Chapter 17
Molecular Methods for Studying Microbial Ecology in the Soil and Rhizosphere

Janice E. Thies

This work is dedicated to my brother, Eric, whose incredible talents are now lost to this world.

17.1 Introduction

As described throughout this book, soil and rhizosphere microorganisms are responsible for a wide range of ecosystem services, including decomposing organic matter, cycling and immobilizing nutrients, aggregating soil, filtering and bioremediating pollutants, suppressing and causing plant disease, and producing and releasing greenhouse gasses. A long-standing challenge for studies in soil and rhizosphere ecology has been developing effective methods that can be used to describe the diversity, function and abundance of soil and plant-associated microbial populations. Enormous advances have been made since the first report by Torsvik (1980) that deoxyribonucleic acids (DNA) could be extracted from soil and subsequently characterized and that there may be as many as 6000–10,000 different genomes in 1 g of soil (Torsvik et al. 1990). A recent analysis based on reassociation kinetics done by Gans et al. (2005) suggests that this number is conservative and that the number of individual genomes per 1 g of soil may approach 277,000. This number far exceeds diversity estimates from any other matrix, making soil the most complex and diverse environment on earth.

Characterizing this diversity is an intricate and difficult task, in which all current methods fall short. Each month better approaches are being developed and published, allowing us to continue to explore this biologically rich environment. These new approaches have enabled us to not only ask who is living in soil, but to also determine how populations respond to management and consider how we might develop better soil management practices to encourage beneficial associations between plants and the soil biota and discourage detrimental ones.

J.E. Thies
Department of Crop and Soil Sciences, Cornell University, Ithaca, NY
e-mail: jet25@cornell.edu

C.S. Nautiyal, P. Dion (eds.) *Molecular Mechanisms of Plant and Microbe Coexistence.* Soil Biology 15, DOI: 10.1007/978-3-540-75575-3

Measures that describe population diversity attempt to capture (i) the genetic variability within a species, (ii) the number (richness) and relative abundance (evenness) of species, and/or (iii) the number of different functional groups within studied communities (Torsvik and Ovreas 2002). Describing diversity at the ecosystem scale often involves (i) identifying the variety of processes occurring, (ii) characterizing the interactions taking place between different organisms and/or (iii) assessing the number of trophic levels represented within the community. A major difficulty with describing diversity for microorganisms is that the species concept, derived from plant and animal community ecology, does not translate well to microbial populations. As yet, there is no satisfactory species concept for bacteria or fungi (Ward 1998; Liu and Stahl 2002), making it somewhat difficult to characterize the diversity of these populations in ecologically meaningful ways. The advent of molecular ecology has not resolved, but rather complicated the picture as more has become known about the lateral transfer of genetic elements between bacteria in the environment (Smalla and Sobecky 2002).

Assessing microbial population function frequently involves measuring rates of different processes, such as organic matter decomposition, respiratory activity or denitrification; or detecting the presence of genes needed to carry out biochemical reactions of ecological relevance, such as nitrogen fixation (e.g., *nif*H), ammonia or methane oxidation (*amo*A or *pmo*A, respectively), or denitrification (*nar*G, *nap*A, *nir*S, *nir*K, *nor*B, *nor*Z, *nos*Z). Molecular biological approaches have contributed substantially to our understanding of how microbial functions vary in relation to space, time and soil management practices (Handelsman and Smalla 2003). Processes that are unique to particular groups of organisms and are catalyzed by well-described enzyme systems and for which sequence information is known, such as those noted immediately above, have been particularly tractable to study with molecular methods. However, key ecosystem processes, such as depolymerization of organic matter, carbon (C) metabolism or sulfur (S) oxidation are so common across diverse lineages of bacteria and fungi (and other soil eukaryotes), and are carried out by such a diversity of enzyme systems, that molecular approaches may cloud, rather than clarify who or what the key system drivers may be.

