9 Advances in Microarray-Based Technologies for Soil Microbial Community Analyses

Christopher W. Schadt, Jizhong Zhou

9.1 Introduction

Studies on the diversity of microbial life within soils indicate that there are several thousand distinct microbial genome types present per gram of soil (Torsvik et al. 1990, 1996), and efforts to further characterise this diversity have revealed significant numbers of previously unknown organisms within the bacterial, archaeal and eukaryal domains of life (Bintrim et al. 1997; Hugenholtz et al. 1998; Dawson and Pace 2002; Schadt et al. 2003; Zhou et al. 2003). Microbial communities in soils also often exhibit extreme spatial and temporal variability; however, the factors that underlie and influence this variation remain largely unknown (Zhou et al. 2003). Perhaps one reason that this goal has remained elusive is a lack of common methodologies that allow for rapid, inexpensive and comprehensive characterisation of these extremely high levels of diversity.

The application of microarrays to the study of gene expression in microbial genomics now regularly allows for comprehensive, high-throughput studies of organisms under a variety of physiological conditions (e.g. Lockhart et al. 1996; Schena et al. 1996; DeRisi et al. 1997; Ye et al. 2000; Thompson et al. 2002; Liu et al. 2003) and this technology is becoming cheaper and easier to implement. Currently, the adaptation of such techniques to problems of microbial ecology in soils and other environments is a major focus effort that is taking place in many research groups. However, the adaptation of these technologies for the detection of microbial genes in complex environmental samples presents many unique challenges (Zhou and Thompson 2002; Zhou 2003; Schadt et al. 2005). This chapter reviews the progress of those efforts to date with an emphasis on the potential application of this technology to questions in soil ecology.

Christopher W. Schadt, Jizhong Zhou: Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831–6038, USA, E-mail: zhouj@ornl.gov

Soil Biology, Volume 8
Nucleic Acids and Proteins in Soil
P. Nannipieri, K. Smalla (Eds.)
© Springer-Verlag Berlin Heidelberg 2006

9.2
Types of Environmental Microarrays

Based on the type of probes arrayed and their potential applications, Zhou and Thompson (2002) classified microarrays of potential use in environmental studies into three different basic groups: (i) *Phylogenetic oligonucleotide arrays* (POAs) primarily contain sequence probes derived from rRNA genes. These types of arrays might be used for the analysis of microbial community composition and structure under a phylogenetic framework (e.g. Loy et al. 2002; Valinsky et al. 2002a,b; El Fantroussi et al. 2003). (ii) *Functional gene arrays* (FGAs) contain probe sequences for genes encoding key proteins involved in various environmental processes such as steps in denitrification or the degradation of naphthalene. These arrays are useful for monitoring the physiological status and functional activities of microbial communities in natural environments (e.g. Wu et al. 2001; Taroncher-Oldenburg et al. 2003; Rhee et al. 2004). (iii) *Community genome arrays* (CGAs) are constructed using whole genomic or large fragment DNA isolated from pure culture microorganisms or even large fragments of genomic DNA originating from fosmid or BAC libraries. These arrays can be useful to describe a microbial community in terms of its cultivable component or screen isolates and metagenomic libraries for discovery of target genes and organisms (e.g. Cho and Tiedje 2001; Wu et al. 2004).

9.3
Important Issues in Microarray Analysis

9.3.1
Specificity

Although gene-based oligonucleotide microarrays have been used successfully for monitoring genome-wide gene expression, the use of POAs and FGAs with complex environmental samples presents unique technical challenges in terms of specificity. The highly conserved nature of the rRNA, and the often high degree of similarity among homologous and orthologous functional genes, makes specific probe design much more difficult to achieve. Additionally, especially in the case of rRNA, the hybridisation of oligonucleotide probes to target nucleic acids possessing stable secondary structure can be particularly challenging. Stable secondary structure of target DNA or RNA must be overcome in order to make complementary sequence regions available for duplex formation with probes and, for this reason, these regions should be avoided. Thus the stable secondary struc-

ture of rRNAs can have serious effects on hybridisation specificity and detection sensitivity.

