

Metagenomic methods for the identification of active micro-organisms and genes in biodegradation processes

Don Cowan and William Stafford

Advanced Research Centre for Applied Microbiology
Department of Biotechnology, University of the Western Cape
Bellville 7535, Cape Town, South Africa

Address for correspondence:

Professor DA Cowan, Department of Biotechnology, University of the Western Cape, Bellville 7535, Cape Town, South Africa (dcowan@uwc.ac.za)

INTRODUCTION

Isolated microorganisms constitute only a minor fraction of the global microbial diversity, which may comprise millions of species (97). The huge discrepancy between species diversity as assessed by culture-dependent and culture-independent methods has led to the concept of ‘unculturables’, more recently revised as ‘uncultured’ (4,40,97). This change in terminology reflects the view that these organisms are probably not unculturable per se, but have merely not been cultured yet! Whatever the causes of the underlying technical difficulties in cultural studies (43,69,85), the appreciation of the uncultured majority has been largely responsible for the rapid development of metagenomic technologies.

The fundamental experimental questions for the analysis and understanding of microbial biodegradation processes are:

- Which organism(s) are responsible for primary degradation processes?
Organisms may be classified as Primary degraders (i.e., producers of the enzymes which undertake the early degradation steps, such as the hydroxylases and dioxygenases in phenol degradation pathways); Secondary degraders (i.e., organisms using the products of the primary degradation) and Tertiary metabolisers (organisms benefiting from the metabolic activities of the first two classes). Of course, a single species may be active simultaneously in all three classes.
- Which gene products (enzyme and pathway activities) are responsible for degradation processes?
- How can the two datasets of phylogeny and function be linked?

There are some cautionary guidelines that should be followed in the attempts to link phylogeny with function:

i. Identification of an organism/phylotype provides little useful information on in vivo function.

At best, the identification of a specific organism, either as an isolate or as a phylotype, provides information of the potential function of that organism; viz. the presence of *Bradyrhizobium* indicates the capacity for dinitrogen fixation, not that it occurs.

ii. A change in the population of any organism in response to a specific action (e.g., addition of a substrate) implies a putative role in the degradation process, but not necessarily as a primary degrader.

It is a reasonable assumption that if addition of a substrate to a microbial community in equilibrium induces a change in population structure, members of the community are affected

by the substrate. It is impossible to infer whether individual members are responsible for primary degradation or the effects are due to secondary degraders and tertiary metabolisers.

iii. The ability of an isolated organism to degrade a substrate cannot be taken as evidence that the organism is an important contributor to the degradation process in a natural community.

Ability to degrade a specific substrate is solid evidence that the organism possesses the appropriate metabolic capability (i.e., its genome harbours the relevant genes, and that these can be adequately expressed under some conditions). To extend this information to a direct implication of *in vivo* function requires information on the presence (and number) of the organism in the natural community, on the expression of the relevant genes *in vivo*, and on the relevant *in situ* activities.

iv. The presence of a specific gene or enzyme activity in a biodegrading community is not *a priori* evidence that either is an important component of the degradation process.

The presence of a specific gene (as determined by metagenomic PCR amplification, for example) provides no information on whether the gene is functionally expressed *in situ*. The detection of a specific enzymic activity is circumstantial evidence for its involvement in a process (e.g., dehalogenase activity in trichloroethylene degradation), but given the multiple possible pathways for degradation of most xenobiotics (see, for example (24)), its relevance *in situ* is difficult to ascertain.

v. Rapid up-regulation of gene expression or increase in measurable enzyme activity in response to addition of a substrate is circumstantial evidence for direct involvement in the degradation pathway.

Due to the rapid response of induction processes, effects which are immediately responsive to substrate additions (i.e., increased levels of specific mRNAs, proteins or enzyme activities) provide good circumstantial evidence that these genes/gene products are directly involved in the utilisation of the substrate. However, mRNA post-transcriptional regulation and protein post-translational modifications can influence the degradation activity.

The following discussion presents some of the techniques employed to address the questions above, and to demonstrate how they have been used to generate more explicit links between diversity and function.

CULTURE-DEPENDENT ENRICHMENT

Culturing techniques

The laboratory isolation of axenic cultures has proved invaluable for the study of degradative activity. The simplest and most commonly used approach to identify organisms involved in biodegradation is to isolate microbial strains capable of utilising the target substrate/xenobiotics/pollutant as a sole C/N source. Subsequent investigations of relevant enzyme pathways, degradation kinetics etc., are typically taken as evidence that the isolate plays a [significant] role in the appropriate *in vivo* biodegradation processes. We now accept that such assumptions are not necessarily soundly based. Examples where organisms that are normally assumed to be 'major players' constitute a very low fraction of the community (80) or where the key activities are linked to previously unknown species (55) all support the need for caution.

Enrichment strategies

In contrast with classical culturing that imposes a high degree of selection during isolation, culture-based enrichments have the potential to capture a larger biodiversity by allowing certain microbial interactions and by maintaining conditions similar to the natural environment. The use of microcosms that simulate the natural environment, or *in situ*

enrichment procedures both reduce negative impact on species abundance and distribution. For example, in the metagenomic screening for novel cellulose genes, a four-fold increase of cellulose genes in a small insert expression library was obtained by selective enrichment culturing on carboxymethylcellulose (79) and enrichment of soil micro-organisms with glycerol under anaerobic conditions prior to library screening yielded novel glycerol dehydratase and alcohol oxidoreductase enzymes (45) .

An alternative strategy for enrichment involves a continuous culture system consisting of encapsulated single cells in agar-microdroplets containing nutrient extracts prepared from the environmental samples. After sorting of microcolonies by flow cytometry, pure cultures could be established in rich medium (112). If the identified microcolonies cannot be cultured axenically, the small quantities of DNA can be amplified by multiple strand displacement with ϕ 29 polymerase (71) so that sufficient quantities are available for genomic analysis.

A relatively unexplored enrichment method uses colloidal chitin (108) bags or traps to recruit active microorganisms in situ (107). Studies of soil chitin degradation have successfully used bags of chitin buried in the soil followed by analysis of metagenomic 16S rRNA and chitinase genes (61)

While enrichment strategies are valuable in increasing the yields of target organisms and genes, such techniques cannot be reliably used to assess the contribution of microbial community members to the in situ degradation process.

COMMUNITY PROFILING

Denaturing Gradient Gel Electrophoresis

Phylogenetic determination of microbial community diversity is the simple analysis available. Common methods of community analysis using the ssu rRNA gene as a phylogenetic marker include sequencing of rRNA gene clone libraries, TRFLP, SSCP and DGGE/TGGE (reviewed (44,47)). Denaturing Gradient Gel Electrophoresis (DGGE) and Temperature Gradient Gel Electrophoresis (TGGE) allow rapid and comprehensive qualitative profiling of microbial community composition. DNA fragments are separated in a polyacrylamide gel according to their melting behaviour, which is sequence-dependent (111). Melting is achieved with a gradient of urea and formamide (DGGE) or temperature (TGGE). The use of GC-clamped primers increases the sensitivity of detection to >99% of sequence variations (65,89).