Measuring abundance or population density normally involves (i) counting individuals within target groups, such as total bacteria, protozoa or nematodes using microscopy or culturing techniques; (ii) using molecular probes combined with microscopy to enumerate target groups of interest, e.g., the alpha-Proteobacteria or Planctomycetes, and/or (iii) measuring the concentration or content of general or unique biochemical markers, such as microbial biomass, adenosine triphosphate (ATP, the total energy charge of soil) or ergosterol (fungi). It is well known that traditional culturing methods detect only a small fraction of the extant microbial abundance and diversity in soil (Torsvik et al. 1990). Molecular techniques in which DNA is extracted from soil, then cloned and sequenced invariably reveal the presence of populations that are not recovered by traditional culturing (Liu and Stahl 2002). The rapidly expanding sequence databases, such as GeneBank (http://www.ncbi.nlm.nih.gov/), the Ribosomal Database Project II (http://rdp.cme.msu.edu/), EMBL-EBI (http://www.ebi.ac.uk/) and TIGR (http://www.tigr.org/tdb/) assist users in designing probes for use in detecting and quantifying specific popu-

lations in environmental samples. However, these approaches require that definitive sequence information is available; hence, populations that may have significant ecological relevance may still be overlooked when using targeted molecular approaches for abundance estimates.

Despite increased access to soil biodiversity by use of molecular methods, community members detected by these approaches may not necessarily correspond to populations responsible for significant biogeochemical processes in situ, especially when these populations constitute only a small proportion of the total community. Likewise, knowing the taxonomic identity or phylogenetic affiliation of a cloned sequence does not necessarily mean that we will know the function of the organism in situ. Even if we can confirm the presence of an organism in a sample whose function is known, it does not necessarily mean that the organism is active. Recently developed RNA-based techniques (e.g., Aneja et al. 2004) and use of stable isotope labeling and tracing (Radajewski et al. 2003) have helped to address this latter point because they can be used to identify those members of a community that are most active under a given set of environmental conditions (see below for approaches and applications).

In studying soil and rhizosphere ecology, one must recognize that organisms residing in these environments are physiologically and phylogenetically diverse. A holistic understanding of microbial communities and their interactions with plant roots requires a polyphasic approach; one that employs culturing and activity measures combined with molecular approaches. Describing and discussing the variety of methods used for polyphasic analysis of rhizosphere communities is beyond the scope of this chapter. Here I focus on recent molecular methods that are being used to characterize soil and rhizosphere microbial community composition and, in some cases, identify the functions of select members of these communities.

The relationship between many of the techniques described and how each is used in microbial community studies is shown in Fig. 17.1. Amplifying and analyzing rRNA genes present in DNA extracted from soil samples forms the basis for many of these techniques. Figure 17.1 also includes references to traditional techniques used in soil microbiology and biochemistry and illustrates how the new molecular approaches support and augment the traditional approaches. A general introduction to soil and rhizosphere ecology is given in Thies and Grossman (2006). An introduction to soil molecular ecology, nucleic acid structure and basic molecular methods is given in Thies (2007a). In this chapter, I describe more advanced analytical methods and how they are being used to better understand soil and rhizosphere microbial ecology.

17.2 Analyzing Nucleic Acids

Nucleic acid sequences define an organism's typical morphology and what activities it can carry out (its genotype). The organism's interaction with its environment and how the genotype is expressed define the organism's phenotype. As more has become known about nucleic acid sequences, particularly the sequences of the rRNA genes, a new phylogeny of the living world has emerged (Woese 1987). This new phylogeny

