

## Chapter 2

# Molecular Techniques to Assess Microbial Community Structure, Function, and Dynamics in the Environment

Gurdeep Rastogi and Rajesh K. Sani

**Abstract** Culture-based methods are important in investigating the microbial ecology of natural and anthropogenically impacted environments, but they are extremely biased in their evaluation of microbial genetic diversity by selecting a particular population of microorganisms. With recent advances in genomics and sequencing technologies, microbial community analyses using culture-independent molecular techniques have initiated a new era of microbial ecology. Molecular analyses of environmental communities have revealed that the cultivable fraction represents <1% of the total number of prokaryotic species present in any given sample. A variety of molecular methods based on direct isolation and analysis of nucleic acids, proteins, and lipids from environmental samples have been discovered and revealed structural and functional information about microbial communities. Molecular approaches such as genetic fingerprinting, metagenomics, metaproteomics, metatranscriptomics, and proteogenomics are vital for discovering and characterizing the vast microbial diversity and understanding their interactions with biotic and abiotic environmental factors. This chapter summarizes recent progress in the area of molecular microbial ecology with an emphasis on novel techniques and approaches that offer new insights into the phylogenetic and functional diversity of microbial assemblages. The advantages and pitfalls of commonly used molecular methods to investigate microbial communities are discussed. The potential applications of each molecular technique and how they can be combined for a greater comprehensive assessment of microbial diversity has been illustrated with example studies.

---

G. Rastogi (✉)

Department of Chemical and Biological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA

and

Department of Plant Pathology, University of California, Davis, CA 95616, USA

e-mail: grastogi@ucdavis.edu

## 2.1 Introduction

The biosphere is dominated by microorganisms and contains about  $4\text{--}6 \times 10^{30}$  prokaryotic cells (Whitman et al. 1998). This number represents at least two to three orders of magnitude more than all of the plant and animal cells combined. Thus, microorganisms are highly diverse group of organisms and constitute about 60% of the Earth's biomass (Singh et al. 2009). In aquatic environments, such as the oceans, the number of microbial cells has been estimated to be approximately  $1.2 \times 10^{29}$ , while in terrestrial environments, soil sustains as many as  $4\text{--}5 \times 10^{30}$  microbial cells (Singh et al. 2009). Owing to such enormous numbers, microorganisms are essential components of the Earth's biota and represent a large unexplored reservoir of genetic diversity. Understanding this unexplored genetic diversity is a high-priority issue in microbial ecology from perspectives such as global climate change and the greenhouse effect.

Microorganisms are key players in important ecological processes such as soil structure formation, decomposition of organic matter and xenobiotics, and recycling of essential elements (e.g., carbon, nitrogen, phosphorous, and sulfur) and nutrients. Thus, microbes play a critical role in modulating global biogeochemical cycles and influence all lives on Earth (Garbeva et al. 2004). In fact, all organisms in the biosphere either directly or indirectly depend on microbial activities. In soil ecosystems, microorganisms are pivotal in suppressing soil-borne plant diseases, promoting plant growth, and in promoting changes in vegetation (Garbeva et al. 2004). An understanding of microbial dynamics and their interactions with biotic and abiotic factors is indispensable in bioremediation techniques, energy generation processes, and in biotechnological industries such as pharmaceuticals, food, chemical, and mining.

The three fundamental questions that exist while discovering and characterizing any natural or artificial ecosystem are the following: (1) what type of microorganisms are present? (2) what do these microorganisms do? and (3) how do the activities of these microorganisms relate to ecosystem functions (e.g., energy flow, biogeochemical cycling, ecological resilience)? Microbial ecology aims to answer these central questions and deals with the study of microorganisms and their interactions with each other and with their environment. A plethora of biochemical and molecular methods have been applied to reveal the microbial community composition over time and space in response to environmental changes. These new approaches allow linkage between ecological processes in the environment with specific microbial populations and help us to answer important questions in microbial ecology such as what factors and resources govern the enormous genetic and metabolic diversity in an environment. This chapter presents an overview of the potentials and limitations of current molecular approaches used in microbial ecology. Although these techniques have been discussed with special emphasis on soil and plant microbial ecosystems, these are equally applicable to many other environments as well, such as oceans and sediments.

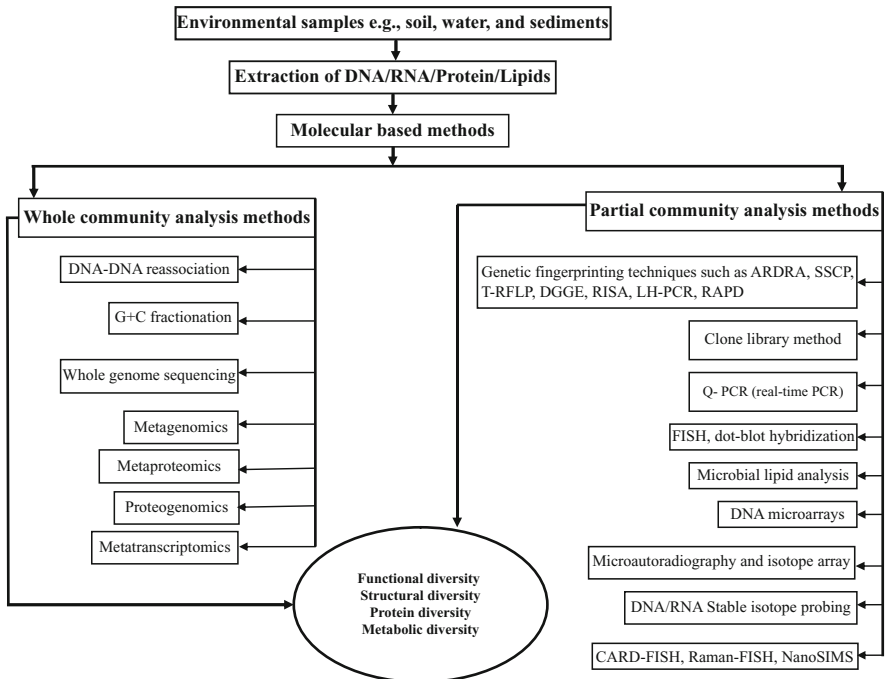
## 2.2 Culture Methods in Microbial Ecology: Applications and Limitations

Standard culture techniques to characterize microbial ecology involve isolation and characterization of microorganisms using commercial growth media such as Luria–Bertani medium, Nutrient Agar, and Tryptic Soy Agar (Kirk et al. 2004). The major limitation of culture-based techniques is that >99% of the microorganisms in any environment observed through a microscope are not cultivable by standard culturing techniques (Hugenholtz 2002). Several improved cultivation procedures and culture media have been devised that mimic natural environments in terms of nutrients (composition and concentration), oxygen gradient, pH, etc. to maximize the cultivable fraction of microbial communities. For example, a technique has been devised for the cultivation of uncultured microorganisms from different environments including seawater and soil that involved encapsulation of cells in gel microdroplets for large-scale microbial cultivation under low nutrient flux conditions (Zengler et al. 2005). Nonetheless, not all “uncultured” organisms are cultivable, and many of them remain “unculturable.” These organisms, although viable in their natural environments, do not grow under laboratory conditions and remain in a “viable but nonculturable” (VBNC) stage (Oliver 2005). Such VBNC organisms could represent completely novel groups and may be abundant or very active but remain untapped by standard culture methods.

Molecular microbial surveys based on 16S rRNA genes reveal that candidate bacterial divisions such as BRC1, OP10, OP11, SC3, TM7, WS2, and WS3 have no cultured representatives and are known only by their molecular sequences (Schloss and Handelsman 2004). These division-level clades, such as OP11, are highly diverse and widely distributed in different environments and are considered as “candidate divisions” to reflect our limited knowledge due to the lack of any cultured representative. Studies suggest the existence of at least 50 bacterial phyla with half represented entirely by molecular sequences (Schloss and Handelsman 2004). Additionally, microorganisms retrieved using common culture methods are rarely numerically abundant or functionally significant in the environment from which they were cultured. These cultured microorganisms are considered as the “weeds” of the microbial world and constitute <1% of all microbial species (Hugenholtz 2002). For example, most of the isolates cultured from soil samples belong to one of four phyla (the “big four”), Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria, primarily due to their ease of cultivation under laboratory conditions. Although Acidobacteria constitutes on average 20% of soil bacterial communities, these organisms are difficult to culture and are represented by few genera (Schloss and Handelsman 2004). These findings suggest that molecular techniques that circumvent the need for isolation and cultivation are highly desirable for in-depth characterization of environmental microbial communities.

### 2.3 Molecular Methods of Microbial Community Analyses

The vast majority of microbial communities in nature have not been cultured in the laboratory. Therefore, the primary source of information for these uncultured but viable organisms is their biomolecules such as nucleic acids, lipids, and proteins. Culture-independent nucleic acid approaches include analyses of whole genomes or selected genes such as 16S and 18S rRNA (ribosomal RNA) for prokaryotes and eukaryotes, respectively. Based on the comparative analyses of these rRNA signatures, cellular life has been classified into three primary domains: one eukaryotic (Eukarya) and two prokaryotic (Bacteria and Archaea) (Hugenholtz 2002). Over the last few decades, the field of microbial ecology has seen tremendous progress, and a wide variety of molecular techniques have been developed for describing and characterizing the phylogenetic and functional diversity of microorganisms (Fig. 2.1). Broadly, these techniques have been classified into two major categories depending on their capability of revealing the microbial diversity structure and function: (1) partial community analysis approaches and (2) whole community analysis approaches.



