriction enzymes and used some of the resulting ~200- to 1500-bp fragments (~50 for each organism) for the array. The array was then tested with pure cultures, mixed cultures, and environmental samples. Most of the tested combinations had <5% cross-hybridization to nontarget probes in mixed culture, although *M. peregrinum* genomic DNA had 29 and 36% cross-hybridization to *G. amarae* and *Z. ramigera* probes, respectively, when applied alone. How-

ever, expansion of the array to include more probes may reduce the probability that all probes would bind to a given nontarget organism and would thus likely improve the specificity. The array detected *G. amarae*, which can cause foaming and bulking in wastewater treatment plants at high populations [19], in activated sludge at populations as low as 10^3 /mL. Cho and Tiedje [14] used a similar approach to differentiate bacteria using an array composed of 60–96 ~1-kb genomic fragments from four fluorescent *Pseudomonas* spp. Hybridization profiles of 12 well-characterized *Pseudomonas* spp. indicated that the array could achieve species- to strain-level resolution.

Randomly selected oligonucleotide probes have also been used to fingerprint bacteria. Kingsley *et al.* [41] developed a prototype array consisting of 47 nonamer probes randomly generated based on the *E. coli* K-12 genome. The array was tested using 14 closely related *Xanthomonas* pathogens. Ten of the 47 probes had diagnostic value, based on statistical tests, and were used to generate fingerprints that revealed differences in the bacteria, including two strains that could not be distinguished using traditional gel electrophoresis of REP-PCR products. Since this method is based on random nonamers, it could potentially be used to fingerprint any microorganism.

Use of Microarrays with Complementary Analyses

Numerous other techniques can be combined with microarray analysis not only to validate results but also to produce powerful synergistic tools for investigating microbial interactions and processes. For example, Loy *et al.* [55] used clone libraries of sulfite reductase genes (*dsrAB*) along with a POA to study sulfate-reducing prokaryote communities in acidic, low sulfate fens in Germany. The clone libraries corroborated POA results that indicated the sulfate-reducing communities at the two tested sites were different. The clone libraries also identified additional sulfate reducers that were not detected by the POA.

The integration of isotope and microarray technologies produces one of the potentially most powerful combined approaches for microbial ecology research. Microarray analysis of DNA or RNA labeled with isotopes can differentiate between active and inactive organisms in a sample and/or identify those organisms that metabolize a labeled substrate. These isotopes can be either radioisotopes such as ^{14}C or stable isotopes such as ^{13}C [1, 67, 68]. Adamczyk *et al.* [1] used ^{14}C -labeled bicarbonate and a POA to study ammonia-oxidizing bacterial communities in two samples of nitrifying activated sludge. Scanning for radioactivity in the rRNA hybridized to the POA enabled detection of populations that consumed the [^{14}C]bicarbonate. The approach detected populations that composed less than 5–10% of the community. This technique could potentially be

applied to ^{13}C -labeled materials also, but it may be more difficult to obtain enough ^{13}C -labeled DNA for microarray analyses since the labeled DNA would have to be separated from nonlabeled DNA before microarray analysis unless the array could be directly scanned for ^{13}C .

Challenges for Microarray Applications

Specificity. The highly conserved nature of many genes and the vast amount of unknown sequence data in environmental samples make it difficult to design and validate microarray probes that are specific to a given target sequence. As mentioned earlier, a major advantage of oligonucleotide probes is the ability to avoid conserved regions of genes or areas containing stable secondary structure. Furthermore, shorter oligonucleotide probes (~20-mer) can differentiate a single mismatch in a probe–target hybridization making them ideal for use in POAs [98, 103, 114]. This level of specificity can also be achieved by using similarly designed WGAs. A common format for these arrays includes sets of probes that perfectly match a target sequence and corresponding sets of probes containing one mismatched nucleotide, usually at a central position, with greater signal intensity for the perfectly matched probes indicating detection of the target sequence. Zhou *et al.* [114] systematically tested a prototype POA consisting of 19-mer probes for different bacterial 16S rRNA gene sequences with one to five mismatches in the mismatched probes. A single mismatch at the central position of the probes decreased the signal to only 15–25% the intensity of the perfectly matched probes. Three to five mismatches reduced the signal to undetectable levels. Although these results demonstrate the potential specificity of this approach, it is still difficult to achieve complete discrimination of rRNA genes using only a single mismatch. Chandler and Jarrell [13] summarized this well when they stated “... because only a small portion of the natural microbial diversity has been identified and because microarray hybridization specificity is not perfect, it is practically and theoretically difficult to know if and when hybridization signals in a new environment result from a perfectly matched or a mismatched probe–target combination.”

One common approach to address this problem is to design multiple perfectly matched and mismatched probe combinations for each organism of interest and then to compare the probe pairs statistically. Unpredictable probes or those providing abnormal results (higher signal intensity for the mismatched probe) are removed from the array or discarded during data analysis. Other researchers have improved the discrimination of matched and mismatched probes by determining the thermal dissociation curve for each probe–target duplex on an array [26, 50, 98]. This has more commonly been done using three-

dimensional array platforms but has been demonstrated to work with planar arrays [45]. Li *et al.* [45] found that this approach could discriminate hybridization to short oligonucleotide probes (18- to 20-mers) with one or two internal (but not terminal) mismatches from hybridization to perfectly matched probes on a planar, rRNA-based array.

