

73 Functional Marker Gene Assays for Hydrocarbon Degrading Microbial Communities: Aerobic

H. Junca^{1,2} · D. H. Pieper¹

¹Biodegradation Research Group, Department of Microbial Pathogenesis, Helmholtz-Zentrum für Infektionsforschung GmbH – HZI, Braunschweig, Germany

dpi@helmholtz-hzi.de

²Centro Colombiano de Genómica y Bioinformática de Ambientes Extremos (GeBiX) – Grupo de Genética Molecular, Corporación CorpoGen, Bogotá, Colombia

1	<i>Introduction</i>	4291
2	<i>Knowledge and Gaps in the Microbial Ecology of Biodegradation</i>	4292
3	<i>Approaches to Study the Microbial Ecology of Biodegradation</i>	4293
4	<i>Key Catabolic Activities and Associated Gene Families in Bacterial Aerobic Hydrocarbon Degradation</i>	4294
5	<i>Sequence-Based Methods to Detect and to Assess Catabolic Functions: PCR-Based Applications</i>	4294
5.1	DNA Template	4295
5.2	Primer Design	4297
5.3	Quantifying Functional Gene Copies in Contaminated Environments	4299
5.4	Resolving Sequence Complexity: PCR Clone Libraries	4300
5.5	Genetic Fingerprinting Methods to Analyze Metabolic Gene Composition	4300
6	<i>Hybridization-Based Detection of Functional Genes</i>	4303
7	<i>Activity-Based Assays to Detect Functional Proteins or Gene Families Involved in Hydrocarbon Metabolism</i>	4303
7.1	Metagenomics to Assess Catabolic Functions	4304
7.2	Genetic Traps to Catch Previously Hidden Biomineralization Genes and Enzymes	4305
7.3	Stable Isotope Probing (SIP)	4305

- 8 *Community Structure Insights* 4306
- 9 *Future Perspectives to Assess Microbial Metabolic Functions in the Environment: Combination of Techniques and New Fields of Research* 4306

Abstract: In this chapter we discuss the state of the art on functional marker gene assays for aerobic hydrocarbon-degrading microbial communities, and the different approaches available to detect genes or activities associated with biodegradative processes in the environment.

1 Introduction

Latest research efforts attempting to characterize the diversity of microorganisms on Earth indicate that the description and the understanding of the dynamics and functions of environmental microbial communities, as a new kind of *Terra Incognita*, is a quest conforming one of the biggest frontiers and scientific challenges in biological and environmental sciences (Quince et al., 2008).

Recent studies and calculations show, that the microbial diversity and the contribution of microbial activities to biogeochemical processes are so immense that we could feel overwhelmed by the microbial complexity awaiting to be properly explored (Quince et al., 2008; Rappe and Giovannoni, 2003; Schloss and Handelsman, 2006). As an example, the total number of approved cultured bacterial species is less than 10,000, while recent calculations indicated that already a single gram of soil may contain such number of species (Schloss and Handelsman, 2006), numbers recently confirmed by pyrosequencing efforts (Roesch et al., 2007). Regarding global impacts, it is currently estimated that the process of anaerobic ammonium oxidation, only discovered about 1 decade ago, is responsible for at least 50% of nitrogen removal from marine ecosystems (Devol, 2003). Ammonium oxidation by Crenarchaeota has been postulated only this century (Francis et al., 2005), but it is now assumed that these organisms are the numerically dominant ammonia oxidizers in the ocean and in soils (Francis et al., 2007).

Consequently, new approaches to allow a more detailed, efficient, and representative detection and assessment of the main functions played by microorganisms in the environment are crucial to reach an adequate understanding of their ecology (Paerl and Steppe, 2003) and their contribution to environmental processes. The main limitations are currently, on the one hand, due to the inefficient techniques available to culture the majority of environmental bacteria, a fact recognized since a long time ago (Amann et al., 1995; Conn, 1918; Razumov, 1932). On the other hand, there are technical limitations to retrieve and process the information of the aforementioned enormous microbial diversity in either the cultured or not yet cultured microbial fractions.

Concerning aerobic hydrocarbon biodegradation, several key functions have been identified in the degradation pathways of aliphatic and aromatic compounds (See ► Chapters 3–6, Vol. 2, Part 2). As with many other processes and functions known in bacteria, it can, however, not be assessed currently, how representative and environmentally relevant these aerobic biodegradation pathways are, as knowledge has been accumulated almost entirely from analyses of cultured bacteria. Thus, we may expect that the large not yet cultured bacterial fraction in environmental samples is contributing to the recycling and removal of hydrocarbons in magnitudes and with mechanisms that are not yet described, and that can be radically different to the mechanisms already known.

For several ecosystems a strong discrepancy between the bacterial types recovered by culturing techniques and the genotypes predominantly found by culture-independent analyses has been described. For instance, Enterobacteria comprise only a small fraction of the human gut microbiota (Andersson et al., 2008), however, culturing techniques proposed them, for a

long time, as predominant. Soil systems, as described above, comprise thousands of species per gram of soil, however, only of some bacterial phyla observed by culture-independent tools, cultured representatives are available (Rappé and Giovannoni, 2003). Members of the phylum Acidobacteria typically represent about 10% of soil bacterial communities but can contribute up to 50% or even 80% (Koch et al., 2008; Roesch et al., 2007) and culture-independent studies indicate that the diversity of this phylum is nearly as great as the diversity of the phylum Proteobacteria, however, only a few species have so far been described.

In situations where a certain hydrocarbon is turning out to be a strong selector on the ecosystem affected, as in contamination events, the conditions may mimic to some extent the artificial and restrictive environment that we use in the laboratory, and thus, such discrepancy may not to be as radical as in pristine environments or other ecosystems with a very high microbial diversity (Schloss and Handelsman, 2006), or in environments where microbial life forms have been existing under strong selectors relatively stably maintained for long (geological) periods of time (Nakagawa and Takai, 2008).

2 Knowledge and Gaps in the Microbial Ecology of Biodegradation

While a large fraction of aliphatic and aromatic compounds on Earth are derived from biogenic anabolic processes, it is generally accepted that such hydrocarbons were also formed by chemical reactions, and that they existed since the very early times of the planet, probably well before the first living forms (Ehrenfreund et al., 2006). The presence of aromatic compounds has been even detected in interstellar clouds and they seem to be widely distributed in the known universe (Henning and Salama, 1998). Thus, it is very likely that the emergence of heterotrophic bacteria occurred in the presence of such compounds and that the mechanisms to obtain carbon and energy from such sources have evolved very early in evolution (Kobayashi et al., 1998). At the beginning of the industrial era, the exploitation and use of petroleum and its derivatives increased dramatically, with a simultaneous increase in contamination of various ecosystems. Almost a century ago, microbiological studies reported the capability of microorganisms to grow on aromatic and aliphatic hydrocarbons as a sole carbon and energy source (Söhngen, 1913) and subsequent studies described the mechanisms used by bacterial isolates to use hydrocarbons such as alkanes, benzene, phenol, or toluene (described in detail in Chapters 2–4, Vol. 2, Part 2). Initially, such information was retrieved exclusively from bacterial isolates capable of aerobic aromatic degradation, due to the ease to culture aerobic bacteria, and also because biodegradation of petroleum hydrocarbons is usually rapid in the presence of oxygen as terminal electron acceptor. However, the recent decades have seen major breakthroughs in the identification of microorganisms and elucidation of pathways responsible for degradation of aromatics such as toluene or naphthalene under anaerobic conditions (Boll et al., 2002), and it became clear that aromatic degradation occurs under nitrate-, iron-, or sulfate-reducing conditions and even under methanogenic conditions (See Chapter 9, Vol. 2, Part 3). Interestingly, whereas biodegradation is usually assumed to be more rapid in the presence of oxygen as electron acceptor, some pollutants seem to be exclusively metabolized under anaerobic conditions. Highly chlorinated hydrocarbons such as tetrachloroethene, or polychlorinated benzenes, dibenzo-*p*-dioxins, or biphenyls can function as terminal electron acceptors in anaerobic respiration processes, resulting in a gain of energy concomitant with elimination of chloride (Janssen et al., 2005; Smidt and de Vos, 2004).

Several bioremediation technologies to clean up the numerous contaminated sites present worldwide have been developed and their efficiency depends crucially on the selection of the appropriate microbial communities or creation of favorable environmental conditions, i.e., the degradation of highly chlorinated biphenyls will be enhanced by stimulating anaerobic microbial communities with the genetic potential of performing reductive dehalogenation rather than by aerobic communities, whereas unchlorinated hydrocarbons can be quickly degraded if sufficient sources of nutrients and oxygen are available for the aerobic microbial community with the appropriate genetic potential. However, rational approaches to optimize bioremediation process performance need to know about the key players in environmental microbial communities, their interactions and specifically on metabolic functions important for the process under optimization.

It is currently still difficult to characterize or even predict the evolution of the functions or the metabolic net of environmental microbial communities. Nevertheless, despite enormous limitations, the information on the mechanisms and functions performed by cultured bacteria has been very helpful to draw analogies and to hypothesize what mechanisms environmental bacteria might perform.