The identification of taxa mediating a biodegradation process by DGGE of 16S rRNA gene fragments may be useful for the rational design of further isolation attempts in order to access the target genomes. The aerobic biodegradation of haloacetic acids (HAAs) that are found in surface waters and in drinking water distribution systems was investigated using enrichment cultures containing either mono- or tri-chloroacetic acid as the sole carbon and energy source. Radiolabeled HAAs indicated that the ^{14}C was primarily converted to $^{14}\text{CO}_2$ with minor incorporation into cell biomass. The community structure of the enrichment cultures was analyzed by both classical isolation and DGGE of the PCR-amplified 16S rRNA gene fragments. Each of the two enrichment cultures had multiple bacterial populations, none of which corresponded to HAA degrading bacteria cultivated on HAA-supplemented agar plates (58). A similar study of squalene degrading denitrifying bacteria identified twelve dominant phylotypes, of which seven corresponded to α -Proteobacteria isolates utilizing squalene as the sole carbon source (12). The Bacterial and Archaeal taxa mediating biodegradation of oil

both in laboratory enrichment culture and in situ has been determined (54,81,82). Functional genes such as ammonia monoxygenase (*amoA*) (67,76), a key enzyme in the chemolithoautotrophic oxidation of ammonia to nitrogen gas, and multicomponent phenol hydroxylase (*LmPh*), mediating oxidation of phenol, (105) have been analysed by DGGE. This can be informative in dissecting community function, but many functional genes have been subjected to lateral gene transfer and therefore are poor phylogenetic markers to identify taxa.

Community profiling techniques generate little reliable quantitative data due to the bias of PCR amplification (95,96). Furthermore, in complex metagenomic samples, DGGE only reveals populations that account for more than 0.1 % of the total community (31) and different DNA fragments with the same melting profile may co-migrate. Another limitation of the DGGE technique is the prerequisite for a single melting domain and therefore the short DNA fragment length (200-500 bp) that limits phylogenetic information. Computer programs are useful in assessing the DNA melting behaviour of a region of DNA of interest (MELT94 and Primo Melt).

Fluorescence in Situ Hybridization (FISH) and Microautoradiography (MAR)

Fluorescence in Situ Hybridization (FISH) uses microscopy to detect cellular nucleic acid sequences hybridized with fluorescently labeled probes. This technique has the considerable strength that it can be applied directly to samples taken from natural environments (e.g., such as cells recovered by filtration, or dissociated from soil particles and recovered by differential centrifugation). The primary limitation of FISH lies in the fact that the technique is restricted to the simultaneous identification of very few phylotypes, for which some phylotypic sequence data must already exist.

16S rRNA is the molecule most commonly targeted in FISH analyses: oligonucleotide probes for each taxonomic level can be designed and tested using bioinformatics tools and experimental hybridization to target and non-target type strains (2,4,50). The critical experimental parameters are specificity, sensitivity, ease of sample penetration and accessibility of the probe to rRNA targets. A typical oligonucleotide probe is a 15-30 bases with the fluorescent dye molecules incorporated into the oligonucleotide during synthesis or post synthesis by chemical or enzymatic coupling (63). Commonly used dyes are fluorescein-derivates (Fluorescein–Isothiocyanate (FITC), 5-(6-)carboxyfluorescein–N-hydroxysuccimide-ester (FluoX)) and rhodamine-derivates (Tetramethyl–Rhodamine–Isothiocyanate (TRITC), Texas Red), and cyanine dyes such as Cy3 and Cy5. Fluorochromes with different excitation and emission maxima allow simultaneous detection of two or more micro-organisms. The sensitivity of FISH can be greatly increased using probes labeled with several fluorochrome molecules (23) and by enzymatic signal amplification using a digoxigenin- or biotin- labeled oligonucleotide coupled to an enzyme catalyzing fluorescent substrates. Tyramide Signal Amplification (TSA) can further increase the sensitivity 10–20-fold (87), but the number of positive cells may be reduced compared to mono-labeled probes, perhaps due to insufficient penetration of the high molecular weight probe into the cells. Optimizing the fixation conditions and washing stringency can improve problems of poor cell penetration and specificity (3). However, a promising approach is the use of peptide nucleic acid (PNA) probes that contain an uncharged polyamide backbone (19) and allow increased cell penetration with specific, sensitive detection (75). Poor accessibility of the probe to target rRNA molecules due to higher-order structure may be overcome with ‘helper’ oligonucleotide probes that bind adjacent to the labeled detection probe (27). FISH signals

can also be recorded using flow cytometry, coupling automated, quantitative analysis with a unique potential for sorting bacteria in suspensions (93).

The use of dual oligonucleotide probes labeled with different dyes (Dual-colour FISH) offers scope to simultaneously identify both a target organism and a target gene, and therefore to ask specific questions about the genetic content of a target species. This technique does not, however, indicate whether a gene is expressed, but may be extended by the use of antisense oligonucleotides designed to identify mRNAs(74). The latter approach thus links an organism with expression of a target gene, although it falls short of being definitive proof that a specific gene product (e.g., a biodegradative enzyme) is functional in the microbial community.

Finally, microautoradiography (MAR) can be used to visualize metabolic activity and, when combined with FISH, can be used to definitively identify taxa carrying out a particular biodegradation. A sample is incubated with radiolabeled substrates and, after analysis by FISH, radioactivity can be visualized by application and development of an autoradiographic emulsion. Using MAR-FISH, radiolabeled substrate uptake was shown to be confined to certain bacterial species and could be monitored within activated sludge under aerobic versus anaerobic (46).

IDENTIFYING GENES AND GENOMES MEDIATING BIODEGRADATIONS

Stable isotope probing (SIP) and 5-Bromo-2-deoxyuridine labeling

Stable-isotope probing (SIP) uses a stable isotope (such as ^{13}C -, ^{18}O - or ^{15}N) labelled substrate to enrich for the genomes responsible for substrate biotransformation. In pioneering studies, $^{13}\text{C}\text{H}_4$ and $^{13}\text{C}\text{H}_3\text{OH}$ were fed to forest soil microcosms and purified ^{13}C -labelled DNA was separated from unlabelled DNA by buoyant density gradient centrifugation. PCR amplification of the 16S rRNA and methanol dehydrogenase (mxoF) genes identified methylotrophic Acidobacterium species with novel mxoF gene variants (77,78). Several other stable isotope-labelled substrates have been used, including $^{13}\text{C}\text{O}_2$ to identify ammonia oxidizing bacteria and [^{13}C]-toluene (73), [^{13}C]-phenol (55) and [^{13}C]-naphthalene (42) to identify taxa responsible for pollutant biodegradation. The SIP study using [^{13}C]-naphthalene identified Polaromonas vacuolata as the major naphthalene-metabolizing bacterium in the environmental sample. The phylogenetic results were confirmed by subsequent isolation of the organism and identification of dioxygenase gene homologues involved in naphthalene degradation (42).