The longer oligonucleotide probes (~40–70-mers) typically used for FGAs are less specific than those used for POAs. However, since most functional genes are more variable than rRNA genes, longer oligonucleotide probes can be used to increase detection sensitivity while still achieving species-level specificity. Rhee *et al.* [72] reported that 50-mer probes could discriminate sequences less than 88–94% similar to the probes with hybridization at 50°C and 50% formamide. Taroncher-Oldenburg *et al.* [88] reported a similar value of 87% for a 70-mer probe FGA. This is slightly higher than the 80–85% sequence identity discrimination power of a 400- to 800-bp PCR product-based FGA [106]. The MGAs based on shorter inserts would likely behave similar to the PCR-based FGAs. In addition to percent similarity, long stretches of a probe that are complementary to a nontarget sequence can lead to substantial nonspecific hybridization and should be considered during probe design [36, 39]. The position of mismatches (those distributed across a probe, rather than localized to a select region, produce more specific binding) [44] and the amount of free energy of probe–target duplexes can also affect specificity [35, 46, 88]. A recent study [48] found that by simultaneously considering multiple probe–target characteristics during the design process, specific probes could be produced using more relaxed design criteria than was possible when each factor was examined separately. The results indicated that specific hybridization could be achieved using 50-mer probes with a free energy release of ≤ -35 kcal/mol and $\leq 90\%$ similarity and ≤ 20 bp continuous stretches to nontarget sequences. The ability to even slightly relax design criteria should increase the percentage of genes, in a given data set, for which probes can be designed. This could be extremely valuable when designing probes from very similar sequence data such as that generated from environmental clone libraries.

Depending on the probe design and objectives of the research, specificity can also be increased or decreased, to a point, by adjusting the stringency of the hybridization conditions (temperature, formamide concentration, salt concentration, etc.) [34, 44, 106]. For example, Wu *et al.* [107] developed a CGA that could distinguish bacteria at the species level when hybridized at 55°C or at the strain level at 65 or 75°C. This helps to illustrate the need for caution in conducting and interpreting microarray analyses. Use of an array under more or less stringent conditions than that for which it was designed can lead

to inaccurate conclusions based on overestimated or underestimated results. Inclusion of control DNA in the hybridization solution that has varying similarity to corresponding control probes on the array can potentially ensure that the correct hybridization stringency is achieved.

Several software programs are currently available for the design of oligonucleotide probes for microarrays. These include ArrayOligoSelector [8], OligoArray [74], OligoArray 2.0 [75], Oligopicker [102], OligoWiz [59], PRIMEGENS [108], PROBEmer [27], ProbeSelect [46], and ROSO [70]. Most of these programs work well for designing probes from whole-genome sequences. However, research by Li *et al.* [47] found that a considerable portion of probes designed by some of these software programs from groups of orthologous functional gene sequences (such as those produced by clone libraries) were not specific to the target sequence (based on experimentally determined values [72]). With the growing database of environmental sequences, especially the highly similar sequences which are often obtained from clone libraries, it can be difficult to design probes that will not cross-hybridize to related sequences.

To address these issues, Li *et al.* [47] designed a new software tool called CommOligo. The program uses a new global alignment algorithm to design single or multiple unique probes for each gene using multiple, simultaneously implemented, user-specified criteria, such as the maximal sequence similarity, the maximal length of continuous perfectly matched nucleotides, free energy, self-binding, melting temperature, and GC content. A major advantage of CommOligo is that it can also design single or multiple group-specific probes for related groups of genes that are too similar for the design of unique probes—analogueous to the design of PCR primers from conserved regions of grouped genes. CommOligo is currently undergoing additional prerelease testing. Until CommOligo or other improved probe design software is available, researchers should be cautious when using software originally developed for use on whole-genome data for designing oligonucleotide probes from environmental sequence data.

Sensitivity. Although oligonucleotide probes have many advantages for probe design, they are typically ~10- to 100-fold less sensitive than longer PCR-based or CGA probes [20, 72, 106, 107]. The sensitivities of all of the different array formats have not been directly compared, but examples from the literature have reported limits of 0.2 ng of target genomic DNA for a CGA [107], 1 ng for a PCR-based FGA [106], and 5–8 ng for 50-mer oligonucleotide FGAs [72, 90] in the absence of background DNA. These sensitivities were about 10-fold lower in the presence of background DNA simulating environmental samples [72, 90, 107]. For the 50-mer FGA, this detection limit

corresponded to ~ 10 cells or 5% of the total community, which agreed with other published studies [15]. He *et al.* [34] recently compared the detection sensitivity of PCR amplicon and oligonucleotide probes. The PCR amplicon probes had a detection limit of 5 ng of genomic DNA, and the 70-, 60-, and 50-mer oligonucleotide probes had detection limits of 25, 100, and 100 ng of genomic DNA, respectively. These limits equaled approximately 1.9×10^6 , 9.2×10^6 , 3.7×10^7 , and 3.7×10^7 gene copies for the PCR amplicon, 70-mer, 60-mer, and 50-mer probes, respectively. The probes were also used to evaluate gene expression of *Shewanella oneidensis* MR-1 under different conditions with the 70-mer probe results again being most comparable to those from the PCR amplicons. The detection sensitivities of MGAs and WGs are likely to resemble PCR fragment-based FGAs and POAs, respectively, depending on the probe design.