3 Approaches to Study the Microbial Ecology of Biodegradation

Molecular ecology approaches have significantly increased our knowledge on environmental microbial communities. Two general lines of research can here be distinguished. On the one hand physical, chemical, and biological parameters (such as soil parameters, degradation rates, community structure) are determined and statistical analyses are performed to find correlations between different variables to deduce how the physical and chemical properties of a given environment influence the structure and functions of the microbial communities. Alternatively, a defined working hypothesis is rejected or supported by detailed data collected, which is tested by experiments under defined laboratory conditions.

In both scenarios marker genes can be important variables to study. Various key genes and enzymes for aerobic metabolism of hydrocarbons are known from studies with isolates (but also communities) which comprise, in case of aerobic aromatic degradation, enzymes carrying out crucial functions such as the activation of aromatics through oxygenations or CoA ligation or the cleavage of the aromatic ring system (See also [Chapter 4, Vol. 2, Part 2](#)). It can be hypothesized that an active microbial community is enriched for such functions. Technically, such functions could be screened for by genotype (sequence) based or phenotype (activity) based approaches.

Sequence-based approaches depend on previously accumulated information and, thus, will only detect known genes or variations of genes considered to encode proteins with a certain metabolic function. Such analyses can give information on differences in gene copy numbers between different samples that can be interpreted as the biological fitness of the gene sequence in the microbial community of the source ecosystem. Moreover, expression of the target gene or gene family can be followed by molecular methods as well as the variation in the composition of a certain gene family.

In the case of an activity-based approach, there is no need of previous knowledge on related sequences. The whole metagenome of an environment is screened for a predefined function of interest, which, however, necessitates an appropriate high-throughput screening method (Committee on Metagenomics: Challenges and Functional Applications, 2007).

Both approaches are complementary, and provide different subsets of information of the system under study. The sequence-based assessment of a catabolic gene target might be interpreted in a more evolution-oriented way, as the presence, selection, diversity, and relative increase of a certain catabolic gene type is an indication of its biological fitness and significance in the studied environment. However, it should also be taken into consideration that detecting a sequence does not necessarily mean that it actually carries out the assumed function in the cellular context.

Function-based screens were initially developed for analyzing cultured bacteria, but have also been used to detect activities encoded in the DNA of the non-cultured microbial fraction, by directly extracting and cloning environmental DNAs (see below) in heterologous hosts expressing the activities. This is a very useful approach to mine new enzymes and activities (Ferrer et al., 2005). However, such detection will be significantly affected by the bacterium used as a host, as effective induction, transcription, translation, folding, and posttranslational modifications are necessary to obtain the requested metabolic signal. Thus, the functions found are not necessarily the main activities exerted by the microbial community in a certain ecosystem and there is a need to integrate the function-based metagenomic information with genomic, transcriptomic, and proteomic disciplines, between others (Warnecke and Hugenholtz, 2007).

The main features of sequence-based and activity-based approaches are largely depending on the kind of gene or activity of interest and, thus such analyses require an integrative custom methodological development (🔗 Fig. 1).

4 Key Catabolic Activities and Associated Gene Families in Bacterial Aerobic Hydrocarbon Degradation

A general key step to approach analysis of a target environment is the definition of the complexity of bacterial genes or activities putatively involved in a given process. The bacterial hydrocarbon degradation mechanisms are indeed greatly dependent on the availability of oxygen as reactive species involved in dearomatization of aromatic systems or activation of aliphatics (🔗 Fig. 2). Similar key reactions can also be defined for anaerobic degradation processes. Catabolic pathways have been selected during evolution to achieve the maximum amount of energy and biomass for the cell (Pazos et al., 2003). For such purpose the bacterial cell require not only enzymes evolved to efficiently process and degrade the compounds, but also the cellular machinery for efficient transport and sensing of the nutrition source as well as efficient regulatory mechanisms and mechanisms to allow survival at high substrate concentrations, among others. All these specific components or signatures can be useful in efforts to track the contribution of bacterial functions to hydrocarbon transformation in the environment.

5 Sequence-Based Methods to Detect and to Assess Catabolic Functions: PCR-Based Applications

Sequence-based methods relying on PCR amplifications have been tailored to detect the presence of certain gene types, such as genes encoding enzymes of a given subfamily carrying out a common function by the use of PCR primers annealing to conserved gene regions. They have been further adapted to quantify the relative amounts of those genes, in assays

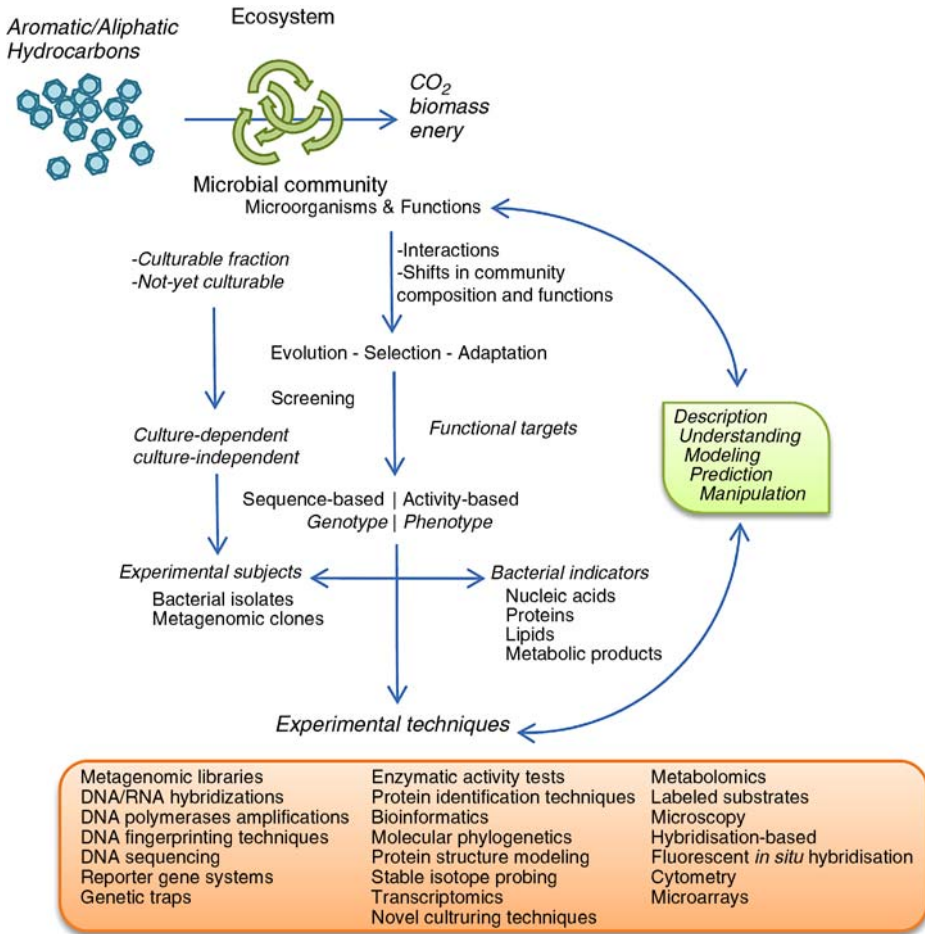


Figure 1 Strategies to analyze the community potential for hydrocarbon biodegradation. General overview of the interrelations of culture-dependent–culture-independent approaches and sequence-based or activity-based techniques to assay functions in microbial communities degrading aromatic and aliphatic hydrocarbons.

aiming at the culture-independent genetic detection and quantification of key biodegradation functions. The most obvious and direct method is the detection of discrete (positive/negative) signals of the targeted gene sequence(s) in DNA extracted from an environmental sample. Even though it is a basic and common approach, there are several points to take into account for a successful and representative amplification of the target in question.

