Environmental Whole-Genome Amplification to Access Microbial Diversity in Contaminated Sediments

Carl B. Abulencia[1,4], Denise L. Wyborski[1,4], Joe Garcia[1,4], Mircea Podar[1,4], Wenqiong Chen[1,4], Sherman H. Chang[1,4], Hwai W. Chang[1,4], David Watson[2], Eoin L. Brodie[3,4], Terry C. Hazen[3,4], and Martin Keller[1,4,*]

[1]Diversa, San Diego, CA 92121, [2]Oak Ridge National Laboratory, Oak Ridge, TN 37831, and [3]Lawrence Berkeley National Laboratory, Berkeley, CA 94720
* Corresponding author. Mailing address: Diversa, 4955 Directors Place, San Diego, CA 92121. Phone: (858) 526-5162. Fax: (858) 526-5662. E-mail: mkeller@diversa.com.
[4]Affiliated with the Virtual Institute for Microbial Stress and Survival, Berkeley, CA 94720. (http://vimss.lbl.gov)

ABSTRACT

Analyses of microbial communities are made possible by the isolation of metagenomic DNA from environmental sources. Direct access to all the genomes present in a sample enables a more complete study of microbial diversity and the discovery of native species and genes. Low biomass samples from nitrate and heavy metal contaminated soils yield DNA amounts, which have limited use for direct, native analysis and screening. A whole genome amplification method was validated, and used to amplify the genomes from environmental, contaminated, subsurface sediments. By first amplifying the gDNA, biodiversity analysis, and genomic DNA library construction of microbes found in contaminated soils were made possible. Whole genome amplification of metagenomic DNA offers access to genomic information from very minute microbial sources.

INTRODUCTION

Terminal electron accepting (TEA) processes (e.g. oxygen, nitrate, iron(III), sulfate, carbon dioxide) have been found increasingly to play a critical and enabling role

in nearly all biogeochemical processes in the subsurface (Koenigsberg et al. 2005; Hazen and Tabak, 2005). This has led to the realization that natural attenuation and bioremediation of metals, radionuclides, and organic contaminants cannot be effectively applied at many sites until we have a better understanding of the physiology, ecology and phylogeny of microbial communities at contaminated sites [10,24,36]. However, the success of many monitored natural attenuation and bioremediation approaches largely depends on our understanding of regulatory mechanisms and cellular responses to different environmental factors affecting the contaminant degradation or metal reduction activity *in situ*. Microorganisms are often exposed to multiple stress conditions *in situ*, and the microbial community structure is most likely affected by many different abiotic and biotic variables in a non-linear fashion [37]. Given the extreme stresses and TEA limitations that the microbial community is under at these contaminated sites an accurate assessment of the microbial community structure and the ecogenomics is critical (Koenigsberg et al. 2005)

Despite the dominance of microorganisms in the biosphere, relatively little is known about the majority of environmental microorganisms, largely because of their resistance to culture under standard laboratory conditions [38]. As such, alternative approaches are required to access the large amount of information in the environmental metagenome. Environmental sequencing projects targeted at 16S ribosomal RNA are a popular method to assess phylogenetic diversity of uncultured organisms [3,17,21]. Cloning and sequencing of PCR amplified functional genes from environmental samples are also powerful tools for investigating the ecology and role of microorganisms [4,12,25]. More recently, direct sequencing of environmental genomic DNA has furthered our

understanding of the metabolic potential of microorganisms occupying various environmental niches. Sequencing efforts include individual large-insert bacterial artificial chromosome (BAC) clones, small-insert libraries made directly from environmental DNA, and high-throughput shotgun sequencing which provides a global view of the environmental community [12,14,23,26,32,34,35]. These techniques have advanced our understanding of the types of microorganisms and biodegradation/biotransformation capabilities found in various habitats.

To understand how biogeochemical processes affect microbial community structure and bioremediation, the U.S. Department of Energy's (DOE) Natural and Accelerated Bioremediation Research (NABIR) program has established a Field Research Center (FRC) on the DOE Oak Ridge Reservation (http://www.esd.ornl.gov/nabirfrc/) in eastern Tennessee. The FRC is heavily contaminated with nitrate, heavy metals, radionuclides and halogenated organics [37]. Characterization of indigenous subsurface microbial populations from this site has mainly focused on microbes collected from from groundwater [18,37] or following biostimulation [17,21]. However, it is recognized that the planktonic microbial population in groundwater may not be representative of the active population residing as biofilms on sediment surfaces (Haglund et al 2002). In fact the biofilm mode of growth confers resistance to environmental stresses such as heavy metals (Teitzel and Parsek, 2003).

Extremely low cell densities in combination with high clay content and heavy metal/radionuclide contamination generally inhibit many standard molecular approaches including shotgun and BAC library generation. Isolation of DNA for library construction is complicated and would require large quantities of subsurface material. The current

minimum of DNA to construct a library used for shotgun sequencing is around 0.5-4 μg of DNA which can be obtained from a minimum of 0.5 g of microbial rich material [33]. In low biomass subsurface environments with cell densities as low as 10^4 per gram, 11-88 kg of sediment would be required assuming an average cell contains 4.5 fg DNA. Considering the quantity of DNA needed for BAC library construction is 20 fold greater, it is clear that current approaches are not viable for such low biomass environments.

One technique used for molecular surveys of microorganisms from environmental samples is to amplify DNA by polymerase chain reaction (PCR) using primers to highly conserved positions in specific rRNA genes [9,14]. However, low cell count environments often do not yield enough DNA for PCR [21]. In addition, DNA amplification by PCR is inherently biased because not all rRNA genes amplify with the same "universal primers", reducing the effectiveness of this approach to survey such environments by a broad shotgun sequencing. Therefore, new methods are required to combine environmental whole genome amplification (WGA) with library construction for metagenome analyses of low cell density environments.

Here we report the construction of whole-genome amplified environmental libraries for biodiversity assessment of low biomass contaminated subsurface sediment cores.

MATERIAL AND METHODS

DNA amplification. Genomic DNA (gDNA) was amplified by Multiple Displacement Amplification (MDA) using the GenomiPhi DNA Amplification Kit (Amersham Biosciences, Piscataway, New Jersey). Amplification was carried out

according to the protocol with a modification in reaction incubation time. One ⌠l of template was added to 9 ⌠l of sample buffer and heated to 95°C for 3 min to denature the template DNA. The sample was cooled and mixed with 9 ⌠l of reaction buffer and 1 ⌠l of enzyme mix, and incubated at 30°C for 3 to 6 h. After amplification the DNA polymerase was heat-inactivated during a 10 min incubation at 65°C. Each sample was amplified in triplicate. The three MDA reaction products per sample were then combined before further processing.