Different nucleic acid labeling methods may increase sensitivity [20, 85, 115]. Denef *et al.* [20] used tyramide signal amplification labeling to increase the signal intensity of a 70-mer FGA ~ 10 -fold over the commonly used Cy dye-labeling techniques. This approach reduced the detection limit to 1% of cells in the total community. Although the above methods are sufficient for detecting the dominant members of relatively high biomass communities, new approaches are needed for investigating less abundant, but ecologically important, populations.

One currently available option to detect less dominant microorganisms within a community is to PCR-amplify these specific populations, although this has its own set of well-documented limitations [18, 28, 71, 87, 101]. Bodrossy *et al.* [5] used this approach with an FGA containing 15- to 26-mer probes for particulate methane monooxygenase (*pmoA*). Although very short oligonucleotide probes were used, populations were detected that comprised as little as 5% of the total community. The use of magnetic beads or other capture techniques may also be useful for enriching certain populations [96]. An option for low-biomass environments that do not produce sufficient quantities of DNA for FGA analysis is the nonspecific amplification of whole community DNA before FGA analysis. Wu *et al.* [unpublished] have developed a whole community genome amplification (WCGA) procedure that can amplify ng quantities of DNA to microgram quantities with a linear relationship between starting template and final concentration ($r^2 = 0.96$ – 0.98). An FGA analysis of DNA amplified by WCGA from low-biomass groundwater samples, which were contaminated with high levels of nitrate and uranium, revealed a correlation between microbial diversity and groundwater geochemistry and contaminant levels.

Many microarrays are currently printed on planar glass slide platforms because this enables high-density printing and high-throughput analyses. The trade-off for this capacity is reduced detection sensitivity. Membrane-

based hybridizations are several orders of magnitude more sensitive than microarray hybridizations on nonporous surfaces most likely due to the limited amount of probe material that can be attached to the nonporous surfaces [15]. Researchers are developing new slide chemistries, including ultrathin three-dimensional platforms, which have increased binding capacities but maintain the high-throughput characteristics that make microarray analyses advantageous [31, 98, 114].

Quantitation and Data Analysis. There has been some concern regarding the quantitative ability of microarrays given the potential variability in steps including DNA extraction, labeling, hybridization, and analysis. However, recent research indicates that FGAs and CGAs can be quantitative within a range of concentrations. Wu *et al.* [106] found a strong linear relationship ($r^2 = 0.96$) between the amount of pure culture DNA hybridized to a PCR-product FGA and the signal intensity within a range of 1–100 ng. The authors also found a good linear relationship ($r^2 = 0.94$) for a mixture of 11 different genes varying in concentration from 1 pg to 1 ng. Likewise, Rhee *et al.* [72] found a strong linear relationship for both DNA and mRNA using a 50-mer oligonucleotide FGA. Various amounts of *Thauera aromatica* K172 genomic DNA was mixed with 1 μg of *S. oneidensis* MR-1 DNA as a background. The FGA signal intensity was linear ($r^2 = 0.95$ – 0.99) for all detected genes over a range of 75–1000 ng genomic DNA. For mRNA analysis, 3.0×10^6 to 1.6×10^9 *Pseudomonas putida* PpG7 cells that had been incubated with naphthalene were mixed with 1.9×10^9 *S. oneidensis* MR-1 cells as background RNA. The FGA signal intensity was again linear for all detected genes ($r^2 = 0.96$ – 0.99) over a range of at least 5.0×10^7 to 1.6×10^9 cells. Wu *et al.* [107] indicated that CGA detection was also linear ($r^2 = 0.98$) from 25 to 1,000 ng of DNA. The MGAs have not been tested for quantitation but likely resemble FGAs and CGAs. It is unknown if POAs and WGs based on perfect match–mismatch probes are quantitative.

It can be difficult to compare data between, and even within, microarray experiments due to the use of different analysis methods and variability in printing, labeling, and hybridization. The two-color dye-swap techniques that are commonly used in gene expression experiments work well to determine relative levels of gene expression in pure cultures, but they do not facilitate comparison between experiments and laboratories unless the same control DNA is used. Researchers have developed modified approaches where known amounts of labeled oligonucleotides or DNA fragments are spiked into the hybridization solution as a control [5, 15, 24]. Microarray results are then normalized based on the signal intensity resulting from hybridization of this control DNA with corresponding control probes on the

array. The oligonucleotide approach is especially promising for standardizing microarray results since the probes could be synthesized directly from sequence data and thus would be readily available to any researcher; however, additional research is needed to optimize this method for different array formats.