5.1 DNA Template

Standard optimizations or changes in PCR conditions are significantly influencing the overall performance of the desired amplifications (LaMontagne et al., 2002). In case the goal is the

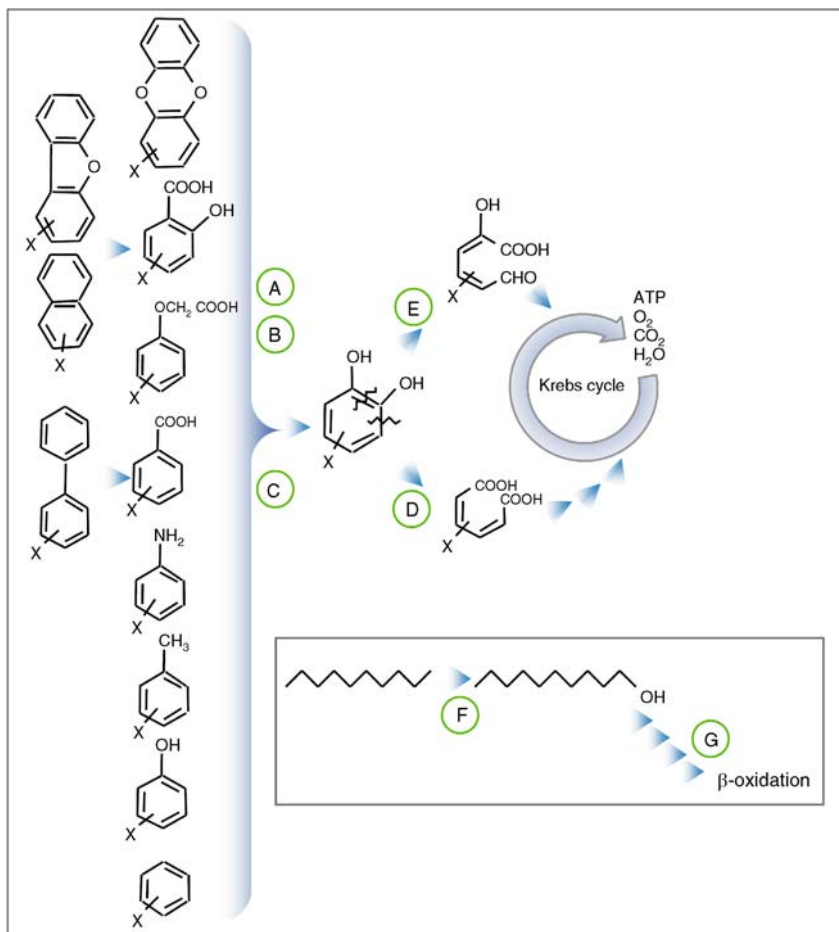


Figure 2

Schematic overview of generic aerobic biodegradation pathways for aromatic and aliphatic compounds. Various aromatic compounds, under aerobic conditions, are activated through oxygenations and metabolized via a few central di- (or tri-)hydroxylated central intermediates, which are subject to ring-cleavage and further channeled to Krebs cycle intermediates. Catabolic enzymes, defined as members of different gene families, are performing successive metabolic steps to achieve the recovery of energy and carbon from hydrocarbons. Crucial genes, proteins, or activities have been or could be used as targets in culture-dependent or culture-independent studies applying sequence-based or activity-based functional gene assays to assess aerobic aromatic/aliphatic hydrocarbon biodegradation potential of microbial communities. Some crucial enzyme families are (letter in a circle localized in the relevant step): A, flavoprotein monooxygenases or soluble diiron monooxygenases; B, Rieske nonheme iron oxygenases (mainly dioxygenases, however, various members of this superfamily also catalyze monooxygenations); C, ferredoxins (involved in various crucial multicomponent enzyme complexes); D, intradiol dioxygenases; E, extradiol dioxygenases; F, alkane monooxygenases (members of the integral membrane monooxygenase family, or the cytochrome P450 subfamily CYP153A); G, alcohol, and aldehyde dehydrogenases.

detection of functional genes in environmental DNA extracts, additional factors influencing the experimental outcome have to be considered. As the nucleic acids used as template in amplifications are extracted and purified from environmental samples (soil, water, sludge, organic matrixes, etc.), the DNA extraction method has a strong influence on the ulterior performance of the PCR amplification and how meaningful and representative the results will be. It is a common observation that different preparations (different extraction and purification methods) of the same sample vary significantly in the amount and purity of the obtained DNA (Sagova-Mareckova et al., 2008). A low efficiency of extraction may affect the representation of gene types present close to the threshold of amplification. However, a high yield not necessarily results in an optimal detection, as impurities (coextracted inhibitors) may severely interfere with amplification, thus increasing the minimum amount of gene copies necessary to obtain amplification. Ideally, the DNA template should contain at least 100–1,000 copies of the gene type targeted, in order to identify predominant gene polymorphisms and variations inside the gene group analyzed. Thus, it is important calculate the threshold of detection by PCR, e.g., in serial dilutions of samples containing defined copies of gene target, and later on, use the maximum amount of DNA template practicable in PCR assays.

However, in various scenarios only minute amounts of DNA are available, restricting the possibilities to detect genes, or to perform several tests. To overcome such limitations, the DNA replication system of the phage phi 29 (Blanco and Salas, 1984) can be recruited in a process termed whole genome amplification. The *phi29* DNA polymerase performs an isothermal multiple displacement amplification that allows obtaining exponential increases of DNA amounts in a way much more accurate and unspecific than other DNA polymerase systems (Abulencia et al., 2006) (See Chapter 70, Vol. 5, Part 5). At the same time, however, saturation of the PCR reaction has to be avoided, as an excess of DNA template, may produce undesirable effects, such as low specificity and low efficiency of primer annealing to the target genes, resulting in false positive results or low efficiency of amplification.

In the case of genes of function in aromatic metabolism, their abundance in DNA extracts from bacterial communities can not be estimated a priori, in contrast to targets that are expected to be present at least once on every bacterial genome, such as 16S rRNA genes, genes involved in DNA replication or basal housekeeping genes. Genes encoding proteins acting in aromatic metabolism are expected to be present only in a fraction of the bacterial genomes present in a sample, but on the other hand may be present in multiple copies per genome equivalent (e.g., if they are encoded on multiple copy plasmids). Thus, the assessment of such number can be only done experimentally.

5.2 Primer Design

PCR based methods are important tools to detect functional gene targets in the environment. For such detection, the availability of appropriate primers is of crucial importance. While primer design has been a key aspect for PCR amplification, the concept that is generally applied in pure cultures studies, where the gene target sequence is usually known, is not useful in microbial ecology studies. In pure cultures, the aim is to design primers based on the optimal thermodynamics, but this is not applicable to environmental samples, as multiple variations of the target gene are expected, all immersed in an unknown metagenomic complexity. For environmental DNA samples it can be assumed that due to the high loads of certain hydrocarbons a selective pressure is exerted and the capability to degrade these

hydrocarbons is of advantage, resulting in the accumulation of degraders and the respective catabolic genes in the sample. In the simplest case, only one bacterial strategy to degrade a certain pollutant is known. However, in other cases such as for the aerobic degradation of toluene, multiple strategies are established. Moreover, a certain type of reaction such as the extradiol cleavage of dihydroxylated aromatics can be catalyzed by members of at least three distinct protein families. Even if only a single gene family is targeted, this family consists of a diverse group of sequences encoding proteins that are able to perform the function of interest. These sequences have been found in different isolates, and are similar in sequence. They will contain highly variable regions, encoding protein stretches not necessary for function, and share some conserved regions as those crucial for functionality. These regions are identified by multiple sequence alignments. After such cognate conserved regions are identified, primer sets are designed, which commonly include degenerative positions to allow all possibilities of variation in the non-conserved sites at the conserved region. Tests with reference strains hosting the target genes as positive controls, negative controls and tests directly on environmental samples are then performed.

This process of primer design requires a sequence database of the group of sequences targeted that can be aligned. One principle of multiple sequence alignments is that the sequences are derived from a common evolutionary ancestor and form so-called gene families. In order to build these databases, one straightforward approach is to start with a subset of protein reference sequences with common domains, ideally proteins characterized biochemically and structurally, having in common that they display the same generic activity. These protein sequences are used as seeds for similarity searches using appropriate databases to find potential homologues. Other methods used to find protein homologues in sequence databases, are Hidden Markov Model methods, which are useful to find remote homologues. However, in the case of catabolic gene families, the rather artificial concepts of gene/protein family and protein sequence space (Nelson, 2005), have to be delimited by pragmatic views, such as experimental evidence of the function by a member belonging to a cluster of sequences that, by similarity, may be considered as members of the gene/protein family. From the similarity search results, protein sequences with high similarity and the corresponding coding DNA sequences (CDS) are collected to build the database that is actually used to find the conserved regions where the primers sets could be designed to anneal. Depending of the conservation of these regions, primers are designed with degenerated positions to cover the complete diversity of the conserved regions. However, care should be taken to apply a conserved sequence in the final 5' stretch (4–8 bases), as this improves the specificity. Standard complete thermodynamic calculations of the suitability for PCR of the oligos designed are usually of limited value due to the presence of degenerations, and as conservation is typically a stronger decision factor than these considerations. If it is not possible to find a region conserved enough, the design of subgroup specific primers should be considered. Most importantly, due to the variable and unknown complexity of the environmental DNA to be analyzed and the complexity of the interactions that might occur in a mix of different oligos (which the degenerate primers actually are), there is no way to predict the actual performance of a newly designed primer, which has to be tested experimentally using positive controls and trials in samples. We should also always bear in mind that wherever PCR is involved, we are obtaining results with the limitations and advantages this technique implies, and primer design will heavily depend on the current representation of a gene group in the databases, which is a dynamic and evolving factor to count with.

The mere detection of a catabolic gene fragment is indicating the overall weight of the encoded function for the fitness of the bacteria hosting it, but it is certainly not discerning selection of polymorphisms inside the group targeted. However, it is well documented that slight changes in protein amino acid sequence (and thus gene sequence) can have severe consequences on the catalytic properties of enzymes. As examples, single amino acid residue differences were reported to result in drastically different substrate specificities of biphenyl 2,3-dioxygenases for chlorinated biphenyl congeners (Pieper and Seeger, 2008) or of toluene dioxygenases for chlorobenzenes, and also in differences in the regioselectivity of attack. Thus, it is important to analyze the relative quantities as well as the composition of the expected PCR product mixture, to characterize changes in diversity, or selection of specific polymorphisms, which can indicate an evolutionary advantage at the site under study.