GeneChip analysis. MDA genome coverage was analyzed using the Affymetrix *E. coli* genome GeneChip array (Affymetrix, Inc., Santa Clara, Calif.). The chip contains 7231 probe sets spanning the entire *Escherichia coli* genome. The *E. coli* strain used was XL1-Blue MR (Stratagene, La Jolla, Calif.). Genomic DNA extracted from an overnight *E. coli* culture was used as a positive control (approx how many cells?). This was compared to MDA products amplified from gDNA extracted from two dilutions of the culture (5000 and five cells). Cell culture concentration was estimated by a spectrophotometer reading (OD 600) and serial dilutions were made to 5000 and five cells. Aliquots of the dilutions were grown on solid agar plates to verify cell counts. The gDNA was extracted from the overnight culture and the dilutions by first encasing the cells in agarose, then completing the extractions as described above. The three sets of DNA were concentrated by ethanol precipitation and fragmented with DNase I (0.6 U per μg of genomic DNA) at 37°C for 10 min. After inactivating DNase I at 100°C for 10 min, the fragmented DNA was end-labeled with biotin-ddATP and hybridized to the *E. coli* genome array using Affymetrix standard protocols for RNA. The probe array was scanned twice and the intensities were averaged with a GeneArray Scanner (Hewlett-

Packard, Palo Alto, CA). Scanned images were processed and quantified using GeneChip Suite 5.0 (Affymetrix).

The data were normalized by setting the mean hybridization signal for each sample equal to 100. The absolute call represents a qualitative indication of whether or not a transcript is detected within a sample. These calls are determined using the following metrics: 1) the ratio of the number of positive probe pairs to the number of negative probe pairs (known as the Positive/Negative Ratio), 2) the fraction of positive probe pairs (Positive Fraction), and 3) the average across the probe set of each probe pair's log ratio of positive intensity over negative intensity (Log Average Ratio) [16].

Amplification and analysis of mixed isolates. Eight isolates with fully sequenced genomes were chosen to represent a range of genome sizes (2.5Mb to 8.7Mb) and G/C content (32% to 72%). The isolates were *Deinococcus radiodurans* ATCC 13939, *Desulfovibrio vulgaris* ATCC 29579, *Geobacter sulfurreducens* ATCC 51573, *Mesorhizobium loti* ATCC 35173, *Nitrosomonas europaea* ATCC 19718, *Shewanella oneidensis* ATCC 700550, *Staphylococcus epidermidis* ATCC 35984, and *Streptomyces coelicolor* ATCC 10147 (ATCC, Manassas, VA). The gDNA from the eight isolates was mixed at an equimolar ratio based on genome size. The resulting mix had a concentration of 60 ng/µl. A small insert library was made from four micrograms (66 µl) of the mixed DNAs, as described below. The mix was then diluted 100-fold and 10,000-fold to 600 pg/µl and 6 pg/µl, respectively. One µl of the diluted gDNA was then amplified by MDA. The DNA from each dilution was amplified to greater than four micrograms. Small insert libraries were created from 4 µg of the un-amplified mixed DNA and from the amplified DNA as described below. Random clones from each of the three libraries

were end-sequenced. Isolate representation within each library was determined by a BLASTN search of the end-sequences against the NCBI genome database.

Site and Sample description. Soil core samples were collected from contaminated subsurface sediments at the U.S. Department of Energy NABIR FRC, located at the Y-12 plant within the Oak Ridge Reservation in Oak Ridge, TN. Approximately 10 million liters of liquid nitric acid and uranium bearing wastes were discarded at this site per year for 30 years, until it was closed in 1984. The site groundwater plumes originate from the former waste disposal ponds at the Y-12 plant. Nine soil samples were taken from five areas surrounding the S-3 waste ponds. The samples were from sites of varying distances to the S-3 ponds, and from core depths ranging from 3.5 - 9 m, with each sample containing different levels of contamination of U(VI), nitrate, plutonium, technetium, toxic metals (nickel, aluminum, barium, chromium, mercury), chelating agents (EDTA), chlorinated hydrocarbons (trichloroethylene and tetrachloroethylene), polychlorinated biphenyls, and fuel hydrocarbons (toluene, benzene). A full description of the area can be obtained at the FRC website (http://www.esd.ornl.gov/nabirfrc/).

Of the nine contaminated samples, the following three were analyzed in more detail: a) FB075, Area 3, which is adjacent to the west side of the ponds, core segment depth 8.4 - 9 m; (Sample 1, Library 1) b) FB076, Area 3, core segment depth 3.9 - 4.5 m; (Sample 3, Library 3) c) FB078, Area 2, which is >200 m to the southwest of the ponds, core segment depth 6.1 -6.4 m (Sample 5, Library 5). Sediment cores were sampled with an Acker Drill Co. Holegator track drill equipped with polyurethane sleeves lining the

corer. The cores were anaerobically sealed in argon, shipped on ice within 24 h of sampling, and DNA extractions were done on arrival in a radiation control area at LBNL.

DNA isolation. Each soil sample was removed from the core sleeve, mixed manually, and 50 g of soil per sample was used for gDNA isolation. High quality, high-molecular-weight DNA was isolated directly from soil samples following separation of cells from the environmental matrix [1,22,27,27,28,28]. Highly purified suspensions of microbial consortia were obtained by isopycnic density gradient centrifugation with Nycodenz. The resulting cell pellet was immobilized in an agarose plug and lysed by enzymatic and chemical digestions [30]. The isolated gDNA (all of it – how much) was then used directly in the amplification reaction, as described above.

16S rRNA gene analysis. Bacterial 16S rRNA genes were amplified from the MDA DNA products using the universal primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3') and the Roche Expand Long Template PCR system (Roche Applied Science, Indianapolis, IN). 16S clone libraries were generated using the TOPO TA Cloning kit (Invitrogen, Carlsbad, CA). Sequencing was performed using an Applied Biosystems 3730xl DNA analyzer (Applied Biosystems, Foster City, CA) and the individual clone reads assembled using Sequencher (Gene Codes Corporation, Ann Arbor, MI).

The assembled sequences were checked for potential chimeric artifacts using the Bellerophon program and the chimeric sequences were discarded (Huber et al 2004). The final sequence datasets were aligned by Clustal W with the closest sequence relatives from ARB and GenBank databases. The alignments were manually curated using BioEdit.