The potential advantage of oligonucleotide probes is that target sequences containing single base mismatches can be differentiated by microarray hybridisation. However, in reality, with the high number of possible probe and target combinations, this has been difficult to systematically demonstrate. To systematically determine whether single mismatch discrimination can be achieved for 16S rRNA genes using microarray hybridisation, a model oligonucleotide microarray consisting of probes derived from three different regions of the 16S rRNA molecule corresponding to different bacterial taxa was generated in our laboratory (X. Zhou and J. Zhou, unpubl. data). The probes had 1–5 mismatches in different combinations along the length of the oligonucleotide probe with at least one mismatch at the central position. Hybridisation signal intensity with a single-base mismatch was decreased by 10–30% depending on the type and position of the mismatched nucleotide base. The signal intensity of probes with two base mismatches was 5–25% of that of the perfect match probes. Probes with three or four base-pair mismatches yielded signal intensities that were 5% of that of the perfect match probes. Maximum discrimination and signal intensity was achieved with 19-base probes. These results indicated that single base discrimination for small subunit (SSU) rRNA genes can be achieved with glass slide-based array hybridisation, but complete discrimination appears to be problematic with SSU rRNA genes and most likely not possible using only single mismatches. Letowski et al. (2004) recently published similar studies showing optimal discrimination was achieved with mismatches distributed nearly evenly across the oligos. Many others have reported similar findings (Liu et al. 2001; Loy et al. 2002). This incomplete discrimination presents great difficulties when DNA derived from diverse, uncharacterised environmental samples is hybridised against POAs and FGAs. Interpretation of signals will obviously be confounded by the fact that, in environmental samples, lower than expected signals could be due either to low abundance of the target or imperfect matches between probe and target.

Compared to the short oligonucleotides commonly used for rDNA in POAs, the protein-coding genes used in FGAs are often designed using longer probes (40–70mer). While these probes may provide less hybridisation specificity, they provide better sensitivity levels. Additionally, the higher nucleotide variability of such genes often allows for species-specific design even at such probe lengths. Our recent studies indicated that genes having less than 88% nucleotide sequence identity could most often be discriminated with FGAs under hybridisation conditions of high stringency (50% formamide and 50 °C; Rhee et al. 2004). Additionally, a further characterisation of probe specificity in our laboratory has shown that by

limiting non-target hybridisation free energy to > −35 kcal/mol and limiting long stretches of continuously matched non-target bases to less than 20 bp can provide for further predictability in specificity, when considered simultaneously with percent similarity. Taroncher-Oldenburg et al. (2003) have reported similar results for 70mer oligonucleotide FGAs designed for nitrogen cycle genes.

Our past results have also shown that DNA/DNA hybridisation on CGAs in the presence of 50% (v/v) formamide at 55 °C could discriminate between microbial genomes of different species within a genus, whereas genomes, in many cases, could not be clearly distinguished at the subspecies level (Wu et al. 2004). By raising the hybridisation temperature to 65–75 °C, CGA-based discrimination between closely related bacterial strains could be achieved. Although some genes within bacterial genomes, such as rRNA genes, are highly similar even among different strains or species, the presence of such highly conserved sequences does not appear to affect the overall hybridisation specificity in CGAs.

9.3.2 Sensitivity

Since very small hybridisation volumes are used in microarray experiments, it is generally thought that the sensitivity of microarray hybridisation is higher than that of conventional membrane-based hybridisation. However, the sensitivity of hybridisation with glass-based microarrays may still be on the order of 100- to 10,000-fold less than with polymerase chain reaction (PCR) amplification and 100-fold less than with membrane-based hybridisation (Taniguchi et al. 2001). One of the main reasons that the sensitivity of glass-based microarray hybridisation is lower than membrane-based hybridisation could be because the probe-binding capacity on glass surfaces is much lower than that on porous membranes (Cho and Tiedje 2001). Increasing the DNA-binding capacity of glass slide microarrays could be one way to enhance microarray hybridisation sensitivity, and efforts to accomplish this using new types of chemical compounds for coating glass slides are ongoing.

Often detection limits are assessed using a known amount of target DNA for a given probe and progressively reducing target DNA concentration until the signal cannot be distinguished reliably from background fluorescence [e.g. signal-to-noise ratio (SNR) <2]. Using a 50mer biodegradation FGA and temperature of 50 °C in the presence of 50% formamide, Rhee et al. (2004) showed gene-dependent detection sensitivities to be between 5–10 ng in the absence of other DNA. However, DNA targets of environmental samples are present in a complex background of non-target DNA.

When assessed in a background of 2 μg of non-target *Shewanella* DNA, it was found that the detection limit for the same genes was about 10 times lower (∼50–100 ng for a range of gene/probe targets). Assuming 4 Mbp genome size, this corresponds to a detection limit of $\sim 10^7$ cells being required for a gene to be detected in an environmental sample. These results correspond well with other published studies of detection limits. Using *nir*S gene PCR products as probes about 1 ng of labeled pure genomic DNA in the absence of background DNA could be detected (Wu et al. 2001). With PCR product targets of 500–900 bp, Cho and Tiedje (2002) showed detection limits to be ∼10 pg in the presence of background DNA. The detection limit of CGAs is much lower and was estimated to be ∼0.2 ng with pure labeled genomic DNA in the absence of background DNA, and about 5 ng of genomic DNA or 2.5×10^5 cells in the presence of background DNA (Wu et al. 2004). While such limits should be sufficient for detection of at least the more dominant members of a microbial community, this is still not sensitive enough to detect less abundant microbial populations in complex communities. To detect rare populations in natural environments, other approaches for increasing hybridisation sensitivity are needed. Our and other groups are currently exploring ways to enhance the levels of detection sensitivity.