The success of DNA-SIP is dependent on the level of isotopic enrichment achieved during isotope feeding. Since RNA synthesis can occur without DNA replication RNA has a high turnover, RNA-SIP can be a better marker to detect active community members. For example, ^{13}C -labelled phenol was fed to an industrial phenol degrading bioreactor and the microbial community monitored by RT-PCR and DGGE of the isotopically enriched fraction. 16S rRNA gene sequence analysis demonstrated that phenol degradation was dominated by a member of the genus Thauera, a group previously unknown as phenol degraders (55).

Limitations of these methods include dilution, cross-feeding and recycling of the label within the community, resulting in loss of specific enrichment. The isotopic labeling should be $>20\text{atom}\%$ for RNA-SIP and $>50\text{atom}\%$ for DNA-SIP so that definitive separation from unlabelled DNA/RNA can be achieved by buoyant density gradient and effects of cross feeding minimized (77). Cross-feeding is further reduced by shortening the labelling pulse

time. The community profile across all density gradient fractions can be observed by DGGE to determine isotopic enrichment of community members (55). Other SIP based approaches have been used to link microbial phylogeny with community function. Particular phylogenetic groups of isotope labelled rRNA can be affinity isolated by the hybridisation with oligonucleotide probes covalently bound to streptavidin magnetic beads, followed by isotope ratio mass spectrometry (IRMS) (53,72)). However, the wide application of SIP is limited by the commercial availability of complex labelled compounds, that otherwise may require expensive custom synthesis.

In a recent comprehensive study of anaerobic methanol utilisation, SIP and full cycle rRNA analysis with FISH-MAR was used to study the microbial consortium in a ^{13}C -methanol-fed batch reactor containing nitrate as the electron acceptor (32). 16S rRNA clone libraries identified Methylophilales as the dominant members of the isotopically labelled fraction. Oligonucleotide probes targeting the Methylophilales taxa were designed and the application of FISH-MAR established that they were the dominant members of ^{13}C - methanol uptake(32).

In contrast to DNA –SIP and RNA-SIP, methods that exploit the natural abundance of isotopes in the substrate of choice avoid the requirement for isotopic enrichment. For example, the natural low abundance of $^{13}\text{CH}_4$ enabled the identification of methanotrophs in marine sediments by secondary ion mass spectrometry (SIMS) after the microorganisms had been identified by FISH (70).

Actively growing micro-organisms can also be labelled with the thymine analogue, 5-bromo-2-deoxyuridine (BrdU), and the labelled DNA or RNA separated by immunocapture or density gradient centrifugation (101). The commercial availability of anti-BrdU antibodies makes immunocapture an effective method with a high degree of specificity. Addition of substrates with BrdU selects among the members of the microbial community for enhanced growth on the specific substrate. This approach has been successfully applied to identify bacterial community members responding to soil phosphate addition (13) and identify a *Bacillus cereus* strain (VA1) associated with arbuscular mycorrhizae (5). The method assumes that there are universal uptake mechanisms for BrdU. This premise that is largely unfounded since several bacteria have been shown to be unable to incorporate BrdU(101).

Suppressive Subtraction Hybridisation (SSH) and Differential Expression Analysis (DEA)

Suppressive Subtraction Hybridisation (SSH) identifies genetic differences between microorganisms and is therefore a powerful technique for gene enrichment. Samples of DNA or cDNA from the organisms to be compared are divided in two fractions and a different adaptor is ligated to each fraction (driver and tester). An excess of driver DNA is denatured and hybridized with the tester DNA pool (this subtractive hybridization may be carried out several times). The result is a mixture of single and double-stranded products. The adaptors are only partially complimentary so the only DNA fragments possessing both linkers will be subsequently amplified by PCR. The specificity of this approach is usually confirmed by labelling of the tester DNA/cDNA and carrying out Southern Analysis with candidate gene fragments. This approach has been used to discover the genetic elements contributing to pathogenesis between two closely related bacteria (6), but has only recently been used to identify differences in metagenomic samples (30).

Several differential expression technologies that target transcriptional differences in gene expression have been developed (reviewed in (34)). Representational difference analysis is a method analogous to SSH and is a particularly effective enrichment tool (7). This type of approach was successfully applied to identify bacterial genes up-regulated in the absence of iron (14). Both SSH and DEA are sensitive and selective approaches; by applying multiple rounds of subtraction, small differences in expression of single copy genes can be detected (1,52). These methods offer powerful opportunities to study the biodegradation or bioremediation of an environmental pollutant. A potential in situ limitation is the need for a suitable reference metagenome prior to pollutant impact so that the genetic differences identified by SSH or DEA are relevant to the pollutant of interest. The use of defined microcosms can clearly resolve some of these limitations.

Gene specific PCR

Sequence-dependent approaches to identify genes are largely limited by an a priori knowledge of gene(s) mediating the biodegradation processes, but have proven invaluable for an understanding the key enzymes in these metabolic pathways. Gene-specific PCR of metagenomic samples has been widely used to identify microorganisms with specific biodegradative and/or metabolic capacities in enrichment cultures or environmental samples. For example, the biodegradative potential of microbial communities has been probed by screening metagenomic DNA extracts for the presence of catechol 2,3-dioxygenase, chlorocatechol dioxygenase and phenol hydroxylase genes (28,60,105). Similarly, the PCR-dependent targeting of methane monooxygenase, methanol dehydrogenase and ammonia monooxygenase genes has been used to identify methanotrophic chemolithotrophic ammonium-oxidizing (39,57), denitrifying (15,37) and polyhydroxyalkanoate producing bacteria (90). In each example, identification was based on the application of gene-specific consensus primers, a process that incorporates both substantial strengths and weaknesses.

Primer Design

The use of consensus primers for the identification of genes in metagenomic samples offers a fast and efficient approach to the analysis of gene diversity. Assisted by powerful and user-friendly alignment and primer-prediction software packages (e.g., DNAMAN; DNA Star, Amplicon) consensus primer design is a routine process. There are also a number of widely applied 'rules of thumb' in designing degenerate consensus primers:-

- Primers should preferably be in the 18mer to 30mer range.
- Primers should be designed to maximise complementarity at the 3' end.
- Where one or two mismatches occur, degenerate bases are included. For more than three different bases in a specific position, an inosine residue is included.
- No more than 25% total degeneracy and less than 10% of inosine residues are allowed per primer.
- Primers are designed with annealing temperatures of between 50°C and 60°C.
- The quality of the subsequent phylogenetic analysis is roughly proportional to the length of the amplified region.

The design of primers from amino acid sequence motifs that are highly conserved between members of a protein family have proven to be highly effective in the identification and characterization of distantly related family members. The CODEHOP (Consensus-Degenerate Hybrid Oligonucleotide Primer) program designs a pool of primers containing all possible 11- or 12-mers for the 3' degenerate core region and having the most probable nucleotide predicted for each position in the 5' non-degenerate clamp region (84). This approach limits primer degeneracy while maintaining primer specificity.

