High-Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library When Sampling the Environment

Todd Z. DeSantis[1,2], Eoin L. Brodie[1], Jordan P. Moberg[1], Ingrid X. Zubieta[1], Yvette M. Piceno[1] and Gary L. Andersen[1,2]

(1) Lawrence Berkeley National Laboratory, Center for Environmental Biotechnology, 1 Cyclotron Road, Mail Stop 70A-3317, Berkeley, CA 94720, USA
(2) Lawrence Berkeley National Laboratory, Virtual Institute for Microbial Stress and Survival, Berkeley, CA 94720, USA

Abstract

Molecular approaches aimed at detection of a broad-range of prokaryotes in the environment routinely rely on classifying heterogeneous 16S rRNA genes amplified by polymerase chain reaction (PCR) using primers with broad specificity. The general method of sampling and categorizing DNA has been to clone then sequence the PCR products. However, the number of clones required to adequately catalog the majority of taxa in a sample is unwieldy. Alternatively, hybridizing target sequences to a universal 16S rRNA gene microarray may provide a more rapid and comprehensive view of prokaryotic community composition. This study investigated the breadth and accuracy of a microarray in detecting diverse 16S rRNA gene sequence types compared to clone-and-sequencing using three environmental samples: urban aerosol, subsurface soil, and subsurface water. PCR products generated from universal 16S rRNA gene-targeted primers were classified by using either the clone-and-sequence method or by hybridization to a novel high-density microarray of 297,851 probes complementary to 842 prokaryotic subfamilies. The three clone libraries comprised 1391 high-quality sequences. Approximately 8% of the clones could not be placed into a known subfamily and were considered novel. The microarray results confirmed the majority of clone-detected subfamilies and additionally demonstrated greater amplicon diversity extending into phyla not observed by the cloning method. Sequences matching operational taxonomic units within the phyla Nitrospira, Planctomycetes, and TM7, which were uniquely detected by the array, were verified with specific primers and subsequent amplicon sequencing. Subfamily richness detected by the array corresponded well with nonparametric richness predictions extrapolated from clone libraries except in the water community where clone-based richness predictions were greatly exceeded. It was concluded that although the microarray is unreliable in identifying novel prokaryotic taxa, it reveals greater diversity in environmental samples than sequencing a typically sized clone library. Furthermore, the microarray allowed samples to be rapidly evaluated with replication, a significant advantage in studies of microbial ecology.

Introduction

Molecular approaches designed to describe prokaryotic diversity routinely rely on classifying heterogeneous nucleic acids amplified via universal 16S rRNA gene polymerase chain reaction (PCR). The resulting mixed amplicons can be quickly, but coarsely, categorized using terminal-restriction fragment length polymorphism (t-RFLP), single-strand conformation polymorphism (SSCP), or temperature/denaturing gradient gel electrophoresis (T/DGGE) [39]. Association of taxonomic nomenclature to each group may be accomplished through sequencing, but this requires additional labor to physically isolate each gene and does not scale well for large comparative studies such as environmental monitoring. Species richness may be predicted from a few hundred sequences, but reproducible discovery of species composition may require $>10^4$ sequencing reactions per sample [14].

To increase the throughput of detection of microorganisms within complex samples, multiple DNA probes have been arrayed onto solid surfaces to allow for parallel, multispecies detection. Successful differen-

tiation of specific collections of bacteria has been achieved by using 16S rRNA gene microarrays containing tens to hundreds of probes for Enterococcus [37], Cyanobacteria [2], nitrifying bacteria [33], fish pathogens [58], and other bacterial groups. Alternatively, protein-encoding genes have been targeted to survey environments using microarrays [63, 64]. Recently, high-density 16S rRNA gene microarrays have emerged in efforts to detect any bacterial type without *a priori* knowledge of the community structure. Two major challenges have impeded this goal: Probes must be designed that are sensitive to only a specified branch of the prokaryotic tree, and hybridization scoring algorithms are required to interpret probe responses into reliable identifications. If a single unique probe for a taxon cannot be found, several probes can be utilized in combination with rule-based scoring. By vastly increasing the total number of probes within a microarray, more taxa can be queried and detection confidence can be improved. Using this approach, it was shown that organisms from environmental samples were accurately classified into their respective orders using an array with 62,358 probes [12, 61].

In this study, a novel microarray containing 297,851 probes targeted to 16S rRNA genes was tested by using amplicons derived from soil, water, and aerosols. The community profiles derived from the hybridizations were compared to the results from cloning-and-sequencing the same amplicons. A fraction of the clones (8%) were sufficiently divergent from database sequences to be considered novel and were not identified by the array. The microarray results confirmed the majority of clone-detected subfamilies, but additionally showed greater amplicon diversity. Importantly, the microarray detected phyla that would have otherwise been overlooked if relying solely on the clone library. Three of these phyla have been confirmed with specific PCR amplification. The results illustrate the consequence of relying only on clone libraries or high-density 16S rRNA gene micro-arrays when profiling a microbial community.

Methods

Microarray Design. The microarray probe design approach previously described for differentiating Staphylococcaceae [9] was applied to all known 16S rRNA gene sequences containing at least 600 nucleotides. Briefly, sequences (*Escherichia coli* bp positions 47 to 1473) were extracted from a multiple sequence alignment composed of more than 30,000 records within the 15 March 2002 release of the 16S rRNA gene database, greengenes.lbl.gov [11]. This region was selected because it is bounded on both ends by universally conserved segments that can be used as PCR priming sites to amplify bacterial or archaeal [13] genomic material using only two to four primers. Putative chimeric sequences were filtered from the data set by using the software package Bellerophon [25], preventing them from being misconstrued as novel organisms [28]. Filtered sequences were clustered to enable each sequence of a cluster to be complementary to a set of perfectly matching (PM) probes. Putative amplicons were placed in the same cluster as a result of common 17-mers found in the sequence. The resulting 8935 clusters, each containing approximately 3% sequence divergence, were considered operational taxonomic units (OTUs) representing all 121 demarcated prokaryotic orders. The taxonomic family of each OTU was assigned according to the placement of its member organisms in Bergey's Taxonomic Outline [22]. The taxonomic outline as maintained by Hugenholtz [26] was consulted for phylogenetic classes containing uncultured environmental organisms or unclassified families belonging to named higher taxa. The OTUs comprising each family were clustered into subfamilies by transitive sequence identity according to a previously described method [9]. Altogether, 842 subfamilies were found. The taxonomic position of each OTU as well as the accompanying NCBI accession numbers of the sequences composing each OTU can be viewed at http://greengenes.lbl.gov/Download/Clones_v_Array/.