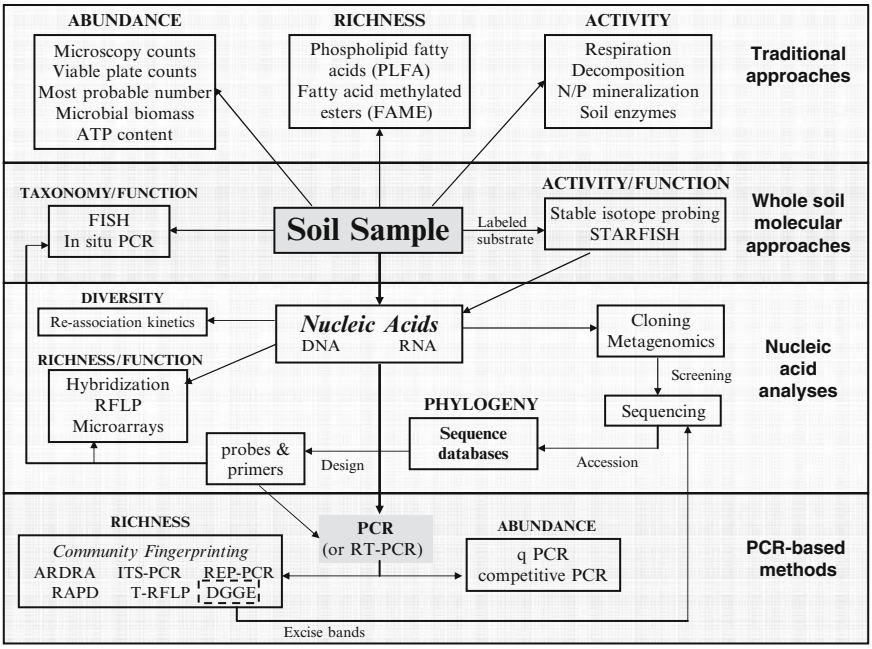

Fig. 17.1 Overview of approaches used to characterize soil microbial communities (adapted from Thies 2007a)

aids in understanding the evolutionary history and relatedness of different organisms to each other (Woese 1987). The ability to predict function from known sequences and to place organisms within a phylogenetic framework make the nucleic acid complement of cells particularly information rich targets for analysis.

A wide range of techniques are now available for analyzing nucleic acids. These techniques fall into three basic categories: (i) methods used to analyze nucleic acids in situ; (ii) those used to analyze extracted DNA/RNA directly; and (iii) those that employ polymerase chain reaction (PCR) to amplify, and subsequently analyze, target DNA sequences or RNA that has been reverse-transcribed to copy DNA (cDNA) (see Fig. 17.1). Applying these techniques to microbial community ecology studies has enabled us to overcome the limitations inherent with traditional enrichment and isolation techniques, thereby allowing us to detect organisms yet to be cultivated and, in some cases, infer their ecological functions.

17.2.1 Extracting DNA and RNA

Most molecular approaches require that nucleic acids be extracted from the soil matrix before analysis. A variety of methods have been developed to extract nucleic acids from soils of varying texture and these have been summarized recently by

Bruns and Buckley (2002). Two main approaches are used: (i) cell fractionation and (ii) direct lysis. In cell fractionation, intact microbial cells are released and separated from the soil matrix. After extraction, the cells are subsequently lysed and the DNA separated from the cell debris. In the direct lysis methods, microbial cells are lysed directly in the soil and then the nucleic acids are separated from the soil matrix. The main considerations when choosing a suitable protocol are extraction efficiency, obtaining a sample that is representative of the resident community, and obtaining an extract free of contaminants that could interfere with either PCR or probe hybridization.

Extraction efficiency of both cell fractionation and direct lysis procedures can be assessed by direct microscopy using vital stains, where extracted soil is examined for intact, viable microbial cells. Alternatively, soil samples may be spiked with a known quantity of DNA (or bacterial cells) and then the recovery of the added DNA assessed. In general, DNA recovery is generally much higher when using direct lysis as compared to cell fractionation protocols (Courtois et al. 2001).

DNA is normally extracted from very small quantities of soil, typically 500–1000 mg. This alone makes obtaining a representative sample difficult. In addition, cell walls of different organisms lyse with varying efficiencies. Cell walls of high G+C Gram-positive bacteria are often difficult to lyse, whereas those of Gram-negative bacteria lyse more readily. Hence, DNA or rRNA recovered may contain an artificially greater amount of DNA derived from Gram-negative bacteria. In characterizing microbes colonizing bulk or rhizosphere soil, the rhizoplane and the endorhizosphere, care must be taken to attribute extracted nucleic acids to their associated habitat. For rhizosphere communities, soil adhering to roots can be removed by soaking the roots in buffer with moderate agitation. Roots are removed and nucleic acids extracted from the soil remaining. Roots are then subjected to several rounds of sonication to remove microbes colonizing the rhizoplane. Finally, enzymatic hydrolysis in an appropriate buffer can be used to enrich extracts for nucleic acids from endophytes (Jiao et al. 2006).

Contaminants, such as humic and fulvic acids, have a similar solubility to nucleic acids and hence are often co-extracted. These contaminants interfere with PCR amplification and hybridization experiments. Co-extracted contaminants can be reduced or removed by use of a post-PCR DNA clean-up kit, such as the QIAquick® PCR purification kit (Qiagen, Chatsworth, CA) or GENECLEAN Spin kit (Qbiogene, Inc., Carlsbad, CA); or by washing the nucleic acid extract with dilute EDTA or passing it through a Sephadex G-75 column (Bruns and Buckley 2002). While improving PCR amplification, extra cleaning steps can also lead to a loss of nucleic acids and hence sparsely represented members of the community may be lost from subsequent analyses. In addition, all post-extraction clean-up procedures add cost and processing time, and thus reduce the number of samples that can be analyzed within the scope of any experiment.