Even if microarray experiments are meticulously designed and conducted, it could be difficult to quantitatively correlate differences in hybridization signals with changes in specific populations due to the large amount of unknown nucleic acid sequences in environmental samples. It is typically assumed that hybridization signal intensity is directly proportional to the abundance of the target organism, but nonspecific hybridization due to uncharacterized microorganisms in environmental samples may occur and confound interpretation. Analysis of key genes with other methods such as real-time PCR may help to validate the quantitative accuracy of major results and strengthen the conclusions drawn from microarray data [72].

Future Perspectives

An ultimate goal of microarray analysis, with respect to microbial ecology research, is to simultaneously measure the activity of multiple microbial populations in relation to different environmental factors. This has traditionally been accomplished by measuring mRNA expression. However, it is still difficult to recover sufficient high-quality mRNA for microarray analysis from many environmental samples. Further advances in mRNA extraction and amplification methods are needed to make microarray analysis of mRNA possible for a broader range of samples. Since mRNA levels are only an indirect measurement of activity, with the translated proteins actually being responsible for most biological processes, it would be best if the actual protein levels could ultimately be quantified. With advances in proteomics, researchers are beginning to develop protein arrays for identifying proteins and studying protein expression and protein–ligand interactions [23, 73]. Protein arrays have not yet been applied to the study of complex environmental samples, but if the technology is further developed and can be successfully adapted, it could be very useful for investigating enzymatic expression in environmental samples.

One of the greatest needs for microarray analysis of microbial communities is the development of standardized methods for data analysis and interpretation. It is continually becoming more difficult to analyze microarray results as more comprehensive arrays, which are necessary to understand many complex communities, are developed. Statistical methods developed for functional genomics may not be appropriate for analyzing the complex data sets often produced from microarray

analysis of environmental samples. New statistical methods need to be devised and/or existing methods adapted to meet the specific challenges posed by these types of arrays. The development of improved universal standards would also enhance data analysis and enable comparison of array data between experiments and laboratories.

In order for microarray technology to reach its full high-throughput potential and provide real-time information on microbial populations in environmental samples, it will be necessary for the technology to eventually be automated and field deployable [13]. With advances in microfabrication and microfluidic technologies, it is now becoming possible to assemble all of the chambers, pumps, valves, mixers, heaters, and detectors that are required for microarray analysis on a single chip [49, 51]. These “laboratories-on-a-chip” still face the same analytical challenges as encountered with manual microarrays and are just in early stages of development, but they have the potential to revolutionize microarray analysis of environmental microbial populations. Ultimately, for whichever array format is used, more comprehensive, broad-scale applications are necessary to further validate and demonstrate the analytical power of microarrays for investigating various biological questions.

Conclusions

Researchers have invested considerable effort over the last few years to adapt microarray technology for the analysis of microbial communities. It is now becoming possible to produce microarrays capable of simultaneously characterizing the dynamics and activities of most, if not all, of the microbial populations even in complex samples such as soils and sediments. Several recent studies have successfully used microarrays to investigate aspects of major ecological issues. However, most of these studies were of limited scope with very few yet utilizing the full high-throughput potential of microarray analysis. Further technological advances are needed to improve methods for data analysis in order for microarrays to be applicable to a broader range of samples and for results to be comparable across experiments and between laboratories. Continued development may ultimately allow microarray technology to achieve its promise for comprehensive high-throughput, near-real-time monitoring of microbial populations within ecological communities.

Acknowledgments

The authors would like to thank E. Hollister for valuable comments on this manuscript. The authors’ efforts in preparing this review were supported by the United States Department of Energy Office of Science as part of its Environmental Research Programs in Natural and

Accelerated Bioremediation Research, Biotechnology Investigations Ocean Margins Program, Microbial Genome Program, Genomics:GTL program, and Carbon Sequestration [as part of the consortium on research to enhance Carbon Sequestration in Terrestrial Ecosystems (CSiTE)]. Oak Ridge National Laboratory is managed by UT-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