To analyze the diversity of the different sequence libraries, a Jukes-Cantor corrected distance matrix was calculated using DNADIST (PHYLIP). The matrix was used as input into the program DOTUR (Schloss and Handelsman, 2005) using the furthest neighbor algorithm to obtain a variety of diversity richness estimators at different genetic distance values (rarefaction curve, bias corrected Chao1 richness, abundance-based coverage estimator ACE, Shannon-Weaver diversity index). Phylogenetic analysis was conducted using PAUP* with the distance criterion (Jukes Cantor) followed by bootstrapping (1000 replicates). Trees were visualized using Tree Explorer.

Library construction, sequencing and analysis. Un-amplified gDNA and MDA amplified DNA was mechanically sheared and used to generate libraries in ZAP-based lambda phage cloning vectors according to the manufacturer's protocol. Phagemid libraries were produced from the parental lambda clones through the *in vivo* excision properties of ZAP-based cloning vectors and used to infect *E. coli* host cells [29]. Average insert sizes were 2-4 KB. Plasmid inserts from randomly picked colonies were end-sequenced using an Applied Biosystems 3730xl DNA analyzer with primers T3 (5'-AATTAACCCTCACTAAAGGG-3') and T7 (5'-GTAATACGACTCACTATAGGGC-3'). The end-sequences of the random clones were analyzed against the NCBI protein database using BLASTX. Only hits with e-values $<10^{-9}$ were considered for further protein analysis. Sequences were assembled into contigs using Vector NTI Contig Express. For the MDA bias analysis, paired clone-end singletons were discarded.

COG analysis. COG functional assignment of proteins predicted from DNA sequences from different genomic libraries was done using the genomic DNA sequences to blast against COG database from http://string.embl.de/, which covers 26,201 protein

families, using BLASTX. Filtering criteria are set to get rid of the non-specific blast hits: 1) If the aligned amino acid sequence length is greater or equal to 100 aa, the blast e-value should be less than (10^{-9}), 2) If the aligned amino acid sequence length is greater or equal to 30 aa but less than 100 aa, amino acid percentage of identity should be greater than 30%, and 3) If the aligned amino acid sequence length is less than 30 aa, percentage of identity should be greater than 25%.

The filtered blast results were parsed to associate the COG IDs with the individual genomic DNA sequences using a custom Perl script. In the case of a particular sequence which has multiple hits that belong to different COG classification, we allowed multiple COG assignment only when the bit scores from the 2nd or 3rd COG classification were no less than 3-fold different than the bit score from the top COG classification.

RESULTS

GeneChip analysis of *E. coli* MDA coverage. The amount of *E. coli* gDNA extracted from both 5000 cells and five cells is very minute and can not be detected by the Affymetrix *E. coli* genome GeneChip following hybridization. This low number of cells was used to simulate the low abundance of organisms found in contaminated soils. Genomic DNA extracted from an overnight *E. coli* culture (how much culture – or about how many cells - 10^9 ?) elicited positive signals for all probe sets on the *E. coli* GeneChip. After amplification by MDA, gDNA extracted from both 5000 cells and five cells resulted in 99.94 percent and 99.2 percent of the probe sets called “Present”, respectively.

We also utilized the quantitative information from the GeneChip experiments to assess if there is over- or under-amplification of regions of the gDNA during the MDA

reaction. Over- or under-amplification is defined as a detection value greater or less than three times the positive control value. For the 5000 cell amplification, 0.4% of probe set regions was over-amplified, and 0.9 % was under-amplified. For the five-cell amplification, 0.6 percent was over-amplified and 4.6% was under-amplified. The amplification bias appears to be random, with regions of over- or under-amplification distributed across the entire genome.

Amplification of Mixed Isolates to test for MDA bias. In order to have complete and adequate representation of genomes in an environmental sample, the whole genome amplification must amplify all the genomes present with minimal bias. To determine if any MDA bias exists when amplifying from the environment, gDNAs from eight different known isolates were mixed to simulate an environmental sample. 510, 421, and 359 end-sequences from the un-amplified library, 100-fold dilution amplified library, and 10,000-fold dilution amplified library, respectively, were analyzed and binned to the corresponding isolate. The comparison of the different libraries showed that there is an MDA bias. Blast analyses of the end-sequences from random clones from each of the three libraries revealed that *Shewanella oneidensis, Nitrosomonas europaea*, and *Geobacter sulfurreducens* were amplified preferentially. *Deinococcus radiodurans, Desulfovibrio vulgaris, Mesorhizobium loti,* and *Streptomyces coelicolor* all had less representation in the amplified DNA library compared to the un-amplified DNA library, with *S. coelicolor* having the least representation. *Staphylococcus epidermidis* showed no bias between un-amplified and amplified DNA. However, in this small sample set of sequences, all isolates were represented in the MDA amplified libraries regardless of bias (Table 1).

Microbial diversity analysis of the soil libraries. DNA was isolated from soil sample cores according to the described protocol. Three different samples were used for further studies: Samples FB075 (sample 1), FB076 (sample 3), and FB078 (sample 5). Extractions from sample 1 and sample 3 resulted in DNA concentrations, which were not sufficient to obtain 16S rRNA gene PCR products. The only sample that contained a sufficient level of DNA template for 16S rRNA gene PCR was sample 5 (<5 ng/⌈l). A native 16S rRNA library (native library 5) was constructed from this sample. In addition, the DNA from this sample was amplified by MDA with greater than 6 ⌈g DNA yield. A 16S rRNA library was constructed from the amplified product (amplified library 5). The same MDA amplification method was used on the DNA obtained from sample 1 and 3 to successfully yield greater than 6 µg DNA each, even though PCR attempts were negative. 16S rRNA libraries were constructed from these amplified products (amplified library 1, amplified library 3).

To compare the microbial diversity before and after MDA, 47 16S rRNA gene clones from the native library 5 were sequenced and compared to 125 16S rRNA gene clones derived from the amplified library 5. Microbial diversity between deep (8.4-9 m, sample 1) and shallow (3.9-4.5 m, sample 3) soil samples from Area 3 was analyzed by comparing the 16S rRNA libraries from amplified samples 1 and 3. To estimate the microbial diversity within these samples, rarefaction curves using a 2% difference in SSU rDNA were used for species separation. The two libraries prepared from sample 5, native and amplified, indicated a significantly higher diversity of the amplified library (149 species versus 88 species for the native library – mention Chao1?), with upper estimates

reaching an excess of 200 species. The Shannon-Weaver index value also indicates higher species diversity in the amplified library prepared from sample 5 (Fig.1, Table 2).