Commercial soil DNA/RNA extraction kits based on direct lysis by bead-beating, such as the FastDNA® SPIN Kit for Soil and the FastRNA® Pro Soil-Direct Kit (Qbiogene, Inc., Carlsbad, CA) and the Ultraclean™ and PowerSoil™ DNA isolation kits (MoBio Laboratories, Solana Beach, CA) have recently become

available. The DNA/RNA extracted is of high molecular weight and of sufficient quality to be used in PCR or nucleic acid hybridization experiments for most soils. The PowerSoil™ DNA isolation kit (MoBio Laboratories) has been specifically recommended, by the manufacturer, for use with high organic matter samples.

For any given study, the type of molecule(s) that will be extracted, e.g., DNA, RNA, both types of nucleic acids and/or PLFAs, must be determined prior to molecular microbial analysis. DNA is extracted and analyzed most commonly because it is more stable and easier and less costly to extract from soil. Post-extraction analyses are straight-forward and information obtained reflects the whole community at the time of sampling. The key issue with DNA analysis is that it does not reflect the abundance of viable organisms or their level of activity. DNA that is free in soil is readily hydrolyzed by nucleases; however, it can be protected from hydrolysis when present in dead cells or protected within soil aggregates. Protected, free DNA is extracted along with that from moribund and active cells. RNA, on the other hand, is highly labile and more difficult to extract. Methods for extracting rRNA from soil are given in Felske et al. (1999) and Sessitsch et al. (2002) and methods for the simultaneous extraction of DNA and RNA are given in Griffiths et al. (2000) and Hurt et al. (2001). Commercial kits for extracting RNA from soil are also now available (Qbiogene, Inc.). Extracting mRNA is still fraught with difficulty, but some success has been reported (Hurt et al. 2001; Sessitsch et al. 2002). Most post-extraction analyses require that RNA is first reverse-transcribed (RT) into cDNA and then the cDNA is used in downstream analyses. The advantage of extracting and analyzing RNA is that it is generally only present in high amounts in actively metabolizing cells. As substrate becomes limiting, cell processes slow down, along with rDNA transcription. Thus, rRNA analysis is more reflective of the portion of the soil microbial community that is either active at the time of sampling or has recently been active. When mRNA can be recovered, insights into genes that are being actively transcribed under a given set of environmental conditions can be obtained and, hence, is most desirable for studies of microbial community function.

17.2.2 Re-Association Kinetics

For assessing total diversity of microorganisms in an environmental sample, re-association kinetics is considered the ‘gold standard’. Yet it is rarely performed in research laboratories because the equipment needed is very costly and sample processing times are high, thus it does not lend itself to high sample throughput. In addition, it does not provide any information on identity or function of any member of the microbial community.

For the analysis, DNA is denatured by either heating or use of a denaturant (e.g., urea). Under highly controlled conditions, the denaturant is removed or the temperature is lowered, and complementary DNA strands are allowed to re-anneal. When genome complexity is low, the time it takes for all single DNA strands to find their complement is brief. As complexity increases, the time it takes

for complementary strands to re-anneal increases. Experimentally, this is referred to as a $C_o t$ curve, where C_o is the initial molar concentration of nucleotides in single-stranded DNA and t is time. This measure reflects both the total amount of information in the system (richness or number of unique genomes) and the distribution of that information (evenness or the relative abundance of each unique genome) (Liu and Stahl 2002), thus making it among the more robust methods for estimating extant diversity in a given sample.

The genetic complexity or genome size of several soil microbial communities was assessed using re-association kinetics by Torsvik et al. (1990, 1998). They estimated that the community genome size in undisturbed organic soils was equivalent to 6,000–10,000 *Escherichia coli* genomes, while a heavy metal-polluted soil contained 350–1500 genome equivalents. Culturing yielded less than 40 genome equivalents. These data and studies employing epifluorescence microscopy to obtain direct cell counts, are what verify that culturing methods capture only the tip of the iceberg of the diversity within soil microbial communities. Gans et al. (2005) recently reported re-association kinetics data analyzed by an improved analytical approach, which yielded an estimate of the extant diversity contained in an undisturbed soil sample of 8.3 million distinct genomes in 30 g of soil, an order of magnitude greater than that reported by Torsvik et al. (1990). In contrast, a heavy metal-polluted soil was estimated to contain only 7900 genome equivalents, 99.9% fewer than in the undisturbed soil.

