Q155: 16S rDNA microarray method reveals broader diversity than typical clone library when sampling the environment

Todd Z. DeSantis, Eoin L. Brodie, Jordan P. Moberg, Sonya R. Murray, Ingrid X. Zubieta, Yvette M. Piceno, Gary L. Andersen
Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley CA 94720

Abstract

Background: Molecular approaches aimed at broad prokaryotic environmental detection routinely rely upon classifying heterogeneous nucleic acids amplified by universal 16S rDNA PCR. The general method of sampling DNA types has been to clone and sequence the PCR products, however, this does not scale well for large studies. Instead, hybridizing PCR products to a universal 16S rDNA microarray allows a more rapid evaluation. This study investigated the breadth and accuracy of the microarray in detecting diverse 16S rDNA sequence types compared to clone-and-sequencing. **Methods:** DNA was extracted from three environmental samples; urban aerosol, subsurface soil, and subsurface water and amplified using universal 16S primers. The amplicons were classified using either the clone-and-sequence method (duration: 4 days) or by microarray hybridization (duration: 1.5 days for 4 replicates). **Results:** In general, each clone library produced over 400 high-quality sequences. Approximately 5% of the clones could not be placed into a known sub-family and were considered novel. The microarrays typically confirmed the majority of clone-detected sub-families but additionally reported greater amplicon diversity extending into phyla missed by the cloning method. As an example, the majority of sub-families documented by cloning aerosol amplicons were also detected in 4 of 4 replicate microarray hybridizations. Conversely, the array detected over twice the number of sub-families than did the clone library. The phyla Nitrospira and Spirochaetes were uniquely detected by the array and were verified with specific PCR primers and subsequent amplicon sequencing. **Conclusion:** Compared with sequencing a 16S rDNA clone library, the microarray was unable to recognize novel prokaryotic families but could identify greater diversity from organisms with similarity to existing sequence. Furthermore, the microarray allowed samples to be rapidly evaluated with replication.

Introduction

Molecular approaches aimed at broad prokaryotic environmental detection routinely rely upon classifying heterogeneous nucleic acids amplified by universal 16S rDNA PCR. The resulting mixed amplicons can be quickly, but coarsely, typed into anonymous groups using RFLP, SSCP or T/DGGE. Subsequent sequencing allows application of taxonomic nomenclature to the groups but requires additional labor to physically isolate each 16S rDNA type and does not scale well for large studies. Instead, hybridizing PCR products to a high-density universal 16S rDNA microarray allows rapid taxonomic classification of community members (Wilson, 2002; DeSantis, 2003; DeSantis, 2005).

In this study, we asked, "Which method is more comprehensive in cataloging an environmental prokaryotic 16S amplicon community, clone-and-sequencing or microarray?"

Materials and Methods

Microarray Design: Aligned sequences from 30,627 16S rDNA genes, (from E. coli positions 47 to 1473) were obtained from greengenes.lbl.gov (DeSantis, 2003). The sequences were clustered into 8,988 Operational Taxonomic Units (OTUs). For each OTU, a set of 11 or more specific 25-mers (targets) that were prevalent in members of a given OTU but were dissimilar from sequences outside the given OTU were sought. Probes presumed to cross-hybridize were those 25-mers that contained a central 17-mer matching sequences in more than one OTU (Urakawa, 2002). As each perfectly matching (PM) probe was chosen, it was paired with a control 25-mer, identical in all positions except the 13th base (mismatching probe, MM). The MM probe did not contain an internal 17-mer complimentary to sequences in any OTU. The 297,850 chosen oligonucleotides were synthesized upon an Affymetrix GeneChip.
Extraction: DNA was extracted from soil samples using a BIO-101 kit and from aerosols and water using a silica bead-beating method (Miller, 1999).
Amplification: Universal primers (Dojka et al., 1998) were used to generate >1400 bp products using a TaKaRa enzyme system. The amplicons were pooled as described below.
Cloning and Sequencing: Amplicons were cloned using a TOPO cloning Kit (Invitrogen, CA) and sequenced at the Joint Genome Institute, CA. Chimeras were removed using Bellerophon (Huber, 2004). Remaining sequences (Air: 422; **Water:** 242; Soil: 432) were placed into subfamilies using ARB's parsimony insertion function (Ludwig, 2004). Rarefaction analysis was performed by DOTUR (Schloss, 2005)
Hybridization: Amplicons were fragmented to 50-200 bp, labeled with biotin and hybridized in quadruplicate to the microarrays overnight at 48 °C.
Chip Processing: Microarrays were washed and scanned using standard Affymetrix equipment.
Probe Set Scoring: A probe pair (PM, MM) was scored as positive if the intensity from the PM probe was significantly greater than that from the MM (PM/MM>1.3 and PM-MM>130N). An OTU was considered present when >90% of its probe pairs were positive. Array results were summarized as a list of sub-families that were present in replicate arrays.
Primer Design: Subfamily-specific primer pairs were chosen using ARB's Probe Match function (Ludwig, 2004).