**Fig. 2.1** Culture-independent molecular toolbox to characterize the structural and functional diversity of microorganisms in the environment

### 2.3.1 Partial Community Analysis Approaches

These strategies generally include polymerase chain reaction (PCR)-based methods where total DNA/RNA extracted from an environmental sample is used as a template for the characterization of microorganisms. In principle, the PCR product thus generated reflects a mixture of microbial gene signatures from all organisms present in a sample, including the VBNC fraction. PCR amplification of conserved genes such as 16S rRNA from an environmental sample has been used extensively in microbial ecology primarily because these genes (1) are ubiquitous, i.e., present in all prokaryotes, (2) are structurally and functionally conserved, and (3) contain variable and highly conserved regions (Hugenholtz 2002). In addition, the suitable gene size (~1,500 bp) and growing number of 16S rRNA sequences available for comparison in sequence databases make it a “gold standard” choice in microbial ecology. By estimating the phylogenetic relatedness to known microorganisms based on the homology of 16S rRNA sequences, the closest affiliation of a new isolate or molecular sequence is assigned. Other conserved genes such as RNA polymerase beta subunit (*rpoB*), gyrase beta subunit (*gyrB*), recombinase A (*recA*), and heat shock protein (*hsp60*) have also been used in microbial investigations and to differentiate some bacterial species (Ghebremedhin et al. 2008). The PCR products amplified from environmental DNA are analyzed primarily by (1) clone library method, (2) genetic fingerprinting, (3) DNA microarrays, or by a combination of these techniques.

#### 2.3.1.1 Clone Library Method

The most widely used method to analyze PCR products amplified from an environmental sample is to clone and then sequence the individual gene fragments (DeSantis et al. 2007). The obtained sequences are compared to known sequences in a database such as GenBank, Ribosomal Database Project (RDP), and Greengenes. Typically, cloned sequences are assigned to phylum, class, order, family, subfamily, or species at sequence similarity cut-off values of 80, 85, 90, 92, 94, or 97%, respectively (DeSantis et al. 2007). While clone libraries of 16S rRNA genes permit an initial survey of diversity and identify novel taxa, studies have shown that environmental samples like soil may require over 40,000 clones to document 50% of the richness (Dunbar et al. 2002). However, typical clone libraries of 16S rRNA genes contain fewer than 1,000 sequences and therefore reveal only a small portion of the microbial diversity present in a sample. A cloning-and-sequencing method was used to decipher the microbial community composition in mining-impacted deep subsurface soils of the former Homestake gold mine of South Dakota, USA (Rastogi et al. 2009). Phylogenetic analysis of 230 clone sequences could reveal only a partial view of phylogenetic breadth present in soil samples. Rarefaction analyses of clone libraries generated nonasymptotic plots, which indicated that diversity was not exhaustively sampled due to insufficient clone sequencing, a common problem when assessing environmental microbial

diversity using cloning approaches. Despite its limitations (e.g., labor-intensive, time-consuming, and cost factor), clone libraries are still considered the “gold standard” for preliminary microbial diversity surveys (DeSantis et al. 2007). With the advent of newer and inexpensive sequencing methods, great progress is expected in this method of microbial diversity analysis.

### 2.3.1.2 Genetic Fingerprinting Techniques

Genetic fingerprinting generates a profile of microbial communities based on direct analysis of PCR products amplified from environmental DNA (Muyzer 1999). These techniques include DGGE/TTGE, SSCP, RAPD, ARDRA, T-RFLP, LH-PCR, RISA, and RAPD and produce a community fingerprint based on either sequence polymorphism or length polymorphism. In general, genetic fingerprinting techniques are rapid and allow simultaneous analyses of multiple samples. Fingerprinting approaches have been devised to demonstrate an effect on microbial communities or differences between microbial communities and do not provide direct taxonomic identities. The “fingerprints” from different samples are compared using computer-assisted cluster analysis by software packages such as GelCompar, and community relationships are inferred. Community fingerprints are scored as present or absent, and the similarities among samples are determined using Jaccards’ coefficient.

#### Denaturing- or Temperature-Gradient Gel Electrophoresis

In denaturing-gradient gel electrophoresis (DGGE), the PCR products are obtained from environmental DNA using primers for a specific molecular marker (e.g., 16S rRNA gene) and electrophoresed on a polyacrylamide gel containing a linear gradient of DNA denaturant such as a mixture of urea and formamide (Muyzer et al. 1993). Temperature-gradient gel electrophoresis (TTGE) is based on the same principle of DGGE except that a temperature gradient rather than chemical denaturant is applied. Sequence variation among different PCR amplicons determines the melting behavior, and therefore amplicons with different sequences stop migrating at different positions in the gel. Both DGGE and TTGE involve the use of a 5'-GC clamped (30–50 nucleotides) forward primer during the PCR step. This is essential to prevent the two DNA strands from complete dissociation into single strands during electrophoresis. For determining the phylogenetic identities from DGGE/TGGE fingerprints, the bands can be excised from the gel, reamplified, and sequenced or blotted onto nylon membranes and hybridized to molecular probes specific for different taxonomic groups. DGGE profiles generated using universal bacterial primers from soil microbial communities are generally very complex. In order to overcome this problem, group-specific PCR-DGGE with primers targeting only specific physiological/phylogenetic groups has been used (Mühling et al. 2008). The other problems associated with DGGE/TGGE are as follows: (1) limited sequence information (<500 bp) obtained for phylogenetic analysis from DNA

bands, (2) different DNA fragments may have similar melting points, (3) number of different DNA fragments, which can be separated by polyacrylamide gel electrophoresis (PAGE), and (4) sequence heterogeneity among multiple rRNA operons of one bacterium, leading to multiple bands in DGGE, which might overestimate the diversity. DGGE analysis has been used to screen the unique clones in clone libraries based on distinct patterns and determining the number of operational taxonomic units (OTUs). In a microbial community investigation, DGGE was applied to soils collected from different agricultural fields in Norway and the USA that were under different agronomic treatments (crop rotation and tillage) (Nakatsua et al. 2000). Of these soil samples, one was also highly contaminated by polyaromatic hydrocarbons (PAH, 700 mg kg<sup>-1</sup>). DGGE profiles were generated using primers based on V3 and V6/V9 regions for the bacterial population and V3 region of 16S rRNA for archaeal communities. Results showed that bacterial diversity was far greater than archaeal diversity except for the PAH-contaminated soil sample.

### Single-Strand Conformation Polymorphism

In single-strand conformation polymorphism (SSCP), the environmental PCR products are denatured followed by electrophoretic separation of single-stranded DNA fragments on a non-denaturing polyacrylamide gel (Schwieger and Tebbe 1998). Separation is based on subtle differences in sequences (often a single base pair), which results in a different folded secondary structure leading to a measurable difference in mobility in the gel. Unlike DGGE, SSCP technology does not require any GC clamped primers, gradient gels, or specialized electrophoretic apparatus; therefore, it is a more simple and straightforward technique than DGGE. Similar to DGGE, the DNA bands can be excised from the gel, reamplified, and sequenced. However, SSCP is well suited only for small fragments (between 150 and 400 bp) (Muyzer 1999). A major limitation of the SSCP method is the high rate of reannealing of DNA strands after an initial denaturation during electrophoresis, which can be overcome using a phosphorylated primer during PCR, followed by specific digestion of the phosphorylated strand with lambda exonuclease. SSCP has successfully been employed to differentiate the pure cultures of *Bacillus subtilis*, *Pseudomonas fluorescens*, and *Sinorhizobium meliloti* isolated from soil samples (Schwieger and Tebbe 1998). These authors have also applied SSCP for the analysis of rhizosphere bacterial communities associated with two different plant species, *Medicago sativa* and a common weed *Chenopodium album*. Their results showed that each plant harbored distinct rhizosphere bacterial communities despite the fact that both plants were growing in the same soil.

### Random Amplified Polymorphic DNA and DNA Amplification Fingerprinting

Random amplified polymorphic DNA (RAPD) and DNA amplification fingerprinting (DAF) techniques utilize PCR amplification with a short (usually ten nucleotides)

primer, which anneals randomly at multiple sites on the genomic DNA under low annealing temperature, typically  $\leq 35^{\circ}\text{C}$  (Franklin et al. 1999). These methods generate PCR amplicons of various lengths in a single reaction that are separated on agarose or polyacrylamide gel depending on the genetic complexity of the microbial communities. Because of the high speed and ease of use, RAPD/DAF has been used extensively in fingerprinting overall microbial community structure and closely related bacterial species and strains (Franklin et al. 1999). Both RAPD and DAF are highly sensitive to experimental conditions (e.g., annealing temperature,  $\text{MgCl}_2$  concentration) and quality and quantity of template DNA and primers. Thus, several primers and reaction conditions need to be evaluated to compare the relatedness between microbial communities and obtain the most discriminating patterns between species or strains. A RAPD profiling study was used with 14 random primers to assess changes in microbial diversity in soil samples that were treated with pesticides (triazolone) and chemical fertilizers (ammonium bicarbonate) (Yang et al. 2000). RAPD fragment richness data demonstrated that pesticide-treated soil maintained an almost identical level of diversity at the DNA level as the control soil (i.e., without contamination). In contrast, chemical fertilizer caused a decrease in the DNA diversity compared to control soil.