Primer design can be a highly flexible tool in the analysis of different gene hierarchies. For example, bacterial multicomponent phenol hydroxylases, which are key components of the degradation pathways of many aromatic pollutants, form three separate phylogenetic groupings, corresponding closely to the groups established on the basis of kinetic analyses (28). Primer sets designed for the amplification of all three groups ('Universal' primer set) and for each of the individual groups were used to demonstrate population changes in mixed cultures in response to different phenol feeding regimes (28).

It is very important to experimentally confirm primer specificity. The inclusion of one or more positive control amplifications (i.e., genomic DNA from an organism known to harbour the relevant gene/express the relevant activity, or a recombinant plasmid harbouring an example of the target gene) that yields an amplicon band of the predicted size is a minimum requirement.

Classical PCR

Two-primer approaches, (i.e., where both primers are specific to internal regions of the target gene) are most commonly employed in assessments of metagenomic gene diversity e.g., (28,60,103,106). The products of such amplifications are generally of a predictable size, although some PCR products are subject to variations due to the presence of insertions or deletions within the sequence targeted by the primer pair. Typically, the heterogeneous amplicon DNA is extracted, purified and cloned into a standard vector system. Quantitative data may be obtained by competitive PCR (60) or by Real-Time PCR (33). For more in depth diversity analysis, a number of single colonies from the transformation are selected and the inserts re-amplified (or excised from bulk plasmid preparations after cultivation of the clones). While gene diversity could be assessed immediately by sequencing each insert, it is more common to reduce the sequencing burden by RFLP analysis (e.g., by ARDRA, TRFLPs etc). The choice of the number and type of restriction enzymes is important in determining the quality of the RFLP clade structure. Since a single restriction site constitutes at best a few percent of any amplified sequence, multiple cut sites (using several enzymes and/or frequent cutters) enhance the ability to discriminate between non-identical sequences (64). Confidence in the quality of the clade structure is enhanced by sequencing more than example from each clade.

As a tool for both gene identification and gene discovery, gene-specific PCR of metagenomic samples has two major drawbacks. Firstly, the design of primers is dependent on known sequence information and strongly skews the results in favour of known sequences. For instance, functionally similar genes resulting from convergent evolution will almost certainly not be detected by a single gene-family-specific set of PCR primers. Secondly, only a fragment of a structural gene will typically be amplified by gene-specific PCR. This is, of course, of less significance where gene identity and diversity is the primary objective. However, where access to full-length genes is required, as in 'gene mining' (10), additional experimental steps are necessary. The third limitation of gene-specific PCR is that it provides neither direct information on which organisms harbour the identified genes, nor on the level of gene expression. Both these issues are important in the context of biodegradation process where, for example, knowledge of which genes are actively involved in the process and the organisms in whose genome they reside may be critical for any knowledge-based (i.e., non-empirical) control of the biodegradation process.

The targeting of genes by PCR does not address if these genes are actually expressed. The majority of control in prokaryotes is thought to occur at the transcriptional level (17) so that the presence of a mRNA is strong evidence for the expression of that gene product. Furthermore, mRNA has a high turnover compared to DNA so that the detection of a mRNA target can provide additional evidence of microorganism activity. The use of mRNA as a template (RT-PCR) is therefore a distinct advantage (109), that can yield additional insights such as the diversity of actively transcribed naphthalene dioxygenase genes in microorganisms catabolizing naphthalene at a coal tar waste-contaminated site (109). The experimental difficulties in working with RNA have limited the widespread use of this approach, although several advances have been made in this area (26,35,41).

Single-primer gene targeting

Methods for amplifying genes requiring only one gene-specific primer impose less sequence-dependent bias compared to standard two-primer PCR amplification procedures. These PCR based strategies have been used for the recovery of the up- or down-stream regions flanking a single PCR primer for the recovery of full-length genes. For example, universal fast walking (62,66) inverse PCR and adaptor ligation PCR (68) have all been employed successfully to access full length genes in metagenomic extracts. Related techniques such as panhandle PCR and random primed PCR, which are widely used for gene cloning (48,59) have not yet been applied to the metagenome. It is important to note, however, that these single primer methods are technically more difficult to apply to metagenomic DNA preparations than to single genome DNA samples due to the hugely increased heterogeneity of the former.

Affinity capture

A new and elegant method for gene recovery is the use of immobilised oligonucleotides (94) designed to target a specific gene fragment or consensus sequence by affinity binding. For example, an oligonucleotide probe containing a poly-d(GGGT) tail is hybridized to RNA in solution. Simultaneously, an aliquot of oligo-dT paramagnetic beads is hybridized to an oligonucleotide made of poly-d(CCCA) with a poly-dA tail. The two solutions are combined and a high-affinity GCAT complex is formed. The magnetic beads are captured and differential melting of all three hybrids permits the release gene target fragment while leaving the majority of the oligonucleotides bound to the beads. Authenticity of the captured product is determined by reverse-transcriptase (RT)-PCR on a sub-sample. Although designed for recovering taxon-specific rRNA sequences for isotopic analysis (72), the technique is potentially adaptable to any target RNA or DNA sequence by appropriate probe design.

Microarrays

Microarray technologies provide a powerful high throughput tool for the study of biological processes, but have only recently been applied to the field of metagenomics (88). Depending on the size and origin of the probes arrayed, microarrays are grouped as follows (see Table 1). The majority of environmental studies (for example, (51,92)) have focussed on analysis of in situ diversity using phylogenetic oligonucleotide arrays. However, functional gene arrays may be exploited for the identification of either the presence (by DNA-DNA hybridisation) or the expression (by DNA-mRNA/cDNA hybridisation) of specific functional genes in a metagenome. In the most comprehensive study of this type reported to date, Rhee et al (80) prepared oligonucleotide arrays containing 1661 50-mer probes specific for known genes involved in aromatic and alkane degradation pathways and in heavy metal resistance. These arrays were effective in showing changes in population distribution in soils amended with naphthalene and during enrichment incubation. Surprisingly, microarray data showed that *Ralstonia*, *Comamonas* and *Rastonia*, rather than *Pseudomonas* naphthalene-degrading genes

were prevalent in soils, a result at odds with the common perception of *Pseudomonas* as a primary degrader of polyaromatics in soil systems (36). The authors acknowledge that the detection limit (about 107 cells in the presence of background RNA) would need to be reduced substantially in order to detect rare genomes.