5.3 Quantifying Functional Gene Copies in Contaminated Environments

If the minimum gene copy number giving a signal in the PCR amplification has been experimentally tested, the relative numbers of gene copies in different environmental DNAs can be calculated (quantified).

Most Probable Number (MPN)-PCR assays are using the principle of probability of presence of particles in serial dilutions, to quantify the gene copy number of a sample. One of the first studies quantifying catabolic genes targeted *nahAc* genes, which are encoding the α -subunit of naphthalene dioxygenases, a subgroup of the Rieske nonheme iron oxygenases catalyzing the activation of naphthalene, and *xylE* genes, a subgroup of type I extradiol dioxygenases catalyzing the ring-cleavage of catechol (a central intermediate of aerobic aromatic metabolism) at a site contaminated with jet-fuel and observed those genes in relatively high quantities over the total community DNA (Chandler and Brockman, 1996). Another study using MPN-PCR, showed that the *nahAc* copy number in oxic layers of petroleum-contaminated sites correlated with ^{14}C -naphthalene mineralization rates (Tuomi et al., 2004). Similarly, catechol 2,3 dioxygenase encoding genes related to *xylE* from *Pseudomonas* sp. mt-2 (Junca and Pieper, 2003) were found to be increased by 2–3 orders of magnitude in abundance in a site highly contaminated with BTEX, compared to the copy number observed in a neighboring pristine site, indicating a strong selection for this function by BTEX contamination at microaerophilic conditions.

Real-Time PCR is a quantitative PCR (qPCR) technique that employs a standard PCR reaction where the increase in PCR product (dsDNA) can be monitored over time (by cycle number) through the use of fluorescent labeled nucleotides or modified DNA oligonucleotide probes that are excited when hybridizing with complementary DNA strands (See ► Chapter 48, Vol. 5, Part 3). This method has also been used to quantify gene copy numbers of catabolic genes involved in aerobic hydrocarbon biodegradation and the numbers of *nagAc*-type genes (related to *Ralstonia* sp. strain U2 naphthalene dioxygenase) were shown to be correlated with in situ naphthalene concentrations (Dionisi et al., 2004), indicating that strains harboring such genes have been selected for by the environmental conditions. An increase in *nagAc* gene copy number (concomitant with an increase in *nidA*-type genes encoding so-called naphthalene inducible dioxygenases, mainly observed in *Rhodococci* and *Mycobacterium* spp.) in correlation with the concentrations of PAH (Debruyne et al., 2007) was

also observed in a *Mycobacterium* dominated community of a coal tar contaminated site. Similarly, this technique was applied to quantify genes involved in aliphatic hydrocarbon biodegradation (*alkB*-like genes encoding part of the integral membrane monooxygenases involved on alkane hydroxylation), the quantity of which was found to be correlated with *n*-alkane concentrations in petroleum-contaminated soils (Powell et al., 2006). Also other studies analyzing the abundance of key catabolic gene families involved in activation of aromatic or aliphatic pollutants, and in the ring-cleavage of metabolites of aromatic degradation observed direct correlations between such abundance and the concentrations of aromatic or aliphatic hydrocarbons (Ringelberg et al., 2001; Salminen et al., 2008).

Competitive PCR assays use a competitor target, usually of modified size, and added in known copy amounts to the PCR reaction. By identifying the ratio at which both targets are amplified with similar yield, and thus present in identical concentration in the mixture, the relative concentration of the target gene (family) can be calculated. This technique has also been applied to quantify changes in gene copy numbers of alkane hydroxylases or catechol 2,3-dioxygenases (Heiss-Blanquet et al., 2005; Mesarch et al., 2000).

5.4 Resolving Sequence Complexity: PCR Clone Libraries

The most straightforward, but expensive and time consuming method to resolve the sequence complexity of PCR amplifications derived from environmental DNA is by preparing PCR clone libraries and screening them by random sequencing. There are several examples in the literature reporting the construction and screening of PCR clone libraries of amplifications targeting functional genes such as those encoding ring-cleavage enzymes like catechol 1,2-dioxygenases and catechol 2,3-dioxygenases (Junca and Pieper, 2004; Kasuga et al., 2007). However, recent advances in sequencing methods may resolve disadvantages of library screening.

To validate that the fragment amplified encodes part of a functional protein, it may be cloned in an appropriate vector that contains such gene kind, but disrupted in a way that it can be completed by insertion of PCR products. In such approaches catabolic gene fragments of catechol 2,3-dioxygenases (amplified from environmental DNA derived from mixed cultures of phenol-degrading or crude oil-degrading bacteria), cytochromes (P450 CYP153), or alkane hydroxylases (amplified from DNA extracted from petroleum-contaminated soil and ground-water) have been integrated into truncated genes to produce functional chimeric genes, supporting the role of the amplified fragments as part of functional genes (Kubota et al., 2005; Okuta et al., 1998).

5.5 Genetic Fingerprinting Methods to Analyze Metabolic Gene Composition

An alternative to random sequencing to resolve the complexity of PCR products of very close or identical size, but differences in sequence, was the adaptation of PCR fingerprinting techniques, initially developed to discern the complexity of 16S rRNA gene amplifications. They allow to compare a larger number of samples in faster and less expensive way compared

to PCR clone library sequencing. Amplified DNA fragments are typically “sorted” based on inherent properties, such as melting behavior or folding giving distinguishable band or peak patterns after appropriate resolution. These techniques, however, have lower resolution than PCR libraries and random clone sequencing, and could have additional lack of resolution inherent to the samples analyzed (e.g., identical migration behavior of different fragments). They also need trained personnel, skilled in such techniques. Despite these possible disadvantages, molecular fingerprinting techniques have been powerful and widely used in microbial ecology for the aforementioned advantages.

Among the techniques to generate fingerprints from complex PCR amplicon mixtures, almost all those developed previously for resolving 16S rRNA genes mixtures have been adapted and applied to analyze PCR amplicons comprising fragments of catabolic gene families. As genes involved in pollutant degradation are often localized on mobile genetic elements such strategy does not directly track the phylogeny or the taxonomical groups harboring such activities, but detects and follows the variation in polymorphisms inside a given gene family.

Denaturing (Temperature) Gradient Gel Electrophoresis (DGGE/TGGE) relies on the difference in melting behavior of different double-stranded DNA strands upon heat (TGGE) or chemical denaturant (DGGE) application (See Chapter 60, Vol. 5, Part 3). If mixtures of homologous DNA fragments are subjected to electrophoresis on gels applying gradients of the denaturant, specific melting behaviors depending on the sequence (G + C content) will be obtained. A so-called GC-clamp, a GC rich terminal region, is artificially introduced by PCR at one end of the amplification mixture, which keeps connection of the strands and prevents further migration as soon as the strands are partially melted. These techniques have been used predominantly to analyze the diversity of 16S rRNA gene fragments amplified from environmental samples (Muyzer, 1999) and more recently, adapted to analyze diversity of functional genes (Felske et al., 2003; Watanabe et al., 1998). They have also been successfully applied to determine the diversity of subgroups of catabolic gene families such as *nahAc* α -subunits of naphthalene dioxygenases (Rieske nonheme iron oxygenase superfamily), *bmoA* large subunits of benzene monooxygenases (soluble diiron monooxygenase superfamily), catechol 1,2 dioxygenases (intradiol dioxygenase superfamily) or catechol 2,3 dioxygenases (extradiol dioxygenases of the vicinal chelate superfamily) (Gomes et al., 2007; Hendrickx et al., 2006; Sei et al., 2004).

These techniques are very useful in case the potential targets are expected to have significant differences in G + C content. However, in case of highly similar sequences these techniques have only low resolution power, and other fingerprinting alternatives methods should be considered.

Single Strand Conformation Polymorphism (SSCP) is another powerful fingerprinting technique, which takes advantage of the fact that single stranded DNA (ssDNA) under non-denaturant conditions acquires a secondary conformation of intramolecular loops and foldings, due to complementarity of the bases on the same strand. As two DNA fragments of the same size but of different sequence will fold differently, these sequence-dependent conformations can be resolved on non-denaturant polyacrylamide gels. While for sequences with significant differences in sequence the resolution is uncertain and would need optimization of the electrophoretic conditions (i.e., two ssDNA fragments of the same size and very different sequence can migrate, theoretically, at the same speed), SSCP is one of the most effective and widely used methods to identify single nucleotide polymorphisms (SNPs). SSCP was optimized for applications in microbial ecology through the use of a phosphorylated primer,

which allows the selective digestion of the phosphorylated strand by lambda exonuclease (Schwieger and Tebbe, 1998). The possibility to analyze only one of the complementary single strands avoids heteroduplex formation or overlapping of forward and reverse strands from different amplicons during separation. Application was initially used to identify shifts in composition of microbial communities by targeting 16S rDNA genes (Schwieger and Tebbe, 2000), and later optimized to resolve the complexity of amplified catabolic gene targets amplified from contaminated sites, like catechol 2,3-dioxygenases (Junca and Pieper, 2004) and α -subunits of different Rieske nonheme iron oxygenases (Witzig et al., 2006, 2007) in BTEX contaminated sites. Predominant polymorphisms in different sample could be relatively rapidly defined. Also single amino acid residues could be identified which obviously are of selective advantage in the site under study and which are crucial in modulating enzyme activity and specificity. Taking into account the knowledge gained through protein engineering or modeling studies, catabolic sequence phylotypes can be analyzed for possible structural determinants shaping substrate specificities in catabolic enzymes (Parales et al., 2000). The prediction of the meaning of polymorphisms found in catabolic genes amplified directly from environmental DNA (eDNA) is a promising field of research.