#### Amplified Ribosomal DNA Restriction Analysis

Amplified ribosomal DNA restriction analysis (ARDRA) is based on DNA sequence variations present in PCR-amplified 16S rRNA genes (Smit et al. 1997). The PCR product amplified from environmental DNA is generally digested with tetracutter restriction endonucleases (e.g., *AluI*, and *HaeIII*), and restricted fragments are resolved on agarose or polyacrylamide gels. Although ARDRA provides little or no information about the type of microorganisms present in the sample, the method is still useful for rapid monitoring of microbial communities over time, or to compare microbial diversity in response to changing environmental conditions. ARDRA is also used for identifying the unique clones and estimating OTUs in environmental clone libraries based on restriction profiles of clones (Smit et al. 1997). One of the major limitations of ARDRA is that restriction profiles generated from complex microbial communities are sometimes too difficult to resolve by agarose/PAGE. The ARDRA technique was applied for assessing the effect of copper contamination on the microbial communities in soil. Whole community ARDRA profiles showed a lower diversity in copper-contaminated soil compared with control soil with no contamination (Smit et al. 1997).

#### Terminal Restriction Fragment Length Polymorphism

Terminal restriction fragment length polymorphism (T-RFLP) is similar to ARDRA except for one major difference, which is the use of one 5' fluorescently labeled primer during the PCR reaction. The resulting PCR products are digested with



restriction enzyme(s), and terminal restriction fragments (T-RFs) are separated on an automated DNA sequencer (Thies 2007). Only the terminally fluorescent labeled restriction fragments are detected, thus simplifying the banding pattern and allowing analysis of complex microbial communities. Community diversity is estimated by analyzing the size, numbers, and peak heights of resulting T-RFs. Each T-RF is assumed to represent a single OTU or ribotype. With recent developments in bioinformatics, several Web-based T-RFLP analysis programs have been developed, which enable researchers to rapidly assign putative identities based on a database of fragments produced by known 16S rDNA sequences. Similar to ARDRA, a T-RFLP pattern is characteristic of the restriction enzyme(s) used, and more than two enzymes should typically be applied. One pitfall of T-RFLP method is that it underestimates community diversity because only a limited number of bands per gel (generally <100) can be resolved, and different bacterial species can share the same T-RF length (OTU overlap or OTU homoplasy). Nonetheless, the method does provide a robust index of community diversity, and T-RFLP results are generally very well correlated with the results from clone libraries (Fierer and Jackson 2006). Fierer and Jackson (2006) applied the T-RFLP technique to understand the biogeographical patterns in soil bacterial communities and to investigate the biotic and abiotic factors that shape the composition and diversity of bacterial communities. They collected 98 soil samples from across North and South America representing a wide range of temperature, pH, and other geographical conditions. Their results demonstrated that bacterial diversity was higher in neutral soils compared to acidic soils and was unrelated to factors such as site temperature, latitude, and other variables that typically act as good predictors of animal and plant diversity.

### Length Heterogeneity PCR

Length heterogeneity PCR (LH-PCR) analysis is similar to the T-RFLP method except that the latter detects amplicon length variations that are produced after restriction digestion, whereas in LH-PCR different microorganisms are discriminated based on natural length polymorphisms that occur due to mutation within genes (Mills et al. 2007). Amplicon LH-PCR interrogates the hypervariable regions present in 16S rRNA genes and produces a characteristic profile. LH-PCR utilizes a fluorescent dye-labeled forward primer, and a fluorescent internal size standard is run with each sample to measure the amplicon lengths in base pairs. The intensity (height) or area under the peak in the electropherogram is proportional to the relative abundance of that particular amplicon. One advantage of using LH-PCR over the T-RFLP is that the former does not require any restriction digestion and therefore PCR products can be directly analyzed by a fluorescent detector. The limitations of LH-PCR technique include inability to resolve complex amplicon peaks and underestimation of diversity, as phylogenetically distinct taxa may produce same-length amplicons (Mills et al. 2007). LH-PCR was used in combination with fatty acid methyl ester (FAME) analysis to investigate the microbial communities in soil samples that differed in terms of type and/or crop management practices

(Ritchie et al. 2000). LH-PCR results strongly correlated with FAME analysis and were highly reproducible, and successfully discriminated different soil samples. The most abundant bacterial community members, based on cloned LH-PCR products, were members of the  $\beta$ -Proteobacteria, Cytophaga–Flexibacter–Bacteriodes, and the high-G+C-content Gram-positive bacterial group.

### Ribosomal Intergenic Spacer Analysis

Ribosomal intergenic spacer analysis (RISA) involves PCR amplification of a portion of the intergenic spacer region (ISR) present between the small (16S) and large (23S) ribosomal subunits (Fisher and Triplett 1999). The ISR contains significant heterogeneity in both length and nucleotide sequence. By using primers annealing to conserved regions in the 16S and 23S rRNA genes, RISA profiles can be generated from most of the dominant bacteria existing in an environmental sample. RISA provides a community-specific profile, with each band corresponding to at least one organism in the original community. The automated version of RISA is known as ARISA and involves use of a fluorescence-labeled forward primer, and ISR fragments are detected automatically by a laser detector. ARISA allows simultaneous analysis of many samples; however, the technique has been shown to overestimate microbial richness and diversity (Fisher and Triplett 1999). Ranjard et al. (2001) evaluated ARISA to characterize the bacterial communities from four types of soil differing in geographic origins, vegetation cover, and physicochemical properties. ARISA profiles generated from these soils were distinct and contained several diagnostic peaks with respect to size and intensity. Their results demonstrated that ARISA is a very effective and sensitive method for detecting differences between complex bacterial communities at various spatial scales (between- and within-site variability).

#### 2.3.1.3 DNA Microarrays

DNA microarrays have been used primarily to provide a high-throughput and comprehensive view of microbial communities in environmental samples. The PCR products amplified from total environmental DNA is directly hybridized to known molecular probes, which are attached on the microarrays (Gentry et al. 2006). After the fluorescently labeled PCR amplicons are hybridized to the probes, positive signals are scored by the use of confocal laser scanning microscopy. The microarray technique allows samples to be rapidly evaluated with replication, which is a significant advantage in microbial community analyses. In general, the hybridization signal intensity on microarrays is directly proportional to the abundance of the target organism. Cross hybridization is a major limitation of microarray technology, particularly when dealing with environmental samples. In addition, the microarray is not useful in identifying and detecting novel prokaryotic taxa. The ecological importance of a genus could be completely ignored if the genus does not have a corresponding probe on the microarray. DNA microarrays used in microbial

ecology could be classified into two major categories depending on the probes: (1) 16S rRNA gene microarrays and (2) functional gene arrays (FGA).

### 16S rRNA gene Microarrays (PhyloChip)

The universal high-density 16S microarray contains about 30,000 probes of 16S rRNA gene targeted to several cultured microbial species and “candidate divisions” (DeSantis et al. 2007). These probes targets all 121 demarcated prokaryotic orders and allow simultaneous detection of 8,741 bacterial and archaeal taxa. PhyloChip technology has been used for rapid profiling of environmental microbial communities during bioterrorism surveillance, bioremediation, climate change, and source tracking of pathogen contamination (Brodie et al. 2007; DeSantis et al. 2007). PhyloChips were used to investigate the indigenous soil bacterial communities in two abandoned uranium mine sites, the Edgemont and the North Cave Hills in South Dakota (Rastogi et al. 2010). PhyloChip analysis revealed greater diversity than corresponding clone libraries at each taxonomic level and indicated the existence of 1,300–1,700 bacterial species in uranium mine soil samples. Most of these species were members of the phylum Proteobacteria and contained lineages that were capable of performing uranium immobilization and metal reduction.

### Functional Gene Arrays

Unlike PhyloChips that are useful in detecting microbial community composition and contain 16S rRNA genes as probes, FGA are designed primarily to detect specific metabolic groups of bacteria. Thus, FGA not only reveal the community structure, but they also shed light on the in situ community metabolic potential. FGA contain probes from genes with known biological functions; therefore, they are also useful in linking microbial community composition to ecosystem functions. For instance, an FGA termed GeoChip contains >24,000 probes from all known metabolic genes involved in various biogeochemical, ecological, and environmental processes such as ammonia oxidation, methane oxidation, and nitrogen fixation (He et al. 2007). GeoChips have been used to interrogate the role of Antarctica soil microbial communities in the global biogeochemical cycling of carbon and nitrogen (Yergeau et al. 2009). Their study demonstrated a significant correlation between the distribution of key genes and soil temperature, chemical characteristics, and vegetation cover. For example, the relative detection of cellulose degradation genes was correlated with temperature, and microbial carbon-fixation genes were found in greater abundance in samples without vegetation.

#### 2.3.1.4 Quantitative PCR

Quantitative PCR (Q-PCR), or real-time PCR, has been used in microbial investigations to measure the abundance and expression of taxonomic and functional gene

markers (Bustin et al. 2005; Smith and Osborn 2009). Unlike traditional PCR, which relies on end-point detection of amplified genes, Q-PCR uses either intercalating fluorescent dyes such as SYBR Green or fluorescent probes (TaqMan) to measure the accumulation of amplicons in real time during each cycle of the PCR. Software records the increase in amplicon concentration during the early exponential phase of amplification which enables the quantification of genes (or transcripts) when they are proportional to the starting template concentration. When Q-PCR is coupled with a preceding reverse transcription (RT) reaction, it can be used to quantify gene expression (RT-Q-PCR). Q-PCR is highly sensitive to starting template concentration and measures template abundance in a large dynamic range of around six orders of magnitude. Several sets of 16S and 5.8S rRNA gene primers have been designed for rapid Q-PCR based quantification of soil bacterial and fungal microbial communities (Fierer et al. 2005). Q-PCR has also been successfully used in environmental samples for quantitative detection of important physiological groups of bacteria such as ammonia oxidizers, methane oxidizers, and sulfate reducers by targeting *amoA*, *pmoA*, and *dsrA* genes, respectively (Foti et al. 2007). Kolb et al. (2003) estimated the abundance of total methanotrophic population and specific groups of methanotrophs in a flooded rice field soil by Q-PCR assay of the *pmoA* genes. The total population of methanotrophs was found to be  $5 \times 10^6$  *pmoA* molecules  $g^{-1}$ , and *Methylosinus* ( $2.7 \times 10^6$  *pmoA* molecules  $g^{-1}$ ) and *Methylobacter/Methylosarcina* groups ( $2.0 \times 10^6$  *pmoA* molecules  $g^{-1}$ ) were the dominant methanotrophs. The *Methylocapsa* group was below the detection limit of Q-PCR ( $1.9 \times 10^4$  *pmoA* molecules  $g^{-1}$ ).