References

- Adamczyk, J, Hesselsoe, M, Iversen, N, Horn, M, Lehner, A, Nielsen, PH, Schlöter, M, Roslev, P, Wagner, M (2003) The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function. *Appl Environ Microbiol* 69: 6875–6887
- Barnett, MJ, Toman, CJ, Fisher, RF, Long, SR (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. *Proc Natl Acad Sci USA* 101: 16636–16641
- Behr, MA, Wilson, MA, Gill, WP, Salamon, H, Schoolnik, GK, Rane, S, Small, PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284: 1520–1523
- Berry, AE, Chiocchini, C, Selby, T, Sosio, M, Wellington, EM (2003) Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiol Lett* 223: 15–20
- Bodrossy, L, Stralis-Pavese, N, Murrell, JC, Radajewski, S, Weilharter, A, Sessitsch, A (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* 5: 566–582
- Bodrossy, L, Sessitsch, A (2004) Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol* 7: 245–254
- Botero, LM, D’Imperio, S, Burr, M, McDermott, TR, Young, M, Hassett, DJ (2005) Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Appl Environ Microbiol* 71: 1267–1275
- Bozdech, Z, Zhu, J, Joachimiak, MP, Cohen, FE, Pulliam, B, DeRisi, JL (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* 4: R9
- Burgmann, H, Widmer, F, Sigler, WV, Zeyer, J (2003) mRNA extraction and reverse transcription–PCR protocol for detection of *nifH* gene expression by *Azotobacter vinelandii* in soil. *Appl Environ Microbiol* 69: 1928–1935
- Cannone, JJ, Subramanian, S, Schnare, MN, Collett, JR, D’Souza, LM, Du, Y, Feng, B, Lin, N, Madabusi, LV, Muller, KM, Pande, N, Shang, Z, Yu, N, Gutell, RR (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3: 2
- Castiglioni, B, Rizzi, E, Frosini, A, Sivonen, K, Rajaniemi, P, Rantala, A, Mugnai, MA, Ventura, S, Wilmotte, A, Boutte, C, Grubisic, S, Balthasart, P, Consolandi, C, Bordoni, R, Mezzelani, A, Battaglia, C, De Bellis, G (2004) Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. *Appl Environ Microbiol* 70: 7161–7172
- Chandler, DP, Newton, GJ, Small, JA, Daly, DS (2003) Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays. *Appl Environ Microbiol* 69: 2950–2958
- Chandler, DP, Jarrell, AE (2004) Automated purification and suspension array detection of 16S rRNA from soil and sediment extracts by using tunable surface microparticles. *Appl Environ Microbiol* 70: 2621–2631
- Cho, JC, Tiedje, JM (2001) Bacterial species determination from DNA–DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* 67: 3677–3682
- Cho, JC, Tiedje, JM (2002) Quantitative detection of microbial genes by using DNA microarrays. *Appl Environ Microbiol* 68: 1425–1430
- Cole, JR, Chai, B, Farris, RJ, Wang, Q, Kulam, SA, McGarrell, DM, Garrity, GM, Tiedje, JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33: D294–D296
- Cook, KL, Saylor, GS (2003) Environmental application of array technology: promise, problems and practicalities. *Curr Opin Biotechnol* 14: 311–318
- Crosby, LD, Criddle, CS (2003) Understanding bias in microbial community analysis techniques due to *rnn* operon copy number heterogeneity. *Biotechniques* 34: 790–798
- de los Reyes, FL, 3rd, Raskin, L (2002) Role of filamentous microorganisms in activated sludge foaming: relationship of mycolata levels to foaming initiation and stability. *Water Res* 36: 445–459
- Denef, VJ, Park, J, Rodrigues, JLM, Tsoi, TV, Hashsham, SA, Tiedje, JM (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, EA, Liss, SN, Fulthorpe, R (2003) Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Appl Environ Microbiol* 69: 769–778
- Dong, Y, Glasner, JD, Blattner, FR, Triplett, EW (2001) Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames. *Appl Environ Microbiol* 67: 1911–1921
- Duburcq, X, Olivier, C, Malingue, F, Desmet, R, Bouzidi, A, Zhou, F, Auriault, C, Gras-Masse, H, Melnyk, O (2004) Peptide–protein microarrays for the simultaneous detection of pathogen infections. *Bioconjug Chem* 15: 307–316
- Dudley, AM, Aach, J, Steffen, MA, Church, GM (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci USA* 99: 7554–7559
- Dziejman, M, Balon, E, Boyd, D, Fraser, CM, Heidelberg, JF, Mekalanos, JJ (2002) Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc Natl Acad Sci USA* 99: 1556–1561
- El Fantroussi, S, Urakawa, H, Bernhard, AE, Kelly, JJ, Noble, PA, Smidt, H, Yershov, GM, Stahl, DA (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl Environ Microbiol* 69: 2377–2382
- Emrich, SJ, Lowe, M, Delcher, AL (2003) PROBEmer: A web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res* 31: 3746–3750
- Farrelly, V, Rainey, FA, Stackebrandt, E (1995) Effect of genome size and *rnn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl Environ Microbiol* 61: 2798–2801
- Gans, J, Wolinsky, M, Dunbar, J (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* 309: 1387–1390
- Greene, EA, Voordouw, G (2003) Analysis of environmental microbial communities by reverse sample genome probing. *J Microbiol Methods* 53: 211–219