Rarefaction curves calculated for the 2% distance shows signs of leveling of the number of novel OTUs (species) identified by increasing sequence sampling of sample 3 and especially sample 1. The Chao 1 estimator predicts a minimum number of 24 species (17 to 92 at the 95% confidence interval) for sample 1 and 45 species (33 to 102 at the 95% CI) for sample 3, similar estimates being obtained using ACE (Table 2).

The taxonomic diversity at higher levels is revealed by the phylogenetic trees constructed based on the sequences from the three different samples as well as the histogram figure summarizing the diversity at phylum or class level for each 16S rRNA library (Figs. 2, 3, 4). The availability of sequences from both native and amplified community gDNA for sample 5 allowed us to investigate possible biases introduced by the amplification step. In general, there is good agreement between the taxonomic groups identified in the two samples with some differences, especially for groups that have low representation (e.g. alpha proteobacteria, *Verrucomicrobia* and some of the uncultured groups). Overall, however, no single phylum appears to be strongly dominating compared to the other sample (what does "other sample" refer to here?). Acidobacteria and the candidate division OP11 are over represented in both samples in terms of species.

Sample 3, while of higher diversity in comparison to sample 1, reveals a relatively even distribution of species across several major phyla, including proteobacteria, actinobacteria, planctomyces, verrucomicrobia, and the candidate division OP11. In sample 1 however, there is a dominance of gamma proteobacteria (~35%), equally

distributed between a close relative of *Pseudomonas synxantha* (>50% of sequences) and a *Legionella*-type species. The next most represented groups (25% each) are *Cytophaga/Bacteroidetes*, represented by 27 sequences and approximately 4 species-level OTUs and beta proteobacteria (several taxa including *Spirillum, Aquaspirillum* and *Cenibacterium* relatives). (do any of the clones correspond to sequences found at the Oak Ridge FRC before?)

Partial end sequencing and COG analysis of environmental clones from genomic libraries 1, 3, and 5. Although a 16S rRNA gene PCR product was successfully amplified from the sample 5 gDNA, the amount of gDNA obtained was not sufficient for phagemid library construction. Extracted DNA from all three samples was amplified by MDA to create a sufficient quantity of DNA (2 μg) to construct genomic libraries. To check the quality and diversity of environmental libraries 1, 3 and 5, constructed from amplified DNA, 960, 864, and 864 random clones from each were sequenced, respectively. A total of 4021 sequences of 400 nt or longer were generated from both ends of the clones. Of these sequences, 2759 (68.6%) showed similarities to known genes in the database. Further analysis showed that 2180 (54.2%) sequences had similarity to bacterial proteins, 133 (3.3%) to archaeal proteins, and 51 (1.3%) to eukaryal proteins. A contig assembly of the end-sequences was used to analyze the extent of MDA bias on the amplified DNA used to construct the libraries. If a high proportion of sequences form contigs, a bias in amplification of these sequence regions is possible. Libraries 1, 3, and 5 had 370 sequences that formed 101 contigs, 152 sequences that formed 53 contigs, and 141 sequences that formed 54 contigs, respectively (Table 3).

The size of the largest contig formed in library 1, 3, and 5, was 1799 nt, 2373 nt, and 2749 nt, respectively.

In order to explore the possible functions of the predicted proteins from the sequences collected in each soil library, COG (cluster of ortholog) analysis with the random sequence reads was performed. More than half of the sequences from each library showed homology to the entries in the STRING COG database. For example, for library 1, among the 1394 random sequences, 1154 of them were assigned to 674 distinct COG IDs. For library 3, among the 1118 sequences obtained, 782 of them were assigned to 561 distinct COG IDs. For library 5, among the 1509 sequences obtained, 1126 of them were assigned to 800 COG IDs. As shown in Fig. 5 (Table 4?), the 3 environmental libraries contain similar distribution for most of the COG functional categories, except for carbohydrate transport and metabolism and secondary metabolites biosynthesis, transport and catabolism, which are significantly lower in library 5. The COG IDs which are present at higher frequencies in each library mostly belong to hypothetical proteins. In library 1, the top 5 most frequent COGs are COG3762: hypothetical protein (35), COG0642: sensor histidine kinase (27), COG0438: hypothetical protein PF1364 (19), COG0745: putative two-component regulator (17), and COG3696:cation efflux system protein (16). In library 3, the top 3 most frequent COGs are COG0642: sensor histidine kinase (14), COG0463: putative glycosyl transferase (14), and KOG1869: annotation not available (14). In library 5, the top 5 COGs which are present most frequently are COG0642: sensor histidine kinase (32), KOG1869: annotation not available (24), KOG1216: annotation not available (16), COG0457: tetratricopeptide

repeat domain 13 (15), and KOG2146: Ser/Arg-related nuclear matrix protein (plenty of prolines 101-l; Ser/Arg-related nuclear matrix (14) (Table 4).

DISCUSSION

The economics of sequencing is rapidly changing, due to the improvement of tools for sequencing and assembly. This has provided a significant boost to the field of environmental genomics [15]. Shotgun sequencing provides a wealth of biomarkers that can be used to assess the phylogenetic diversity of a sample with more power than conventional PCR-based rRNA studies allow [35]. In addition, the direct sequencing of environmental samples has provided valuable insights into the lifestyles and metabolic capabilities of uncultured organisms occupying various environmental niches [33]. Information derived from comparative metagenomic analyses could be used to predict features of the sampled environments such as elemental recycling, conversion of biomass, bioremediation, and stress response [15]. So far however, only environments with relatively high biomass such as biofilms, open ocean water, agricultural soils and forest soils have been studied [32-35]. Samples with extremely low cell counts such as contaminated subsurface sediment cores are usually not accessible for an environmental sequencing study. A significant, and typically impractical, amount of contaminated sediment would be necessary to isolate enough DNA for traditional library construction, and in the case of radionuclide contaminated soils, would pose the added problem of secondary waste generation. Possible solutions include the development of new methods to decrease the amount of DNA needed for library construction or methods to amplify environmental DNA to obtain enough for library construction. PCR-based methods have

been used for whole genome amplification, however they exhibit a high amplification bias and do not amplify genomes in their entirety [5,13].