METAGENOMIC GENE LIBRARIES

Metagenomic DNA libraries can theoretically provide access to the entire metagenomic sequence space (38). While this technology has been developed primarily as a tool for the discovery (and commercialisation) of novel genes (21) it has considerable potential for contributing to the analysis of environmental processes. For example, a combination of a suitable labelling method (such as SIP) or techniques of subtraction (SSH/DEA or subtractive cDNA libraries) with metagenomic library construction and screening could potentially provide datasets of multiple genes involved in specific biodegradation processes. A prerequisite for the construction of metagenomic libraries is the efficient and non-biased extraction of high-quality DNA from environmental samples (see this volume, Reisenfeld et al.). The minimum number of clones (N) that need to be screened in order to find a gene of interest in the metagenomic library (18) increases with the presence of eukaryotic DNA (eukaryotic genome is 3-140 000 Mbp compared to prokaryotic organisms of 0.6 - 9.5 Mbp (102)). This is exacerbated by the presence of non-coding introns that reduce valuable sequence information and the ability to express functional gene products when screening expression libraries. Size selection may be particularly useful preliminary step for reducing the total metagenome size, avoiding a high load of 'junk' DNA, or merely to focus on a specific metagenome fraction (e.g., eukaryotic, prokaryotic or viral). A crude separation of eukaryotic and prokaryotic cells can be simply carried out by filtration (104). Alternatively, differential and density gradient centrifugation can be used. For example, differential centrifugation has been used to enrich *Buchnera aphidicola* and *Cenarchaeum symbiosum* symbionts from their hosts (86) and to prepare community viruses in preparation for whole genome sequencing (16,110). Full representation of the metagenome in a library is also limited the diversity of the sample (perhaps several thousand species) (11,22) and by the unequal distribution of species. This may be partially resolved by means of experimental normalization. Fragmented genomic DNA is heat denatured and allowed to re-annealing under stringent conditions (e.g., 68°C for 12-36 hours). Abundant ssDNA will more rapidly anneal to generate double-stranded nucleic acids than rare DNA, and separated from the double-stranded nucleic acids enabling an enrichment of rarer sequences within the environmental sample (25,91). Metagenomic cloning often assumes that *E.coli* is globally suitable as a host for all environmental DNA fragments. For DNA libraries low copy number BAC or fosmid systems offer the best means for stable library propagation and maintenance (8). For expression libraries, systems allowing expression from native promoters offer the best chance for recovery of heterologously expressed genes. Although the *E. coli* transcriptional machinery is known to be relatively promiscuous in recognising foreign promoters, a bias in favour of Firmicutes genes has been noted (29). The development of *Streptomyces*- shuttle systems and *Pseudomonas* hosts offer new possibilities in this regard (20,56).

Depending upon the insert size, 106 - 108 library clones will be required in order to represent the collective genomes of the thousands of different species that are typically present in an environmental sample (49). Although metagenomic sequencing approaches have revealed functional insights into communities and their biotransformations (9,83,98,99,104), the enormous diversity and current limitations on library screening and gene annotation limits

this approach. For example, the sequencing of 1.045 Gbp from the Sargasso Sea metagenome identified 1.2 million putative genes, but the diversity of species (approx. 1800) and their unequal distribution enabled only one complete genome to be assembled (104). Due to limitations in gene assignment and database homology, relatively little functional information was gained although many new rhodopsin-like photoreceptors were identified; indicating an important role of bacterial photoautotrophy in marine populations (8,9,104). The acquisition of very large volumes of metagenomic sequence from different environments has opened the prospect of a new field of comparative metagenomics. A recent study demonstrated that data from shotgun sequencing of metagenomic small-insert libraries can be used to generate environmental fingerprints (98). The core data for these fingerprints are Environmental Genome Tags (EGTs), short individual sequence reads yielding putative gene identities. A comparison of gene distribution across eight different metagenomic libraries revealed substantial, predictable system differences; a predominance of photoautotrophy in Sargasso Sea samples versus carbohydrate metabolism in soils.

Technologies for screening expression libraries are have been reviewed elsewhere (38). However, a high throughput substrate-induced gene expression screening (SIGEX) method has recently been developed that is particularly relevant to the identification of catabolic genes. This technique uses an operon-trap gfp-expression vector to capture catabolic operon DNA fragments relevant to the target compounds. An environmental metagenome library (152,000 clones with an average insertion size of 7 kb) was constructed from groundwater and screened with benzoate or naphthalene substrates. Non-recombinants were selected against by cell sorting after IPTG induction since the vector used, a puC18 derivative, contains lacZ in the multiple cloning site to allow for λ -complementation. The screening was accomplished within 4 days and yielded 62 positive clones. Many of the identified ORFs were found to be homologs of genes in known benzoate-degradative and catechol-degradative operons (100).

In summary, metagenomic techniques have begun to address both the identity of active organisms and their gene products that mediate biodegradation. Advances in isotopic labeling, fluorescent microscopy and gene targeting has enabled a directed approach to uncovering some of the active microorganisms carrying out degradative processes in complex microbial communities. The application of high-throughput methods such as microarrays and comparative genomic sequencing offers the opportunity to uncover novel pathways and to study complex microbial populations. A combination of approaches will prove the most valuable in the linking of phylogeny with function. Ultimately, a clearer picture of the identity, activity, distribution, and abundance of active organisms will lead us to a more complete understanding of biodegradation processes.

Acknowledgments

The authors gratefully acknowledge the National Research Foundation HBU program and the University of the Western Cape for financial support

Table 1. Classification of microarrays used in environmental studies (80)

Class of microarray	Source of probe	Application	Reference
Functional gene	PCR-amplified DNA and cDNA fragments Oligonucleotides (gene specific)	Physiological status and functional activities of microbial communities	(114)
Community genome	Whole genomic DNA from pure cultures	Identification of individual species in microbial communities	(113)
Phylogenetic oligonucleotide	Oligonucleotides (ssu rRNA gene)	Phylogenetic analyses of microbial communities	(50,51,92)

Reference List

1. Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, Lukyanov SA, Sverdlov ED, and Berg DE. 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. **Proc Natl Acad Sci U S A** 95:13108-13113.
2. Amann, R. I., B. J. Binder, O. R. J. , S. W. Chisholm, R. Devereux, and D. A. Stahl. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. **Appl Environ Microbiol** 56:1919-1925.
3. Amann, R. I., L. Krumholz, and D. A. Stahl. 1990. Fluorescent oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. **J Bacteriol** 172:762-770.
4. Amman, R. L., W. Ludwig, and K. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. **Microbiol. Rev.** 59:143-169.
5. Artursson, V. and J. K. Jansson. 2003. Use of Bromodeoxyuridine Immunocapture To Identify Active Bacteria Associated with Arbuscular Mycorrhizal Hyphae. **Applied and Environmental Microbiology** 69:6208-6215.
6. Bart, A., J. Dankert, and A. van der Ende. 2000. Representational difference analysis of *Neisseria meningitidis* identifies sequences that are specific for the hyper-virulent lineage III clone. **FEMS Microbiol. Lett.** 188:111-114.
7. Becker, P., W. Hufnagle, G. Peters, and M. Herrmann. 2001. Detection of Differential Gene Expression in Biofilm-Forming versus Planktonic Populations of *Staphylococcus*

aureus Using Micro-Representational-Difference Analysis. **Applied and Environmental Microbiology** 67:2958-2965.