9.3.3 Quantitation

We have evaluated the quantitative potential of microarray hybridisation for both FGA (Wu et al. 2001; Rhee et al. 2004; Tiquia et al. 2004) and CGA formats (Wu et al. 2004). Linear quantitative relationships are most often observed between signal intensity and target DNA concentration over concentration ranges from 1–1,000 ng for both pure cultures and mixed target DNA populations. These results suggest that DNA microarrays may potentially be used for quantitative analysis of environmental samples. However, the difficult challenge in quantifying microbial populations in natural environments is that the community composition of the environmental samples is largely unknown. It is assumed that hybridisation signal intensity is directly related to population abundance of the target organism. However, non-specific hybridisation due to unknown diverse members in a sample may contribute to signals.

One way to assess quantitative accuracy of microarray hybridisations is the use of real-time PCR with sequence-specific primers. While it is not possible to verify an entire large microarray data set because these techniques are labour and time intensive, verification of microarray findings for several key genes is achievable and is common practice in microbial

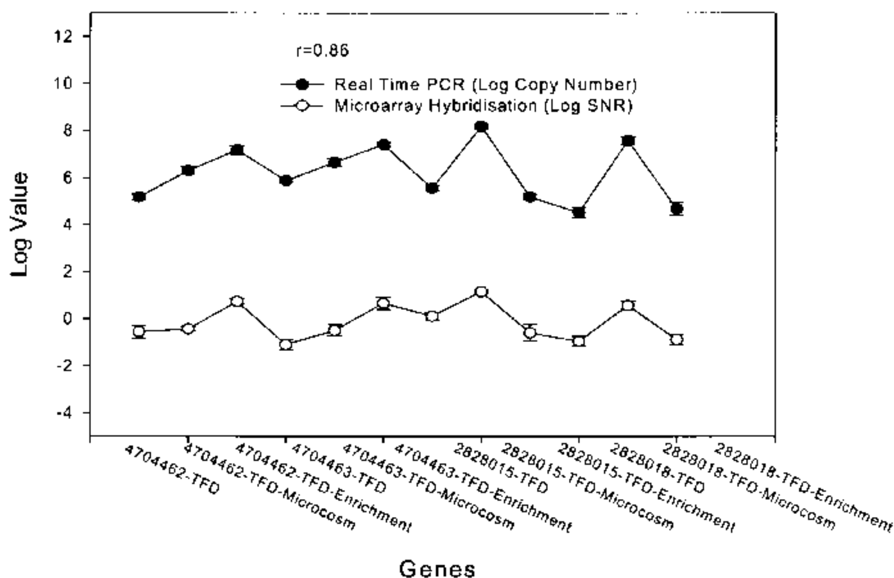


Fig. 9.1. Correspondence of microarray-based hybridisation signals with quantitative PCR for genes associated with naphthalene degradation (Rhee et al. 2004). Quantitative PCR may be used for independent verification of key results associated with microarray studies

gene expression studies. Rhee et al. (2004) used such an approach to verify microarray hybridisation results associated with naphthalene degradation (Fig. 9.1). Overall, significant correlations between the microarray hybridisation signals and the gene copy number determined by real-time PCR were obtained with $r^2 = 0.74$ for all genes tested, and $r^2 = 0.96$ for the genes with SNR > 3. Such results showing consistency between microarray hybridisation data and real-time PCR data are encouraging for the use of microarray data in quantitative interpretations.

9.4 Applications of Different Formats of Microarrays

9.4.1 Phylogenetic Oligonucleotide Arrays (POAs)

POAs primarily contain probes/sequences derived from rRNA genes; however, efforts have been made to incorporate other markers such as gyrB (Fukushima et al. 2003; Kakinuma et al. 2003) or rpoB (Dahllöf et al. 2000). The probes incorporated into POAs may be designed on a species-specific basis, a hierarchical manner, or some combination thereof, such

