Challenges in applying microarrays to environmental studies

Jizhong Zhou* and Dorothea K Thompson

Although DNA microarray technology has been used successfully to analyze global gene expression in pure cultures, it has not been rigorously tested and evaluated within the context of complex environmental samples. Adapting microarray hybridization for use in environmental studies faces several challenges associated with specificity, sensitivity and quantitation.

Addresses
Environmental Sciences Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831, USA
*e-mail: zhouj@ornl.gov

Current Opinion in Biotechnology 2002, **13**:204–207

0958-1669/02/$ – see front matter
Published by Elsevier Science Ltd.

Abbreviations

CGA	community genome array
FGA	functional gene array
PCR	polymerase chain reaction
rRNA	ribosomal ribonucleic acid
SSU	small subunit

Introduction

DNA- or oligonucleotide-based microarray technology is a powerful functional genomics tool that allows researchers to view the physiology of a living cell from a comprehensive and dynamic molecular perspective (e.g. [1–7]). Compared with traditional nucleic acid hybridization with porous membranes, glass-slide-based microarrays offer the additional advantages of high density, high sensitivity, rapid ('real-time') detection, lower cost, automation, and low background levels [8]. As a result, microarray-based technology is potentially well suited for identifying populations of microorganisms in natural environments. Target functional genes in the environment tend to be highly diverse, and it is difficult, sometimes even experimentally impossible, to identify conserved DNA sequence regions for designing oligonucleotide probes for hybridization or primers for polymerase chain reaction (PCR) amplification. The microarray-based approach does not require such sequence conservation, however, because all of the diverse gene sequences from different populations of the same functional group can be fabricated on arrays and used as probes to monitor their corresponding distributions in environmental samples.

In theory, microarray-based genomic technology provides the advantages necessary for comprehensive and quantitative characterization of complex microbial communities [9]; however, the performance of microarray hybridization has not been rigorously tested and validated with diverse environmental samples. In contrast with studies using pure cultures, microarray-mediated analysis of environmental nucleic acids presents several technical challenges that must be addressed. First, in environmental studies the target and probe sequences are very diverse. It is not clear whether the performance of microarrays with diverse environmental samples is similar to that with pure culture samples or how sequence divergence is reflected in hybridization signal intensity. Second, environmental samples are generally contaminated with substances such as humic matter, organic contaminants, and metals, which may interfere with DNA hybridization on microarrays. Third, in contrast to pure cultures, the retrievable biomass in environmental samples is generally low; consequently, it is not clear whether microarray hybridization is sensitive enough to detect microorganisms in all types of environmental samples. Finally, it is uncertain whether microarray-based detection can be quantitative. Environmental and ecological studies require experimental tools that not only detect the presence or absence of particular groups of microorganisms, but which also provide quantitative data on their *in situ* biological activities. The objective of this commentary is to discuss the recent work on the development and application of microarray formats for environmental microbial studies, with an emphasis on the key microarray hybridization issues of specificity, sensitivity and quantitation.

Specificity

On the basis of the type of probe arrayed, the microarrays used in environmental studies can be divided into three major classes [10]. First, phylogenetic oligonucleotide arrays (POAs) contain sequence probes derived from ribosomal ribonucleic acid (rRNA) genes and are used primarily for the phylogenetic analysis of microbial community composition and structure. Second, functional gene arrays (FGAs) contain genes encoding key enzymes involved in various biogeochemical cycling processes (e.g. denitrification and nitrification) and are useful for monitoring the physiological status and functional activities of microbial communities in natural environments [11]. Third, community genome arrays (CGAs) are constructed using whole genomic DNA isolated from pure culture microorganisms and can be used to describe a microbial community in terms of its cultivable component.

Unlike non-rRNA gene-based oligonucleotide microarrays, which have been used successfully to monitor genome-wide gene expression (e.g. [12,13]) and to detect genetic polymorphisms (e.g. [14,15]), rRNA gene-based oligonucleotide arrays present some unique technical challenges [10]. Because the rRNA gene is highly conserved at the nucleotide sequence level and is present in all microorganisms, specific detection with rRNA-targeted oligonucleotide microarrays can be difficult. In addition, the hybridization of oligonucleotide probes to target nucleic acids possessing stable secondary structure can be particularly challenging [16]. Any stable secondary structure of the target DNA or RNA must be overcome to make complementary sequence regions available for duplex formation. The

stable secondary structure of small subunit (SSU) rRNA will have serious effects on hybridization specificity and detection sensitivity.