### 2.3.1.5 Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) enables in situ phylogenetic identification and enumeration of individual microbial cells by whole cell hybridization with oligonucleotide probes (Amann et al. 1995). A large number of molecular probes targeting 16S rRNA genes have been reported at various taxonomic levels (Amann et al. 1995). The FISH probes are generally 18–30 nucleotides long and contain a fluorescent dye at the 5' end that allows detection of probe bound to cellular rRNA by epifluorescence microscopy. In addition, the intensity of fluorescent signals is correlated to cellular rRNA contents and growth rates, which provide insight into the metabolic state of the cells. FISH can be combined with flow cytometry for a high-resolution automated analysis of mixed microbial populations. The FISH method was used to follow the dynamics of bacterial populations in agricultural soils treated with s-triazine herbicides (Caracciolo et al. 2010). A variety of molecular probes were used to target specific phylogenetic groups of bacteria such as  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  subdivisions of Proteobacteria and Planctomycetes. Results demonstrated that  $\gamma$ -Proteobacteria populations diminished sharply after 14 days of incubation in treated soil compared to control soil with no s-triazine treatment. In contrast,  $\beta$ -Proteobacteria populations remained higher than that of the control soils throughout the incubation period (70 days). Other bacterial groups, e.g.,  $\alpha$ -Proteobacteria and Planctomycetes were not significantly affected by the presence of the herbicide.

Low signal intensity, background fluorescence, and target inaccessibility are commonly encountered problems in FISH analysis. In the last few years, extensive improvements have been made to solve some of these problems which include the use of brighter fluorochromes, chloramphenicol treatment to increase the rRNA content of active bacterial cells, hybridization with probes carrying multiple fluorochromes, and signal amplification with reporter enzymes (Rogers et al. 2007). In a modified FISH method known as *catalyzed reporter deposition* (CARD) FISH, the hybridization signal is enhanced through the use of tyramide-labeled fluorochromes (Pernthaler et al. 2002). This allows the accumulation of several fluorescent probes at the target site, which ultimately increases the signal intensity and sensitivity. Li et al. (2008) developed an advanced imaging technique by combining FISH to secondary-ion mass spectrometry (SIMS). In principle, the technique uses 16S rRNA probes for in situ hybridization; however, the probes are labeled with a stable isotope or element (e.g., fluorine or bromine atoms) rarely present in biomass. Once the probe is hybridized, the microbial identities of stable isotope-labeled cells are simultaneously determined in situ by NanoSIMS imaging. With next-generation SIMS instruments, spatial resolution of ~50 nm (NanoSIMS) was achieved, which allowed quantifying the isotopic composition at single-cell level.

### 2.3.1.6 Microbial Lipid Analysis

Microbial community characterization by biomolecules other than nucleic acids such as lipids has been used without relying on culturing (Banowetz et al. 2006). Fatty acids are present in a relatively constant proportion of the cell biomass, and signature fatty acids exist in microbial cells that can differentiate major taxonomic groups within a community. The fatty acids are extracted by saponification followed by derivatization to give the respective FAMES, which are then analyzed by gas chromatography. The emerging pattern is then compared to a reference FAME database to identify the fatty acids and their corresponding microbial signatures by multivariate statistical analyses. FAME profiling and multivariate statistical methods were used to identify the sources of soil that were contaminating surface waters (Banowetz et al. 2006). A variety of reference soils collected from land with contrasting uses in different seasons was used to generate FAME fingerprints for reliable classification of soils. FAME fingerprints generated from different soil samples were capable of discriminating reference soils. Results showed that FAME analysis can successfully classify sediment samples provided soil FAME profiles are developed for reference soils collected at the same time as surface water samples.

### 2.3.2 Whole Community Analysis Approaches

Sequence analysis of 16S rRNA genes is commonly used in most microbial ecological surveys. However, being a highly conserved molecule, the 16S rRNA gene does not provide sufficient resolution at species and strain level (Konstantinidis

et al. 2006). Whole-genome molecular techniques offer a more comprehensive view of genetic diversity compared to PCR-based molecular approaches that target only a single or few genes. These techniques attempt to analyze all the genetic information present in total DNA extracted from an environmental sample or pure culture.

### 2.3.2.1 DNA–DNA Hybridization Kinetics

Whole-genome DNA–DNA hybridization (DDH) offers true genome-wide comparison between organisms. A value of 70% DDH was proposed as a recommended standard for bacterial species delineation (Goris et al. 2007). Typically, bacterial species having 70% or greater genomic DNA similarities usually have >97% 16S rRNA gene sequence identity. Although DDH techniques have been originally developed for pure culture comparisons, they have been modified for use in whole microbial community analysis. In DDH technique, total community DNA extracted from an environmental sample is denatured and then incubated under conditions that allow them to hybridize or reassociate. The rate of DNA reassociation is correlated with the genomic complexity (diversity) present in the sample. If the sample has high sequence diversity, the rate of DNA reassociation will decrease. Under defined conditions, the time needed for half of the DNA to reassociate (the half association value  $C_0t$ , where  $C_0$  is the concentration of single-stranded DNA at time zero and  $t$  is time) is proportional to genomic diversity and can be used as a diversity index. Based on DDH data, 6,000–10,000 different prokaryotic genomes per gram of soil have been suggested (Torsvik and Øvreås 2002). This number could be much higher as genomes representing rare and unrecovered species might have been overlooked in the analysis.

### 2.3.2.2 Guanine-Plus-Cytosine Content Fractionation

Different prokaryotic groups differ in their guanine-plus-cytosine (G+C) content of DNA, and phylogenetically related bacterial groups only vary by 3–5% in their G+C content (Nüsslein and Tiedje 1999). Thus, the fractionation of total community DNA can be achieved by density-gradient centrifugation based on G+C content. The technique generates a fractionated profile of the entire community DNA and indicates relative abundance of DNA (hence taxa) as a function of G+C content. The total community DNA is physically separated into highly purified fractions, each representing a different G+C content that can be analyzed by additional molecular techniques such as DGGE/ARDRA to better assess total community diversity. However, G+C content fractionation technique provides a coarse level of phylogenetic resolution as different phylogenetic groups may have the same G+C range. Additionally, it requires a large amount of DNA (about 50 µg) and a total time of about 4 days for completion. G+C fractionation has been widely applied in investigation of soil microbial communities to evaluate the effect of different treatments or management practices (e.g., change in vegetation, grazing, application of

pesticides, and compost application). Nüsslein and Tiedje (1999) applied G+C fractionation together with ARDRA and 16S rRNA gene sequence analyses to investigate the influence of forest versus pasture vegetation in Hawaiian soil microbial communities. All three techniques demonstrated that plants are an important determinant of microbial community structure and shift in vegetative cover to pasture resulted in about 50% change in the microbial community composition.

### 2.3.2.3 Whole-Microbial-Genome Sequencing

Exploring microbial systems through whole-genome analysis is a comprehensive and integrated approach to understand microbial ecology and function. Whole microbial genomes are sequenced using a shotgun cloning method that involves (1) extraction of DNA from pure cultures, (2) random fragmentation of obtained genomic DNA into small fragments of ~2 kb, (3) ligation and cloning of DNA fragments into plasmid vectors, and (4) bidirectional sequencing of DNA fragments. Once the sequences are obtained, they are aligned and assembled into finished sequences using specialized computer programs such as MEGAN (*MEtaGenome ANalyzer*) (Huson et al. 2007). The sequences are annotated in open reading frames (ORFs) to predict the encoded proteins (functions). Whole-genome sequencing has provided unprecedented insights into microbial processes at the molecular level and has potential applications in individual and community ecology, bioenergy production, bioremediation, human and plant health, and various industries (Ikeda et al. 2003). Several institutions and laboratories such as The Institute of Genome Research, the U.S. Department of Energy's Joint Genome Institute, Lawrence Berkeley National Laboratory, and J. Craig Venter Institute have completed sequencing of whole genomes of several important microorganisms such as *Pseudomonas syringae* DC3000 (a plant pathogen), *Desulfovibrio desulfuricans* G20 (bioremediation capabilities), and *Methanosaeta thermophila* (a thermophilic aceticlastic methanogen). The genome sequence of *Desulfovibrio desulfuricans* G20, a model sulfate-reducing  $\delta$ -proteobacterium, demonstrated the existence of metabolic pathways by which these bacteria are able to reduce toxic metals such as uranium(VI) and chromium(VI) to less water-soluble species (Li et al. 2009). These molecular insights were highly crucial for the use of sulfate-reducing bacteria in bioremediation of metal-contaminated groundwater or soils. Recent developments in short-read sequencing techniques such as pyrosequencing have dramatically reduced the time and cost needed for whole-microbial-genome sequencing projects (Metzker 2010). The enormous amount of data gathered from genome sequencing programs is deposited in searchable databases that could be mined with various powerful bioinformatic tools available at the Integrated Microbial Genomes (IMG) Web server (Markowitz et al. 2010) for evolutionary studies, comparative genomics, and proteomics. For example, Microbial Genomes Resources at the National Center for Biotechnology Information (NCBI) is a public database for prokaryotic genome sequencing projects and has now 1,000 complete prokaryotic genomes (<http://www.ncbi.nlm.nih.gov/genomes/> [verified on 15th May, 2010]). The Genomes

Online Database (GOLD) is another database resource for comprehensive information regarding complete and ongoing genome projects, as well as metagenomes and metadata, around the world (<http://www.genomesonline.org>). As of 15th May, 2010, the GOLD database held 1,284 completed and published genomes and 4,289 ongoing bacterial, 199 archaeal, and 1,338 eukaryotic sequencing projects.