ϕ 29 DNA polymerase is an enzyme which is widely used for Rolling Circle Amplification of plasmids and circular DNA templates [6,8]. The strand displacing enzyme has proof reading activity, is extremely sensitive and has been shown to amplify DNA up to 70kB [2]. This polymerase has also been used for whole genome amplification of bacterial isolates [8,20]. To evaluate ϕ 29 DNA polymerase MDA for whole genome amplification from low biomass samples we first performed MDA on as few as five *E. coli* cells. Gene chip data demonstrated that 99.2% of the *E. coli* genome was detectable showing no significant hot spots (areas of over- or under-representation). Previous analysis of *Xyella fastidiosa* libraries constructed from amplified and un-amplified genomic DNA also showed similar genome coverage from both DNA sources [8]. However, the mixing of eight sequenced bacterial strains, normalized according to their genome size, revealed preferred amplification of certain strains. This bias is not thought to be based on genome accessibility due to G+C content or secondary structures [2,13]. The introduction of this bias could be due to stochastic effects of amplifying from very low concentrations of template [20]. Another possibility could be the differential cloning efficiency of the amplified DNA in comparison to the un-amplified DNA. Analysis of the library constructed from the un-amplified DNA also demonstrated some bias, possibly introduced through different cloning efficiencies, which can be affected by G+C content, repeats native to the genome sequence, and toxicity, among others. However, even from the limited number of MDA- amplified library clones sampled by sequencing, clones from all the bacterial strains could be identified within the constructed library.

It has been demonstrated that core samples obtained from the NABIR site are very low in bacterial cell counts [17]. Most of the microbial diversity studies on samples from these sites have been performed on groundwater or on biostimulated samples [18,21,37]. One study that attempted 16S rRNA PCR from DNA extracts of contaminated sediments was unsuccessful [21]. In the current study, a 16S rRNA gene PCR product from native gDNA was obtained only from one sample out of nine. Therefore, we used this sample to study the bias on microbial diversity introduced through MDA. The comparison of the microbial diversity of the native and amplified 16S rRNA library showed relatively small changes within the taxonomic groups. Distinguishing traditional taxonomic units (e.g. species) based on rRNA gene sequences is controversial (Mircea). Typically, 2-3% difference at the level of SSU rDNA is used to define operational taxonomic units (OTU) equivalent to bacterial species, while 5% could be used to distinguish genera (Mircea). The rarefaction curves calculation used for these samples revealed a higher proposed species count for the MDA amplified library (149 vs. 88 species). This can be due to a more sensitive amplification of underrepresented species through the MDA process in comparison to the PCR amplification directly from environmental DNA. This might be in agreement with Gonzalez et al. (11) who demonstrated a more effective amplification of species through MDA combined with 16S rRNA specific PCR, versus direct 16S rRNA specific amplification from environmental samples. The differences between the taxonomic groups identified in the two libraries, which occur for groups that have low representation (e.g. alpha proteobacteria, *Verrucomicrobia* and some of the uncultured groups), may be due to limited sequence sampling, especially for the native DNA library.

The sequence quality of the 16S rRNA amplified library was, in general, good, and there was no evidence for chimera formation introduced through the MDA amplification.

The16S rRNA library from sample 3, while of lower diversity, reveals a relatively even distribution of species across several major phyla, including proteobacteria, actinobacteria, planctomyces, verrucomicrobia and the candidate division OP11. The lower diversity in library 1 compared to library 3 could be explained by the increased depth of sample 1 and differences in the contamination composition. However, we cannot exclude that the extremely low amount of DNA obtained from these samples biases the total number of predicted species. Both Library 1 and Library 3 contain sequences closely related to *Pseudomonas synxantha*, which is a known reducer of Cr(VI) as well as a hydrocarbon degrader, pointing toward a potential involvement in bioremediation [19,31]. Overall, our data shows that MDA can be used to obtain enough DNA from low biomass samples to evaluate their microbial diversity. For samples 1 and 3, microbial diversity could be analyzed only after MDA. Trace amounts of DNA template in an environmental sample may not be enough to generate a 16S rRNA gene PCR product, and PCR may also be inhibited by chemical inhibitors found in the soils. Only after first MDA amplifying the gDNA in the soil samples is PCR possible[11].

A shotgun sequencing approach was used to further test the quality of the libraries constructed from MDA amplified gDNA. The amplification of environmental DNA through MDA is a very powerful technology. Because the extreme sensitivity of this method is problematic, reactions must be performed under exceptionally clean conditions. In addition, it is important to perform negative control experiments, which help to evaluate potential contamination of the reagents through foreign DNA.

From our validation of the MDA protocol, we have observed very high genome coverage from an isolate, and a biased but representative (can something be both biased and representative at the same time? – do you mean genome coverage of each organism was similar but relative abundances were biased) DNA library from a mix of known genomes. From the sampling of sequences of library clone ends, a number of sequences were assembled into contigs (sequences that formed paired clone-end singletons were discarded). The fact that contigs were revealed within a shallow sampling of library sequences might suggest that there is some bias introduced by MDA. The assembled sequences possibly represent the gDNA areas of over-amplification. The great majority of the contigs were comprised of only two sequences. While some contigs were comprised of more than two sequences, the longest of these contigs contained 10 sequences and was 2700 nt long (data not shown). This indicates that while there is some bias introduced, the bias is random (as observed in the GeneChip data) and there is no major area of over-amplification. To further decrease the effect of MDA bias on library cloning, each sample was amplified in triplicate. Each sample's amplified products were then mixed before any further processing or analysis. In this way, any random bias in one reaction should be balanced out by the random biases in the other two reactions, the end-result being representative genome coverage. Only by further sequencing to a far greater depth can areas of under-amplification be analyzed. Analysis by BLAST showed that 31.4% of the environmental library sequences had no sequence similarity to any of the known proteins in the database (e-values $>10^{-14}$). These sequences are common in normal MDA-independent genomic libraries (data not shown) [32] and could contain non-coding, intergenic regions, or possibly novel genes. The sequences that are found in

MDA libraries could also be the result of MDA artifacts such as primer-derived multimers (similar to PCR primer-dimer artifacts), and direct and inverted sequence repeats. Because the sensitivity of the MDA reaction is known to create DNA products even in the absence of added cells [20], it can be expected to find sequences with low sequence-similarity and sequences containing combinations of sequence-repeats in MDA amplified libraries. The data from the limited number of clones sequenced reveals diverse environmental libraries that can be readily screened for native or novel sequences. Analysis of the environmental amplified-DNA libraries showed that 80.5% (but there were 31.4% with no similarity? – I only counted 58.8% between the bacterial, archaeal and eukaryal) of the sequences had significant similarities to known protein-encoding genes.