The objective of the probe selection strategy was to obtain an effective set of probes capable of correctly categorizing mixed amplicons into their proper OTU. For each OTU, a set of 11 or more specific 25-mers (probes) was sought to be prevalent in members of a given OTU but dissimilar from sequences outside the given OTU. The average number of probes chosen for each OTU was 24. In the first step of probe selection for a particular OTU, each of the sequences in the OTU was separated into overlapping 25-mers, the potential targets. Then each potential target was matched to as many sequences of the OTU as possible. It was not adequate to use a text pattern search to match potential targets and sequences because partial gene sequences were included in the reference set. Therefore, the multiple sequence alignment provided by Greengenes was necessary to provide a discrete measurement of group size at each potential probe site. For example, if an OTU containing seven sequences possessed a probe site where one member was missing data, then the site-specific OTU size was only six. In ranking the possible targets, those having data for all members of that OTU were preferred over those found only in a fraction of the OTU members. In the second step, a subset of the prevalent targets was selected and reverse-complimented into probe orientation, avoiding those capable of mishybridization to an unintended amplicon. Probes presumed to have the capacity to mishybridize were those 25-mers that contained a central 17-mer matching sequences in more than one OTU [56]. Thus, probes that were unique to an

OTU solely due to a distinctive base in one of the outer four bases were avoided. Also, probes with mishybridization potential to sequences having a common tree node near the root were favored over those with a common node near the terminal branch. Probes complementary to target sequences that were selected for fabrication are termed PM probes. As each PM probe was chosen, it was paired with a control 25-mer [mismatching probe (MM)], identical in all positions except the thirteenth base. The MM probe did not contain a central 17-mer complimentary to sequences in any OTU. The probe complementing the target (PM) and MM probes constitute a probe pair analyzed together. Sets of probes for each OTU can be viewed at: http://greengenes.lbl.gov/cgi-bin/nph-show_probes_2_otu_alignments.cgi.

The chosen oligonucleotides were synthesized by a photolithographic method at Affymetrix Inc. (Santa Clara, CA, USA) directly onto a 1.28×1.28 cm glass surface at an approximate density of 10,000 molecules/μm^2 [6]. The entire array of 506,944 probe features was arranged as a grid of 712 rows and columns. Thus, each unique probe sequence (feature) on the array occupied a square with an 18-μm side and had a copy number of roughly 3.2×10^6. Of these features, 297,851 were oligonucleotide PM or MM probes with exact or inexact complementarity, respectively, to 16S rRNA genes. The remaining were used for image orientation, normalization controls, or for pathogen-specific signature amplicon detection using additional targeted regions of the chromosome [62]. Each high-density 16S rRNA gene microarray was designed with additional probes that: (1) target amplicons of prokaryotic metabolic genes spiked into the 16S rRNA gene amplicon mix in defined quantities just before fragmentation and (2) are complimentary to prelabeled oligonucleotides added into the hybridization mix. The first control collectively tests the fragmentation, biotinylation, hybridization, staining, and scanning efficiency. It also allows the overall fluorescent intensity to be normalized across all the arrays in an experiment. The second control directly assays the hybridization, staining, and scanning.

Environmental Sampling and DNA Extraction. Air samples were collected at a flow rate of approximately 10 L/min onto 1.0 μm polyethylene terephthalate (Celanex) filters (Hoechst-Celanese, Dallas, TX) over a 24-h period simultaneously from six locations in San Antonio, TX, USA. Sample filters were washed in 10 mL buffer (0.1 M sodium phosphate, 10 mM EDTA, pH 7.4, 0.01% Tween 20), and the suspension was stored frozen until needed. One 0.6-mL aliquot of wash was taken from each thawed filter wash and combined in a "day pool". DNA was extracted by using a modification of a soil technique [46]. After centrifugation of the day pool at $16,000 \times g$ for 25 min, the pellets were resuspended in

400 μL sodium phosphate buffer (100 mM NaH_2PO_4, pH 8) and transferred into two 2-mL silica bead lysis tubes containing 0.9 g of zirconia/silica lysis bead mix (0.3 g of 0.5 mm and 0.6 g of 0.1 mm). For each lysis tube, 300 μL buffered SDS [100 mM NaCl, 500 mM Tris pH 8, 10% (wt/vol) SDS] and 300 μL phenol/chloroform/isoamyl alcohol (25:24:1) were added. Lysis tubes were inverted and finger-flicked three times to mix the buffers before bead mill homogenization with a Bio101 Fast Prep 120 machine (Qbiogene, Irvine, CA, USA) at 6.5 m/s for 45 s. The bead-beating duration was selected for its ability to release DNA from spores while not overfragmenting genomes [12]. Following lysate centrifugation at $16,000 \times g$ for 5 min, the aqueous supernatant was removed to a new 2-mL tube and maintained at –20°C for 1 h to overnight. An equal volume of chloroform was added to the thawed supernatant prior to vortexing for 5 s and centrifugation at $16,000 \times g$ for 3 min. The supernatant was then combined with two volumes of binding Solution 3 (MoBio, Carlsbad, CA, USA). Genomic DNA (gDNA) from the mixture was isolated on a MoBio spin column, washed with Solution 4, and eluted in 60 μL of $1\times$ TE according to the manufacturer's instructions. The gDNA was further purified by passage through a Sephacryl S-200 HR spin column (Amersham, Piscataway, NJ, USA) and stored at 4°C. Each of the gDNA preparations from four different "day pools" from the week of July 14, 2003 was independently PCR-amplified. PCR products were combined to constitute the sample for the week.

Subsurface water was collected during polylactate-stimulated bioremediation of a chromate-contaminated aquifer at the Hanford 100H site, WA (http://www-esd.lbl.gov/ERT/hanford100h/). Water, approximately 150 mL, was filtered through sterile 0.22-μm anodisc filters (Whatman, Florham Park, NJ, USA) and DNA was extracted by using a modification of the procedure described for air samples. Anodisc filters were manually fragmented in a sterile whirlpak bag, and 1 mL of phosphate buffer was added. Filter fragments in buffer were transferred to a bead lysis tube. Tubes were centrifuged at $16,000 \times g$ for 5 min and 700 μL of buffer was removed. Next, 300 μL of buffered SDS solution and 300 μL of phenol/chloroform/isoamyl alcohol (25:24:1) were added and bead beating was performed at 5.5 m/s for 30 s. After centrifugation, the aqueous phase was mixed with an equal volume of chloroform in a phase-lock gel tube (Eppendorf, Westbury, NY, USA) and further extracted. The top phase containing nucleic acids was purified as for the air samples without the need for additional Sephacryl purification. DNA was eluted in 50 μL sterile water and stored at –20°C until needed.

Subsurface soils were obtained from a uranium-contaminated soil (area 2) at the NABIR Field Research

Center at Oak Ridge, TN, USA (http://www.esd.ornl.gov/nabirfrc/). More information about the soil characteristics is available at http://public.ornl.gov/nabirfrc/other/FRCSummary.pdf DNA was extracted from triplicate 500-mg (wet weight) subsamples of soil using a BIO101 soil DNA extraction kit (Qbiogene) according to the manufacturer's protocol.