that several different taxonomic or phylogenetic levels of information are targeted. While species-specific probes can be easily generated on a one-by-one basis using relatively computationally simple algorithms or searches such as those available through the ribosomal database project (RDP; Cole et al. 2003), ARB (Ludwig et al. 2004) and ProbeBase (Loy et al. 2003), comprehensive and effective hierarchical probe design remains more difficult. These efforts require full and accurate alignments as well as robust phylogenetic information in order to accurately target specific groups of organisms and predict hierarchical probe behaviour. However, recently, significant progress has been made towards these fundamental issues. Using new search methods and quality score calculations, Zhang et al. (2002) were able to identify 10,487 perfect-match 15–25mers that together covered more than 86% of the phylogenetic nodes in a representative bacterial tree that included 929 operational taxonomic units (OTUs). This represented a core set from the RDP of highly reliable and complete sequences. Additionally, a new potential resource for comprehensive probe design called CASCADE-P is also under development (http://greengenes.llnl.gov/16S/). Through the incorporation of new global alignment algorithms (DeSantis et al. 2003), this group hopes to offer comprehensive data mining and probe design tools to the scientific community in the near future. While still under evaluation, these tools have already been used to build a 500,000 feature Affymetrix microarray containing sets of hierarchical probes (up to 25) for each OTU in the RDP database. Such phylogenetic arrays, once fully developed and evaluated, will be extremely useful for the analysis of microbial community composition and structure under a phylogenetic framework. Additionally, these methods offer much more quantitative potential than PCR-based methods. Unfortunately, because of low sensitivity levels, at present most studies still rely on a pre-hybridisation PCR amplification. This is feasible with POAs because of the highly conserved nature of the targeted 16S rDNA, but greatly complicates the quantitative interpretation of results due to the potential for biases in PCR amplifications. Additionally, as expression of rRNA has been shown to be highly correlated with growth rate and activity level (Asai et al. 1999a,b), rRNA targets could potentially also provide information on relative activity levels of the different target organisms on POAs if the sensitivity of hybridisations could be improved.

The actual application of POAs to the study of microbial ecology has come thus far from only a few example studies. While no comprehensive studies have yet been published from any environment, several groups have made great progress in targeting specific groups of organisms. For example, Loy et al. (2002) have developed a POA that includes probes claimed to distinguish all known sulfate-reducing lineages based on 16S rDNA sequences. In evaluative tests, this array of 132 probes was able to distinguish at least 41 species of sulfate-reducing taxa. In some cases these probes were

specific at even the strain and isolate level. Additionally, in application to mixed environmental samples from periodontal and hypersaline mat communities, this array returned results suggesting the presence of dominant taxa that were consistent with results from PCR-based studies targeting 16S rDNA and dissimilatory sulfite reductase (dsrAB) genes. Another recent study used a promising new combination of techniques to track active ammonium-oxidising bacteria in activated sludge samples (Adamczyk et al. 2003). The authors combined isotope labeling to track ammonia-oxidising bacteria incorporating ^{14}C-labeled bicarbonate, with a POA for the detection of specific rRNA associated with these organisms. The total rRNA in the sample was then labeled with Cy3 dyes, and, by using sequential fluorescent and autoradiographic scans of the resulting hybridizations, the authors could estimate the proportion of the population that was actively incorporating the labeled bicarbonate. Both of these studies demonstrate the promise of array technologies. However, both are extremely limited in depth of application and do not fully realise the potential of microarrays for comprehensive, high-throughput analyses.