The similarity of the COG analysis from Area 3 (Libraries 1 and 3) suggests the environment may influence the "functional" profile of a community, which had been hypothesized previously [33]. The significance of the sensor histidine kinase and two-component regulator COGs present in all contaminated sediments examined and their relationship to stress response is being analyzed further. Overall, the sequence analysis demonstrates that MDA introduces some artifacts but allows the creation of libraries for shotgun sequencing from these extremely low biomass-containing samples. Further studies are needed to reveal if it would be possible to contig whole pathways or genomes, although significant information about the environment can be obtained by identifying environment-specific genes [33]. Our limited sequencing analysis of these contaminated environments is not intended to define the soil metagenome, but to give insight to the possibilities provided by WGA using ϕ29 DNA polymerase. By amplifying the extracted

DNA from a sample, it is now possible to access the genome information from contaminated environments that was not previously accessible. Downstream analysis could involve functional screening for active clones or sequence-based screening using probes homologous to known genes.

Genomic analyses of uncultured microbes can provide significant insight into the biological properties of individuals within microbial populations [7]. By examining convergent and divergent sets of proteins and regulatory elements, within niches and across niches, we can better understand subsurface mobilization and immobilization of radionuclides and metals. This will help to manipulate, stabilize, and predict long-term stability of these contaminants and their relative risk.

ACKNOWLEDGEMENTS

This work was supported at Lawrence Berkeley National Laboratory from the U.S. Department of Energy, Genomics:GTL program under Contract No. DE-AC02-05CH11231, and by the Office of Science (BER), U.S. Department of Energy, Genomics:GTL program, Grant No. DE-FG02-04ER63771.

We thank the sequencing team at Diversa for excellent sequencing support. We also thank Sharon Borglin, Dominique Joyner, Rick Huang, and Tamas Torok at LBNL for help with initial sample processing. We also thank Mel Simon and Keith Kretz for critical reading of the manuscript.

REFERENCES

FIGURES

TABLE 1. Amplification of mixed isolates to test for MDA bias

TABLE 1. Comparison of random end sequences of pre- and post-MDA libraries

Isolate	Size (Mb)	% G+C	Pre MDA		1e-2 dilution, post MDA		1e-4 dilution, post MDA	
			Number of Sequences [a]	%	Number of Sequences	%	Number of Sequences	%
Deinococcus radiodurans	3.3	66	47	9.2%	5	1.2%	12	3.3%
Desulfovibrio vulgaris	3.8	60	44	8.6%	5	1.2%	8	2.2%
Geobacter sulfurreducens	3.8	61	29	5.7%	54	12.8%	58	16.2%
Mesorhizobium loti	7.6	62	123	24.1%	15	3.6%	12	3.3%
Nitrosomonas europaea	2.8	50	4	0.8%	42	10.0%	46	12.8%
Shewanella oneidensis	5.1	45	97	19.0%	277	65.8%	192	53.5%
Staphylococcus epidermidis	2.5	32	26	5.1%	21	5.0%	30	8.4%
Streptomyces coelicolor	8.7	72	140	27.5%	2	0.5%	1	0.3%
Total			510		421		359	

a. Classification of sequences to their respective isolate determined by BLASTN searches against the NCBI genome database

Figure 1. Rarefaction curves for SSU rRNA genes from amplified samples 1, 3 and 5 and native sample 5, at 98% sequence identity level

Figure 2. Phylogenetic tree of SSU rRNA genes from sample 1, calculated using Jukes-Cantor corrected distances. For comparison, the closest relatives (known species and environmental sequences) were included in the analysis. The numbers in circles indicate multiple independent clones with highly similar sequences (>99%) to the sequence indicated on the tree. Numbers at nodes indicate bootstrap support (if <50, indicated by a small circle).

Figure 3. Phylogenetic tree of SSU rRNA genes from sample 3, calculated using Jukes-Cantor corrected distances. For comparison, the closest relatives (known species and environmental sequences) were included in the analysis. The numbers in circles indicate

multiple independent clones with highly similar sequences (>99%) to the sequence indicated on the tree. Numbers at nodes indicate bootstrap support (if <50, indicated by a small circle).

Figure 4. Phylogenetic tree of SSU rRNA genes from sample 5, calculated using Jukes-Cantor corrected distances. Sequences obtained using non-amplified DNA material are indicated by filled circle, sequences obtained from MDA-amplified DNA are indicated by open circles. For comparison, the closest relatives (known species and environmental sequences) were included in the analysis. The numbers in circles indicate multiple independent clones with highly similar sequences (>99%) to the sequence indicated on the tree. Numbers at nodes indicate bootstrap support (if <50, indicated by a small circle).