8. Beja, O. 2004. To BAC or not to BAC: marine ecogenomics. **Curr. Opin. Biotech.** 15:187-190.
9. Beja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. **Environ. Microbiol.** 2:516-529.
10. Bell, P., A. Sunna, M. D. Gibbs, N. C. Curach, H. Nevalainen, and P. L. Bergquist. 2002. Prospecting for novel lipase genes using PCR. **Microbiology.** 148:2283-91. 2283-2291.
11. Bohannan, B. J. M. and J. Hughes. 2003. New approaches to analyzing microbial biodiversity data. **Curr. Opin. Microbiol.** 6:282-287.
12. Bonin P.C., Michotey V.D., Mouzdahir A., and Rontani J.-F. 2002. Anaerobic biodegradation of squalene: Using DGGE to monitor the isolation of denitrifying Bacteria taken from enrichment cultures. **FEMS Microbiology Ecology**, October 2002, 42:37-49.
13. Borneman, J. 1999. Culture-Independent Identification of Microorganisms That Respond to Specified Stimuli. **Applied and Environmental Microbiology** 65:3398-3400.
14. Bowler, L. D., M. Hubank, and B. G. Spratt. 1999. Representational difference analysis of cDNA for the detection of differential gene expression in bacteria: development using a model of iron-regulated gene expression in *Neisseria meningitidis*. **Microbiology** 145:3529-3537.
15. Braker, G., A. Fesefeldt, and K. Witzel. 1998. Development of PCR primer systems for amplification of nitrite reductase genes (*nirk* and *nirs*) to detect denitrifying bacteria in environmental samples. **Appl. Environ. Microbiol.** 64:3769-3775.
16. Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. **Proceedings of the National Academy of Sciences** 99:14250-14255.
17. Cases I. and V. de Lorenzo. 2005. Promoters in the Environment: Transcription Regulation in its natural context. **Nature Microbiology Reviews** 3:105-118.
18. Clarke, L. and J. Carbon. 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. **Cell** 9:91-100.
19. Corey, D. R. 1997. Peptide nucleic acids: expanding the scope of nucleic acid recognition. **Tibtech** 15. 15:224-229.
20. Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helyncx, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, and J. L. Pernodet. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. **Appl. Environ. Microbiol.** 69:49-55.
21. Cowan, D. A. 2000. Microbial genomes - the untapped resource. **Trends Biotechnol.** 18:14-16.
22. Curtis, T. P., W. T. Sloan, and J. W. Scannell. 2002. Estimation prokaryotic diversity and its limits. **PNAS** 99:10494-10499.
23. DeLong, E. F., L. T. Taylor, T. L. Marsh, and C. M. Preston. 1999. Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent in situ hybridization. **Appl. Environ. Microbiol.** 65, pp. 5554-5563..
24. Diaz, E. 2004. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. **Intl. Microbiol.** 7:173-180.
25. Eastwood, R. F., E. S. Lagudah, and R. Appels. 1994. A directed search for DNA sequences tightly linked to cereal cyst nematode resistance in *Triticum tauschii*. **Genome** 311-319.

26. Frischer, M. E., J. M. Danforth, M. A. Newton Healy, and F. M. Saunders. 2000. Whole-cell versus total RNA extraction for analysis of microbial community structure with 16S rRNA-targeted oligonucleotide probes in salt marsh sediments. **Appl. Environ. Microbiol.** 66:3037-3043.
27. Fuchs, B. M., G. Wallner, W. Beisker, I. Schwippl, W. Ludwig, and R. Amann. 1998. Flow cytometric analysis of the in situ accessibility of Escherichia coli 16S rRNA for fluorescently labeled oligonucleotide probes. **Appl. Environ. Microbiol.** 64:4973-4982.
28. Futamata, H., S. Harayama, and K. Watanabe. 2001. Group-specific monitoring of phenol hydroxylase genes for a functional assessment of phenol-stimulated trichloroethylene bioremediation. **Appl. Environ. Microbiol.** 67:4671-4677.
29. Gabor, E. M., Alkema W.B.L., and Janssen D.B. 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. **Environ. Microbiol.** 6:879-886.
30. Galbraith, E. A., D. A. Antonopoulos, and B. A. White. 2004. Suppressive subtractive hybridisation as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. **Environ. Microbiol.** 6:928-937.
31. Gelsomino, A., A. C. Keyzer-Wolters, G. Cacco, and J. D. van Elsas. 1999. Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electro-phoresis . **J. Microbiol. Methods** 38:15.
32. Ginige, M. P., P. Hugenholtz, H. Daims, M. Wagner, J. Keller, and L. L. Blackall. 2004. Use of Stable-Isotope Probing, Full-Cycle rRNA Analysis, and Fluorescence In Situ Hybridization-Microautoradiography To Study a Methanol-Fed Denitrifying Microbial Community. **Applied and Environmental Microbiology** 70:588-596.
33. Ginzinger, D. G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. **Experimental Hematology** 30:503-512.
34. Green, C. D., J. F. Simons, B. E. Taillon, and D. A. Lewin. 2001. Open systems: panoramic views of gene expression. **J. Immun. Methods** 250:67-79.
35. Griffiths, R. I., A. S. Whiteley, A. G. O'Donnell, and M. J. Bailey. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-and rRNA-based microbial community composition. **Appl. Environ. Microbiol.** 66:5488-5491.
36. Habe, H. and Omori T. 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. **Biosci. Biotechnol. Biochem.** 67:225-243.
37. Hallin, S. and P. Lindgren. 1999. PCR detection of genes encoding nitrite reductase in denitrifying bacteria. **Appl. Environ. Microbiol.** 65:1652-1657.
38. Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chem. Biol.** 5:R245-R249.
39. Henckel, T., U. Jackel, S. Schnell, and R. Conrad. 2000. Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. **Appl. Environ. Microbiol.** 66:1801-1808.
40. Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. **J. Bacteriol** 180:4765-4774.
41. Hurt, R. A. 2001. Simultaneous recovery of RNA and DNA from soils and sediments. **Appl. Environ. Microbiol.** 67:4495-4503.
42. Jeon, C. O., W. Park, P. Padmanabhan, C. DeRito, J. R. Snape, and E. L. Madsen. 2003. Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. **Proceedings of the National Academy of Sciences** 100:13591-13596.