16S rRNA Gene Amplification. The 16S rRNA gene was amplified from the DNA extracts using universal primers 27f.1 (5′-AGRGTTTGATCMTGGCTCAG) and 1492R (5′-GGTTACCTTGTTACGACTT). PCR for air and soil samples was carried out by using the TaKaRa *Ex Taq* system (Takara Bio Inc., Japan) as follows, with at least three replicate PCR reactions performed per sample and pooled before analysis. Each PCR reaction mix contained 1× buffer, 0.8 mM *TaKaRa* dNTP mixture, 0.02 U/μL *Ex Taq* polymerase, 0.4 mg/mL bovine serum albumin (BSA), and 1.0 μM of each primer. PCR conditions were 1 cycle of 3 min at 95°C, followed by 35 cycles of 95°C (30 s), 53°C (30 s), 72°C (60 s), and a final extension at 72°C for 7 min. DNA extracts from water samples were amplified by using a slightly different protocol using a range of eight different annealing temperatures between 48°C and 58°C. Only 30 cycles were performed for amplification from water samples and amplicons from the eight different annealing temperatures were combined.

Cloning-and-Sequencing. Amplicon pools from the three environments were subjected to cloning as follows: Amplicons were ligated and cloned by using the TOPO-TA pCR2.1 kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Individual clones containing organism-specific 16S rRNA gene fragments were purified by using magnetic beads [54], and sequenced from each terminus using an ABI3700 (Applied Biosystems, Foster City, CA), assembled using Phred and Phrap [16, 17], and were required to pass quality tests of Phred 20 (base call error probability $< 10^{-2.0}$) to be included in the analysis. Sequencing was performed at the DOE Joint Genome Institute (JGI; http://www.jgi.doe.gov/). Putative chimeric sequences were obtained by using Bellerophon [25]. Sequences were aligned to the Greengenes 7682-character format by using the NAST [10] web server (http://greengenes.lbl.gov/NAST). Similarity to public database records was calculated with DNADIST [19], by using the DNAML-F84 option assuming a transition/transversion ratio of 2.0, and an A, C, G, and T 16S rRNA gene base frequency of 0.2537, 0.2317, 0.3167, and 0.1979, respectively. This was empirically calculated from all records of the Greengenes 16S rRNA gene multiple sequence alignment over 1250 nucleotides in length. The lane mask [35] was used to restrict similarity observations to 1287 conserved columns

(lanes) of aligned characters. Cloned sequences from this study were rejected from further analysis when less than 1000 characters could be compared to a lane-masked reference sequence. Sequences were assigned to a taxonomic node by using a sliding scale of similarity thresholds [52]. Phylum, class, order, family, subfamily, or OTU placement was accepted when a clone surpassed similarity thresholds of 80%, 85%, 90%, 92%, 94%, or 97%, respectively. When similarity to nearest database sequence was below 94%, the clone was considered to represent a novel subfamily and a novel class was denoted when similarity was less than 85%.

Accumulation curves, diversity estimates (Shannon–Weaver index [43]), and nonparametric richness estimations (Chao1 and ACE [4, 5]) were calculated by using the software DOTUR [51] with the clone distance matrix as input and a nearest-neighbor clustering algorithm. Dominance in clone libraries was calculated as $1 - $ Shannon evenness index ($1 - E$), where evenness (E) is represented as follows: $E = H/\ln S$ ($H = $ Shannon–Weaver diversity index; $S = $ total richness in a sample).

Accession Numbers. Sequences generated in this study have been deposited in Genbank as accession numbers DQ125500–DQ125935 (soil), DQ129237–DQ129656 (air), and DQ264398–DQ264650 (water). Fasta formatted records can also be obtained at http://greengenes.lbl.gov/Download/Clones_v_Array/.

Microarray Processing. Identical amplicon pools used for cloning were also used for array analysis. For air samples, 2 μg amplicons was concentrated to a volume less than 40 μL with a Microcon YM100 spin filter (Millipore, Billerica, MA, USA). For soil samples and water samples, 2 μg ($\sim 10^{12}$ gene copies) and 500 ng ($\sim 3 \times 10^{11}$ gene copies) of amplicons, respectively, were concentrated using a PCR clean up kit (MoBio). The PCR products were spiked with known concentrations of amplicons derived from prokaryotic metabolic genes. This mix was fragmented to 50–200 bp using DNase I (0.02 U/μg DNA; Invitrogen) and One-Phor-All buffer per the Affymetrix protocol. The complete mixture was incubated at 25°C for 10 min, 98°C for 10 min, and then labeled. Biotin labeling was accomplished using an Enzo Bioarray Terminal Labeling Kit (Affymetrix) as per the manufacturer's instructions. Next, labeled DNA was denatured (99°C for 5 min) and hybridized to the DNA microarray at 48°C overnight (>16 h) at 60 rpm. The arrays were subsequently washed and stained. Reagents, conditions, and equipment are detailed elsewhere [44].

Scanning and Probe Set Scoring. Arrays were scanned using a GeneArray Scanner (Affymetrix). The scan was captured as a pixel image using standard

Affymetrix software (GeneChip Microarray Analysis Suite, version 5.1) that reduced the data to an individual signal value for each probe. Background probes were identified as those producing intensities in the lowest 2% of all intensities. The average intensity of the background probes was subtracted from the fluorescence intensity of all probes. The noise value (N) was the variation in pixel intensity signals observed by the scanner as it read the array surface. The standard deviation of the pixel intensities within each of the identified background cells was divided by the square root of the number of pixels comprising that cell. The average of the resulting quotients was used for N in the calculations described below.

Probe pairs scored as positive were those that met two criteria: (1) the fluorescence intensity from the perfectly matched probe (PM) was at least 1.3 times greater than the intensity from the mismatched control (MM), and (2) the difference in intensity, PM minus MM, was at least 130 times greater than the squared noise value ($>130N^2$). The positive fraction (PosFrac) was calculated for each probe set as the number of positive probe pairs divided by the total number of probe pairs in a probe set. An OTU was considered "present" when its PosFrac was greater than 0.92 in all three replicates. Present calls were propagated upwards through the taxonomic hierarchy by considering any node (subfamily, family, order, etc.) as "present" if at least one of its subordinate OTUs was present.

Validation of Array-Detected OTUs Not Detected by the Clone Library. PCR primers targeting specific OTUs within the Nitrospiraceae and Planctomycetaceae (Table 3) were generated by ARB's probe design feature [41] and Primer3 [50]. Melting temperatures were constrained from 45°C to 65°C and G + C content between 40% and 70% was preferred. The primers were chosen to contain 3′ bases non-complimentary to sequences outside of the sub-family. TM7 phylum-specific primers [29] were obtained from the literature. DNA sequences were generated by PCR as described above with the necessary adjustments in annealing temperatures. Amplicons were purified (PureLink PCR Purification Kit, Invitrogen), sequenced, and matched to an OTU in the same manner as described in the cloning-and-sequencing method except the minimum count of base comparisons was not used to exclude data.