9.4.2 Functional Gene Arrays (FGAs)

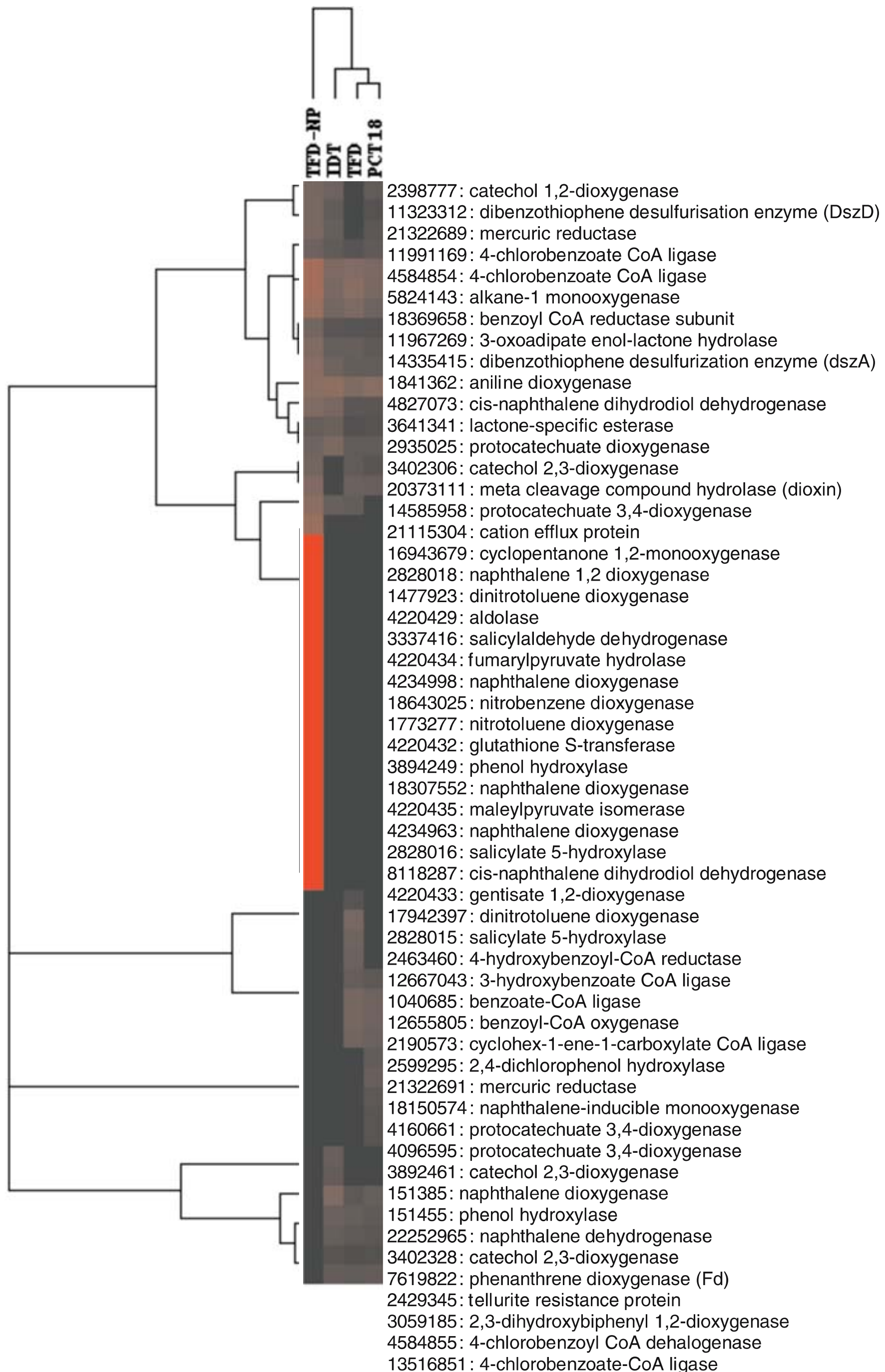
FGAs contain genes encoding key proteins involved in various environmental processes. Microbial involvement in processes important to ecosystem biogeochemistry such as nitrification, denitrification and sulfate reduction in ocean sediments (Wu et al. 2001; Tiquia et al. 2004), nitrogen fixation in picoplankton communities (Jenkins et al. 2004; Steward et al. 2004), and naphthalene degradation in soils (Rhee et al. 2004) have recently been targets of studies using FGAs. Our laboratory has been actively developing FGAs designed to be comprehensive for the known diversity of microbial functional genes involved in N, C (including methane) and S cycling, as well as genes involved in the biodegradation of organic contaminants and in the resistance to metal toxicity (Wu et al. 2001; Zhou 2003; Rhee et al. 2004; Schadt et al. 2005; Tiquia et al. 2004).

A 50mer-based oligonucleotide microarray was developed in our laboratory based on known genes and pathways involved in: biodegradation, metal resistance and reduction, denitrification, nitrification, nitrogen fixation, methane oxidation, methanogenesis, carbon polymer decomposition, and sulfate reduction. Rhee et al. (2004) have demonstrated the applicability of this type of array design in studies of degradation of naphthalene from both enrichments and soil microcosms. In this study it was demonstrated using both unamplified DNA and mRNA as targets for probes that different sets of organisms were primarily responsible for naphthalene

degradation in each system. While the naphthalene-degrading genes from *Rhodococcus*-type microorganisms were found to be dominant in enrichments, the genes involved in naphthalene degradation from Gram-negative microorganisms, such as *Ralstonia*, *Comamonas*, *Burkholderia*, were most abundant in the soil microcosms (Fig. 9.2). Although naphthalene degradation is widely known and studied primarily in *Pseudomonas*, genes from these organisms were not detected in either system. Comparatively little is known about the organisms and pathways found to dominate this process in both systems. To verify these results four genes detected in these hybridisations were also amplified and quantified using real-time PCR and results showed good correlation between the two methods (Fig. 9.1).