31. Guschin, D, Yershov, G, Zaslavsky, A, Gemmell, A, Shick, V, Proudnikov, D, Arenkov, P, Mirzabekov, A (1997) Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal Biochem* 250: 203–211
32. Gutell, RR, Lee, JC, Cannone, JJ (2002) The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* 12: 301–310
33. Handelsman, J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669–685
34. He, Z, Wu, L, Fields, MW, Zhou, J (2005) Comparison of microarrays with different probe sizes for monitoring gene expression. *Appl Environ Microbiol* 71: 5154–5162
35. Held, GA, Grinstein, G, Tu, Y (2003) Modeling of DNA microarray data by using physical properties of hybridization. *Proc Natl Acad Sci USA* 100: 7575–7580
36. Hughes, TR, Mao, M, Jones, AR, Burchard, J, Marton, MJ, Shannon, KW, Lefkowitz, SM, Ziman, M, Schelter, JM, Meyer, MR, Kobayashi, S, Davis, C, Dai, H, He, YD, Stephaniants, SB, Cavet, G, Walker, WL, West, A, Coffey, E, Shoemaker, DD, Stoughton, R, Blanchard, AP, Friend, SH, Linsley, PS (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19: 342–347
37. Hurt, RA, Qiu, X, Wu, L, Roh, Y, Palumbo, AV, Tiedje, JM, Zhou, J (2001) Simultaneous recovery of RNA and DNA from soils and sediments. *Appl Environ Microbiol* 67: 4495–4503
38. Jenkins, BD, Steward, GF, Short, SM, Ward, BB, Zehr, JP (2004) Fingerprinting diazotroph communities in the Chesapeake Bay by using a DNA macroarray. *Appl Environ Microbiol* 70: 1767–1776
39. Kane, MD, Jatkoa, TA, Stumpf, CR, Lu, J, Thomas, JD, Madore, SJ (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28: 4552–4557
40. Kim, BC, Park, JH, Gu, MB (2004) Development of a DNA microarray chip for the identification of sludge bacteria using an unsequenced random genomic DNA hybridization method. *Environ Sci Technol* 38: 6767–6774
41. Kingsley, MT, Straub, TA, Call, DR, Daly, DS, Wunschel, SC, Chandler, DP (2002) Fingerprinting closely related *Xanthomonas* pathovars with random nonamer oligonucleotide microarrays. *Appl Environ Microbiol* 68: 6361–6370
42. Koizumi, Y, Kelly, JJ, Nakagawa, T, Urakawa, H, El-Fantroussi, S, Al-Muzaini, S, Fukui, M, Urushigawa, Y, Stahl, DA (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
43. Larsen, N, Olsen, GJ, Maidak, BL, McCaughey, MJ, Overbeek, R, Macke, TJ, Marsh, TL, Woese, CR (1993) The ribosomal database project. *Nucleic Acids Res* 21: 3021–3023
44. Letowski, J, Brousseau, R, Masson, L (2004) Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods* 57: 269–278
45. Li, ES, Ng, JK, Wu, JH, Liu, WT (2004) Evaluating single-base-pair discriminating capability of planar oligonucleotide microchips using a non-equilibrium dissociation approach. *Environ Microbiol* 6: 1197–1202
46. Li, F, Stormo, GD (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* 17: 1067–1076
47. Li, X, He, Z, Zhou, J (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* 33: 6114–6123
48. Liebich, J, Schadt, CW, Chong, SC, He, Z, Rhee, SK, Zhou, J (2006) Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl Environ Microbiol* 72: 1688–1691
49. Liu, RH, Yang, J, Lenigk, R, Bonanno, J, Grodzinski, P (2004) Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. *Anal Chem* 76: 1824–1831
50. Liu, WT, Mirzabekov, AD, Stahl, DA (2001) Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ Microbiol* 3: 619–629
51. Liu, WT, Zhu, L (2005) Environmental microbiology-on-a-chip and its future impacts. *Trends Biotechnol* 23: 174–179
52. Liu, Y, Zhou, J, Omelchenko, MV, Beliaev, AS, Venkateswaran, A, Stair, J, Wu, L, Thompson, DK, Xu, D, Rogozin, IB, Gaidamakova, EK, Zhai, M, Makarova, KS, Koonin, EV, Daly, MJ (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA* 100: 4191–4196
53. Loy, A, Lehner, A, Lee, N, Adamczyk, J, Meier, H, Ernst, J, Schleifer, KH, Wagner, M (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* 68: 5064–5081
54. Loy, A, Horn, M, Wagner, M (2003) probeBase: an online resource for rRNA-targeted oligonucleotide probes. *Nucleic Acids Res* 31: 514–516
55. Loy, A, Kusel, K, Lehner, A, Drake, HL, Wagner, M (2004) Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages. *Appl Environ Microbiol* 70: 6998–7009
56. Loy, A, Schulz, C, Lucker, S, Schopfer-Wendels, A, Stoecker, K, Baranyi, C, Lehner, A, Wagner, M (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order “*Rhodocyclales*.” *Appl Environ Microbiol* 71: 1373–1386
57. Mitra, RM, Shaw, SL, Long, SR (2004) Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume–rhizobia symbiosis. *Proc Natl Acad Sci USA* 101: 10217–10222
58. Murray, AE, Lies, D, Li, G, Neelson, K, Zhou, J, Tiedje, JM (2001) DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc Natl Acad Sci USA* 98: 9853–9858
59. Nielsen, HB, Wernersson, R, Knudsen, S (2003) Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays. *Nucleic Acids Res* 31: 3491–3496
60. Noller, HF, Kop, J, Wheaton, V, Brosius, J, Gutell, RR, Kopylov, AM, Dohme, F, Herr, W, Stahl, DA, Gupta, R, Waese, CR (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 9: 6167–6189
61. Noller, HF (1984) Structure of ribosomal RNA. *Annu Rev Biochem* 53: 119–162
62. Ochman, H, Lerat, E, Daubin, V (2005) Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci USA* 102(Suppl 1): 6595–6599
63. Ochman, H, Santos, SR (2005) Exploring microbial microevolution with microarrays. *Infect Genet Evol* 5: 103–108
64. Olsen, GJ, Lane, DJ, Giovannoni, SJ, Pace, NR, Stahl, DA (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* 40: 337–365
65. Olsen, GJ, Overbeek, R, Larsen, N, Marsh, TL, McCaughey, MJ, Maciukenas, MA, Kuan, WM, Macke, TJ, Xing, Y, Woese, CR (1992) The Ribosomal Database Project. *Nucleic Acids Res* 20(Suppl): 2199–2200
66. Peplies, J, Glockner, FO, Amann, R (2003) Optimization strategies for DNA microarray-based detection of bacteria with