Results

Cloning. Sequences from the clone libraries of the three environments were assembled into contigs from two sequencing reactions initiated at the 5′ and 3′ termini of the 16S rRNA gene. Initially, 1391 contigs with low base call error probability were accepted for analysis. After removing contigs that did not contain sufficient data to allow at least 1000 characters to be compared to a lane-masked reference sequence and filtering the sequences for putative chimeras, 1155 remained. The vetted libraries from air, soil, and water contained 417, 485, and 253 clones, respectively. Figure 1 shows the class level (85% database sequence homology) distributions of clones within each of the three ecosystem samples; not shown are 1, 8, and 0 clones (air, soil, water) considered novel at the class level compared to existing database entries. For air, soil, and water clone libraries, 9.8%, 14.4%, and 1.6% of clones, respectively, were outside the subfamily level (94%) assignment threshold and were considered novel sequences.

Clone library analysis indicated that Firmicutes dominated the air sample mostly within the class Bacilli, whereas Actinobacteria dominated the soil sample. The water sample consisted solely of three classes: flavobacteria (Bacteroidetes), β-proteobacteria, and γ-proteobacteria, each with similar distribution. Figure 2 shows the accumulation curves for the three samples where the cumulative number of subfamilies observed is plotted against the sampling effort. All communities were incompletely sampled as evidenced by the nonasymptotic curves [30]; however, the water community appeared relatively well sampled compared to air and soil.

Comparison of Cloning with Microarray Analysis. A comparison between clone library and microarray assessment of community composition is shown for each taxonomic level in Table 1. The breadth of 16S rRNA gene sequence types was expressed as a count of distinct groups detected in each environment, by each method, at six levels of taxonomic resolution. It is clear that even at the phylum level cloning underestimated richness compared with the microarray. Among all three environments, the amplicons categorized by cloning were in general concordance with a subset of taxonomic categories reported from the arrays. This trend continued as the resolution of the comparisons increased from phylum to subfamily. Hybridization of aerosol amplicons produced "present" calls in 238 subfamilies, 178 of which were not found in the corresponding clone library. Subfamily richness appeared to be greater in the soil sample with 279 subfamilies detected; of these, 239 were not encountered in the clone library. The water sample was shown to have the lowest richness by both methods, but of the 99 subfamilies detected by microarray, only 6 were detected by cloning. For all sample types, very few subfamilies deemed present by the clone libraries were not reflected in the microarray hybridization results.

A synopsis of the phylum-level community composition is assembled in Table 2, allowing comparison of the 34 phyla reported in at least one of the environments

by at least one of the methods. By cloning, 10 phyla were detected in the aerosol and soil samples, whereas only 2 phyla were detected in the water sample. In contrast, the array method detected all of the phyla detected by cloning and an additional 17, 23, and 17 phyla in the air, soil, and water samples, respectively.

PCR and Sequencing Confirmation of Additional Phyla Detected by Microarray. To determine whether the additional phyla detected by microarray were true positives, and not the product of unforeseen cross-hybridization, three OTUs from diverse phyla, detected in aerosols by microarray only, were chosen for further investigation (Table 3). OTU 864 (OTU numbers correspond to 16S rRNA microarray probe sets) within the phylum Nitrospira comprised sequences discovered in sludge, soil, reservoirs, and in an aquifer. All 13 probe pairs in the probe set for OTU 864 were positive in 3 of 3 arrays. The sample was interrogated with primers designed from known sequences in the OTU and the resulting amplicons were sequenced, revealing similarity to OTU 864. Similarly, all 11 probe pairs of a Planctomycetes OTU (OTU 4948) were consistently positive despite this OTU being unrepresented among the 417 aerosol clones. Primers were designed from the five 16S rRNA gene sequences generated from a municipal wastewater plant [7] that defined OTU 4948. Taxon-specific PCR and sequencing confirmed that a sequence matching this OTU was present in the sample. Evidence for the presence of phylum TM7 came from the probe set complementary to OTU 8155. General TM7 phylum-specific primers [29] produced sequences attributed to a related TM7 OTU identified as 3664.

Diversity Estimates. Table 4 lists diversity estimates and richness predictions based on the clones sampled and also compares predicted richness values to those observed by both cloning and array methods.

Shannon–Weaver diversity estimates for the clone libraries indicate that sample diversity is of the order air > soil >>> water. Both Chao1 and ACE nonparametric richness estimators predicted that the subfamily level richness of the air and soil samples is far greater than that observed through clone sampling and these estimates were in strong agreement with the subfamily counts reported by the array analysis. The water sample produced a large discrepancy in subfamily richness between the cloning and array methods regardless of whether direct clone observations or nonparametric predictors were used for the comparison. In Table 4, a trend was observed—the greater the dominance encountered in the clone library, the greater the difference between the cloning and array observed counts of subfamilies.

Discussion

The postulated complexity of each microbial community that has been isolated from the environment, combined with the number of potentially unique ecosystems, has hindered efforts to sufficiently catalog microbial biodiversity. There are estimates of thousands [14, 55] to millions [21] of unique bacterial genomes present in a gram of soil. Outdoor aerosols may be equally complex, composed of organisms released from multiple habitats, both locally and over long distances. Furthermore, microbial communities can change temporally as environmental conditions vary. The need for high-throughput accurate biological monitoring is clear.