9.4.3 Community Genome Arrays (CGAs)

The CGA is conceptually analogous to membrane-based reverse sample genome probing (RSGP; Greene and Voordouw 2003). In contrast to RSGP, CGAs use a non-porous surface, a microarray, for fluorescence-based detection. This development enables the utility of RSGP to be scaled up in the number of hybridisations that can be achieved simultaneously allowing high-throughput analyses. These arrays have been used to describe a microbial community in terms of its cultivable component or screen isolates for identification (Cho and Tiedje 2001; Wu et al. 2004). Because of the ease of construction and greater sensitivity, such arrays may be particularly useful for monitoring the abundance of known species that are important in a given environment. Like RSGP, the main disadvantage of the CGAs has been that only the cultured components of a community could be monitored, because the construction required the availability of large amounts of high-molecular-weight genomic DNA from individual pure sources. However, CGA-based hybridisation itself does not require culturing and, with the recent advances in environmental genomics, high-molecular-weight DNA from uncultivated microorganisms can also be accessed through bacterial artificial chromosome (BAC)-based or fosmid-based cloning approaches (e.g. metagenomes). BAC or fosmid clones could then be used to fabricate CGAs, thus allowing the further investigation of uncultivated components of a complex microbial community. Additionally, these types of arrays could potentially be very useful for the screening of metagenomes for specific genes of interest. For example, preselected 16S rDNA targets or functional gene targets involved in C cycling, N cycling, etc., could be hybridised against arrayed environmental BAC or fosmid clones to better identify those clones of interest for further study, such as high-throughput genome-sequencing efforts of inserts.


TFD-NP
IDT
TFD
PCT18
2398777: catechol 1,2-dioxygenase
11323312: dibenzothiophene desulfurisation enzyme (DszD)
21322689: mercuric reductase
11991169: 4-chlorobenzoate CoA ligase
4584854: 4-chlorobenzoate CoA ligase
5824143: alkane-1 monooxygenase
18369658: benzoyl CoA reductase subunit
11967269: 3-oxoadipate enol-lactone hydrolase
14335415: dibenzothiophene desulfurization enzyme (dszA)
1841362: aniline dioxygenase
4827073: cis-naphthalene dihydrodiol dehydrogenase
3641341: lactone-specific esterase
2935025: protocatechuate dioxygenase
3402306: catechol 2,3-dioxygenase
20373111: meta cleavage compound hydrolase (dioxin)
14585958: protocatechuate 3,4-dioxygenase
21115304: cation efflux protein
16943679: cyclopentanone 1,2-monooxygenase
2828018: naphthalene 1,2 dioxygenase
1477923: dinitrotoluene dioxygenase
4220429: aldolase
3337416: salicylaldehyde dehydrogenase
4220434: fumarylpyruvate hydrolase
4234998: naphthalene dioxygenase
18643025: nitrobenzene dioxygenase
1773277: nitrotoluene dioxygenase
4220432: glutathione S-transferase
3894249: phenol hydroxylase
18307552: naphthalene dioxygenase
4220435: maleylpyruvate isomerase
4234963: naphthalene dioxygenase
2828016: salicylate 5-hydroxylase
8118287: cis-naphthalene dihydrodiol dehydrogenase
4220433: gentisate 1,2-dioxygenase
17942397: dinitrotoluene dioxygenase
2828015: salicylate 5-hydroxylase
2463460: 4-hydroxybenzoyl-CoA reductase
12667043: 3-hydroxybenzoate CoA ligase
1040685: benzoate-CoA ligase
12655805: benzoyl-CoA oxygenase
2190573: cyclohex-1-ene-1-carboxylate CoA ligase
2599295: 2,4-dichlorophenol hydroxylase
21322691: mercuric reductase
18150574: naphthalene-inducible monooxygenase
4160661: protocatechuate 3,4-dioxygenase
4096595: protocatechuate 3,4-dioxygenase
3892461: catechol 2,3-dioxygenase
151385: naphthalene dioxygenase
151455: phenol hydroxylase
22252965: naphthalene dehydrogenase
3402328: catechol 2,3-dioxygenase
7619822: phenanthrene dioxygenase (Fd)
2429345: tellurite resistance protein
3059185: 2,3-dihydroxybiphenyl 1,2-dioxygenase
4584855: 4-chlorobenzoyl CoA dehalogenase
13516851: 4-chlorobenzoate-CoA ligase

◀ **Fig. 9.2.** Cluster analysis of naphthalene-enriched (TFD-NP) and control DNA microarray hybridisations using a 50mer oligonucleotide FGA (Rhee et al. 2004). Genes involved in naphthalene biodegradation show significantly greater average hybridisation signals (as indicated by more intense red coloration) in naphthalene-enriched soils. Cluster analyses allow for straight forward visualisation and analysis of differences between microarray hybridisations

Wu et al. (2004) recently tested a multispecies and strain CGA using mixtures of known amounts of DNA or known number of cells from 16 different species. This CGA was used to examine differences in microbial community composition in soil, river, and marine sediments. Consistent with expectations, principal components analysis of the microarray results revealed samples from within each habitat to contain communities more similar to one another than between habitat types (Fig. 9.3). The microarray data also revealed considerable differences for some abundant microbial populations among these samples. For instance, *Marinobacter* species similar to the strains D3-15, D5-10, and E1-7, and *Shewanella* species similar to *S. algae* and *S. oneidensis*, were abundant in the top layer of a marine sediment sample; however, they were not observed in microarray hybridisations of soil and river sediments. Similarly, bacteria similar to the *Pseudomonas* sp. G179, a typical *nir*K-containing soil denitrifier, were abundant in soil samples but not in the river and marine sediment samples. Such results suggest that these arrays may be especially useful for monitoring the abundance of microbial targets of known interest in a given environment.