17.2.3 Cloning, Sequencing and Metagenomics

DNA (or RNA) sequence information can be obtained from environmental samples in two main ways: (i) cloning DNA extracted from soil directly or (ii) cloning PCR-amplified DNA (or reverse-transcribed RNA), followed in both cases by sequencing of the cloned DNA (or cDNA). In direct cloning, purified DNA extracted from soil is ligated into a vector, most frequently a self-replicating plasmid. The vector is then transformed into a competent host bacterium, such as commercially available *E. coli* competent cells, where it is maintained and multiplied (Lane 1991). Recombinant DNA clone libraries are produced in this way. Once a clone library is obtained, DNA inserts contained in the clones can be re-isolated from the host cells, purified and sequenced. The clone library can also be screened for biological activity expressed directly in *E. coli* or probed for sequences of interest using various genomics applications. This approach circumvents the need to culture microorganisms from environmental samples, although cloning itself is subject to its own inherent biases (Handelsman 2004).

Recently it became possible to clone large (100–300-kb) fragments of genomic DNA isolated directly from soil into bacterial artificial chromosome (BAC) vectors (Handelsman et al. 1998; Rondon et al. 2000). BAC vectors are low-copy number plasmids that can readily maintain large DNA inserts. When Rondon et al. (2000) analyzed their two BAC libraries, sequences homologous to the low-G+C Gram-positive *Acidobacterium*, *Cytophagales*, and *Proteobacteria* were found. They also

identified clones that expressed amylase, nuclease, lipase, hemolytic and antibacterial activities. This study heralded in the field of metagenomics, which is the genomic analysis of a population of microorganisms (Handelsman 2004). Metagenomic libraries are useful for phylogenetic studies, analyses of microbial function and as a tool for natural product discovery (Handelsman et al. 1998, Handelsman 2004, 2005). When Treusch et al. (2005) probed metagenomic libraries derived from a range of environments, they discovered that uncultivated members of the Crenarchaeota contained gene sequences homologous to the ammonia monooxygenase (*amo*A) gene in nitrifying bacteria. In a follow-on study, Leininger et al. (2006) examined 12 soils from different climatic zones from both agricultural and unmanaged systems and demonstrated that the number of Crenarchaeota *amo*A sequences (AOA) was consistently higher than those from ammonia oxidizing Bacteria (AOB), with the ratio of AOA to AOB ranging from 1.5 to 232. These results suggest that the Crenarchaeota could be playing a more significant role in global N cycling than thought previously (Nicol and Schleper 2006). Metagenomic libraries are powerful tools for exploring soil microbial diversity and will form the basis for future genomic studies that link phylogenetic information with soil microbial function (Handelsman 2004).

An alternative method for creating large clone libraries from soil sequences that allows subsequent profiling of microbial communities is called serial analysis of ribosomal sequence tags (SARST). In this approach, a region of the 16S rRNA gene is amplified by PCR, such as the V1-region. Through a series of enzymatic and ligation (linking) steps, the various V1 region amplicons are joined together. The resulting concatemers are then purified, cloned, screened and sequenced. The sequences (RSTs) of the individual V1 amplicons are deduced by ignoring the linking sequences and analyzing each sequence tag individually (Neufeld et al. 2004). Neufeld and Mohn (2005) used SARST to analyze arctic tundra and boreal forest soils and found that overall diversity was higher in the arctic tundra soil. They suggested that the high carbon flux and low pH characteristic of the boreal forest soils might contribute to lower bacterial diversity or that the high diversity in the arctic soils may be influenced by allochthonous organisms coming in via air currents and being preserved by low temperature. Yet, the comparative diversity between the two systems did not change when singleton sequence tags were eliminated from the analysis, suggesting that the arctic may serve as an unrecognized reservoir microbial diversity and biochemical potential.

Several other PCR-based community analysis methods described below, such as denaturing gradient gel electrophoresis (DGGE) and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) allow DNA fragments to be retrieved in a selective manner and these can then be cloned and sequenced using the methods described above.