43. Kaerberlein T, Lewis K, Epstein SS. 2005. Isolating "uncultivable" microorganisms in pure culture using a simulated natural environment. **Science** 296(5570):1127-9
44. Kirk, J. L., L. A. Beaudette, M. Hart, P. Moutoglis, J. N. Klironomos, H. Lee, and J. T. Trevors. 2004. Methods of studying soil microbial diversity. **Journal of Microbiological Methods** 58:169-188.
45. Knietsch, A., T. Waschkowitz, S. Bowien, A. Henne, and R. Daniel. 2003. Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. **J. Mol. Microbiol. Biotechnol.** 5:46-56.
46. Lee, N., P. H. Nielsen, K. H. Andreasen, S. Juretschko, J. L. Nielsen, K.-H. Schleifer, and M. Wagner. 1999. Combination of fluorescent in situ hybridization and microautoradiography - a new tool for structure-function analyses in microbial ecology. 7. **Appl. Environ. Microbiol.** 65:1289-1297.
47. Liu, W.-T. and D. A. Stahl. 2005. Manual of Environmental Microbiology. **Manual of Environmental Microbiology**, ASM Press, Washington, DC. 2nd edition:114-134.
48. Liu, Y. and R. F. Whittier. 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from pi and yac clones for chromosome walking. **Genomics** 25:674-681.
49. Lorenz, P. and Schleper C. 2002. Metagenome - a challenging source of enzyme discovery. **J. of Mol. Catalysis B: Enzymatic**, 13-19.
50. Loy, A. and M. W. M. Horn. 2003. probeBase - an online resource for rRNA-targeted oligonucleotide probes. **Nucleic Acids Res.** 31:514-516.
51. Loy, A. A., N. Lehner, J. Lee, H. Adamczyk, J. Meier, H. Schleifer, and M. Wagner. 2002. Oligonucleotide microarray for 16S rRNA-based detections of all recognized lineages of sulphate-reducing prokaryotes in the environment. **Appl. Environ. Microbiol.** 68:5064-5081.
52. Lukyanov SA, Gurskaya NG, Lukyanov KA, Tarabykin VS, and Sverdlov ED. 1994. Highly efficient subtractive hybridisation of cDNA. **Bioorg. Khim.** (Russ.). 20:701-704.
53. MacGregor BJ, Brüchert V, Fleischer S, and Amann R. 2002. Isolation of small-subunit rRNA for stable-isotopic characterization. **Environ Microbiol** 2002, 4:451-464.
54. MacNaughton SJ, Stephen JR, Venosa AD, Davis GA, Chang YJ, and White DC. 1999. Microbial population changes during bioremediation of an experimental oil spill. **Appl Environ Microbiol.** 65:3566-3574.
55. Manefield, M., A. S. Whiteley, R. I. Griffiths, and M. J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. **Appl. Environ. Microbiol.** 68:5367-5373.
56. Martinez, A., S. J. Kolvek, C. L. T. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, and M. S. Osburne. 2004. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. **Applied and Environmental Microbiology** 70:2452-2463.
57. McDonald, I. R., E. M. Kenna, and J. C. Murrell. 1995. Detection of methanotrophic bacteria in environmental samples with the PCR. **Appl. Environ. Microbiol.** 61:116-121.
58. McRae, B. M., T. M. LaPara, and R. M. Hozalski. 2004. Biodegradation of Haloacetic Acids by Bacterial Enrichment Cultures. **Chemosphere** 55:915-925., p. 915-925.
59. Magonigal, M. D., E. F. Rappaport, R. B. Wilson, D. H. Jones, J. A. Whitlock, J. A. Ortega, D. J. Slater, P. C. Nowell, and C. A. Felix. 2000. Panhandle PCR for cDNA: a rapid method for isolation of MLL fusion transcripts involving unknown partner genes. **PNAS** 97:9597-9602.

60. Mesarch, M. B., C. H. Nakatsu, and L. Nies. 2000. Development of catechol 2,3-dioxygenase-specific primers for monitoring bioremediation by competitive quantitative PCR. **Appl. Environ. Microbiol.** 66:678-683.
61. Metcalfe, A. C., M. Krsek, G. W. Gooday, J. I. Prosser, and E. M. H. Wellington. 2002. Molecular Analysis of a Bacterial Chitinolytic Community in an Upland Pasture. **Applied and Environmental Microbiology** 68:5042-5050.
62. Mishra, R. N., S. L. Singla-Pareek, S. Nair, S. K. Sopory, and M. K. Reddy. 2002. Directional genome walking using PCR. **Biotechniques** 33:830-834.
63. Moter, A. and U. B. Global. 2000. Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. **Journal of Microbiological Methods** 41:85-112.
64. Moyer, C. L., J. M. Tiedje, F. C. Dobbs, and D. M. Karl. 1996. A computer-simulated restriction fragment length polymorphism analysis of bacterial small-subunit rRNA genes: efficacy of selected tetrameric restriction enzymes for studies of microbial diversity in nature. **Applied and Environmental Microbiology** 62:2501-2507.
65. Myers, R. M., S.G.Fisher, L.S.Lerman, and T.Maniatis. 1985. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. **Nucleic Acids Res.** 13:3131-3145.
66. Myrick, K. V. and W. M. Gelbart. 2002. Universal fast walking for direct and versatile determination of flanking sequence. **Gene** 284:125-131.
67. Nicolaisen, M. H. and N. B. Ramsing. 2002. Denaturing gradient gel electrophoresis (DGGE) approaches to study the diversity of ammonia-oxidizing bacteria. **Journal of Microbiological Methods** 50:189-203.
68. Ochman, H., F. J. Ayala, and D. L. Hartl. 1993. Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. **Methods Enzymol.** 218:309-321.
69. Oliver, J. Nilsson L, and Kjelleberg S. 1991. Formation of nonculturable *Vibrio vulnificus* cells and its relationship to the starvation state. **Appl Environ Microbiol.** 57(9):2640-2644.
70. Orphan, V. J., C. H. House, K. U. Hinrichs, K. D. McKeegan, and E. F. DeLong. 2002. From the Cover: Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. **Proceedings of the National Academy of Sciences** 99:7663-7668.
71. Paez, J. G., M. Lin, R. Beroukhim, J. C. Lee, X. Zhao, D. J. Richter, S. Gabriel, P. Herman, H. Sasaki, D. Altshuler, C. Li, M. Meyerson, and W. R. Sellers. 2004. Genome coverage and sequence fidelity of {phi}29 polymerase-based multiple strand displacement whole genome amplification. **Nucl. Acids Res.** 32:e71.
72. Pearson, A., A.L.Sessions, K.J.Edwards, and J.M.Meyers. 2004. Phylogenetically specific separation of rRNA from prokaryotes for isotopic analysis. **Marine Chem** 92, p. 295-306.
73. Pelz O, Chatzinotas A, Andersen N, Bernasconi SM, Hesse C, Abraham WR, and Zeyer J. 2001. Use of isotopic and molecular techniques to link toluene degradation in denitrifying aquifer microcosms to specific microbial populations. **Arch Microbiol** 175:270-281.
74. Pernthaler, A. and R.Amann. 2004. Simultaneous fluorescence in situ hybridization of mRNA and rRNA in environmental bacteria. **Appl. Environ. Microbiol.** 70:5426-5433.
75. Prescott, A. M. and C. R. Fricker. 1999. Use of PNA oligonucleotides for the in situ detection of *Escherichia coli* in water samples. **Mol. Cell. Probes** 13:261-268.
76. Purkhold, U., M. Wagner, G. Timmermann, A. Pommerening-Roser, and H. P. Koops. 2003. 16S rRNA and amoA-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. **Int J Syst Evol Microbiol** 53:1485-1494.