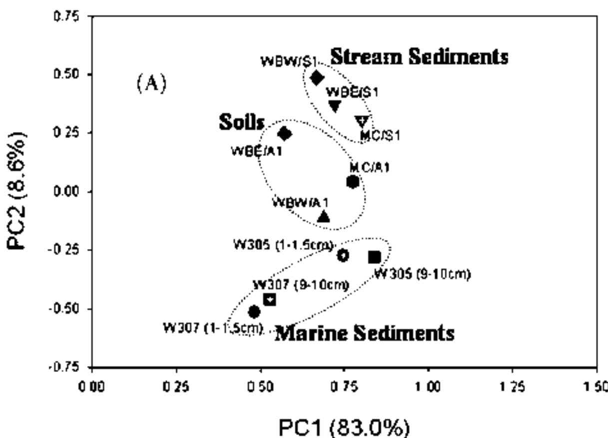


Fig. 9.3. Principal components analysis (PCA) showing the ability of CGAs to distinguish among environmental microbial DNA samples of different origins (Wu et al. 2004). PCA allows for systematic reduction and visualisation of patterns within the large and complex volumes of data generated by microarray analysis

9.5 Conclusions and Future Perspectives

Several challenges still remain for environmental microarrays to reach their full potential. Firstly, inherent experimental variation is a critical issue in microarray hybridisation and novel experimental designs and strategies for evaluating such variation are needed for improving the accuracy of microarray-based studies. Standardisation procedures and technologies need to be developed so that microarray results can be easily compared across laboratories and microarray platforms. Second, microarray hybridisation is still not sensitive enough for some environmental studies when the amount of recoverable biomass is very low and community complexity is very high. Novel methods for increasing detection sensitivity are urgently needed. Third, the quantity of data generated by microarray-based studies of environmental samples will likely be enormous, but rapid processing and mining of hybridisation data still remain difficult endeavours. Finally, although microarrays may provide a rapid means of characterising a microbial community once they are constructed, preparing high-quality samples and probes suitable for microarray analysis appears to be the bottleneck. Automation and improvements in sample processing and probe design are necessary for alleviating this bottleneck.

Despite these remaining issues, the usefulness of this technology is rapidly improving. Once arrays are fully developed, they will be very useful for microbial ecology studies. With fully developed and comprehensive arrays it will become possible to assess simultaneously in a single assay all or most of the constituents of a complex natural community, which could allow us to begin to build a comprehensive view of microbial community dynamics. Although some of the studies reviewed here have shown that microarrays hold promise as valuable tools for analysing environmental samples, the specificity, sensitivity, and quantitative capabilities of microarray technology for environmental applications require close attention and further refinement. Additionally, studies have yet to take full advantage of the high-throughput capabilities of microarrays. More rigorous and systematic assessment, development and applications are needed to realise the full potential of microarrays for microbial ecology studies.

Acknowledgements. The authors participation in this chapter was funded by the US DOE Office of Science, Biological and Environmental Research Programs for: Carbon Sequestration (as part of the consortium on research to enhance Carbon Sequestration in Terrestrial Ecosystems–CSiTE); Biotechnological Investigations – Ocean Margins Program (BI-OMP); Natural and Accelerated Bioremediation Research (NABIR); and Genomes to Life (GTL)

programs. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725.

References

Adamczyk J, Hesselsoe M, Iversen N, Horn M, Lehner A, Nielsen PH, Schloter 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

Asai T, Zaporojets D, Squires C, Squires CL (1999a) An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc Natl Acad Sci 96:1971–1976

Asai T, Condon C, Voulgaris J, Zaporojets D, Shen B, Al-Omar M, Squires C, Squires CL (1999b) Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons. J Bacteriol 181:3803–3809

Bintrim SB, Donohue TJ, Handelsman J, Roberts GP, Goodman RM (1997) Molecular phylogeny of archaea from soil. Proc Natl Acad Sci USA 94:277–282

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, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res 31:442–443

Dahllöf I, Baillie H, Kjelleberg S (2000) rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Appl Environ Microbiol 66:3376–3380

Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. Proc Natl Acad Sci USA 99:8324–8329

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

DeSantis TZ, Dubosarskiy I, Murray SR, Andersen GL (2003) Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics 19:1461–1468

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

Fukushima M, Kakinuma K, Hayashi H, Nagai H, Ito K, Kawaguchi R (2003) Detection and identification of *Mycobacterium* species isolates by DNA microarray. J Clin Microbiol 41:2605–2615