Pyrosequencing is a very recent innovation that is making a big splash in the large-scale analysis of bacterial genomes (Margulies et al. 2005) and potentially soil metagenomes. Margulies et al. (2005) have developed an entirely new approach to DNA sequencing that employs fragmenting genomic DNA, ligating the fragments to adapters and separating them into single strands. The single-stranded

fragments are then bound to beads, with one fragment per bead, and the beads captured in droplets of a "PCR-reaction-in-oil emulsion". DNA amplification takes place inside each droplet, such that the bead contained in the droplet has attached to it 10 million copies of a unique DNA template. The emulsion is then dispersed, the DNA strands denatured and the beads carrying the single-stranded DNA copies placed into individual wells on a fiber-optic slide. Smaller beads that have the enzymes needed for pyrophosphate sequencing immobilized on them are then added to the wells. The prepared fiber-optic slide is placed into a chamber through which sequencing reagents flow. The base of the slide comes into optical contact with a second fiber-optic bundle that is fused to a charge-coupled device (CCD) sensor. Reagents are delivered cyclically to the chamber and flow into the wells of the fiber-optic slide. Simultaneous extension reactions occur on the template-carrying beads, such that each time a nucleotide is incorporated, inorganic pyrophosphate is released and photons are generated. Raw signals are captured, background subtracted, normalized and corrected. Post run analysis are used for base calling and sequence alignments. After an individual nucleotide is pulsed into the chamber, a wash containing apyrase is used to prevent nucleotides from remaining in the wells before the next nucleotide is introduced. Using this approach, Margulies et al. (2005) shotgun sequenced and de novo assembled the genome of *Mycoplasma genitalium* (580,069 bases) with 96% coverage and 99.96% accuracy in a single 4-h run. The average read length in their study was 110 bases per fragment. The method completely circumvents cloning of DNA fragments into bacterial vectors and handling individual clones in any way. The implications of using this approach for exploring soil metagenomes are fantastic. The capacity for sequencing 25 million bases in one run means that bioinformatics approaches are now under pressure to manage the quantity of information that can potentially be generated in meaningful ways that will allow data mining on a massive scale.

Edwards et al. (2006) used pyrosequencing to explore microbial community genomics in a deep mine borehole and in water seeping from it. The borehole water and water emerging from the borehole was described as an anoxic "black" environment with a pH of 6.7 and a redox potential of −142 mV. The oxygenated seepage water a few cm from the borehole orifice was characterized as a "red" environment, with a pH of 4.37 and a redox potential of −8.0 mV. Through use of pyrosequencing combined with comparative metagenomics, systems analysis, statistics, chemical analyses and hydrogeology they were able to characterize the differences in the genetic composition of the two communities and derive what their metabolic capacities were in the two environments. Comparing sequences generated with those in the Ribosomal Database II, indicated that the "black" water was dominated by Actinomycetales (*Brevibacterium* and *Corynebacterium*) and the "red" water was dominated by members of the Chromatiales (*Chromatiaceae*, *Thiobacillus* and *Halothiobacillus*). The two communities, just centimeters distant from each other along the same seep, as well as their respective environments were fundamentally different. Additionally, the "red" sample had a much higher species richness than the "black" sample. Sequences from the two pyrosequencing libraries were compared to the SEED database (http://theseed.uchicago.edu/FIG/index.cgi) of microbial

genomes to identify groups of genes (subsystems) that were enriched in the two environments. Subsystems involved in iron uptake and use (siderophores and ABC transporters for ferrichrome) and denitrification were common in the "black" sample, whereas respiratory complexes and cytochrome-C oxidases were commonly found in the "red" sample. This study represents a large step toward linking phylogeny with function in two extreme environments.

17.2.4 Sequence Databases

DNA sequencing, annotating sequences and maintaining sequence databases are important activities for discovery of novel genetic properties, exploring phylogenetic affiliations, and in developing more specific primers and gene probes to address particular ecological questions. Gene sequences, once obtained, are submitted to and maintained within various databases such as GenBank or the Ribosomal Database Project II. GenBank and its collaborating databases, the European Molecular Biology Laboratory (EMBL) and the DNA databank of Japan (DDBJ) reached a milestone recently of containing 100 billion bases (100 gigabases) of sequence information from over 165,000 organisms, including bacteria, fungi, protozoa, nematodes and other fauna. The Ribosomal Database Project II, Release 9 (Cole et al. 2005), update 50 (release 9.50) contains 368,406 aligned and annotated Bacterial small subunit (16S) rRNA gene sequences (as of 5/2/07) with updated on-line analyses.