Restriction Fragment Length Polymorphism (RFLP) is a technique to generate patterns of different sized fragments from an initial dsDNA fragment due to variations in recognition sites for a certain restriction enzyme. Differences in size patterns produced from two different DNA fragments of the same size, normally visualized on gels, are a direct indication of sequence differences between them, and ideally, the similarity of the patterns should reflect the evolutionary relationships of the DNA sequence fragments digested. A variation of this technique to analyze functional genes involved in aerobic hydrocarbon biodegradation pathways is the PCR-RFLP analyses of functional gene fragments (Amplified Functional DNA Restriction Analysis – AFDRA) (Junca and Pieper, 2003).

Terminal (T)-RFLP, is a variation of RFLP where the complexity of fragments that will be generated and visible when complex amplicon mixtures are subject to restriction based analysis is restricted by the use of two (fluorescent) labeled primers. Only terminal labeled fragments are then analyzed on an automatic sequencer (Chapter 58, Vol. 5, Part 3). T-RFLP analysis is a highly reproducible and robust technique that yields high-quality fingerprints consisting of fragments of precise sizes. Like other fingerprinting methods, it has been mainly used to analyze microbial community composition and changes. However, patterns produced from amplified 16S rDNA are simplified representations of the community, and different species may produce identical length terminal fragments with a given restriction enzyme, such that a careful selection of restriction enzymes giving meaningful taxonomic information, is necessary (Osborn et al., 2000). Some developments have been taking place to transfer this technology for fingerprinting of functional genes (Braker et al., 2001), however, the major limitation for its application to a wider range of functional targets is the limited range of suitable target gene families where enough sequence information is available and the absence of phylogenetically informative restriction sites (Lueders and Friedrich, 2003). However, T-RFLP has been applied to study extradiol dioxygenases in PAH contaminated soil samples under rhizoremediation treatment (Sipila et al., 2008) helping to identify the evolutionary cluster predominant in this bioremediation study.

The fingerprinting techniques described above are useful to rapidly determine the sequence type diversity in PCR amplicon mixtures. They are, however, restricted to a sequence space defined by the primers used. Methods based on hybridizations allow the composite detection of a broad range of potential targets in a single assay.

6 Hybridization-Based Detection of Functional Genes

The detection of functional gene targets based on complementary DNA strands used as probes has been developed and miniaturized allowing the detection of multiple gene types in arrays. These developments have been also transferred to analyze catabolic gene targets involved in aerobic hydrocarbon biodegradation.

Miniaturized hybridization devices such as microarrays, consisting of probes of PCR fragments derived from reference genes or oligonucleotides and designed to anneal to sequences representing different catabolic gene families have been developed recently. The advantage of such array systems is the amount of different sequences that can be detected in a single assay, contrasting PCR primer-based detections, where usually only a subset of a catabolic gene family can be targeted with a single primer set. However, arrays require time for careful design, are relatively costly, and require a detailed processing of information. The obtained results also require a validation to confirm the correctness of the signal.

An oligoarray to detect hundreds of functions related to bacterial degradation of pollutants, including catabolic, regulatory, resistance, and stress genes, has been reported (Rhee et al., 2004) and evolved to the so-called GeoChip (He et al., 2007) (▶ Chapter 51, Vol. 5, Part 3).

In the case of microarray studies, apart of the obvious high-throughput advantages, there is the difficulty to report results that can be replicated in other labs, as it is very common that the actual information about sequence of the probes is not provided. Also, as mentioned before, the detection of a signal on a probe targeting a certain gene type shall not be interpreted as an unequivocal assignment of the sequence hybridizing, as it might be derived from an unspecific hybridization. There are some additional interesting approaches in the field of microarrays to detect catabolic functions related to aerobic aromatic biodegradation, such as the oligoarrays specifically targeting aromatic ring hydroxylating dioxygenases or monooxygenases (Iwai et al., 2008).

7 Activity-Based Assays to Detect Functional Proteins or Gene Families Involved in Hydrocarbon Metabolism

Several activity-based methods have been developed initially to study bacterial isolates able to degrade aromatic and aliphatic compounds. The ease to detect activities varies greatly from enzyme to enzyme and depends on factors such as protein localization, stability of the protein or of substrates and products or sensitivity of the enzyme to the substrates. In fact rapid and easy tests that can be applied in a high-throughput format are necessary to screen large isolate collections or large metagenomic libraries.

One of the classical assays to screen genomic libraries of single organisms is the capability of various oxygenases such as toluene dioxygenase to convert indole to blue indigo (Eaton and Timmis, 1986; Ensley et al., 1983). This screening method also has the potential to be used for the activity-based screening of metagenomic libraries, libraries constructed from DNA extracted from environmental samples and composed of various genomes. This method and its target gene exemplify also some of the limitations of activity-based screenings. Rieske type nonheme iron oxygenases are multicomponent enzyme systems where electrons have to flow through the different components, which is very restricted in case phage expression systems are used to construct the libraries. On the other hand, indole is not a universal substrate for all such oxygenases.

A second classical assay is the screening of bacterial isolates, or genomic clone libraries for catechol 2,3-dioxygenase expression, as the 2-hydroxymuconic semialdehyde ring-cleavage product is of bright yellow color (Junca and Pieper, 2004; Kitayama et al., 1996).

7.1 Metagenomics to Assess Catabolic Functions

The study of the eDNA extracted from a sample proved to be a valuable tool to access the hidden diversity in the fraction of non-cultured organisms. Gene libraries prepared from such *metagenomic* DNA, the composite and pooled study of several genomes (Handelsman et al., 1998) allow to maintain and to recover a broad range of sequences and functions. In the last 5 years the amount of reports using metagenomic approaches has rapidly increased as this methodology has proven to offer exciting discoveries practically in any ecosystem or any target activity or gene that has been chosen to be studied (Ferrer et al., 2007) (See also Chapter 38, Vol. 4, Part 5).

Many reports on metagenomic analyses are dealing with massive sequencing efforts of end reads on metagenomic DNA inserts cloned in bacterial artificial chromosomes or fosmids (Handelsman, 2004). The alternative approach to analyze metagenomic libraries is the screening for specific activities easily evidenced in the recombinant clones with metagenomic DNA inserts. Activity-based functional metagenomics is basically an heterologous expression experiment, and therefore, it has all the constraints related with the correct processing by the recipient host of the function encoded, with the additional complexity of ignoring the precise identity of the source host. Typically, *E. coli* has been used as a host for activity-based metagenomics, however, a survey of 32 prokaryotic genomes indicated that only 40% of genes could be expressed in this host (Gabor et al., 2004). It is, thus, not surprising, that in a metagenomic study where three different hosts (*E. coli*, *Pseudomonas putida* and *Streptomyces lividans*) were used to screen a metagenomic library, different positive clones were detected in each of the hosts used (Martinez et al., 2004). This clearly indicates that screening of metagenomic libraries can be improved using a range of hosts such as *E. coli*, *P. putida*, *S. lividans*, *S. meliloti* and *Rhizobium leguminosarum*.

In the case of hydrocarbon biodegradation, there are not many methods developed yet to easily screen metagenomic clones for biodegradative functions, and due to the relative novelty of the metagenomic approach, there are not yet many metagenomic studies using such screenings.

However, the production of a yellow colored metabolite during the extradiol cleavage of catechol has recently been used to screen metagenomic libraries constructed with DNA extracted from activated sludge (Suenaga et al., 2007) and with DNA extracted from samples of a site highly contaminated with BTEX (Brennerova et al., 2009). The recovered extradiol dioxygenases were, despite their divergence in sequence, belonging to the well described vicinal chelate superfamily (type I extradiol dioxygenases) and were similar to members described in cultured bacteria, indicating that the majority of the catechol 2,3-dioxygenases (C23O) gene types found in these sample are members of the same family considered to be important from culture-dependent studies. Aiming at representative coverage, vectors allowing high expression such as phages or large inserts such as fosmids or bacterial artificial chromosomes are being used. The protein release in phage libraries is an advantage in cases where the screening of the activity is facilitated by such localization (Ferrer et al., 2005), but in the case of enzymes requiring tight intracellular assembly or which are inactivated by oxygen,

the conversion to phagemids or the cloning on vectors maintaining the host cell may be more appropriate (Jones et al., 2008).