#### 2.3.2.4 Metagenomics

Metagenomics is the investigation of collective microbial genomes retrieved directly from environmental samples and does not rely on cultivation or prior knowledge of the microbial communities (Riesenfeld et al. 2004). Metagenomics is also known by other names such as environmental genomics or community genomics, or microbial ecogenomics. Essentially, metagenomics does not include methods that interrogate only PCR-amplified selected genes (e.g., genetic fingerprinting techniques) as they do not provide information on genetic diversity beyond the genes that are being amplified. In principle, metagenomic techniques are based on the concept that the entire genetic composition of environmental microbial communities could be sequenced and analyzed in the same way as sequencing a whole genome of a pure bacterial culture as discussed in the preceding section. Metagenomic investigations have been conducted in several environments such as soil, the phyllosphere, the ocean, and acid mine drainage and have provided access to phylogenetic and functional diversity of uncultured microorganisms (Handelsman 2004). Thus, metagenomics is crucial for understanding the biochemical roles of uncultured microorganisms and their interaction with other biotic and abiotic factors. Environmental metagenomic libraries have proved to be great resources for new microbial enzymes and antibiotics with potential applications in biotechnology, medicine, and industry (Riesenfeld et al. 2004; Rondon et al. 2000). Metagenomic library construction involves the following steps: (1) isolation of total DNA from an environmental sample, (2) shotgun cloning of random DNA fragments into a suitable vector, and (3) transforming the clones into a host bacterium and screening for positive clones. Metagenomic libraries containing small DNA fragments in the range of 2–3 kb provide better coverage of the metagenome of an environment than those with larger fragments. It has been estimated that to retrieve the genomes from rare members of microbial communities at least  $10^{11}$  genomic clones would be required (Riesenfeld et al. 2004). Small-insert DNA libraries are also useful to screen for phenotypes that are encoded by single genes and for reconstructing the metagenomes for genotypic analysis. Large-fragment metagenomic libraries (100–200 kb) are desirable while investigating multigene biochemical pathways. Metagenomic libraries could be screened either by sequence-driven metagenomic analysis that involves massive high-throughput sequencing or by functional screening of expressed phenotypes. Sequence-driven massive whole-genome metagenomic sequencing sheds light on many important genomic features such as redundancy of functions in a community, genomic organizations, and traits that are acquired from distinctly related taxa through horizontal gene transfers (Handelsman 2004).



In function-driven metagenomic analysis (functional metagenomics), libraries are screened based on the expression of a selected phenotype on a specific medium. A wide variety of biochemical activities have been discovered in environmental metagenomic libraries. For example, novel antibiotics (e.g., turbomycin, terragine), microbial enzymes (e.g., cellulases, lipases, amylases), and proteins (e.g., antiporters) have been identified in soil metagenomic libraries (Rondon et al. 2000). Function-driven metagenomic approaches require successful expression of a desired gene in a heterologous host such as *E. coli*. Thus, a major limitation is very low level or no expression of the majority of environmental genes in *E. coli*. In some cases, improved gene expression can be achieved by transforming metagenomic DNA into several additional surrogate hosts such as *Streptomyces*, *Bacillus*, *Pseudomonas*, and *Agrobacterium*. Strategies that can enhance heterologous expression of unknown genes in host cells are highly desirable. For example, genetically engineered *E. coli* that can support the translation and transcription of wide diversity of genes, or cloning vectors with strong promoters that can provide additional transcription factors will be highly desirable. In a metagenomic library, the frequency of active gene clones expressing a phenotype is typically very low. For example, in an environmental metagenomic library established from soil, only one in 730,000 clones showed lipolytic activity (Henne et al. 2000). The DNA and inferred protein sequence of a novel lipolytic clone exhibited only a moderate identity (<50%) with known lipases, indicating that it could be from an uncultured organism. Low occurrence of actively expressing clones in metagenomic libraries necessitates improved high-throughput screening and detection assays.

## 2.4 Next-Generation DNA Sequencing Techniques Transform Microbial Ecology

Large-scale sequencing technologies allow us to investigate deeper and deeper layers of the microbial communities and are vital in presenting an unbiased view of phylogenetic composition and functional diversity of environmental microbial communities (Zwolinski 2007). The capability of large-scale sequencing techniques to generate billions of reads at low cost with high speed is useful in many applications such as whole-genome sequencing, metagenomics, metatranscriptomics, and proteogenomics. Recent developments in new sequencing chemistries, bioinformatics, and instruments have revolutionized the field of microbial ecology and genomics. Next-generation sequencing platforms such as Roche/454, Illumina/Solexa, Life/APG, and HeliScope/Helicos BioSciences are much faster and less expensive than traditional Sanger's dideoxy sequencing of cloned amplicons (Metzker 2010). 454Life Sciences commercially developed a 454 pyrosequencing technique, which allows massive parallel high-throughput sequencing of hypervariable regions of 16S rRNA genes and offers two to three orders of magnitude higher coverage of microbial diversity than typical Sanger sequencing of a few hundred 16S rRNA gene clones. The hypervariable regions

targeted are short enough (100–350 bases) but provide sufficient phylogenetic information and are easily covered in the short read lengths generated by pyrosequencing techniques.

One advantage of using the pyrosequencing technique is that multiple environmental samples can be combined in a single run, and after sequencing, the reads can be parsed through their assigned nucleotide barcode, which is added in templates during PCR. The latest release of the third-generation platform 454 Genome Sequencer XLR (GS FLX Titanium) can yield read lengths exceeding >450 bp and approximately 400 million high-quality bases per 10-h instrument run with an accuracy of 99.96% (Metzker 2010). Third-generation sequencing platforms developed by Helicos and Pacific Biosciences are expected to be released in the year 2010 and would be capable of single-molecule sequencing and producing reads exceeding more than 1 kb with an accuracy of >99.99% (Metzker 2010).

Environmental samples such as soil contain huge genetic diversity that encompasses microorganisms from the Eukarya, Bacteria, and Archaea domains. For example, GenBank, the largest database of microbial sequences, provides >686,266 sequence entries when searched for the keyword “soil” (verified on 15 May 2010). This vast genetic information available in databases is the evidence of advances in genomics and increased use of nucleic-acid sequencing. Until recently, first-generation automated Sanger sequencing has been used in most molecular microbial surveys. The major limiting factor in the Sanger technique was the cost and time involved, with the result that most of the studies included sequencing of only few hundred clones. Sequencing of a low number of clones captures only the dominant components of microbial communities that mask the detection of low-abundance microorganisms. These low-abundance microorganisms constitute a highly diverse “rare biosphere” in almost every environmental sample including soil (Lauber et al. 2009). The rare biosphere microbial populations are largely unexplored and offer a potentially inexhaustible genetic reservoir that could be explored only by using next-generation sequencing techniques. In a molecular investigation, spatial changes in soil bacterial communities were explored by targeting V1 and V2 hypervariable regions of 16S rRNA genes using a massive bar-coded pyrosequencing technique (Lauber et al. 2009). Eighty-eight soil samples representing a wide range of ecosystems from across North and South America were collected, and a total of 152,359 high-quality sequences on average of 1,501 sequences per sample were generated. Results suggested enormous phylogenetic diversity in soil microbial communities with an average of at least 1,000 species per soil sample. The dominant phyla in all soil samples were Acidobacteria, Alphaproteobacteria, Actinobacteria, Bacteroidetes, and Beta/Gammaproteobacteria. The Lauber et al. (2009) study demonstrated that even after sequencing more than 1.5 billion 16S rRNA gene amplicons, the full extent of species diversity was not covered. This provided further evidence that soil bacterial communities are extremely diverse and contain a large “rare biosphere” represented by an enormous number of low-abundance unique taxa. Such studies highlight the importance of large-scale sequencing techniques in investigating the highly diverse soil microbial communities.

## 2.5 Functional Microbial Ecology: Linking Community Structure and Function

Understanding how microbial communities function in natural environments is a central goal in microbial ecology. RNA extracted from environmental samples provides more valuable information than DNA in revealing active microbial communities versus dormant microbial communities (Torsvik and Øvreås 2002). This is due to the fact that rRNA and mRNA are considered as indicators of functionally active microbial populations. The amount of rRNA in a cell roughly correlates with the growth activity of bacteria, and mRNA of functional genes allows the detection and identification of bacteria actually expressing key enzyme activities under specific conditions (Wellington et al. 2003). Several genes, e.g., *amoA* (ammonia oxidation), *nifH* (nitrogen fixation), *nirK* and *nirS* (denitrification), and *dsrA* (sulfate reduction), have been amplified from DNA/RNA isolated from microbial communities to obtain insights into key microbial processes (Hansel et al. 2008). Microbial catabolic diversity could also be studied by enzyme-coding genes involved in utilization of specific carbon substrates such as chitin, cellulose, and lipids (Torsvik and Øvreås 2002). The diversity of lipase-producing microorganisms in glacier soil was investigated by the PCR amplification of lipase genes, and sequence analysis showed the existence of several novel lipase-producing organisms in soil (Yuhong et al. 2009). More advanced methods utilizing stable isotopes such as stable isotope probing (SIP), microautoradiography–FISH (MAR–FISH), and Raman–FISH offer more detailed insights into the metabolic activities of microbial communities and are discussed in the following sections.