GenBank holds the data generated by over 400 whole genome shotgun (WGS) sequencing projects (http://www.ncbi.nlm.nih.gov/projects/WGS/WGSprojectlist.cgi). The WGS database contains genomes from individual organisms (more than 250 bacteria and 120 eukaryotes) and environmental metagenomes from over 30 projects (NCBI News 2006/2007). The environmental genomics projects include a farm soil, acid mine drainage biofilm and the symbionts of an ocean sediment-dwelling annelid that has no digestive tract and a reduced excretory system and thus relies on the symbionts to provide its nutritional and excretory needs. As more becomes known about the genetic subsystems dominant in these metagenomes, large leaps will begin to be made in our understanding of the functional significance of different community signatures in different environments – including the rhizosphere.

17.3 Phospholipid Fatty Acids (PLFA)

Phospholipid fatty acid (PLFA) analysis is an alternative technique for studying the soil microbial community without culturing. It is a non-selective method, where phospholipid fatty acid composition of the soil is analyzed by gas chromatography (GC) (Tunlid and White 1992). PLFAs are the basic components of cell membranes and are decomposed rapidly in soil when cells die. Consequently, extracting phospholipids from soil samples provides information about living members present in microbial communities (Fritze et al. 1998; Frostegard et al. 1993).

The entire PLFA profile can be used as a fingerprint of the whole soil community. Since phospholipid-linked branched fatty acids are characteristic of bacterial origin, lipids can be used to indicate specific subgroups within the community and physiological status of those populations (Roslev et al. 1998). For example, sulfate reducers, methane-oxidizing bacteria, mycorrhizal fungi and actinomycetes have unique lipid signatures. Also, environmental changes can induce changes in certain PLFA components, such as the ratio of saturated to unsaturated fatty acids, ratio of trans- to cis-monoenoic unsaturated fatty acids and the proportion of cyclopropyl fatty acids. Such changes herald changes in the microbial community. In addition PLFA profiles may contain information concerning the dynamics of larger groups of organisms such as eukaryotes. However, common fatty acids, e.g., polyenoic fatty acids found in eukaryotes, are less able to distinguish between groups when compared to the number of fatty acids found almost exclusively in bacteria (Tunlid and White 1992).

17.4 Whole Soil Molecular Approaches

17.4.1 Stable Isotope Probing

Nucleic acid methods have recently been coupled with stable isotope labeling and detection to provide a culture-independent means of linking the identity of bacteria with their function in the environment (Manefield et al. 2002a,b; McDonald et al. 2005; Dumont and Murrell 2005). Soil is either incubated after adding a ^{13}C-labeled substrate or a plant is labeled with ^{13}C-CO_2 and rhizosphere soil sampled after labeling. Soil DNA or RNA is then extracted and centrifuged in a density gradient to separate ^{13}C-labeled nucleic acids from those containing ^{12}C. Once separated, labeled DNA can be amplified using PCR and universal primers to Bacteria, Archaea or Eucarya. Analysis of the PCR products, through cloning and sequencing for example, allows the microbes that have assimilated the labeled substrate to be identified (Manefield et al. 2002a; Wellington et al. 2003; Griffiths et al. 2004; Leake et al. 2006).

PFLA-SIP has also been used successfully to analyze active soil communities (Treonis et al. 2004; Lu et al. 2007). Lu et al. (2007) labeled rice plants in mesocosms by incubating them with ^{13}C-CO_2. After 49 pulses of $^{13}CO_2$ over 7 days, PLFAs were extracted from soils taken from different regions of the rhizosphere. By this approach they were able to establish that Gram-negative bacteria and eukaryotes were most active in incorporating ^{13}C-labeled root exudates, whereas Gram-positive bacteria dominated in the bulk soil. Microbial community changes in relation to root depth were also readily observed.

Rangel-Castro et al. (2005) used ^{13}C-CO_2 pulse-labeling, followed by RNA-SIP, to study the effect of liming on the structure of the rhizosphere microbial community metabolizing root exudates in a grassland. Their results indicated that limed soils contained a microbial community that was more complex and more active in using ^{13}C-labeled compounds in root exudates than were those in unlimed soils.