7.2 Genetic Traps to Catch Previously Hidden Biomineralization Genes and Enzymes

There are many biotransformation processes of interest that are not producing metabolites that can be massively and easily detected by simple activity tests (such as reaction color). Moreover, metagenomic library clones are usually in numbers that are not suitable for single chemical analyses. Thus, alternative high-throughput methods to screen such kind of libraries are needed. One approach to overcome these limitations is known as molecular or genetic traps (Galvao and de Lorenzo, 2006) and use a transcriptional regulator that is blind to the reaction substrate but responds to the reaction product, and as a result activates a promoter fused to a reporter gene (See Chapter 96, Vol. 5, Part 6). Respective regulators may be searched for in natural regulatory circuits, but can also be engineered, to recognize the product of the desired activity. Bacteria containing such a regulator/promoter/reporter system are used as receptors of a metagenomic library and only the clones hosting a metagenomic insert encoding an enzyme capable to catalyze the desired reaction will activate the reporter gene.

This is a powerful technology that is capable to detect genes involved in transformation/metabolization of hydrocarbons in metagenomic DNA by the indirect detection of products of the reaction for which a direct detection method is not available. There are some examples of applications of this strategy (Galvao and de Lorenzo, 2006; Mohn et al., 2006; Uchiyama et al., 2005).

7.3 Stable Isotope Probing (SIP)

The use of stable isotopes (^{13}C) to trace carbon from specific substrates into microbes that assimilate carbon from that substrate has significantly advanced our understanding of the relationship between environmental processes and microbial phylogeny (See Chapter 49, Vol. 5, Part 3). Bacteria active on a labeled substrate should incorporate the label into their biomass, and also in molecules that can be used as phylogenetic markers. Previously, such incorporation into lipids (Abraham et al., 1998; Boschker and Middelburg, 2002) was used to identify active community members. More recently, this method was adapted to analyze incorporation into nucleic acids and density-gradient centrifugation was proven to be capable to separate ^{13}C -DNA from ^{12}C -DNA allowing molecular tools described above to be applied exclusively to the fraction of bacteria active on the given compound (Radajewski et al., 2000). Later on, stable-isotope-labeled RNA was used to identify, after amplification and sequencing of PCR clone libraries or via fingerprint generation, metabolically active community members (Radajewski et al., 2003) and led to the precise identification of microorganisms responsible for contaminant degradation in engineered systems, and to applications enhancing our understanding of carbon flow in terrestrial ecosystems. In the first studies there was a great emphasis on describing the microbial community by means of 16S rRNA gene sequences found in the labeled fraction. In recent studies, stable isotope probing (SIP) was also used to analyze for functional genes in the labeled fraction of microbial communities metabolizing

PCBs, indicating that some of the mechanisms described in cultured bacteria are, in fact, environmentally relevant (Leigh et al., 2007). SIP was also used to identify and finally to isolate a bacterium that was the main consumer of ^{13}C -labeled naphthalene in contaminated sediments (Jeon et al., 2003). It should, however, be noted that experiments using SIP need to be carefully designed. A problem arises, particularly for nucleic acids, when substrate incorporation and incubation time are insufficient, generating poorly labeled biomarker molecules that are not distinguishable above a background of relatively abundant unlabeled molecules. Excessive substrate concentrations or incubation times may also be problematic, as they could lead to an enrichment bias that does not reflect the natural process in the environment and/or enhanced cross-feeding due to use of partially metabolized SIP substrates or scavenging of labeled biomass by other community members (Neufeld et al., 2007).

8 Community Structure Insights

Some studies aim on a joint analysis of ribosomal and catabolic gene sequences to link community composition or changes with key functions such as alkane (Harayama et al., 2004) or monoaromatics (Junca et al., 2004) degradation in contaminated environments. Due to the localization of various catabolic genes on mobile genetic elements (Herrick et al., 1997), such linkages are often not meaningful (Siciliano et al., 2003), however, in other cases, catabolic gene sequence divergence was shown to be closely linked to host phylogeny (Witzig et al., 2007). Thus, attempts to draw conclusions from such experiments should consider the available knowledge on localization of the gene functions under study. Correlations between community structure changes and catabolic potential can, however, be used to direct the efforts of isolation towards key community members. As an example, predominant 16S rRNA gene types found by DGGE in enrichments with polycyclic aromatic hydrocarbons (PAH), helped to recover the respective strains capable to mineralize PAHs (Hilyard et al., 2008).

Another way of analyzing community-function relationships, is through the analysis of metagenomic sequence data (Schloss and Handelsman, 2008). The advances in DNA sequencing technologies, with the invention of pyrosequencing and its posterior use in the modified high-throughput 454 adaptation (Andersson et al., 2008; Ronaghi et al., 1998) are modifying completely the magnitude of sequences that can be analyzed. An alternative to whole metagenomic analysis consists in the analysis of total rRNA or specifically mRNA in a sort of meta-transcriptomic approach (Frias-Lopez et al., 2008). This approach selects from the amazing crowd of sequence diversity found in metagenomic DNA, those genes that are being expressed, and sequencing efforts are restricted to this genetic subgroup supposed to be important in the environment analyzed, allowing to draw connections between structure and function of microbial communities (Gilbert et al., 2008; Urich et al., 2008).

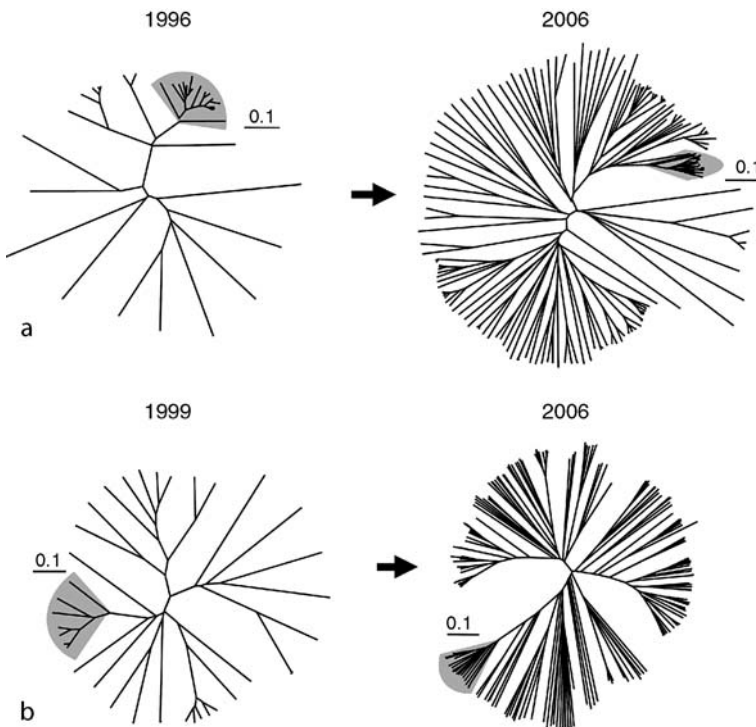
9 Future Perspectives to Assess Microbial Metabolic Functions in the Environment: Combination of Techniques and New Fields of Research

An obvious trend is the composite use of methods described above in single experiments to overcome limitations of single techniques. One example is the use of fluorescent probes targeting specific taxonomically meaningful 16S rRNA gene signatures (fluorescent in situ

hybridization – FISH). Coupled with assays to follow consumption and incorporation of ^{14}C -labeled substrates into the microbial biomass (Lee et al., 1999), by a technique known as Microautoradiography – FISH, functional analysis can be directly linked with taxonomical structure in situ (See Chapter 56, Vol. 5, Part 3).

There are some techniques that are being or can be foreseen to be incorporated in the functional analysis of microbial communities, such as flow cytometry (Czechowska et al., 2008) to study of microbial physiology under environmentally relevant conditions, to identify active microbial populations and to isolate previously uncultured microorganisms (See also Chapter 57, Vol. 5, Part 3), or the recently reported coupling of Raman microscopy and FISH (Huang et al., 2007) for simultaneous cultivation-independent identification and determination of ^{13}C incorporation into microbial cells (See Chapter 50, Vol. 5, Part 3).

The viromes of microbial communities, especially in soil communities, have been largely neglected due to the difficulty to be analyzed in the past. There are recent reports pointing out



■ **Figure 3**

Increased representation of members of two selected catabolic gene families due to genome projects, environmental surveys, and new bacterial isolates. NJ tree dendrograms of multiple sequence alignments of (a) Extradiol Dioxygenase type I (EXDO) protein sequences, and of (b) Rieske nonheme iron oxygenases (α -subunits) (RHDO) as represented in GenBank databases in years 1996 and 2006 (EXDO), and in years 1999 and 2006 for (RHDO). For orientation, the α -subunits of the subfamily of benzene/toluene/isopropylbenzene/biphenyl dioxygenases of Rieske nonheme iron oxygenases (Witzig et al., 2006); and the subfamily I.2.A of catechol 2,3-dioxygenases as members of the extradiol dioxygenases (Eltis and Bolin, 1996), are shaded.

the strong effects they have on microbial communities from diverse ecosystems (Weinbauer and Rassoulzadegan, 2004). Phages and other viruses in microbial communities in contaminated sites can be viewed as gene reservoir and as shuttle for catabolic functions, a perspective that, thanks to the technical advances on genomics, it is now possible to assess, but it is currently awaiting more in-depth investigations.