Samples:
- Urban aerosols were collected using six air samplers on four different days in San Antonio, Texas during the week of July 14th, 2003.
- Subsurface water was collected during bioremediation of a chromate contaminated aquifer in Hanford, WA.
- Subsurface soils were obtained from a Uranium contaminated area at Oak Ridge, TN.

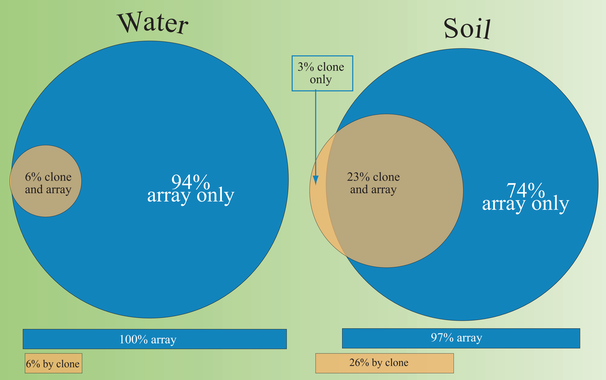
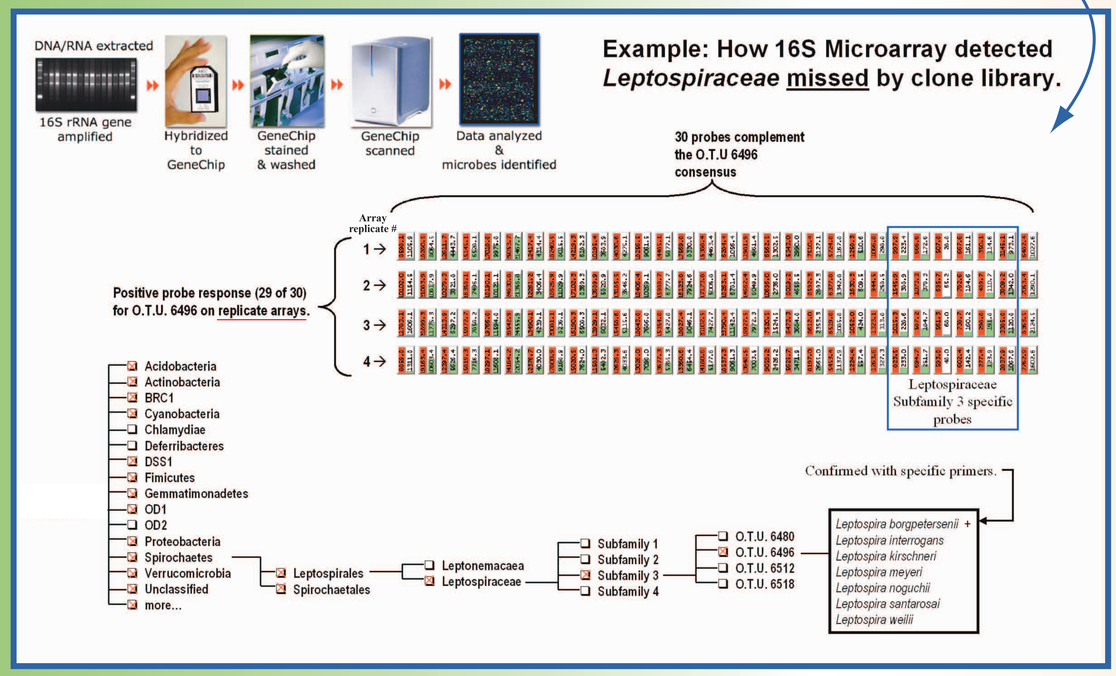
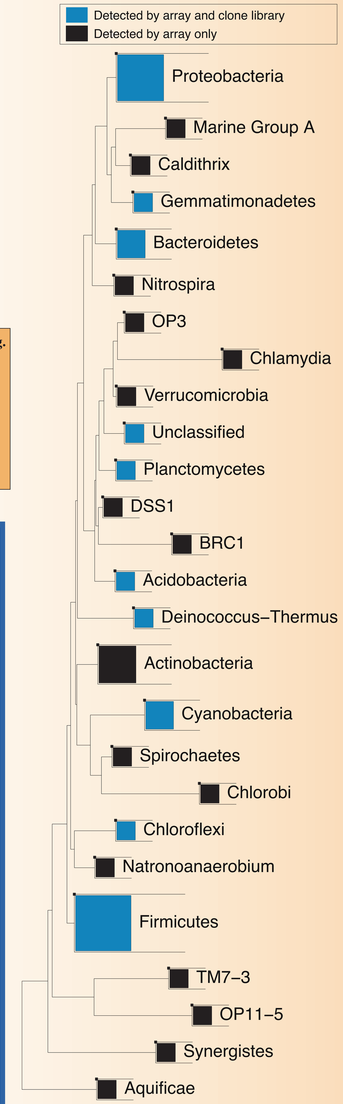