### 2.5.1 Stable Isotope Probing

SIP involves offering a stable isotope (e.g.,  $^{13}\text{C}$ )-labeled substrate to microbial communities whose utilization is of interest to decipher a key biogeochemical process (Wellington et al. 2003). Active microbial communities that utilize the labeled substrate during growth incorporate the isotopes within their biomass. The labeled biomolecules (e.g., DNA, RNA, phospholipid fatty acids [PLFA]) are then separated from biomass by different biochemical methods, and the phylogenetic identity of microorganisms metabolizing the substrate is established using molecular techniques. SIP relying on DNA biomarkers involves labeling of DNA with  $^{13}\text{C}$  that could be separated from  $^{12}\text{C}$  by CsCl equilibrium density-gradient centrifugation. The  $^{13}\text{C}$ -labeled DNA could be analyzed by genetic fingerprinting or clone library techniques, leading to the identification of microorganisms. SIP was applied to decipher the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D)-degrading soil microbial communities (Cupples and Sims 2007). Soil samples were amended with  $^{13}\text{C}$ -labeled 2,4-D and were incubated for 17 days. After incubation, labeled DNA was purified from soil samples and was used to construct 16S rRNA clone libraries.

Phylogenetic analyses of clone sequences revealed that bacteria belonging to  $\beta$ -Proteobacteria such as Comamonadaceae and *Ramlibacter* were responsible for uptake and degradation of the herbicide.

In recent years, with advances in imaging and spectroscopic techniques, SIP has been combined with other techniques such as FISH and Raman microscopy to simultaneously investigate the taxonomic identities and activity of microbial communities at single-cell resolution (Huang et al. 2007). In the Raman–FISH method, environmental samples are incubated with a substrate labeled with  $^{13}\text{C}$  stable isotope. After incorporation, the spectral profiles of uncultured microbial cells at single-cell resolution are generated using Raman microscopy, which measures the laser light scattered by chemical bonds of different cell biomarkers. The proportion of stable isotope incorporation in cells affects the amount of light scattered, resulting in measurable peak shifts for labeled cellular components. The Raman–FISH provides much higher resolution and overcomes many of the limitations associated with conventional SIP/MAR–FISH techniques. Huang et al. (2007) used the Raman–FISH method to investigate naphthalene-degrading *Pseudomonas* communities in groundwater. Their results, based on differences in  $^{13}\text{C}$  content of the various pseudomonad cells, suggested that different *Pseudomonas* species and even members of the same species vary in their capability of naphthalene degradation.

### 2.5.2 *Microautoradiography*

Microautoradiography (MAR) relies on the fact that metabolically active cells utilizing radiolabeled substrate can be visualized by exposure to radiation-sensitive silver halide emulsion (Okabe et al. 2004). The emulsion is placed on the top of cells that are mounted on a microscope slide. After exposure, excited silver ions precipitate as black grains of metallic silver inside or adjacent to the cells that can be observed by transmission electron microscopy. Commonly used radiolabeled substrates include glucose, acetate, and amino acids, which provide a general view of the overall metabolic diversity. More specific substrates along with selective growth (incubation) conditions have been used to identify important physiological processes in situ. For example, radiolabeled iron or sulfate can be provided under controlled anaerobic conditions to identify the iron- and sulfate-reducing microbial communities, respectively. When MAR is used in combination with FISH (MAR–FISH), it allows simultaneous phylogenetic identification of active cells that consume the radioactive substrate (Rogers et al. 2007). MAR–FISH has been modified slightly, leading to other methods such as STAR (substrate tracking autoradiography)–FISH. However, STAR–FISH differs from MAR–FISH only in methodological details, and the basic principle of the technique remains the same. Nielsen et al. (2003) developed a quantitative MAR (QMAR)–FISH approach that can detect even single cells due to its improved fixation protocol and use of an internal standard of bacteria with known specific radioactivity. MAR–FISH technique was used to study the autotrophic nitrifying bacteria in biofilms (Okabe et al. 2005). The uptake by heterotrophic bacteria

of  $^{14}\text{C}$ -labeled products derived from nitrifying bacteria was directly visualized by MAR–FISH. Results revealed that members belonging to Chloroflexi and Cytophaga–Flavobacterium play an important role in scavenging the dead biomass and metabolites of nitrifying bacteria and ultimately preventing the accumulation of organic waste products in the biofilms.

### 2.5.3 Isotope Array

Isotope arrays allow for functional and phylogenetic screening of active microbial communities in a high-throughput fashion. The technique uses a combination of SIP for monitoring the substrate uptake profiles and microarray technology for deciphering the taxonomic identities of active microbial communities (Adamczyk et al. 2003). In principle, samples are incubated with a  $^{14}\text{C}$ -labeled substrate, which during the course of growth becomes incorporated into microbial biomass. The  $^{14}\text{C}$ -labeled rRNA is separated from unlabeled rRNA and then labeled with fluorochromes. Fluorescent labeled rRNA is hybridized to a phylogenetic microarray followed by scanning for radioactive and fluorescent signals. The technique thus allows parallel study of microbial community composition and specific substrate consumption by metabolically active microorganisms of complex microbial communities. The major strengths of the technique lie in the fact that it does not involve any amplification step and is hence free of biases associated with PCR. The limitations of the technique include difficulties in obtaining high-quality rRNA and detecting low abundance but active microbial populations from environmental samples (Adamczyk et al. 2003). Adamczyk et al. (2003) successfully used this technique to demonstrate phylogenetic diversity and  $\text{CO}_2$  fixation activity of ammonia-oxidizing bacteria (AOB) in nitrifying activated sludge samples. Their results suggested that *Nitrosomonas* was the dominant lineage in AOB communities of sludge samples.

## 2.6 Postgenomic Approaches

The recent applications of DNA-based molecular techniques such as metagenomics have revealed new insights into the phylogenetic and functional diversity of microbial communities. However, in the postgenomic era, the limitations of DNA-based molecular approaches have been realized. For example, DNA-based techniques do not provide information on the gene expression (functionality) as it occurs under in situ conditions (Wilmes and Bond 2006). With the availability of comprehensive metagenomic databases, which also includes genomic sequences from uncultured microorganisms, it is now possible to apply postgenomic approaches such as metaproteomics and metatranscriptomics to reveal the link between genetic potential and functionality in microbial communities. In the following sections, these techniques are discussed in detail with their potential applications in investigating functionality of microbial communities.

### 2.6.1 *Metaproteomics*

Metaproteomics, also commonly known as environmental proteomics, deals with the large-scale study of proteins expressed by environmental microbial communities at a given point in time (Wilmes and Bond 2006; Keller and Hettich 2009). Compared to other cell molecules such as lipids and nucleic acids, protein biomarkers are more reliable and provide a clearer picture of metabolic functions than functional genes or even the corresponding mRNA transcripts of microbial communities (Wilmes and Bond 2006). Although methods such as SIP/MAR-FISH have been developed for structure–function analyses of microbial communities, these methods reveal information only on microbial communities associated with a specific biogeochemical process (e.g., nitrification, methane oxidation) and do not reveal an overall picture of microbial functionalities. Compared to these methods, proteomics offers a comprehensive approach to investigate the physiology of microbial communities both qualitatively and quantitatively. For example, proteomic profiling of microbial communities provides critical information on protein abundances and protein–protein interactions, which could not be achieved by DNA/RNA molecular techniques such as metatranscriptomics and metagenomics (Keller and Hettich 2009). The physiological responses of microbial communities due to a stress condition could be identified from an altered proteofingerprint, which reflects changes in the functional status of the communities. Once the proteins are identified, they could be linked to corresponding metagenomic sequences to link metabolic functions to individual microbial species.

Methodologically, metaproteome analysis involves extraction of total proteins from an environmental sample. Although in situ protein lysis methods provide an exhaustive recovery, a significant amount of protein originates from other organisms such as protozoa, fungi, and multicellular organisms, which further complicate the taxonomic characterization of proteins (Keller and Hettich 2009). Therefore, in some cases, microbial cells are first separated from the environmental matrix by ultracentrifugation and then lysed, which allows obtaining much higher quality and quantity of bacterial proteins. Once the total protein is obtained, it is separated by one-dimensional and two-dimensional electrophoresis to generate a community proteofingerprint. After separation, protein spots are digested and are identified by a variety of powerful analytical methods. Currently, high-throughput proteomic profiling of microbial communities is possible due to development of chromatographic and mass spectroscopic techniques (MS-based proteomics). High-efficiency mass spectrometry integrated with liquid chromatography allows a highly sensitive and rapid identification of proteins. The availability of Web-based services such as ExPASy (*Expert Protein Analysis System*; <http://www.expasy.org/>) offers a comprehensive suite of tools that are vital in identification and characterization of protein mass fingerprinting data. A metaproteomic approach was employed to identify proteins that were involved in the biodegradation of chlorophenoxy acid in soil samples (Benndorf et al. 2007). Soil samples were first enriched for chlorophenoxy acid-degrading bacteria by incubating with 2,4-D for a

period of 22 days. After incubation, protein extracts were isolated from soil and separated by SDS-PAGE, and protein bands were identified by liquid chromatography linked to mass spectroscopy. Proteomic analysis identified a major catabolic enzyme 2,4-dichlorophenoxyacetate dioxygenase, membrane transport proteins (porins), and molecular chaperones.