Community fingerprinting has provided methods for rapidly profiling microbial communities with replication. These approaches, based on heterogeneity in amplicon length, endonuclease cleavage sites, melting profiles, or single strand secondary structure, have allowed high-throughput inference of species richness and evenness [39]. However, fingerprinting methods are
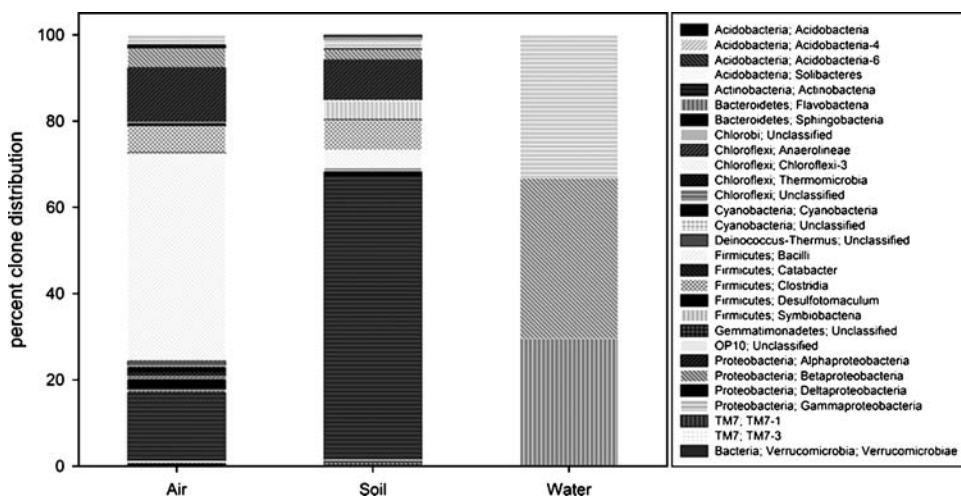
Figure 1. Class-level distribution of clones within libraries from air, soil, and water samples.
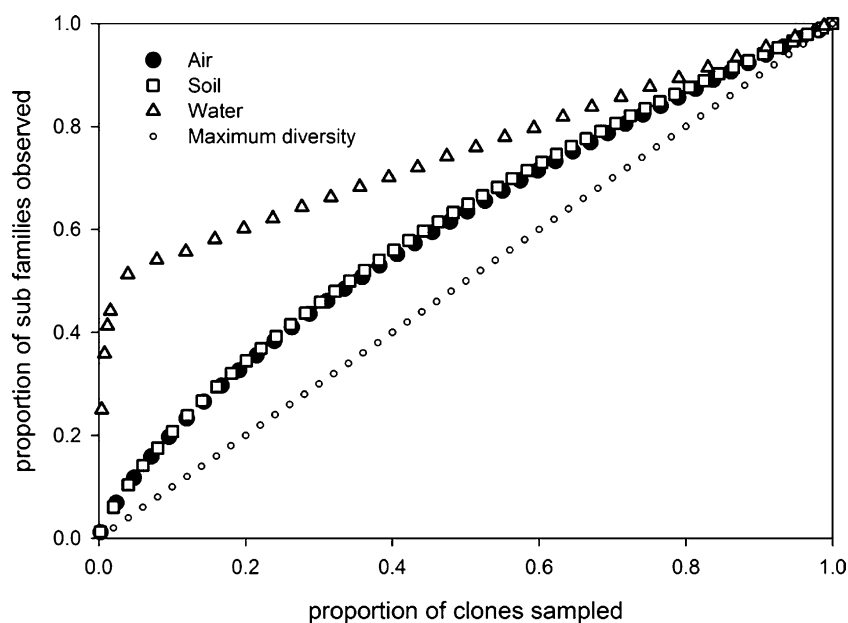
Figure 2. Accumulation curves for air ($\bullet$; $n = 417$), soil ($\square$; n = 485) and water ($\triangle$; $n = 253$) bacterial communities analyzed by clone library. The maximum diversity scenario where every sample is a new observation ($\circ$). Sampling efficiency at subfamily level is presented: subfamily is defined as 94% sequence homology. Data for curves represent an average of 1000 simulations performed using the software DOTUR.

generally deficient in providing taxonomic microbial identity and typically yield less than 100 clearly defined bands, peaks, or products for analysis. T-RFLP offers the greatest taxonomic resolution of the rapid fingerprinting methods, potentially capable of class identification [34] when DNA sequences in the sample are nearly identical to database reference sequences, or when over 10 restriction enzymes are used in parallel for each sample [49]. However, the reliability of T-RFLP for taxonomic assignment is unclear, because the presence of classes reported by T-RFLP—but not found in a corresponding clone library—have been left unverified.

PCR has made it possible to easily obtain composite samples of mixed rRNA genes from natural environments [53, 60]. Although amplification biases have been demonstrated in defined communities due to primer selection, number of cycles, and template concentration [18, 48, 59], this technique has been valuable in increasing our understanding of the complexity of individual communities [27]. Regardless of the method used to limit biases from environmental samples, most of our knowledge on microbial composition of specific communities comes from isolating individual, amplified 16S rRNA genes for cloning and sequencing. The sequences are compared to references in large databases, allowing either specific phylogenetic classification or proposal of novel taxa when a clone is sufficiently divergent from known groups. The limitation becomes the number of clones or PCR products requiring sequencing and analysis. It has been suggested that environmental samples may require over 40,000 sequencing reactions to document 50% of the richness [14]. This approach is laborious, costly, and time-consuming, often taking weeks to complete the analysis of one clone library. Thus, performing studies with sample replication becomes

rapidly overwhelming. Despite the effort required, clone libraries are often the chosen method, and the current "gold standard" for obtaining the greatest estimate of diversity. Typical libraries of cloned 16S rRNA gene fragments include fewer than 1000 sequences [15, 32, 45], well below the suggested quantity.

By making it possible to conduct sequence analyses on the complete pool of 16S rRNA gene fragments at once, high-density photolithography microarrays have the ability to provide microbial identification within complex environmental samples in a high-throughput manner. Hybridization (with replicates) requires only

Table 1. Count of distinct taxonomic groups detected by microarrays and/or clone libraries from three sample matrices

Matrix	Taxonomic level	Array only	Array and clone[a]	Clone only[a]
Air	Phylum	17	10	0
	Class	34	21	1
	Order	82	35	2
	Family	157	62	5
	Subfamily	178	60	6
	OTU	961	54	76
Soil	Phylum	23	10	0
	Class	49	19	1
	Order	105	29	1
	Family	194	46	3
	Subfamily	239	40	3
	OTU	1237	37	28
Water	Phylum	17	2	0
	Class	31	3	0
	Order	60	4	0
	Family	86	6	0
	Subfamily	93	6	0
	OTU	250	17	5

[a] The DNAML-F84 homology cutoff used for clone assignment to each level of taxonomic resolution is as follows: Phylum 80%, Class 85%, Order 90%, Family 92%, Subfamily 94%, OTU 97%.

Table 2. Phyla detected in different sample matrices by cloning and sequencing or by high-density DNA microarray analysis

	Air		Soil		Water	
Phylum	Clone	Array	Clone	Array	Clone	Array
Acidobacteria	×	×	×	×		×
Actinobacteria	×	×	×	×		×
AD3		×		×		
Aquificae		×		×		
Bacteroidetes	×	×	×	×	×	×
BRC1		×		×		×
Caldithrix		×		×		
Chlamydiae		×		×		
Chlorobi	×	×		×		
Chloroflexi	×	×	×	×		×
Coprothermobacteria		×		×		
Cyanobacteria	×	×		×		×
Deinococcus–Thermus	×	×		×		
DSS1		×		×		
Firmicutes	×	×	×	×		×
Gemmatimonadetes	×	×		×		
Lentisphaerae				×		×
marine group A		×		×		×
Natronoanaerobium		×		×		
NC10				×		×
Nitrospira		×[a]		×		×
OD1		×				
OP10			×	×		×
OP8		×		×		
OP9/JS1				×		×
Planctomycetes		×[a]		×		×
Proteobacteria	×	×	×	×	×	×
SPAM				×		
Spirochaetes		×		×		
Termite group 1				×		
TM7		×[a]	×	×		×
Verrucomicrobia		×	×	×		×
WS3		×	×	×		×
WS5				×		×

[a]Phylum confirmed by taxa-specific PCR.