Figure 1. Rarefaction curves for SSU rDNA sequences at 98% identity level

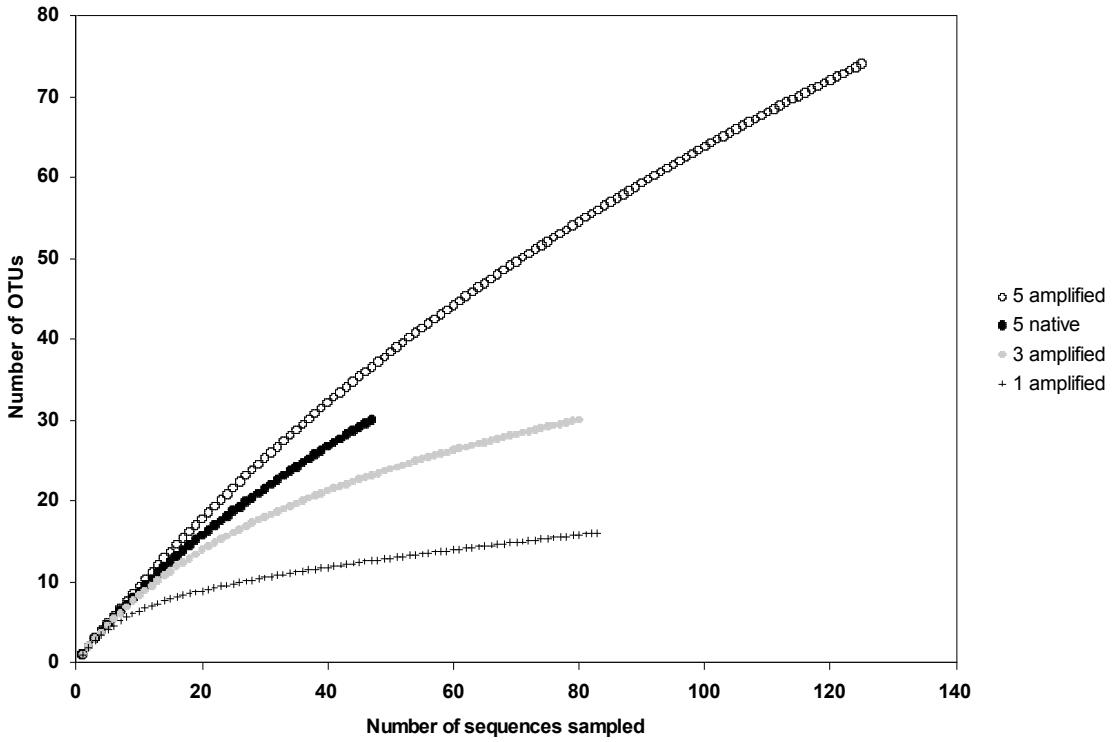

TABLE 2.

	Number of sequences	Chao estimates (95% confidence)	ACE estimate	Shannon diversity index
1	83	24 (17..92)	27	2.30 (2.12..2.49)
3	80	45 (33..102)	47	3.10 (2.91..3.29)
5 amplified	125	149 (107..248)	163	4.10 (3.95..4.25)
5 native	47	88 (41..330)	82	3.21 (2.97..3.46)

Figure. 2: Phylogenetic Tree of Sample 5- amplified and un-amplified.

Figure 3: Phylogenetic Tree of Sample 1.

Figure 4: Phylogenetic Tree of Sample 3.

TABLE 3. Statistics on library end-sequences

Library	1	%	3	%	5	%	Total	%
Number of clones sequenced	960		864		864			
Sequences generated	1,920		1,728		1,728			
Quality Sequences [a]	1,394	100	1,118	100	1,509	100	4,021	100
Sequences that form contigs	370	26.5	152	13.6	141	9.3	663	16.5
Number of contigs assembled	101		53		54		208	
Sequences with similarities <e-10 [b]	977	70.1	703	62.9	1,079	71.5	2,759	68.6
sequences with similarities <e-15	854	61.3	603	53.9	931	61.7	2,388	59.4
Highest similarity to bacterial gene <e-10	880	63.1	629	56.3	671	44.5	2,180	54.2
Highest similarity to archaeal gene <e-10	12	0.9	43	3.8	78	5.2	133	3.3
Highest similarity to eukaryotic gene <e-10	12	0.9	18	1.6	21	1.4	51	1.3

a. Sequences >400nt in length
b. e-values from BLASTX searches against the NCBI protein database

TABLE 4: COG Analysis

Functional Classification	**Library 1**		**Library 3**		**L**
	# seq	**%**	**# seq**	**%**	**# seq**
Amino acid transport and metabolism	64	0.041	47	0.041	76
Carbohydrate transport and metabolism	186	0.118	186	0.162	73
Cell cycle control, cell division, chromosome partitioning	16	0.010	11	0.010	21
Cell motility	4	0.003	4	0.003	7
Cell wall/membrane/envelope biogenesis	117	0.074	90	0.079	122
Coenzyme transport and metabolism	68	0.043	23	0.020	31
Defense mechanisms	42	0.027	28	0.024	43
DNA replication, recombination and repair	86	0.054	61	0.053	84
Energy production and conversion	215	0.136	164	0.143	86
Function unknown	118	0.075	48	0.042	68
General function prediction only	120	0.076	101	0.088	130
Inorganic ion transport and metabolism	89	0.056	57	0.050	44
Intracellular trafficking, secretion, and vesicular transport	13	0.008	12	0.010	15
Lipid transport and metabolism	36	0.023	12	0.010	22
Nucleotide transport and metabolism	30	0.019	21	0.018	47
Posttranslational modification, protein turnover, chaperones	189	0.120	148	0.129	199
RNA processing and modification	1	0.001			3
Secondary metabolites biosynthesis, transport and catabolism	26	0.016	21	0.018	13
Signal transduction mechanisms	80	0.051	31	0.027	90
Transcription	37	0.023	29	0.025	47
Translation, ribosomal structure and biogenesis	43	0.027	52	0.045	84
Total # sequenced	1580		1146		1305

Reference List

1. **Berry, A. E., C. Chiocchini, T. Selby, M. Sosio, and E. M. Wellington**. 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol. Lett. **223**:15-20.

2. **Blanco, L., A. Bernad, J. M. Lazaro, G. Martin, C. Garmendia, and M. Salas**. 1989. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J. Biol. Chem. **264**:8935-8940.

3. **Brown, M. V. and J. P. Bowman**. 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS Microbiol. Ecol. **35**:267-275.

4. **Cottrell, M. T., J. A. Moore, and D. L. Kirchman**. 1999. Chitinases from uncultured marine microorganisms. Appl. Environ. Microbiol. **65**:2553-2557.

5. **Dean, F. B., S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken**. 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. U. S. A **99**:5261-5266.

6. **Dean, F. B., J. R. Nelson, T. L. Giesler, and R. S. Lasken**. 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. **11**:1095-1099.

7. **DeLong, E. F.** 2002. Microbial population genomics and ecology. Curr. Opin. Microbiol. **5**:520-524.

8. **Detter, J. C., J. M. Jett, S. M. Lucas, E. Dalin, A. R. Arellano, M. Wang, J. R. Nelson, J. Chapman, Y. Lou, D. Rokhsar, T. L. Hawkins, and P. M. Richardson**. 2002. Isothermal strand-displacement amplification applications for high-throughput genomics. Genomics **80**:691-698.

9. **Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace**. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl. Environ. Microbiol. **64**:3869-3877.

10. **Edwards, K. J., P. L. Bond, T. M. Gihring, and J. F. Banfield**. 2000. An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. Science **287**:1796-1799.

11. **Gonzalez, J. M., M. C. Portillo, and C. Saiz-Jimenez**. 2005. Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. Environ. Microbiol. **7**:1024-1028.

12. **Henne, A., R. Daniel, R. A. Schmitz, and G. Gottschalk**. 1999. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. **65**:3901-3907.

13. **Lasken, R. S. and M. Egholm**. 2003. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol. **21**:531-535.