Table 1: Subfamilies confirmed by specific PCR and sequencing.

Actinobacteria	Actinosynnemataceae	sf_1
Nitrospira	Nitrospiraceae	sf_1
Clostridia	Syntrophomonadaceae	sf_5
Planctomycetes	Plantomycetaceae	sf_3
Gammaproteobacteria	Pseudoaltermonadaceae	sf_1
Acidobacteria	Ellin6075/11-25	sf_1
Spirochaetes	Spirochaetaceae	sf_1
Spirochaetes	Spirochaetaceae	sf_3
Spirochaetes	Leptospiracea	sf_3



Microarray detects phyla undetected by clone library



View complete taxonomy at http://greengenes.lbl.gov

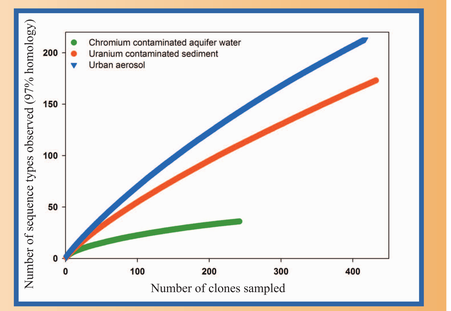
Discussion

The array allows rapid taxonomic classification of 16S rDNA amplicons. The clone-and-sequence method can be completed over 4 days whereas the hybridization (with replicates) requires 1.5 days.

In each sample, the array revealed a greater number of subfamilies than the corresponding clone library. This result was expected since non-asymptotic rarefaction curves demonstrated that the clone libraries were only a partial sample of the total sequence diversity. To validate the presence of the subfamilies that were detected only by the array, subfamily-specific primers were created for the aerosol sample. The resulting sequences verified the array detection of subfamilies listed in Table 1. Entire phyla including Nitrospira and Spirochaetes would have been overlooked if the clone library was the sole source of taxonomic sampling.

The clone libraries produced from the aerosol and soil sample exposed 18 subfamilies not detected by the array. Seven (one from aerosol, six from soil) of these subfamilies had no representatives in the database used to design the chip. In the remaining 11, the probe sets produced an inconsistent hybridization response among the replicate arrays.
The tree diagram demonstrated that the array is capable of detecting twice as many phyla as the typical clone library.

Rarefaction curve



Conclusion

- 16S rDNA array reveals broader diversity than clone library.
- 16S rDNA array processing is expeditious, facilitating sample replication.
- 16S rDNA clone and sequencing allows discovery of novel subfamilies.

References

DeSantis, T.Z., et al. (2003) Bioinformatics 19, 1461.
DeSantis, T.Z., et al. (2005) FEMS Microbiol Lett 245, 271.
Dojka, M.A., et al (1998) Appl Environ Microbiol 64, 3869.
Huber, T., et al (2004) Bioinformatics 20, 2317.
Ludwig, W., et al. (2004) Nucleic Acids Res. 32, 1363.
Miller, D.N., et al (1999) Appl Environ Microbiol 65, 4715.
Schloss, P.D. and Handelsman, J. (2005) Appl Environ Microbiol 71, 1501.
Wilson, K.H., et al (2002) Appl Environ Microbiol 68, 2535.

Acknowledgements

The authors are grateful to Dr. Philip Hugenholtz for ARB consulting. This work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC03-76SF00098 and was funded in part by the Department of Homeland Security under grant number HSSCHQ04X00037.

For application of array in bioremediation, see poster N240 (Eoin Brodie), and presentation by Gary Andersen