Table 3. Phyla detected only by DNA microarray analysis and subsequently verified by PCR

Phyla verified	Confirming amplicon GenBank acc. #	Closest matching database sequence (GenBank acc. #)	Closest matching array OTU	Primer sequences (5′ to 3′)	PCR annealing temperature (°C)
Nitrospira	DQ129656	Sludge clone GC86 (Y14644)	*Nitrospiraceae*; OTU 864[a]	For-TCGAAAAG CGTGGGG Rev-CTTCCTC CCCCGTTC	47.0
Planctomycetes	DQ129666	anoxic basin clone CY0ARA027D12 (BX294763)	*Planctomycetaceae*; OTU 4948[a]	For-GAAACTGC CCAGACAC Rev-AGTAACG TTCGCACAG	60.0
TM7	DQ236249	Sludge clone SBR1071 (AF268996)	TM7-1; OTU 3664[b]	For-AYTGGGCGT AAAGAGTTGC Rev-TACGGYTA CCTTGTTACGACTT	66.3

[a]Closest match of confirming amplicon is contained within array-detected OTU.
[b]Closest match of confirming amplicon is contained within alternate TM7 OTU.
Amplicons confirming the phyla detected by microarray were sequenced and identified by comparison to database sequences.

Table 4. Clone library based estimates of diversity and predicted richness compared with observed richness determined by cloning and array approaches

Matrix	Estimated diversity		Predicted richness[a]		Observed richness		
	Shannon[b]	Dominance[c]	Chao1	ACE	Clone	Array	Difference[d]
Air	2.74	0.38	155–396	165–354	66	238	3.61
Soil	2.18	0.46	119–276	123–250	43	279	6.48
Water	0.65	0.64	4–17	2–17	6	99	16.5

[a]95% confidence intervals are shown both Chao 1 and ACE richness estimators.
[b]Shannon–Weaver index of diversity.
[c]Dominance is expressed as 1 − Shannon's equitability (evenness) index.
[d]Fold difference in richness observation between clone library and array approaches.

1.5 days compared to the clone-and-sequence method, which necessitates 3 days or longer to sequence and analyze a typical library of several hundred clones. It has previously been demonstrated that organisms from complex environmental samples can be accurately classified into their respective orders by using an array with 62,358 probes [12, 61]. The present study investigated the response of a novel microarray containing 297,851 probes when hybridized to 16S rRNA gene amplicons generated from aerosols, soil, and water. The microarray design approach was based on the anticipation that the tool would be used to characterize samples without prior knowledge of their microbial composition. For this reason, more than 30,000 diverse 16S rRNA gene sequences were clustered into 8935 OTUs consisting of 842 subfamilies. Every OTU was interrogated by 24 probes, on average, each adjacent to a control probe used to subtract the effects of nonspecific hybridization. The requirement of a sequence-specific interaction from multiple unique probes to identify the presence of each OTU was implemented to increase the confidence of detection over single probe per OTU methods.

This high-density microarray targets the most unique portions of the 16S rRNA gene for a given cluster, and the results are summarized at higher phylogenetic levels. This is in contrast to scoring probes at each node in a hierarchical tree [40]. Our approach accommodates OTUs that may be divergent from other members of the same encapsulating node (which is often the case with environmental sequences) by not requiring that a single probe solution must be found for the entire node. Although the microbial census from every environment is far from complete [3], the key questions for suitability of this approach are: "Is the sequence variation from all extant prokaryotes unlimited, encompassing every possible nucleotide variation within the 16S rRNA gene?", or conversely, "Can a majority of the organisms be classified on the basis of similarity to identified sequences in the databases?" Unlike estimates of microbial genomic variability [21], sequence variability of the 16S rRNA gene appears more constrained [52], most likely because of the functional necessity of the ribosome. Thus, to some degree, sequences not yet in the

database may share some homology with targets used for probe analysis. This basic assumption of probe design, which has enabled the identification of one or a few OTUs using Southern analysis [20, 36], fluorescence in situ hybridization (FISH) [8, 23], and quantitative PCR (QPCR) [1, 24], was extended to design probes for a substantially greater number of OTUs. To allow for the detection of environmental sequences slightly divergent from those represented on our array, we do not require a sequence specific interaction of 100% of probe pairs defining an OTU.

It was hypothesized that a phylogenetic profile calculated from an array analysis should reflect the composition of sequence types obtained by cloning the same amplicon pool. Specifically, we tested the effectiveness of the novel microarray in detecting and categorizing environmental 16S rRNA genes into taxa with defined nomenclature. Three environments were selected for bacterial community evaluation by means of DNA extraction and universal 16S rRNA gene PCR amplification. Products were split for sampling by cloning-and-sequencing or microarray hybridization.

The three clone libraries produced 253 – 485 sequences each and varied in composition relative to each other. The soil and air were dominated by Actinobacteria and Bacilli, respectively (Fig. 1), and contained over 40 subfamilies each. The water library possessed considerably less richness with only six subfamilies. As predicted, accumulation curves demonstrated that hundreds of clones were insufficient to catalog all subfamilies putatively present, but that the water appeared more thoroughly sampled than the others. The divergent characteristics of the three clone libraries were considered beneficial for testing the array against dissimilar 16S rRNA gene amplicon community structures.

After each amplicon pool was hybridized to replicate arrays, probe responses were matched to OTUs in the database. Detection of an OTU required more than 92% of the probe pairs assigned to the particular probe set to hybridize such that the PM probe had a greater intensity than the MM probe partner. This threshold was chosen to allow sequences with minor divergence from database entries, from which the array was designed, to be

detectable by the array. The OTUs found by the array and/or the cloning method were summarized to the subfamily, family, order, class, and, lastly, phylum to discern the resolution at which the results deviate. Regardless of the resolution considered, the array consistently revealed greater richness than the corresponding clone library (Table 1). This result was expected because nonasymptotic accumulation curves demonstrated that the clone libraries were only a partial sample of the total sequence diversity. The array predicted the presence of every phylum represented in the clone library. The same concurrence held for nearly all classes and orders in all three environmental samples. The atmospheric bacterial samples offered the most relevant example of a previously uncharacterized environment, because very few aerosol-derived 16S rRNA gene sequences are publicly available. In this relatively unstudied environment, OTU-level matching of cloned sequences and array positives was in poor agreement. Yet, the two methods concurred at the subfamily ranking with some exceptions. For example, in air samples, subfamilies within the Myxococcaceae (δ-proteobacteria) and Williamsiaceae (Actinobacteria) were overlooked by the array. Similarly, in the soil, cloned subfamilies within the Opitutaceae (Verrucomicrobia) and Propionibacteriaceae (Actinobacteria) were not found by the arrays using a PosFrac threshold of 0.92. Although an explanation for the reduced PosFrac within Myxococcaceae and Opitutaceae could not be attributed to mismatches between the clone and probe sequences, it was clear that divergence at the loci targeted by the probes would prohibit a sequence-specific response for the Williamsiaceae and Propionibacteriaceae.