77. Radajewski, S., I. R. McDonald, and J. C. Murrell. 2003. Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms. **Curr. Opin. Biotech.** 14:296-302.
78. Radajewski, S., G. Webster, D. S. Reay, S. A. Morris, P. Ineson, D. B. Nedwell, J. I. Prosser, and J. C. Murrell. 2002. Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. **Microbiology** 148:2331-2342.
79. Rees, H. C., W. D. Grant, B. E. Jones, and S. Heaphy. 2004. Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. **Extremophiles** 8:63-71.
80. Rhee, S.-K., X.Liu, L.Wu, S.C.Chong, X.Wan, and J.Zhou. 2004. Detection of genes involved in the biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide arrays. **Appl. Environ. Microbiol.** 70:4304-4317.
81. Roling WF, de Brito Couto IR, Swannell RP, and Head IM. 2004. Response of Archaeal communities in beach sediments to spilled oil and bioremediation. **Appl Environ Microbiol.** 70:2614-2620.
82. Roling WF, Milner MG, Jones DM, Fratepietro F, Swannell RP, Daniel F, and Head IM. 2004. Bacterial community dynamics and hydrocarbon degradation during a field-scale evaluation of bioremediation on a mudflat beach contaminated with buried oil. **Appl Environ Microbiol.** 70:2603-2613.
83. 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.
84. Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. **Nucleic Acids Res** 26:1628-1635.
85. Roszak DB, Grimes DJ, Colwell RR. 1984. Viable but nonrecoverable stage of Salmonella enteritidis in aquatic systems. **Can J Microbiol.** 30(3): 334-338.
86. Schloss, P. D. and J. Handelsman. 2003. Biotechnological prospects from metagenomics., p. 303-310. In .
87. Schönhuber, W., B. Fuchs, , S. Juretschko, and R. Amann. 1997. Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. **Appl. Environ. Microbiol.** 63:3268-3273.
88. Sebat, J. L., F. S. Colwell, and R. L. Crawford. 2003. Metagenomic profiling: microarray analysis of a metagenomic DNA library. **Appl. Environ. Microbiol.** 69:4927-4934.
89. Sheffield, V. C., D.R. Cox, L.S. Lerman, and R.M. Myers. 1989. Attachment of a 40-Base-Pair GC rich sequence (GC-Clamp) to genomic DNA fragments by the polymerase chain-reaction results in improved detection of single-base changes. **Proc. Natl. Acad. Sci. U. S. A.** 86:232-236.
90. Sheu, D., Y. Wang, and C. Lee. 2000. Rapid detection of polyhydroxyalkanoate-accumulating bacteria isolated from the environment by colony PCR. **Microbiology** 146:2019-2025.
91. Short, J. M. and E. J. Mathur. 1999. Production and use of normalized DNA libraries. **US Patent** 6001574 .
92. Small, J., D.R. Call, F.J.Brockman, T.M.Straub, and D.P.Chandler. 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. **Appl. Environ. Microbiol.** 67:4708-4716.

93. Snaidr, J., B. Fuchs, G. Wallner, M. Wagner, K. H. Schleifer, and R. Amann. 1999. Phylogeny and in situ identification of a morphologically conspicuous bacterium, *Candidatus Magnospira bakii*, present at very low frequency in activated sludge. **Environ. Microbiol.** 1:125-135.
94. Stull, D. and J. M. Pisano. 2001. Purely RNA: New innovations enhance the quality, speed, and efficiency of RNA isolation techniques. **The Scientist** 15:29-31.
95. Suzuki, M. T. and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. **Applied and Environmental Microbiology** 62:625-630.
96. Thompson, J., L. Marcelino, and M. F. Polz. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. **Nucleic Acids Res** 30:2083-2088.
97. Torsvik, V. and L. Ovreas. 2002. Microbial diversity and function in soil: from genes to ecosystems. **Curr. Opin. Microbiol.** 5:240-245.
98. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, and Rubin EM. 2005. Comparative Metagenomics of Microbial Communities . **Science**, 308:554-557.
99. Tyson, J. W., J. Chapman, P. Hugenholtz, 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.
100. Uchiyama T., Abe T., Ikemura T, and Watanabe K. 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. **Nat Biotechnol** 23:88-93.
101. Urbach, E., K. L. Vergin, and S. J. Giovannoni. 1999. Immunochemical detection and isolation of DNA from metabolically active bacteria. **Appl. Environ. Microbiol** 65:1207-1213.
102. Ussery, D. W. 2004. Microbiology update: Genome Update: length distributions of sequenced prokaryotic genomes. **Microbiology Comment** 1113-1115.
103. Vallaeys, T. F., Persello-Cartieaux, N. Rouard, C. Lors, G. Laguerre, and G. Soulas. 1997. PCR-RFLP analysis of 16S rRNA, *tfdA* and *tfdB* genes reveals a diversity of 2,4-D degraders in soil aggregates. **FEMS Microbiol. Lett.** 24:269-278.
104. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rush, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Sevy, H. Knap, M. W. Lomas, K. Nealson, O. White, J. D. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. **Science** 304:66-74.
105. Watanabe, K., M. Teramoto, H. Futamata, and S. Harayama. 1998. Molecular detection, isolation, and physiological characterization of functionally dominant phenol-degrading bacteria in activated sludge. **Appl. Environ. Microbiol.** 64:4396-4402.
106. Wawrik, B. L., G. Kerkhof, J. Zylstra, and Kukor J.L. 2005. Identification of unique type II polyketide synthase genes in soil. **Appl. Environ. Microbiol.** 71:2232-2238.
107. Wellington, E. M. H., A. Barry, and M. Krsek. 2003. Resolving functional diversity in relation to microbial community structure in soil; exploiting genomics and stable isotope probing. **Curr. Opin. Microbiol.** 6:295-301.
108. Wen CM, Tseng CS, Cheng CY, and Li YK. 2002. Purification, characterization and cloning of a chitinase from *Bacillus* sp. NCTU2. **Biotechnol Appl Biochem.** 35:213-219.
109. Wilson, M. S., C. Bakermans, and E. Madsen. 1999. In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. **Appl. Environ. Microbiol.** 65:80-87.

110. Wommack, K. E., S.R Bench, K.E.Williamson, and M.Radosevich. 2004. Viruses in soils: The first terrestrial viral metagenome. **10th International Symposium in Microbial Ecology**, Cancun, Mexico, August, 2004.
111. Wu, Y., V. M. Hayes, J. Osinga, I. M. Mulder, M. W. Looman, C. H. Buys, and R. M. Hofstra. 1998. Improvement of fragment and primer selection for mutation detection by denaturing gradient gel electrophoresis. **Nucl. Acids Res.** 26:5432-5440.
112. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M. 2002. Cultivating the uncultured., **Proc Natl Acad Sci U S A.** 99(24):15681-15686..
113. Zhou, J.-Z. 2003. Microarrays for bacterial detection and microbial community analysis. **Curr. Opin. Microbiol.** 6:294.
114. Zhou, J.-Z. and D.K.Thompson. 2002. Challenges in applying microarrays to environmental studies. **Curr. Opin. Biotechnol.** 13:202-204.