Another expected development is the improvement of techniques to culture the bacterial types not cultured yet, as the best holistic experimental model of bacterial behavior is still to have the organism in culture. There are several reports showing different approaches to achieve high-throughput culturing methods to isolate strains representing previously refractory to culture bacterial taxa (Connon and Giovannoni, 2002; Davis et al., 2005; Zengler et al., 2002), despite other optimistic views stating that genomic studies might be enough for our understanding of the microbial communities functioning (Tyson and Banfield, 2005).

This will continue to be a controversy, but very likely, polyphasic approaches, combining complementary and informative strategies will be the way to obtain a *less incomplete* picture of microbial diversity and functioning on the environment. All the aforementioned techniques have been cumulatively helping to better describe and understand environmental processes. As an example, [Fig. 3](#) is depicting the protein sequence phylogenies of the members reported in the GenBank database of two of the most researched and better studied catabolic gene families, the extradiol dioxygenases type I, including the catechol 2,3-dioxygenases (shaded) (phylogenies at 1996 and the status at 2006) and the Rieske nonheme iron oxygenases (α -subunit), including the benzene/toluene/biphenyl/isopropylbenzene dioxygenase (shaded) (phylogenies at 1999 and upgraded version of 2006). Thanks to genome projects, environmental surveys, and new bacterial isolates, the representation in both cases increased substantially, and this certainly would allow a better and more refined development of tools for the detection of these catabolic functions in the environment.

Acknowledgments

We would like to thank the members of Biodegradation Research Group at HZI (formerly known as GBF), for stimulating discussions over the last years regarding functional gene assays on aerobic hydrocarbon biodegradation. We thank the European Commission for the funding kindly granted to AMICO, BIOTOOL, and BACSIN EC-projects.

References

- Abraham WR, Hesse C, Pelz O (1998) Ratios of carbon isotopes in microbial lipids as an indicator of substrate usage. *Appl Environ Microbiol* 64: 4202–4209.
- Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, et al. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* 72: 3291–3301.
- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143–169.
- Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, Engstrand L (2008) Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE* 3: e2836.
- Blanco L, Salas M (1984) Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci USA* 81: 5325–5329.

- Boll M, Fuchs G, Heider J (2002) Anaerobic oxidation of aromatic compounds and hydrocarbons. *Curr Opin Chem Biol* 6: 604–611.
- Boschker HTS, Middelburg JJ (2002) Stable isotopes and biomarkers in microbial ecology. *FEMS Microbiol Ecol* 40: 85–95.
- Braker G, Ayala-del-Rio HL, Devol AH, Fesefeldt A, Tiedje JM (2001) Community structure of denitrifiers, bacteria, and archaea along redox gradients in Pacific Northwest marine sediments by terminal restriction fragment length polymorphism analysis of amplified nitrite reductase (*nirS*) and 16S rRNA genes. *Appl Environ Microbiol* 67: 1893–1901.
- Brennerova M, Josefiova J, Brenner V, Pieper DH, Junca H (2009) Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air sparging bioremediation. *Environ Microbiol*, in press.
- Chandler DP, Brockman FJ (1996) Estimating biodegradative gene numbers at a JP-5 contaminated site using PCR. *Appl Biochem Biotechnol* 57–58: 971–982.
- Committee on Metagenomics: Challenges and Functional Applications NRC (2007) The new science of metagenomics: revealing the secrets of our microbial planet. Washington, DC: National Research Council.
- Conn HJ (1918) The microscopic study of bacteria and fungi in soil. *N Y Agric Exp Sta Tech Bull* 64: 3–20.
- Connon SA, Giovannoni SJ (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 68: 3878–3885.
- Czechowska K, Johnson DR, van der Meer JR (2008) Use of flow cytometric methods for single-cell analysis in environmental microbiology. *Curr Opin Microbiol* 11: 205–212.
- Davis KE, Joseph SJ, Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Appl Environ Microbiol* 71: 826–834.
- Debruyn JM, Chewning CS, Saylor GS (2007) Comparative quantitative prevalence of Mycobacteria and functionally abundant *nidA*, *nahAc*, and *nagAc* dioxxygenase genes in coal tar contaminated sediments. *Environ Sci Technol* 41: 5426–5432.
- Devol AH (2003) Nitrogen cycle – solution to a marine mystery. *Nature* 422: 575–576.
- Dionisi HM, Chewning CS, Morgan KH, Menn FM, Easter JP, Saylor GS (2004) Abundance of dioxxygenase genes similar to *Ralstonia* sp. strain U2 *nagAc* is correlated with naphthalene concentrations in coal tar-contaminated freshwater sediments. *Appl Environ Microbiol* 70: 3988–3995.
- Eaton RW, Timmis KN (1986) Characterization of a plasmid-specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J Bacteriol* 168: 123–131.
- Ehrenfreund P, Rasmussen S, Cleaves J, Chen L (2006) Experimentally tracing the key steps in the origin of life: the aromatic world. *Astrobiology* 6: 490–520.
- Eltis LD, Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* 178: 5930–5937.
- Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222: 167–169.
- Felske AD, Fehr W, Pauling BV, Von Canstein H, Wagner-Dobler I (2003) Functional profiling of mercuric reductase (*merA*) genes in biofilm communities of a technical scale biocatalyzer. *BMC Microbiol* 3: 22.
- Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins Dos Santos VA, Yakimov MM et al. (2005) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* 12: 895–904.
- Ferrer M, Golyshina O, Beloqui A, Golyshin PN (2007) Mining enzymes from extreme environments. *Curr Opin Microbiol* 10: 207–214.
- Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* 102: 14683–14688.
- Francis CA, Beman JM, Kuypers MM (2007) New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J* 1: 19–27.
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, Delong EF (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* 105: 3805–3810.
- Gabor EM, Alkema WBL, Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6: 879–886.
- Galvao TC, de Lorenzo V (2006) Transcriptional regulators a la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr Opin Biotechnol* 17: 34–42.
- Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, Joint I (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* 3: e3042.
- Gomes NC, Borges LR, Paranhos R, Pinto FN, Krogerrecklenfort E, Mendonca-Hagler LC, Smalla K (2007) Diversity of *ndo* genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. *Appl Environ Microbiol* 73: 7392–7399.

- Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669–685.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245–249.
- Harayama S, Kasai Y, Hara A (2004) Microbial communities in oil-contaminated seawater. *Curr Opin Biotechnol* 15: 205–214.
- He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1: 67–77.
- Heiss-Blanquet S, Benoit Y, Marechaux C, Monot F (2005) Assessing the role of alkane hydroxylase genotypes in environmental samples by competitive PCR. *J Appl Microbiol* 99: 1392–1403.
- Hendrickx B, Dejonghe W, Faber F, Boenne W, Bastiaens L, Verstraete W, et al. (2006) PCR-DGGE method to assess the diversity of BTEX mono-oxygenase genes at contaminated sites. *FEMS Microbiol Ecol* 55: 262–273.
- Henning T, Salama F (1998) Carbon in the universe. *Science* 282: 2204–2210.
- Herrick JB, Stuart-Keil KG, Ghiorse WC, Madsen EL (1997) Natural horizontal transfer of a naphthalene dioxygenase gene between bacteria native to a coal tar-contaminated field site. *Appl Environ Microbiol* 63: 2330–2337.
- Hilyard EJ, Jones-Meehan JM, Spargo BJ, Hill RT (2008) Enrichment, isolation, and phylogenetic identification of polycyclic aromatic hydrocarbon-degrading bacteria from Elizabeth River sediments. *Appl Environ Microbiol* 74: 1176–1182.
- Huang WE, Stoecker K, Griffiths R, Newbold L, Daims H, Whiteley AS, Wagner M (2007) Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Environ Microbiol* 9: 1878–1889.
- Iwai S, Kurisu F, Urakawa H, Yagi O, Kasuga I, Furumai H (2008) Development of an oligonucleotide microarray to detect di- and monooxygenase genes for benzene degradation in soil. *FEMS Microbiol Lett* 285(1): 111–121.
- Janssen DB, Dinkla JJ, Poelarends GJ, Terpstra P (2005) Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ Microbiol* 7: 1868–1882.
- Jeon CO, Park W, Padmanabhan P, DeRito C, Snape JR, Madsen EL (2003) Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proc Natl Acad Sci USA* 100: 13591–13596.
- Jones BV, Begley M, Hill C, Gahan CG, Marchesi JR (2008) Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci USA* 105: 13580–13585.
- Junca H, Pieper DH (2003) Amplified functional DNA restriction analysis to determine catechol 2,3-dioxygenase gene diversity in soil bacteria. *J Microbiol Methods* 55: 697–708.
- Junca H, Pieper DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environ Microbiol* 6: 95–110.
- Junca H, Plumeier I, Hecht HJ, Pieper DH (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiology* 150: 4181–4187.
- Kasuga I, Nakajima F, Furumai H (2007) Diversity of catechol 2,3-dioxygenase genes of bacteria responding to dissolved organic matter derived from different sources in a eutrophic lake. *FEMS Microbiol Ecol* 61: 449–458.
- Kitayama A, Achioku T, Yanagawa T, Kanou K, Kikuchi M, Ueda H et al. (1996) Cloning and characterization of extradiol aromatic ring-cleavage dioxygenases of *Pseudomonas aeruginosa* J1104. *J Ferment Bioeng* 82: 217–223.
- Kobayashi K, Kaneko T, Hashimoto H, Kouchi A, Saito T, Yamashita M (1998) Abiotic formation of bioorganic compounds in space—preliminary experiments on ground and future exobiology experiments in space. *Biol Sci Space* 12: 102–105.
- Koch IH, Gich F, Dunfield PF, Overmann J (2008) *Edaphobacter modestus* gen. nov., sp. nov., and *Edaphobacter aggregans* sp. nov., acidobacteria isolated from alpine and forest soils. *Int J Syst Evol Microbiol* 58: 1114–1122.
- Kubota M, Nodate M, Yasumoto-Hirose M, Uchiyama T, Kagami O, Shizuri Y, Misawa N (2005) Isolation and functional analysis of cytochrome P450 CYP153A genes from various environments. *Biosci Biotechnol Biochem* 69: 2421–2430.
- LaMontagne MG, Michel FC, Jr., Holden PA, Reddy CA (2002) Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *J Microbiol Methods* 49: 255–264.
- Lee N, Nielsen PH, Andreasen KH, Juretschko S, Nielsen JL, Schleifer KH, Wagner M (1999) Combination of fluorescent in situ hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. *Appl Environ Microbiol* 65: 1289–1297.
- Leigh MB, Pellizari VH, Uhlik O, Sutka R, Rodrigues J, Ostrom NE, et al. (2007) Biphenyl-utilizing bacteria