14. **Liles, M. R., B. F. Manske, S. B. Bintrim, J. Handelsman, and R. M. Goodman**. 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl. Environ. Microbiol. **69**:2684-2691.

15. **Nelson, K. E.** 2003. The future of microbial genomics. Environ. Microbiol. **5**:1223-1225.

16. **Nimgaonkar, A., D. Sanoudou, A. J. Butte, J. N. Haslett, L. M. Kunkel, A. H. Beggs, and I. S. Kohane**. 2003. Reproducibility of gene expression across generations of Affymetrix microarrays. BMC. Bioinformatics. **4**:27.

17. **North, N. N., S. L. Dollhopf, L. Petrie, J. D. Istok, D. L. Balkwill, and J. E. Kostka**. 2004. Change in bacterial community structure during in situ biostimulation of subsurface sediment cocontaminated with uranium and nitrate. Appl. Environ. Microbiol. **70**:4911-4920.

18. **Palumbo, A. V., J. C. Schryver, M. W. Fields, C. E. Bagwell, J. Z. Zhou, T. Yan, X. Liu, and C. C. Brandt**. 2004. Coupling of functional gene diversity and geochemical data from environmental samples. Appl. Environ. Microbiol. **70**:6525-6534.

19. **Pattanapipitpaisal, P., A. N. Mabbett, J. A. Finlay, A. J. Beswick, M. Paterson-Beedle, A. Essa, J. Wright, M. R. Tolley, U. Badar, N. Ahmed, J. L. Hobman, N. L. Brown, and L. E. Macaskie**. 2002. Reduction of Cr(VI) and bioaccumulation of chromium by gram positive and gram negative microorganisms not previously exposed to Cr-stress. Environ. Technol. **23**:731-745.

20. **Raghunathan, A., H. R. Ferguson, Jr., C. J. Bornarth, W. Song, M. Driscoll, and R. S. Lasken**. 2005. Genomic DNA amplification from a single bacterium. Appl. Environ. Microbiol. **71**:3342-3347.

21. **Reardon, C. L., D. E. Cummings, L. M. Petzke, B. L. Kinsall, D. B. Watson, B. M. Peyton, and G. G. Geesey**. 2004. Composition and diversity of microbial communities recovered from surrogate minerals incubated in an acidic uranium-contaminated aquifer. Appl. Environ. Microbiol. **70**:6037-6046.

22. **Robertson, D. E., J. A. Chaplin, G. DeSantis, M. Podar, M. Madden, E. Chi, T. Richardson, A. Milan, M. Miller, D. P. Weiner, K. Wong, J. McQuaid, B. Farwell, L. A. Preston, X. Tan, M. A. Snead, M. Keller, E. Mathur, P. L. Kretz, M. J. Burk, and J. M. Short**. 2004. Exploring nitrilase sequence space for enantioselective catalysis. Appl. Environ. Microbiol. **70**:2429-2436.

23. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong,**

M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. **66**:2541-2547.

24. **Senko, J. M., J. D. Istok, J. M. Suflita, and L. R. Krumholz**. 2002. In-situ evidence for uranium immobilization and remobilization. Environ. Sci. Technol. **36**:1491-1496.

25. **Seow, K. T., G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, and J. Davies**. 1997. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. J. Bacteriol. **179**:7360-7368.

26. **Short, J. M.** 1997. Recombinant approaches for accessing biodiversity. Nat. Biotechnol. **15**:1322-1323.

27. Short, J. M. Protein activity screening of clones having DNA from uncultivated microorganisms. [5,958,672]. 9-28-1999.
Ref Type: Patent

28. Short, J. M. Gene expression library produced from DNA from uncultivated organisms and methods for making the same. [6,280,926]. 8-28-2001.
Ref Type: Patent

29. **Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse**. 1988. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Res. **16**:7583-7600.

30. **Stein, J. L., T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong**. 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. **178**:591-599.

31. **Thomassin-Lacroix, E. J., Z. Yu, M. Eriksson, K. J. Reimer, and W. W. Mohn**. 2001. DNA-based and culture-based characterization of a hydrocarbon-degrading consortium enriched from Arctic soil. Can. J. Microbiol. **47**:1107-1115.

32. **Treusch, A. H., A. Kletzin, G. Raddatz, T. Ochsenreiter, A. Quaiser, G. Meurer, S. C. Schuster, and C. Schleper**. 2004. Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. Environ. Microbiol. **6**:970-980.

33. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin**. 2005. Comparative metagenomics of microbial communities. Science **308**:554-557.

34. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield**. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**:37-43.

35. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith**. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**:66-74.

36. **Vrionis, H. A., R. T. Anderson, I. Ortiz-Bernad, K. R. O'Neill, C. T. Resch, A. D. Peacock, R. Dayvault, D. C. White, P. E. Long, and D. R. Lovley**. 2005. Microbiological and geochemical heterogeneity in an in situ uranium bioremediation field site. Appl. Environ. Microbiol. **71**:6308-6318.

37. **Yan, T., M. W. Fields, L. Wu, Y. Zu, J. M. Tiedje, and J. Zhou**. 2003. Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate- and uranium-contaminated groundwater. Environ. Microbiol. **5**:13-24.

38. **Zengler, K., G. Toledo, M. Rappe, J. Elkins, E. J. Mathur, J. M. Short, and M. Keller**. 2002. Cultivating the uncultured. Proc. Natl. Acad. Sci. U. S. A **99**:15681-15686.

Hazen, T. C., and H. H. Tabak. 2005. Developments in bioremediation of soils and sediments polluted with metals and radionuclides: 2. Field research on bioremediation of metals and radionuclides. Reviews in Environmental Science and Bio/Technology 4:157–183

Koenigsberg, S. S., T. C. Hazen, and A. D. Peacock. 2005. Environmental Biotechnology: a Bioremediation Perspective. Remediation Journal 15:5-25.

Haglund, A.L., Tornblom, E., Bostrom, B., and L. Tranvik. 2002. Large differences in the fraction of active bacteria in plankton, sediments, and biofilm. Microbial Ecology 43:232-241.

Teitzel G.M. and M.R. Parsek. 2003. Heavy metal resistance of biofilm and planktonic Pseudomonas aeruginosa. Appl. Environ. Microbiol. **69**:2313-2320.

Huber, T., Faulkner, G., and P. Hugenholtz. 2004. Bellerophon; a program to detect chimeric sequences in multiple sequence alignments, Bioinformatics **20:**2317-2319.

Schloss, P.D., and J. Handelsman. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71**:1501-1506.