The greater number of phyla reported by the array, but not represented in the clone libraries, was unanticipated (Table 2). There are two main factors which, individually or combined, may help explain this anomaly: (1) either the array approach overestimates richness as a result of nonspecific hybridization leading to false-positives or (2) cloning does not truly represent sequence distribution because of insufficient sampling or perhaps cloning bias. Three phyla detected only by array analysis (Nitrospira, Planctomycetes, and TM7) in air samples were chosen for further investigation. Amplification using specific PCR primers and sequencing of amplicons confirmed the presence of the phyla and in two cases the exact OTU detected by the array was also confirmed. This demonstrated not only that the microarrays revealed broader diversity than a typical clone library, but also that the additional components could be identified and subsequently verified with a confirmatory third method. It is significant that entire phyla would have been overlooked if the clone library were the sole source of taxonomic sampling.

It was impractical to determine if sequencing to extinction (asymptotic accumulation/rarefaction curves)

would have revealed the additional phyla, because it has been estimated that $>10^4$ sequences may only be sufficient to encounter half of an environment's microbial richness [14]. However, the clone libraries presented in this work reflect the method as it is typically practiced rather than how it would be statistically complete. Nonetheless, it is possible to predict richness within microbial communities by using rarefaction and statistical estimators [4, 5, 51]. As expected for environmental bacterial community sampling efforts, accumulation curves and nonparametric richness estimators demonstrated that no community in this study was sampled to completion. Importantly, the predicted richness extrapolated from cloning observations from the air and soil was quite similar to that enumerated by array analysis. The richness detected by the array for the water sample considerably exceeded the predicted richness due to high dominance.

In a previous study on environmental amplicon sampling, we demonstrated a lack of correlation between the numbers of clones from a phylogenetic taxon and the corresponding hybridization intensity by using a 62,358 probe array [61]. Analogous discrepancies have been documented when sampling environmental PCR products by cloning *versus* SSCP [31], or *versus* T-RFLP [42]—suggesting a cloning bias. It is possible that the cloning process is limited because of nonrandom selection from a heterogeneous pool when amplicons are nonuniform in length [47] or form variable secondary structures [38]. Conversely, the 16S rRNA microarray has the potential advantage of increased sensitivity. Where typical clone libraries must be pruned of sequencing aberrations (including chimeras) usually resulting in only hundreds of amplicons graduating to the final taxonomic assessment, the array accepts the entire mass of PCR products to be exposed to the probes. Using the described method of data analysis, the microarray requires $>10^7$ gene copies for detection (manuscript in preparation). In this study, between 10^{11} and 10^{12} molecules were sampled by the array, whereas only hundreds were analyzed by cloning. Therefore, minority amplicon types, with concentrations 4 orders of magnitude less than those in the majority, will have an increased probability of being detected by the microarray. In fact, high dominance within clone libraries correlated with large differences between the richness detected by array and cloning approaches (Table 4). The trend may predict that underestimation of the true richness can occur when cloning efforts produce only a limited pool of diversity. We acknowledge that further investigation of this trend is necessary, especially to exclude the possibility that simply small sample sizes are the sole cause of underestimation.

The described high-density universal 16S rRNA microarray has been successfully used to monitor metal-reducing bacteria during uranium bioremediation [57] and flux in airborne prokaryote populations in urban

settings (Andersen *et al.*, unpublished data). In this study, we presented the results of applying PCR products to the arrays; however, by interrogating nonamplified rRNA, a significant source of bias can be alleviated. This is the focus of ongoing studies.

In summary, although this microarray is unreliable in classifying novel taxa it, was capable of confirming the majority of clone-detected subfamilies in addition to revealing greater richness, even at the phylum level. Furthermore, richness observations from the array analysis corresponded well with nonparametric richness predictions calculated from clone sampling, indicating a more complete inventory of the sampled ecosystems. A subset of taxa uniquely identified by the array was verified, illustrating the consequence of relying solely on clone libraries when profiling a microbial community. The laborious, costly, and time-consuming nature of clone library analysis diminishes its utility in studies requiring replication and temporal monitoring. The responsiveness of the 16S rRNA microarray to nucleic acids from diverse phyla in complex mixtures and its suitability for investigations requiring replication, demonstrated a necessary advance toward the goal of high-throughput ecological monitoring. For these reasons, we believe the high-density DNA microarray offers a promising approach for studies of microbial ecology.

References

1. Bach, HJ, Tomanova, J, Schloter, M, Munch, JC (2002) Enumeration of total bacteria and bacteria with genes for proteolytic activity in pure cultures and in environmental samples by quantitative PCR mediated amplification. J Microbiol Methods 49: 235–245

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

3. Chandler, DP, Jarrell, AE (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. BioTechniques 38: 591–600

4. Chao, A (1984) Non-parametric estimation of the number of classes in a population. Scand J Statist 11: 265–270

5. Chao, S, Lee, AM (1992) Estimating the number of classes via sample coverage. J Am Stat Assoc 87: 210–217

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

7. Chouari, R, Le Paslier, D, Dauga, C, Daegelen, P, Weissenbach, J, Sghir, A (2005) Novel major bacterial candidate division within a municipal anaerobic sludge digester. Appl Environ Microbiol 71: 2145–2153

8. Crocetti, GR, Hugenholtz, P, Bond, PL, Schuler, A, Keller, J, Jenkins, D, Blackall, LL (2000) Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. Appl Environ Microbiol 66: 1175–1182

9. 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

10. DeSantis, TZ, Hugenholtz, P, Keller, K, Brodie, EL, Larsen, N, Piceno, YM, Phan, R, Andersen, GL (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 34: W394–399

11. DeSantis, TZ, Hugenholtz, P, Larsen, N, Rojas, M, Brodie, EL, Keller, K, Huber, T, Dalevi, D, Hu, P, Andersen, GL (2006) Greengenes, a chimera-checked 16S rRNA genes database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072

12. DeSantis, TZ, Stone, CE, Murray, SR, Moberg, JP, Andersen, GL (2005) Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiol Lett 245: 271–278

13. Dojka, MA, Hugenholtz, P, Haack, SK, Pace, NR (1998) Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl Environ Microbiol 64: 3869–3877

14. Dunbar, J, Barns, SM, Ticknor, LO, Kuske, CR (2002) Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol 68: 3035–3045

15. Elshahed, MS, Senko, JM, Najar, FZ, Kenton, SM, Roe, BA, Dewers, TA, Spear, JR, Krumholz, LR (2003) Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. Appl Environ Microbiol 69: 5609–5621

16. Ewing, B, Green, P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186–194

17. Ewing, B, Hillier, L, Wendl, MC, Green, P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8: 175–185

18. Farrelly, V, Rainey, FA, Stackebrandt, E (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl Environ Microbiol 61: 2798–2801