- 16S rRNA-targeting oligonucleotide probes. *Appl Environ Microbiol* 69: 1397–1407
67. Polz, MF, Bertilsson, S, Acinas, SG, Hunt, D (2003) A(r)Ray of hope in analysis of the function and diversity of microbial communities. *Biol Bull* 204: 196–199
 68. Radajewski, S, Ineson, P, Parekh, NR, Murrell, JC (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* 403: 646–649
 69. Ratushna, VG, Weller, JW, Gibas, CJ (2005) Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC Genomics* 6: 31
 70. Reymond, N, Charles, H, Duret, L, Calevro, F, Beslon, G, Fayard, JM (2004) ROSO: optimizing oligonucleotide probes for microarrays. *Bioinformatics* 20: 271–273
 71. Reysenbach, AL, Giver, LJ, Wickham, GS, Pace, NR (1992) Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol* 58: 3417–3418
 72. Rhee, SK, Liu, X, Wu, L, Chong, SC, Wan, X, 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
 73. Ro, HS, Jung, SO, Kho, BH, Hong, HP, Lee, JS, Shin, YB, Kim, MG, Chung, BH (2005) Surface plasmon resonance imaging-based protein array chip system for monitoring a hexahistidine-tagged protein during expression and purification. *Appl Environ Microbiol* 71: 1089–1092
 74. Rouillard, JM, Herbert, CJ, Zuker, M (2002) OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics* 18: 486–487
 75. Rouillard, JM, Zuker, M, Gulari, E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 31: 3057–3062
 76. Rudi, K, Skulberg, OM, Skulberg, R, Jakobsen, KS (2000) Application of sequence-specific labeled 16S rRNA gene oligonucleotide probes for genetic profiling of cyanobacterial abundance and diversity by array hybridization. *Appl Environ Microbiol* 66: 4004–4011
 77. Salama, N, Guillemin, K, McDaniel, TK, Sherlock, G, Tompkins, L, Falkow, S (2000) A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* 97: 14668–14673
 78. Saleh-Lakha, S, Miller, M, Campbell, RG, Schneider, K, Elahimanes, P, Hart, MM, Trevors, JT (2005) Microbial gene expression in soil: methods, applications and challenges. *J Microbiol Methods* 63: 1–19
 79. Schadt, CW, Liebich, J, Chong, SC, Gentry, TJ, He, Z, Pan, H, Zhou, J (2005) Chapter 11, Design and use of functional gene microarrays (FGAs) for the characterization of microbial communities. In: Savidge, T, Pothulakis, H (Eds.) *Microbial Imaging, Methods in Microbiology*, vol. 34. Academic Press, Inc., London, UK, pp 331–368
 80. Schadt, CW, Zhou, J (2005) Advances in microarrays for soil microbial community analyses. In: Nannipieri, P, Smalla, K (Eds.) *Soil Biology: Nucleic Acids and Proteins in Soil*. Springer-Verlag, Berlin
 81. Sebat, JL, Colwell, FS, Crawford, RL (2003) Metagenomic profiling: microarray analysis of an environmental genomic library. *Appl Environ Microbiol* 69: 4927–4934
 82. Sessitsch, A, Gyamfi, S, Stralis-Pavese, N, Weilharter, A, Pfeifer, U (2002) RNA isolation from soil for bacterial community and functional analysis: evaluation of different extraction and soil conservation protocols. *J Microbiol Methods* 51: 171–179
 83. Small, J, Call, DR, Brockman, FJ, Straub, TM, Chandler, DP (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* 67: 4708–4716
 84. Stahl, DA, Lane, DJ, Olsen, GJ, Pace, NR (1985) Characterization of a Yellowstone hot spring microbial community by 5S ribosomal-RNA sequences. *Appl Environ Microbiol* 49: 1379–1384
 85. Steward, GF, Jenkins, BD, Ward, BB, Zehr, JP (2004) Development and testing of a DNA microarray to assess nitrogenase (*nifH*) gene diversity. *Appl Environ Microbiol* 70: 1455–1465
 86. Stralis-Pavese, N, Sessitsch, A, Weilharter, A, Reichenauer, T, Riesing, J, Csontos, J, Murrell, JC, Bodrossy, L (2004) Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ Microbiol* 6: 347–363
 87. Suzuki, MT, Giovannoni, SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* 62: 625–630
 88. Taroncher-Oldenburg, G, Griner, EM, Francis, CA, Ward, BB (2003) Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* 69: 1159–1171
 89. Thompson, DK, Beliaev, AS, Giometti, CS, Tollaksen, SL, Khare, T, Lies, DP, Neelson, KH, Lim, H, Yates, J, 3rd, Brandt, CC, Tiedje, JM, 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
 90. Tiquia, SM, Wu, L, Chong, SC, Passovets, S, Xu, D, Xu, Y, Zhou, J (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* 36: 664–675
 91. Torsvik, V, Goksoyr, J, Daae, FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56: 782–787
 92. Torsvik, V, Daae, FL, Sandaa, RA, Ovreas, L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* 64: 53–62
 93. Torsvik, V, Ovreas, L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* 5: 240–245
 94. Torsvik, V, Ovreas, L, Thingstad, TF (2002) Prokaryotic diversity—magnitude, dynamics, and controlling factors. *Science* 296: 1064–1066
 95. Tringe, SG, von Mering, C, Kobayashi, A, Salamov, AA, Chen, K, Chang, HW, Podar, M, Short, JM, Mathur, EJ, Detter, JC, Bork, P, Hugenholtz, P, Rubin, EM (2005) Comparative metagenomics of microbial communities. *Science* 308: 554–557
 96. Tsai, YL, Le, JY, Olson, BH (2003) Magnetic bead hybridization to detect enterotoxigenic *Escherichia coli* strains associated with cattle in environmental water sources. *Can J Microbiol* 49: 391–398
 97. Tyson, GW, Chapman, J, Hugenholtz, P, Allen, EE, Ram, RJ, Richardson, PM, Solovyev, VV, Rubin, EM, Rokhsar, DS, Banfield, JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37–43
 98. Urakawa, H, El Fantroussi, S, Smidt, H, Smoot, JC, Tribou, EH, Kelly, JJ, Noble, PA, Stahl, DA (2003) Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl Environ Microbiol* 69: 2848–2856
 99. Van Dyk, TK, DeRose, EJ, Gonye, GE (2001) LuxArray, a high-density, genomewide transcription analysis of *Escherichia coli* using bioluminescent reporter strains. *J Bacteriol* 183: 5496–5505
 100. Venter, JC, Remington, K, Heidelberg, JF, Halpern, AL, Rusch, D, Eisen, JA, Wu, D, Paulsen, I, Nelson, KE, Nelson, W, Fouts, DE, Levy, S, Knap, AH, Lomas, MW, Neelson, K, White, O, Peterson, J, Hoffman, J, Parsons, R, Baden-Tillson, H, Pfannkoch, C,