- and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME J* 1: 134–148.
- Lueders T, Friedrich MW (2003) Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl Environ Microbiol* 69: 320–326.
- Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA, Osburne MS (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. *Appl Environ Microbiol* 70: 2452–2463.
- Mesarch MB, Nakatsu CH, Nies L (2000) Development of catechol 2,3-dioxygenase-specific primers for monitoring bioremediation by competitive quantitative PCR. *Appl Environ Microbiol* 66: 678–683.
- Mohn WW, Garmendia J, Galvao TC, de Lorenzo V (2006) Surveying biotransformations with a *la* crite genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes. *Environ Microbiol* 8: 546–555.
- Muyzer G (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Curr Opin Microbiol* 2: 317–322.
- Nakagawa S, Takai K (2008) Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. *FEMS Microbiol Ecol* 65: 1–14.
- Nelson DR (2005) Gene nomenclature by default, or BLASTing to Babel. *Hum Genomics* 2: 196–201.
- Neufeld JD, Dumont MG, Vohra J, Murrell JC (2007) Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb Ecol* 53: 435–442.
- Okuta A, Ohnishi K, Harayama S (1998) PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene* 212: 221–228.
- Osborn AM, Moore ERB, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* 2: 39–50.
- Paerl HW, Stepe TF (2003) Scaling up: the next challenge in environmental microbiology. *Environ Microbiol* 5: 1025–1038.
- Parales RE, Lee K, Resnick SM, Jiang H, Lessner DJ, Gibson DT (2000) Substrate specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme. *J Bacteriol* 182: 1641–1649.
- Pazos F, Valencia A, De Lorenzo V (2003) The organization of the microbial biodegradation network from a systems-biology perspective. *EMBO Rep* 4: 994–999.
- Pieper DH, Seeger M (2008) Bacterial metabolism of polychlorinated biphenyls. *J Mol Microbiol Biotechnol* 15: 121–138.
- Powell SM, Ferguson SH, Bowman JP, Snape I (2006) Using real-time PCR to assess changes in the hydrocarbon-degrading microbial community in Antarctic soil during bioremediation. *Microb Ecol* 52: 523–532.
- Quince C, Curtis TP, Sloan WT (2008) The rational exploration of microbial diversity. *ISME J* 2(10): 997–1006.
- Radajewski S, Ineson P, Parekh NR, Murrell JC (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* 403: 646–649.
- Radajewski S, McDonald IR, Murrell JC (2003) Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms. *Curr Opin Biotechnol* 14: 296–302.
- Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* 57: 369–394.
- Razumov AS (1932) The direct method of calculation of bacteria in water. Comparison with the Koch method. (Original article in Russian.) *Mikrobiologiya* 2: 131–146.
- Rhee SK, Liu X, Wu L, Chong SC, Wan X, Zhou J (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* 70: 4303–4317.
- Ringelberg DB, Talley JW, Perkins EJ, Tucker SG, Luthy RG, Bouwer EJ, Fredrickson HL (2001) Succession of phenotypic, genotypic, and metabolic community characteristics during in vitro bioslurry treatment of polycyclic aromatic hydrocarbon-contaminated sediments. *Appl Environ Microbiol* 67: 1542–1550.
- Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, Farmerie WG, Triplett EW (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1: 283–290.
- Ronaghi M, Pettersson B, Uhlen M, Nyren P (1998) PCR-introduced loop structure as primer in DNA sequencing. *Biotechniques* 25: 876–878, 880–872, 884.
- Sagova-Mareckova M, Cermak L, Novotna J, Plhachova K, Forstova J, Kopecky J (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl Environ Microbiol* 74: 2902–2907.
- Salminen JM, Tuomi PM, Jorgensen KS (2008) Functional gene abundances (*nahAc*, *alkB*, *xylE*) in the assessment of the efficacy of bioremediation. *Appl Biochem Biotechnol* 151(2–3): 638–652.
- Schloss PD, Handelsman J (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* 2: e92.

- Schloss PD, Handelsman J (2008) A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics* 9: 34.
- Schwieger F, Tebbe CC (1998) A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol* 64: 4870–4876.
- Schwieger F, Tebbe CC (2000) Effect of field inoculation with *Sinorhizobium meliloti* L33 on the composition of bacterial communities in rhizospheres of a target plant (*Medicago sativa*) and a non-target plant (*Chenopodium album*)-linking of 16S rRNA gene-based single-strand conformation polymorphism community profiles to the diversity of cultivated bacteria. *Appl Environ Microbiol* 66: 3556–3565.
- Sei K, Inoue D, Wada K, Mori K, Ike M, Kohno T, Fujita M (2004) Monitoring behaviour of catabolic genes and change of microbial community structures in seawater microcosms during aromatic compound degradation. *Water Res* 38: 4405–4414.
- Siciliano SD, Germida JJ, Banks K, Greer CW (2003) Changes in microbial community composition and function during a polyaromatic hydrocarbon phytoremediation field trial. *Appl Environ Microbiol* 69: 483–489.
- Sipila TP, Keskinen AK, Akerman ML, Fortelius C, Haahntela K, Yrjala K (2008) High aromatic ring-cleavage diversity in birch rhizosphere: PAH treatment-specific changes of I.E.3 group extradiol dioxygenases and 16S rRNA bacterial communities in soil. *ISME J* 2(9): 968–981.
- Smidt H, de Vos WM (2004) Anaerobic microbial dehalogenation. *Annu Rev Microbiol* 58: 43–73.
- Söhngen NL (1913) Benzin, Petroleum, Paraffinöl und Paraffin als Kohlenstoff- und Energiequelle für Mikroben. *Zentr Bacteriol Parasitenk, Abt. II* 37: 595–609.
- Suenaga H, Ohnuki T, Miyazaki K (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* 9: 2289–2297.
- Tuomi PM, Salminen JM, Jorgensen KS (2004) The abundance of *nahAc* genes correlates with the 14C-naphthalene mineralization potential in petroleum hydrocarbon-contaminated oxic soil layers. *FEMS Microbiol Ecol* 51: 99–107.
- Tyson GW, Banfield JF (2005) Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol* 13: 411–415.
- Uchiyama T, Abe T, Ikemura T, Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23: 88–93.
- Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* 3: e2527.
- Warnecke F, Hugenholtz P (2007) Building on basic metagenomics with complementary technologies. *Genome Biol* 8: 231.
- Watanabe K, Teramoto M, Futamata H, Harayama S (1998) Molecular detection, isolation, and physiological characterization of functionally dominant phenol-degrading bacteria in activated sludge. *Appl Environ Microbiol* 64: 4396–4402.
- Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? *Environ Microbiol* 6: 1–11.
- Witzig R, Junca H, Hecht HJ, Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* 72: 3504–3514.
- Witzig R, Aly HA, Strompl C, Wray V, Junca H, Pieper DH (2007) Molecular detection and diversity of novel diterpenoid dioxygenase *ditA1* genes from proteobacterial strains and soil samples. *Environ Microbiol* 9: 1202–1218.
- Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, Keller M (2002) Cultivating the uncultured. *Proc Natl Acad Sci USA* 99: 15681–15686.