19. Felsenstein, J (1989) PHYLIP—Phylogeny Inference Package (Version 3.65). Cladistics 5: 164–166

20. Felske, A, Akkermans, AD (1998) Prominent occurrence of ribosomes from an uncultured bacterium of the Verrucomicrobiales cluster in grassland soils. Lett Appl Microbiol 26: 219–223

21. Gans, J, Wolinsky, M, Dunbar, J (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309: 1387–1390

22. Garrity, GM (2001) Bergey's Manual of Systematic Bacteriology, 2nd edn. Springer-Verlag, New York

23. Ginige, MP, Hugenholtz, P, Daims, H, Wagner, M, Keller, J, Blackall, LL (2004) Use of stable-isotope probing, full-cycle rRNA analysis, and fluorescence in situ hybridization-microautoradiography to study a methanol-fed denitrifying microbial community. Appl Environ Microbiol 70: 588–596

24. Holmes, DE, Finneran, KT, O'Neil, RA, Lovley, DR (2002) Enrichment of members of the family Geobacteraceae associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. Appl Environ Microbiol 68: 2300–2306

25. Huber, T, Faulkner, G, Hugenholtz, P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20: 2317–2319

26. Hugenholtz, P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3: 1–8

27. 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

28. Hugenholtz, P, Huber, T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. Int J Syst Evol Microbiol 53: 289–293

29. Hugenholtz, P, Tyson, GW, Webb, RI, Wagner, AM, Blackall, LL (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Appl Environ Microbiol 67: 411–419

30. Hughes, JB, Hellmann, JJ, Ricketts, TH, Bohannan, BJ (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol 67: 4399–4406

31. Junca, H, Pieper, DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. Environ Microbiol 6: 95–110

32. Kazor, CE, Mitchell, PM, Lee, AM, Stokes, LN, Loesche, WJ, Dewhirst, FE, Paster, BJ (2003) Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients. J Clin Microbiol 41: 558–563

33. Kelly, JJ, Siripong, S, McCormack, J, Janus, LR, Urakawa, H, El Fantroussi, S, Noble, PA, Sappelsa, L, Rittmann, BE, Stahl, DA (2005) DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. Water Res 39: 3229–3238

34. Kent, AD, Smith, DJ, Benson, BJ, Triplett, EW (2003) Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. Appl Environ Microbiol 69: 6768–6776

35. Lane, DJ (1991) Nucleic acid techniques in bacterial systematics. In: Stackebrandt, E, Goodfellow, M (Eds.) Nucleic Acid Techniques in Bacterial Systematics, Vol 1. Wiley, New York, pp 115–175

36. Lanser, JA, Adams, M, Doyle, R, Sangster, N, Steele, TW (1990) Genetic relatedness of Legionella longbeachae isolates from human and environmental sources in Australia. Appl Environ Microbiol 56: 2784–2790

37. Lehner, A, Loy, A, Behr, T, Gaenge, H, Ludwig, W, Wagner, M, Schleifer, KH (2005) Oligonucleotide microarray for identification of Enterococcus species. FEMS Microbiol Lett 246: 133–142

38. Lindstrom, UM, Kool, ET (2002) An orthogonal oligonucleotide protecting group strategy that enables assembly of repetitive or highly structured DNAs. Nucleic Acids Res 30: e101

39. Liu, WT, Stahl, DA (2002) Molecular approaches for the measurement of density, diversity, and phylogeny. In: Hurst, CJ, Crawford, RL, Knudsen, GR, McInerney, MJ, Stetzenbach, LD (Eds.) Manual of Environmental Microbiology, 2nd edn., pp 114–134

40. 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

41. 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, KH (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371

42. Lueders, T, Friedrich, M (2000) Archaeal population dynamics during sequential reduction processes in rice field soil. Appl Environ Microbiol 66: 2732–2742

43. Magurran, AE (1988) Ecological Diversity and Its Measurement. Princeton University Press, Princeton, NJ

44. Masuda, N, Church, GM (2002) Escherichia coli gene expression responsive to levels of the response regulator EvgA. J Bacteriol 184: 6225–6234

45. McCaig, AE, Glover, LA, Prosser, JI (1999) Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Appl Environ Microbiol 65: 1721–1730

46. Miller, DN, Bryant, JE, Madsen, EL, Ghiorse, WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl Environ Microbiol 65: 4715–4724

47. Ohara, O, Temple, G (2001) Directional cDNA library construction assisted by the in vitro recombination reaction. Nucleic Acids Res 29: E22

48. Polz, MF, Cavanaugh, CM (1998) Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol 64: 3724–3730

49. Rosch, C, Bothe, H (2005) Improved assessment of denitrifying, N_2-fixing, and total-community bacteria by terminal restriction fragment length polymorphism analysis using multiple restriction enzymes. Appl Environ Microbiol 71: 2026–2035

50. Rozen, S, Skaletsky, H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–386

51. Schloss, PD, Handelsman, J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71: 1501–1506

52. Schloss, PD, Handelsman, J (2004) Status of the microbial census. Microbiol Mol Biol Rev 68: 686–691

53. Schmidt, TM, DeLong, EF, Pace, NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 173: 4371–4378

54. Skowronski, EW, Armstrong, N, Andersen, G, Macht, M, McCready, PM (2000) Magnetic, microplate-format plasmid isolation protocol for high-yield, sequencing-grade DNA. BioTechniques 29: 786–788, 790, 792

55. 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

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

57. Wan, J, Tokunaga, TK, Brodie, E, Wang, Z, Zheng, Z, Herman, D, Hazen, TC, Firestone, MK, Sutton, SR (2005) Reoxidation of bioreduced uranium under reducing conditions. Environ Sci Technol 39: 6162–6169

58. Warsen, AE, Krug, MJ, LaFrentz, S, Stanek, DR, Loge, FJ, Call, DR (2004) Simultaneous discrimination between 15 fish pathogens by using 16S ribosomal DNA PCR and DNA microarrays. Appl Environ Microbiol 70: 4216–4221

59. Webster, G, Newberry, CJ, Fry, JC, Weightman, AJ (2003) Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. J Microbiol Methods 55: 155–164

60. Wilson, KH, Blitchington, RB, Greene, RC (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction [published erratum appears in J Clin Microbiol 1991 Mar; 29(3): 666]. J Clin Microbiol 28: 1942–1946

61. Wilson, KH, Wilson, WJ, Radosevich, JL, DeSantis, TZ, Viswanathan, VS, Kuczmarski, TA, Andersen, GL (2002) High-density microarray of small-subunit ribosomal DNA probes. Appl Environ Microbiol 68: 2535–2541

62. Wilson, WJ, Strout, CL, DeSantis, TZ, Stilwell, JL, Carrano, AV, Andersen, GL (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol Cell Probes 16: 119–127

63. 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

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