### 2.6.2 *Proteogenomics*

In metaproteomics, protein sequences could be identified with confidence only if they have significant homology to existing proteins in available databases. However, in most of the environmental proteomic surveys, proteins are only distantly related to known database sequences. Therefore, it appears that the majority of short protein sequences retrieved from metaproteomes will remain unidentified and cannot be assigned to their functional and phylogenetic features. However, these limitations have been overcome by combining the metaproteomic and metagenomic approaches together under the name of “proteogenomics” (Banfield et al. 2005). In community proteogenomics, total DNA and proteins are extracted from the same sample, which allows linking of biological functions to phylogenetic identity with greater confidence. The metagenomic part of the proteogenomic approach plays a very significant role and increases the identification of protein sequences by metagenomic analysis of the same sample from which the proteins were extracted. The proteogenomics approach was applied to decipher phyllosphere bacterial communities in a study by Delmotte et al. (2009). Bacterial biomass was harvested from leaf surfaces of soybean, clover, and *Arabidopsis*, and proteins were extracted. This was followed by tryptic digestion and separation of fragments by liquid chromatography and analysis by mass spectrometry. This led to the identification of 2,883 unique proteins from nearly one-half million spectra. The metagenomic data generated from the DNA extracted from the same pool of bacterial biomass significantly increased (up to 74%) the number of identified proteins, indicating that the majority of the bacterial communities present in the phyllosphere were genetically distinct from those currently available in databases. Most identified proteins in the phyllosphere proteome were assigned to the three bacterial genera *Methylobacterium*, *Sphingomonas*, and *Pseudomonas*. Large numbers of proteins involved in methanol oxidation were identified and were assigned to *Methylobacterium* species that can use methanol as a source of carbon and energy.

### 2.6.3 *Metatranscriptomics*

Metatranscriptomics (or environmental transcriptomics) allows monitoring of microbial gene expression profiles in natural environments by studying global transcription of genes by random sequencing of mRNA transcripts pooled from

microbial communities at a particular time and place (Moran 2009). Metatranscriptomics is particularly suitable for measuring changes in gene expression and their regulation with respect to changing environmental conditions. The major challenge in metatranscriptomics is the fact that prokaryotic microbial mRNA transcripts are not polyA tailed, so obtaining complementary DNA (cDNA) is not easy. This results in coextraction of more abundant rRNA molecules in the total RNA pool, which can lead to overwhelming background sequences in a large-scale sequencing analysis. A method for selectively enriching mRNA by subtractive hybridization of rRNA has been developed and evaluated for the gene transcript analysis of marine and freshwater bacterioplankton communities, which revealed the presence of many transcripts that were linked to biogeochemical processes such as sulfur oxidation (*soxA*), assimilation of C1 compounds (*fdh1B*), and acquisition of nitrogen via polyamine degradation (*aphA*) (Poretsky et al. 2005). More recently, a “double-RNA” method has been devised to analyze the total RNA pool of a community, as it is naturally rich in not only functionally but also taxonomically relevant molecules, i.e., mRNA and rRNA, respectively (Urich et al. 2008). This offers a means to investigate both structural and biochemical activity of microbes in a single experiment. Their study combined transcriptomic profiling with massive pyrosequencing techniques to produce 193,219 rRNA tags and 21,133 mRNA-tags from sandy soil samples that were poor in nutrients and neutral in pH. The rRNA tags provided data on the phylogenetic composition of soil microbial communities and showed that Actinobacteria and Proteobacteria were most abundant, while Crenarchaeota were less abundant in soil samples. The mRNA tags provided a glimpse of the in situ expression of several key metabolic enzymes such as ammonia monoxygenase (*amoA* and *amoC*) and nitrite reductase (*nirk*) that were involved in ammonia oxidation. In addition, microbial gene transcripts coding for the enzymes methyl-malonyl-CoA mutase and 4-hydroxybutyryl-CoA dehydratase that play a role in CO<sub>2</sub> fixation pathways in Crenarchaeota were detected.

## 2.7 Bias in Molecular Community Analysis Methods

Like culture methods, molecular techniques have their own pitfalls and are associated with bias at every step (von Wintzingerode et al. 1997). Biases associated with DNA extraction such as incomplete or preferential lysis of certain microbial cells can distort the community composition, richness, and microbial community structure. Feinstein et al. (2009) suggested the use of several validated DNA extraction methods and pooled DNA extracts in PCR-based molecular methods to minimize any risk of bias. Biases associated with PCR could include inhibition by compounds such as humic acids, which are generally coextracted with DNA extracted from soil. Several DNA purification steps have been devised; however, they lead to loss of DNA during purification, which also causes bias in subsequent PCR. Dilution of DNA templates or dialysis can be applied, but it influences the PCR efficiency. Hybridization efficiency and specificity of primers sometimes cause preferential amplification of certain templates, which affects the quantitative assessment of



microbial diversity. Formations of PCR artifacts (e.g., chimeric molecules, deletion mutants, and point mutants) could also lead to misleading results (von Wintzingerode et al. 1997).

## 2.8 Concluding Remarks and Future Directions

With the development and application of molecular genomic tools, the field of microbial ecology is undergoing unprecedented changes. Postgenomic molecular approaches enable us to interrogate the structural and functional diversity of environmental microbial communities and reveal that we have only scratched the surface of the genetic and metabolic diversity present in the most abundant organisms of the Earth, the *Prokaryotes*. Several important questions such as “How many microbial species are there on the Earth?”, “What is the extent of metabolic diversity in natural microbial communities?”, and “How microbial communities are governed by biological, chemical and physical factors?” remain to be understood. Understanding the functional roles of uncultured organisms still remains a daunting task, as most of the genes identified have no homologous representatives in databases. Although considerable progress has been made in the characterization of microbial communities by the application of metagenomic, metatranscriptomic, and proteogenomic approaches, many technical challenges remain including DNA, RNA, and protein extraction from environmental samples, mRNA instability, and low abundance of certain gene transcripts in total RNA. The next-generation sequencing techniques are still developing, and many technological innovations particularly tuned for environmental samples are expected in these techniques. Development in bioinformatics tools is also needed for evaluating the tremendous amount of information generated through whole-genome analysis and metagenomic and metatranscriptomics approaches. Quantitative assessment of microbial communities is the greatest challenge due to significant biases associated with nucleic acid isolation and PCR and requires more advanced DNA/RNA extraction techniques for environmental samples. All of the molecular approaches available for community structure and function analysis have advantages and limitations associated with them, and none provides complete access to the genetic and functional diversity of complex microbial communities. A combination of several techniques should be applied to interrogate the diversity, function, and ecology of microorganisms. Culture-based and culture-independent molecular techniques are neither contradictory nor excluding and should be considered as complementary. An interdisciplinary systems approach embracing several “omics” technologies to reveal the interactions between genes, proteins, and environmental factors will be needed to provide new insights into environmental microbiology. Development of multi-“omics” approaches will be a high-priority area of research in the coming years.

**Acknowledgement** The authors would like to sincerely thank Dr. David E. Cummings, Point Loma Nazarene University, San Diego, CA for providing critical comments and suggestions on this chapter.

## References

- Adamczyk, J., Hesselsoe, M., Iversen, N., Horn, M., Lehner, A., Nielsen, P.H., Schloter, M., Roslev, P., Wagner, M. 2003. The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function. *Appl. Environ. Microbiol.* 69:6875–6887.
- Amann, R.L., Ludwig, W., Schleifer, K.H. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59:143–169.
- Banfield, J.F., Verberkmoes, N.C., Hettich, R.L., Thelen, M.P. 2005. Proteogenomic approaches for the molecular characterization of natural microbial communities. *OMICS* 9:301–333.
- Banowetz, G.M., Whittaker, G.W., Dierksen, K.P., Azevedo, M.D., Kennedy, A.C., Griffith, S.M., Steiner, J.J. 2006. Fatty acid methyl ester analysis to identify sources of soil in surface water. *J. Environ. Qual.* 3:133–140.
- Benndorf, D., Balcke, G.U., Harms, H., von Bergen, M. 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J.* 1:224–234.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., Andersen, G.L. 2007. Urban aerosols harbor diverse and dynamic bacterial populations. *Proc. Natl. Acad. Sci. USA.* 104:299–304.
- Bustin, S.A., Benes, V., Nolan, T., Pfaffl, M.W. 2005. Quantitative real-time RT-PCR – a perspective. *J. Mol. Endocrinol.* 34:597–601.
- Caracciolo, A.B., Bottoni, P., Grenni, P. 2010. Fluorescence *in situ* hybridization in soil and water ecosystems: a useful method for studying the effect of xenobiotics on bacterial community structure. *Toxicol. Environ. Chem.* 92:567–579.
- Cupples, A.M., Sims, G.K. 2007. Identification of *in situ* 2,4 dichlorophenoxyacetic acid-degrading soil microorganisms using DNA-stable isotope probing. *Soil Biol. Biochem.* 39:232–238.
- Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C., Vorholt, J.A. 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc. Natl. Acad. Sci. USA.* 106:16428–16433.
- DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubietta, I.X., Piceno, Y.M., Andersen, G.L. 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb. Ecol.* 53:371–383.
- Dunbar, J., Barns, S.M., Ticknor, L.O., Kuske, C.R. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. *Appl. Environ. Microbiol.* 68:3035–3045.
- Feinstein, L.M., Sul, W.J., Blackwood, C.B. 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* 75:5428–5433.
- Fierer, N., Jackson, R.B. 2006. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA.* 103:626–631.
- Fierer, N., Jackson, J.A., Vilgalys, R., Jackson, R.B. 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl. Environ. Microbiol.* 71:4117–4120.
- Fisher, M.M., Triplett, E.W. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.* 65:4630–4636.
- Foti, M., Sorokin, D.Y., Lomans, B., Mussman, M., Zacharova, E.E., Pimenov, N.V., Kuenen, J.G., Muyzer, G. 2007. Diversity, activity, and abundance of sulfate-reducing bacteria in saline and hypersaline soda lakes. *Appl. Environ. Microbiol.* 73:2093–3000.
- Franklin, R.B., Taylor, D.R., Mills, A.L. 1999. Characterization of microbial communities using randomly amplified polymorphic DNA (RAPD). *J. Microbiol. Methods.* 35:225–235.
- Garbeva, P., van Veen, J.A., van Elsas, J.D. 2004. Microbial diversity in soil: selection microbial populations by plant and soil type and implications for disease suppressiveness. *Annu. Rev. Phytopathol.* 42:243–270.
- Gentry, T.J., Wickham, G.S., Schadt, C.W., He, Z., Zhou, J. 2006. Microarray applications in microbial ecology research. *Microb. Ecol.* 52:159–175.