In a study by Guschin *et al.* [9], gel-pad oligonucleotide microarrays were constructed using oligonucleotides complementary to SSU rRNA sequences from key genera of nitrifying bacteria. In the gel array format, synthetic oligonucleotide probes are immobilized in a matrix of polyacrylamide gel elements, which are affixed to glass slides. The results show that specific detection can be achieved with this type of microarray; however, the probe specificity depends on various factors, such as probe length. Guschin *et al.* [9] showed that as the length of the oligonucleotide probe increases, mismatch discrimination is lost; conversely, as the length of the probe decreases, hybridization signal intensity (i.e. sensitivity) is sacrificed. A recent study showed that gel-pad-based oligonucleotide microarrays can also be used to distinguish *Bacillus* species, namely *Bacillus thuringiensis* and *Bacillus subtilis* [17]. Using glass-based two-dimensional microarrays, Small *et al.* [16] detected metal-reducing bacteria, such as *Geobacter chapellei* and *Desulfovibrio desulfuricans*.

The potential advantage of oligonucleotide probes is that target sequences containing single-base mismatches can be differentiated by microarray hybridization; however, this has not been fully demonstrated with SSU rRNA gene-based microarrays. To systematically determine whether single mismatch discrimination can be achieved for SSU rRNA genes using microarray hybridization, we constructed a model oligonucleotide microarray consisting of probes derived from three different regions of the SSU rRNA molecule corresponding to different bacterial taxa (X Zhou, J Zhou, unpublished data). The probes had one to five mismatches in different combinations along the length of the oligonucleotide probe, with at least one mismatch at the central position. The hybridization signal intensity with a single-base mismatch was decreased by 10–30% depending on the type of 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 indicate that single-base discrimination for SSU rRNA genes can be achieved with glass-slide-based array hybridization, but complete discrimination appears to be problematic with SSU rRNA genes [16–18]. Urakawa *et al.* [18] demonstrated that the single base-pair near-terminal and terminal mismatches have a significant effect on hybridization signal intensity.

Compared with oligonucleotide microarrays, DNA-based microarrays such as FGAs and CGAs provide less hybridization specificity. Recent studies indicated that genes having less than 80–85% sequence identity could be reliably discriminated with FGAs under hybridization conditions of high stringency (65°C) [11]. Furthermore, microarray hybridization conditions can be adjusted to achieve a broad range of detection. At low stringency (e.g. 45°C), for example, genes with 60–70% sequence identity to the arrayed probe can consistently be detected. No substantial difference in fluorescence intensity was measured for probe sequences exhibiting 80–100% identity to the labeled target DNA [11]. An analysis of published sequences and our own unpublished sequences indicates that the genes involved in many important biogeochemical processes such as nitrogen fixation (e.g. *nifH*), denitrification (e.g. nitrite reductase genes *nirS* and *nirK*), and sulfate reduction (e.g. *dsrA/B*) are diverse in natural environments. Hence, the difficulty is how to capture all the sequence diversity present in natural environments and to appropriately interpret hybridization results from environmental samples of unknown diversity.

Our results also showed that DNA–DNA hybridization on CGAs in the presence of 50% (vol/vol) formamide at 55°C could discriminate between microbial genomes of different species within a genus, whereas, in many cases, genomes could not be clearly distinguished at the subspecies level (DK Thompson *et al.*, unpublished data). By raising the hybridization temperature to 65–75°C, CGA-based discrimination between closely related bacterial strains could be improved. Although some genes, 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 hybridization specificity.

Sensitivity

Sensitivity is another critical parameter that impacts the effectiveness of a microarray-based approach for detecting microorganisms. With oligonucleotide microarrays, the *G. chapellei* SSU rRNA gene could be detected using ~0.5 μg of total RNA extracted from soils [16]. Using FGAs, *nirS* genes were detected in only 1 ng of labeled pure genomic DNA and 25 ng of bulk community DNA extracted from surface soil samples [11]. Similarly, the detection limit of CGAs was estimated to be ~0.2 ng with labeled pure genomic DNA (DK Thompson *et al.*, unpublished data). These approximate levels of detection sensitivity should be sufficient for many studies in microbial ecology.