Greene EA, Voordouw G (2003) Analysis of environmental microbial communities by reverse sample genome probing. J Microbiol Methods 53:211–219

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774

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

Kakinuma K, Fukushima M, Kawaguchi R (2003) Detection and identification of *Escherichia coli*, *Shigella*, and *Salmonella* by microarrays using the *gyrB* gene. Biotechnol Bioeng 20:721–728

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 57:369–278

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

Liu Y, Zhou J-Z, Omelchenko M, Beliaev A, Venkateswaran A, Stair J, Wu L, Thompson DK, Xu D, Rogozin IB, Gaidamakova EK, Zhai M, Makarova KS, Koonin EV, Daly MJ (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Proc Natl Acad Sci 100:4191–4196

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675–1680

Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J, Schleifer K-H, 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

Loy A, Horn M, Wagner M (2003) probeBase: an online resource for rRNA-targeted oligonucleotide probes. Nucleic Acids Res 31:514–516

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer K-H (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32:1363–1371

Rhee SK, Liu X, Wu L-Y, 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

Schadt CW, Martin AP, Lipson DA, Schmidt SK (2003) Seasonal dynamics of novel fungal lineages in tundra soils. Science 301:1359–1361

Schadt CW, Liebich J, Chong SC, Gentry TJ, He Z, Pan H, Zhou J-Z (2005) Design and use of functional gene microarrays (FGAs) for the characterization of microbial communities. In: Savidge T, Charalabos P (eds) Methods in microbiology, vol 34: Microbial imaging. Elsevier, Amsterdam, pp 331–368

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

Steward GF, Jenkins BD, Ward BB, Zehr JP (2004) Development and testing of a DNA macroarray to assess nitrogenase (*nif* H) gene diversity. Appl Environ Microbiol 70:1455–1465

Taniguchi M, Miura K, Iwao H, Yamanaka S (2001) Quantitative assessment of DNA microarrays – comparison with northern blot analyses. Genomics 71:34–39

Taroncher-Oldenburg G, Griner EM, Francis CA, Ward BB (2003) Oligonucleotide microarray for study of functional gene diversity in the nitrogen cycle in the environment. Appl Environ Microbiol 69:1159–1171

Thompson DK, Beliaev AS, Giometti CS, Tollaksen SL, Khare T, Lies DP, Nealson KH, Lim H, Yates J III, Brandt CC, Tiedje JM, Zhou J-Z (2002) Transcriptional and proteomic analysis of a ferric uptake regulator (Fur) mutant of *Shewanella oneidensis*: possible

involvement of Fur in energy metabolism, transcriptional regulation, and oxidative stress. Appl Environ Microbiol 68:881–892

Tiquia SM, Wu L, Chong SC, Passovets S, Xu D, Xu Y, Zhou J-Z (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. BioTechniques 36:664–675

Torsvik V, Goksoyr J, Daae FL (1990) High diversity in DNA of soil bacteria. Appl Environ Microbiol 56:782–787

Torsvik V, Sorheim R, Goksoyr J (1996) Total bacterial diversity in soil and sediment communities – a review. J Ind Microbiol 17:170–178

Valinsky L, Della-Vedova G, Jiang T, Borneman J (2002a) Oligonucleotide fingerprinting of ribosomal RNA genes for analysis of fungal community composition. Appl Environ Microbiol 68:5999–6004

Valinsky L, Della-Vedova G, Scupham AJ, Alvey S, Figueroa A, Yin B, Hartin J, Chrobak M, Crowley DE, Jiang T, Borneman J (2002b) Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. Appl Environ Microbiol 68:3243–3250

Wu L, Thompson DK, Li G, Hurt RA, Tiedje JM, Zhou J (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol 67:5780–5790

Wu L, Thompson DK, Liu X, Fields MW, Bagwell CE, Tiedje JM, Zhou J-Z (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

Ye RW, Tao W, Beddzyk L, Young T, Chen M, Li L (2000) Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. J Bacteriol 182:4458–4465

Zhang Z, Wilson RC, Fox GE (2002) Identification of characteristic oligonucleotides in the bacterial 16S ribosomal RNA sequence dataset. Bioinformatics 18:244–250

Zhou J-Z (2003) Microarrays for bacterial detection and microbial community analysis. Curr Opin Microbiol 6:288–294

Zhou J-Z, Thompson D (2002) Microarrays: applications in environmental microbiology. In: Bitton G (ed) Encyclopedia of environmental microbiology, vol. 4. John Wiley & Sons, New York, pp 1968–1979

Zhou J-Z, Xia B, Treves DS, Wu L, Marsh TL, O'Neill RV, Palumbo AV, Tiedje JM (2003) Spatial and resource factors influencing high microbial diversity in soil. Appl Environ Microbiol 68:326–334