- Ghebremedhin, B., Layer, F., König, W., König, B. 2008. Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16 rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* gene sequences. *J. Clin. Microbiol.* 46:1019–1025.
- Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57:81–91.
- Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68:669–685.
- Hansel, C.M., Fendorf, S., Jardine, P.M., Francis, C.A. 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl. Environ. Microbiol.* 74:1620–1633.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., Gu, B., Jardine, P., Criddle, C., Zhou, J. 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* 1:67–77.
- Henne, A., Schmitz, R.A., Bömeke, M., Gottschalk, G., Daniel, R. 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 66:3113–3116.
- Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., 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.
- Hugenholtz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3:Reviews 0003.
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17:377–386.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Omura, S. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* 21:526–531.
- Keller, M., Hettich, R. 2009. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol. Mol. Biol. Rev.* 73:62–70.
- Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglou, P., Klironomos, J.N., Lee, H., Trevors, J.T. 2004. Methods of studying soil microbial diversity. *J. Microbiol. Methods.* 58:169–188.
- Kolb, S., Knief, C., Stubner, S., Conrad, R. 2003. Quantitative detection of methanotrophs in soil by novel *pmoA*-targeted real-time PCR assays. *Appl. Environ. Microbiol.* 69:2423–2429.
- Konstantinidis, K.T., Ramette, A., Tiedje, J.M. 2006. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B.* 361:1929–1940.
- Lauber, C.L., Hamady, M., Knight, R., Fierer, N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75:5111–5120.
- Li, X., Luo, Q., Wofford, N.Q., Keller, K.L., McInerney, M.J., Wall, J.D., Krumholz, L.R. 2009. A molybdopterine oxidoreductase is involved in H<sub>2</sub> oxidation in *Desulfovibrio desulfuricans* G20. *J. Bacteriol.* 191:2675–2682.
- Li, T., Wu, T.D., Mazéas, L., Toffin, L., Guerquin-Kern, J.L., Leblon, G., Bouchez, T. 2008. Simultaneous analysis of microbial identity and function using NanoSIMS. *Environ. Microbiol.* 10:580–588.
- Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Anderson, I., Lykidis, A., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C. 2010. The integrated microbial genomes system: an expanding comparative analysis resource. *Nucl. Acids Res.* 38:382–390.
- Metzker, M.L. 2010. Sequencing technologies – the next generation. *Nat. Rev. Genet.* 11:31–46.
- Mills, D.K., Entry, J.A., Gillevet, P.M. 2007. Assessing microbial community diversity using amplicon length heterogeneity polymerase chain reaction. *Soil Sci. Soc. Am. J.* 71:572–578.
- Moran, M.A. 2009. Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe.* 4:329–335.

- Mühling, M., Woolven-Allen, J., Murrell, J.C., Joint, I. 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J.* 2:379–392.
- Muyzer, G. 1999. Genetic fingerprinting of microbial communities – present status and future perspectives. *Methods of microbial community analysis. Proceedings of the 8th international symposium on microbial Ecology. Atlantic Canada Society for Microbial Ecology, Halifax, Canada.*
- Muyzer, G., Waal, E.C.D., Uitterlinden, A.G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59:695–700.
- Nakatsua, C.H., Torsvik, V., Ovreas, L. 2000. Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. *Soil Sci. Soc. Am. J.* 64:1382–1388.
- Nielsen, J.L., Christensen, D., Kloppenborg, M., Nielsen, P.H. 2003. Quantification of cell-specific substrate uptake by probe-defined bacteria under in situ conditions by microautoradiography and fluorescence in situ hybridization. *Environ. Microbiol.* 5:202–211.
- Nüsslein, K., Tiedje, J.M. 1999. Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. *Appl. Environ. Microbiol.* 65:3622–3626.
- Okabe, S., Kindaichi, T., Tsukasa, I. 2004. MAR–FISH: an ecophysiological approach to link phylogenetic affiliation and in situ metabolic activity of microorganisms at a single-cell resolution. *Microbes Environ.* 19:83–98.
- Okabe, S., Kindaichi, T., Ito, T. 2005. Fate of <sup>14</sup>C-labeled microbial products derived from nitrifying bacteria in autotrophic nitrifying biofilms. *Appl. Environ. Microbiol.* 71:3987–3994.
- Oliver, J.D. 2005. The viable but nonculturable state in bacteria. *J. Microbiol.* 43:93–100.
- Pernthaler, A., Pernthaler, J., Amann, R. 2002. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* 68:3094–3101.
- Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., Hollibaugh, J.T. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* 71:4121–4126.
- Ranjard, L., Poly, F., Lata, J.C., Mougél, C., Thioulouse, J., Nazaret, S. 2001. Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. *Appl. Environ. Microbiol.* 67:4479–4487.
- Rastogi, G., Stetler, L.D., Peyton, B.M., Sani, R.K. 2009. Molecular analysis of prokaryotic diversity in the deep subsurface of the former Homestake gold mine, South Dakota, USA. *J. Microbiol.* 47:371–384.
- Rastogi, G., Osman, S., Vaishampayan, P.A., Andersen, G.L., Stetler, L.D., Sani, R.K. 2010. Microbial diversity in uranium mining-impacted soils as revealed by high-density 16S microarray and clone library. *Microb. Ecol.* 59:94–108.
- Riesenfeld, C.S., Schloss, P.D., Handelsman, J. 2004. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38:525–552.
- Ritchie, N.J., Schutter, M.E., Dick, R.P., Myrold, D.D. 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. *Appl. Environ. Microbiol.* 66:1668–1675.
- Rogers, S.W., Moorman, T.B., Ong, S.K. 2007. Fluorescent in situ hybridization and microautoradiography applied to ecophysiology in soil. *Soil Sci. Soc. Am. J.* 71:620–631.
- Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., Goodman, R.M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66:2541–2547.
- Schloss, P.D., Handelsman, J. 2004. Status of the microbial census. *Microbiol. Mol. Biol. Rev.* 68:686–691.
- Schwieger, F., Tebbe, C.C. 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.

- Singh, B.K., Campbell, C.D., Sorenson, S.J., Zhou, J. 2009. Soil genomics. *Nature Reviews Microbiology* 7:756 doi:10.1038/nrmicro2119-c1.
- Smit, E., Leeftang, P., Wernars, K. 1997. Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiol. Ecol.* 23:249–261.
- Smith, C.J., Osborn, A.M. 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol. Ecol.* 67:6–20.
- Thies, J.E. 2007. Soil microbial community analysis using terminal restriction fragment length polymorphisms. *Soil Sci. Soc. Am. J.* 71:579–591.
- Torsvik, V., Øvreås, L. 2002. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5:240–245.
- Urich, T., Lanzen, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One.* 3:e2527.
- von Wintzingerode, F., Göbel, U.B., Stackebrandt, E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21:213–229.
- Wellington, E.M., Berry, A., Krsek, M. 2003. Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr. Opin. Microbiol.* 6:295–301.
- Whitman, W.B., Coleman, D.C., Wiebe, W.J. 1998. Prokaryotes: the unseen majority. *Proc Natl. Acad. Sci. USA.* 95:6578–6583.
- Wilmes, P., Bond, P.L. 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol.* 14:92–97.
- Yang, Y., Yao, J., Hu, S., Qi, Y. 2000. Effects of agricultural chemicals on DNA sequence diversity of soil microbial community: a study with RAPD marker. *Microb. Ecol.* 39:72–79.
- Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Déjean, S., DeSantis, T.Z., Gonçalves, O., Piceno, Y.M., Andersen, G.L., Kowalchuk, G.A. 2009. Environmental microarray analyses of Antarctic soil microbial communities. *ISME J.* 3:340–351.
- Yuhong, Z., Shi, P., Liu, W., Meng, K., Bai, Y., Wang, G., Zhan, Z., Yao, B. 2009. Lipase diversity in glacier soil based on analysis of metagenomic DNA fragments and cell culture. *J. Microbiol. Biotechnol.* 19:888–897.
- Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E.J., Woodnutt, G., Short, J.M., Keller, M. 2005. High-throughput cultivation of microorganisms using microcapsules. *Methods Enzymol.* 397:124–130.
- Zwolinski, M.D. 2007. DNA sequencing: strategies for soil microbiology. *Soil Sci. Soc. Am. J.* 71:592–600.