As very small hybridization volumes are used in microarray experiments, it is generally thought that the sensitivity of microarray hybridization is higher than that of conventional membrane-based hybridization [8]. However, the sensitivity of hybridization with glass-based microarrays may still be in the order of 100- to 10 000-fold less than with PCR amplification [19], and can be 10- to 100-fold less than with membrane-based hybridization [20]. One of the main reasons for the lower sensitivity of glass-based microarray hybridization as compared with membrane-based hybridization, could be that the probe-binding capacity on glass surfaces is much lower than on porous membranes. Increasing binding capacity could be one way to enhance microarray hybridization sensitivity.

Quantitation

Because of the inherently high variation associated with array fabrication, probe labeling, hybridization, and image processing, the accuracy of microarray-based quantitative assessment is still uncertain. Comparison of microarray hybridization results with previously known results suggested that microarray hybridization appears to be quantitative enough to detect differences in gene expression patterns under various conditions [1,13,21]. DNA microarrays have also been used to measure differences in DNA copy number in breast tumors [22,23]. Single-copy deletions or additions can be detected [23], suggesting that microarray-based detection is potentially quantitative. A recent study in which λ DNA was co-spotted with probe DNA also indicated that microarrays can accurately quantify genes in DNA samples [24].

We have evaluated the quantitative potential of microarray hybridization for both FGA and CGA formats. Linear quantitative relationships (r^2 = 0.89 to 0.95) were observed between signal intensity and target DNA concentration over a fourfold order of concentration range for both pure cultures and mixed target DNA populations [11] (DK Thompson *et al.*, unpublished data). This observation suggests that DNA microarrays may potentially be used for quantitative analysis of environmental samples. The difficult challenge in quantifying microbial populations in natural environments on the basis of hybridization signal intensity, however, is how to distinguish differences in hybridization intensity owing to population abundance from those due to sequence divergence [11].

Future perspectives

Because of their high-density and high-throughput capacity, it is expected that microarray-based genomic technologies will revolutionize the analysis of microbial community structure, function and dynamics. For the first time in history, the potential exists to assess simultaneously in a single assay all, or most, of the constituents of a complex natural community. With this facility it should be possible to begin to build a comprehensive, integrated view of the dynamics of a microbial community. Although several studies have shown that microarrays hold promise as valuable tools for analyzing environmental samples, the specificity, sensitivity, and quantitative capabilities of microarray technology for environmental applications are still in the early stages of evaluation.

More rigorous and systematic assessment and development are needed to realize the full potential of microarrays for microbial ecology studies. Several challenges will need to be addressed and overcome. First of all, inherent experimental variation is a critical issue in microarray hybridization when analyzing environmental samples. Novel experimental designs and strategies for evaluating and eliminating such variation will therefore be needed for improving microarray-based quantitative accuracy. Currently, it is difficult to compare microarray data in a meaningful way among different laboratories and even among different experiments in a single laboratory. Second, microarray hybridization is still not sensitive enough for some environmental studies when the amount of recoverable biomass is very low. PCR amplification is a widely used and powerful tool for increasing detection sensitivity. However, coupling single-gene-based PCR technology with multiple-gene-based microarray hybridization is a great challenge for quantitatively assessing microbial community structure and activity in natural environments, because efficient and non-biased amplification of all target genes with different primer sets is very difficult to achieve with PCR. Third, the quantity of data generated by microarray-based studies of environmental samples is likely to be enormous, but the rapid processing and mining of hybridization data still remain difficult endeavours. Bioinformatic tools developed for analyzing gene expression data can be used to analyze environmental samples to some extent, but may have difficulty in dealing with the complexity of environmental samples. Although the high-density capacity of microarrays is an advantage in dealing with the extreme sequence diversity of environmental samples, interpreting microarray hybridization results is a challenge when analyzing unknown environmental samples because of potential cross-hybridization and/or background. Novel bioinformatic tools are needed to efficiently discriminate positive hybridization signals from background noise. Fourth, the number of genes needed for monitoring on the whole-community scale could extend far beyond the current capacity of microarrays. Thus, advancements in the technology need to be made that lead to higher density microarrays with less variation. Fifth, although microarrays can provide a rapid means of characterizing a microbial community once they are constructed, preparing high-quality samples suitable for microarray analysis appears to be a bottleneck. Automation and improvements in sample processing will also be important for future studies. Finally, microarrays are only tools and, as such, they should be integrated with studies focused on clear ecological and environmental questions and hypotheses that can be addressed. Only in this way can the power of microarray hybridization for microbial community analysis be ascertained.