- Rogers, YH, Smith, HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74
101. von Wintzingerode, F, Gobel, UB, Stackebrandt, E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21: 213–229
102. Wang, X, Seed, B (2003) Selection of oligonucleotide probes for protein coding sequences. *Bioinformatics* 19: 796–802
103. Wilson, KH, Wilson, WJ, Radosevich, JL, DeSantis, TZ, Viswanathan, VS, Kuczmariski, TA, Andersen, GL (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* 68: 2535–2541
104. Woese, CR, Magrum, LJ, Gupta, R, Siegel, RB, Stahl, DA, Kop, J, Crawford, N, Brosius, J, Gutell, R, Hogan, JJ, Noller, HF (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 8: 2275–2293
105. Woese, CR, Gutell, R, Gupta, R, Noller, HF (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* 47: 621–669
106. Wu, LY, Thompson, DK, Li, GS, Hurt, RA, Tiedje, JM, Zhou, JZ (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* 67: 5780–5790
107. Wu, LY, Thompson, DK, Liu, XD, Fields, MW, Bagwell, CE, Tiedje, JM, Zhou, JZ (2004) Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ Sci Technol* 38: 6775–6782
108. Xu, D, Li, G, Wu, L, Zhou, J, Xu, Y (2002) PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis. *Bioinformatics* 18: 1432–1437
109. Yilmaz, LS, Noguera, DR (2004) Mechanistic approach to the problem of hybridization efficiency in fluorescent *in situ* hybridization. *Appl Environ Microbiol* 70: 7126–7139
110. Yilmaz, LS, Okten, HE, Noguera, DR (2006) Making all parts of the 16S rRNA of *Escherichia coli* accessible *in situ* to single DNA oligonucleotides. *Appl Environ Microbiol* 72: 733–744
111. Zhou, J, Xia, B, Treves, DS, Wu, LY, Marsh, TL, O'Neill, RV, Palumbo, AV, Tiedje, JM (2002) Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* 68: 326–334
112. Zhou, J (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* 6: 288–294
113. Zhou, JZ, Thompson, DK (2002) Challenges in applying microarrays to environmental studies. *Curr Opin Biotechnol* 13: 204–207
114. Zhou, X, Wu, L, Zhou, J (2004) Fabrication of DNA microarrays on nanoengineered polymeric ultrathin film prepared by self-assembly of polyelectrolyte multilayers. *Langmuir* 20: 8877–8885
115. Zhou, X, Zhou, J (2004) Improving the signal sensitivity and photostability of DNA hybridizations on microarrays by using dye-doped core-shell silica nanoparticles. *Anal Chem* 76: 5302–5312
116. Zuker, M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415
117. Zverlov, V, Klein, M, Lucker, S, Friedrich, MW, Kellermann, J, Stahl, DA, Loy, A, Wagner, M (2005) Lateral gene transfer of dissimilatory (bi)sulfite reductase revisited. *J Bacteriol* 187: 2203–2208