Acknowledgements

The authors' efforts in preparing this commentary were supported by the United States Department of Energy under the Natural and Accelerated Bioremediation Research Program, Microbial Genome Program, and Microbial Cell Project of the Office of Biological and Environmental Research. Oak Ridge National Laboratory is managed by the University of Tennessee-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

References

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

2. Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO, Yanofsky C: **DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli.*** *Proc Natl Acad Sci USA* 2000, **97**:12170-12175.

3. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of**

cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998, **9**:3273-3297.

4. Tao H, Bausch C, Richmond C, Blattner FR, Conway T: **Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media.** *J Bacteriol* 1999, **181**:6425-6440.

5. Wei Y, Lee JM, Richmond C, Blattner FR, Rafalski JA, LaRossa RA: **High-density microarray-mediated gene expression profiling of *Escherichia coli.*** *J Bacteriol* 2001, **183**:545-556.

6. Wodicka L, Dong H, Mittman M, Ho MH, Lockhart DJ: **Genome-wide expression monitoring in *Saccharomyces cerevisiae.*** *Nat Biotechnol* 1997, **15**:1359-1367.

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

8. Shalon D, Smith SJ, Brown PO: **A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.** *Genome Res* 1996, **6**:639-645.

9. Guschin DY, Mobarry BK, Proudnikov D, Stahl DA, Rittmann BE, Mirzabekov AD: **Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology.** *Appl Environ Microbiol* 1997, **63**:2397-2402.

10. Zhou J, Thompson D: **Microarrays: applications in environmental microbiology.** In *Encyclopedia of Environmental Microbiology*, vol 4. Edited by Bitton G. New York: John Wiley & Sons; 2002:1968-1979.

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

12. de Saizieu A, Certa U, Warrington J, Gray C, Keck W, Mous J: **Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays.** *Nat Biotechnol* 1998, **16**:45-48.

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

14. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA: **Accessing genetic information with high-density DNA arrays.** *Science* 1996, **274**:610-614.

15. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J *et al.*: **Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.** *Science* 1998, **280**:1077-1082.

16. Small J, Call DR, Brockman FJ, Straub TM, Chandler DP: **Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays.** *Appl Environ Microbiol* 2001, **67**:4708-4716.

17. Bavykin SG, Akowski JP, Zakhariev VM, Barsky VE, Perov AN, Mirzabekov AD: **Portable system for microbial sample preparation and oligonucleotide microarray analysis.** *Appl Environ Microbiol* 2001, **67**:922-928.

18. Urakawa H, Noble PA, El Fantroussi S, Kelly JJ, Stahl DA: **Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses.** *Appl Environ Microbiol* 2002, **68**:235-244.

19. Zhou JZ, Davey ME, Figueras JB, Rivkina E, Glichinsky D, Tiedje JM: **Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA.** *Microbiology* 1997, **143**:3913-3919.

20. Voordouw G, Shen Y, Harrington CS, Telang AJ, Jack TR, Westlake DWS: **Quantitative reverse sample genome probing of microbial communities and its application to oil field production waters.** *Appl Environ Microbiol* 1993, **59**:4101-4114.

21. Taniguchi M, Miura K, Iwao H, Yamanaka S: **Quantitative assessment of DNA microarrays — comparison with northern blot analyses.** *Genomics* 2001, **71**:34-39.

22. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y *et al.*: **High-resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207-211.

23. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays.** *Nat Genet* 1999, **23**:41-46.

24. Cho JC, Tiedje JM: **Bacterial species determination from DNA–DNA hybridization by using genome fragments and DNA microarrays.** *Appl Environ Microbiol* 2001, **67**:3677-3682.