SIP-based approaches do hold great potential for linking microbial identity with function, but at present a high degree of labelling is necessary to be able to separate labeled from unlabeled marker molecules. This need for high substrate concentrations may bias community responses. Alternatively, use of long incubation times to ensure that sufficient label is incorporated increases the risk of having cross-feeding of ^{13}C from the primary consumers to the rest of the community, complicating data interpretation. Another complicating factor is identifying enriched nucleic acids within the density gradient. The point at which a given nucleic acid molecule is retrieved from the caesium chloride gradient is a function of both the incorporation of the heavy isotope and the overall G+C content of the nucleic acids. Thus, a means to attribute band position in the gradient to either incorporated label or high G+C content must be devised.

17.4.2 Fluorescence In Situ Hybridization (FISH)

Nucleic acid hybridization involves binding a discrete fragment (a probe) of DNA or RNA to a target sequence. The probe is generally labeled with a radioisotope or fluorescent molecule and the target sequence is bound to a nylon membrane. A positive hybridization signal is obtained when complementary base pairing occurs between the probe and the target sequence. This signal is visualized by exposing the membrane to auto-radiographic film after removing any unbound probe or viewing by fluorescence microscopy with an appropriate filter. The type of probe used and how the probe is labeled determine the range of applications. For example, oligonucleotide probes (up to 30 nucleotides long) may be used under very stringent conditions that resolve single base-pair mismatches but these will have limited sensitivity due to the constraint on the number of labels that may be attached to the probe. In contrast, larger DNA fragments may be labeled to high specific activity but it is difficult to control hybridization conditions sufficiently to guarantee 100% stringency. Both the ARB (http://www.arb-home.de/) and RDPII (http://rdp.cme.msu.edu/) websites have probe design features. Information about probes that have already been designed for specific purposes can be found at (http://www.microbial-ecology.net/probebase/).

Techniques based on nucleic acid colony hybridization (colony blotting) are useful for rapidly screening bacterial isolates to establish identity or uniqueness, for example, identifying specific rhizobia strains occupying root nodules or screening libraries containing DNA clones obtained from a soil community. Nucleic acid probes can also be used to detect specific phylogenetic groups of bacteria in appropriately prepared soil samples. In the latter application, a specific probe is fluorescently labeled and hybridized to target sequences contained within microbial cells in situ using the fluorescence in situ hybridization (FISH) technique. These protocols have been described extensively in reviews by Amann et al. (1995), Amann and Ludwig (2000) and more recently by Zwirglmaier (2005).

In the FISH technique, an oligonucleotide probe is conjugated with a fluorescent molecule (or fluorochrome). The probe is designed to bind to complementary

sequences in the 16S rRNA subunit of the ribosomes within bacterial cells. Because metabolically active cells contain a large number of ribosomes, the concentration of fluorescently labeled probe is relatively high inside these cells causing them to fluoresce under UV light. The final result is high binding specificity and typically low background fluorescence. Early techniques suffered from low signal intensity, however, two new methods, tyramide signal amplification and multiply-labeled polynucleotide probes, increase the signal intensity and allow FISH approaches to be used for a wider range of ecological settings and questions (Zwirglmaier 2005). For simultaneous counting of sub-populations in a given sample, probes can be designed that bind to specific sequences of rRNA that are found only in a particular group of organisms (i.e., Archaea, Bacteria, or sub-divisions these domains) and used in conjunction with each other. FISH can also be combined with microautoradiography to determine specific substrate uptake profiles for individual cells within complex microbial communities in a method called STARFISH, substrate-tracking autoradiographic fluorescence in situ hybridization (Ouverney and Fuhrman 1999; Lee et al. 1999). Because these methods label mainly metabolically active cells, the samples can be labeled simultaneously with dyes that bind to nucleic acids, such as 4,6-diamidine-2-phenylindole, dihydrochloride (DAPI), to facilitate a total cell count using fluorescence microscopy (Li et al. 2004). FISH is particularly useful when used in conjunction with confocal laser scanning microscopy (CLSM, see below) as it allows the relative position of diverse populations to be visualized in three dimensions, even within complex communities such as biofilms and the surfaces of soil aggregates (Binnerup et al. 2001). The ability to visualize and identify organisms on a microscale in their natural environment is the key advantage of FISH. Such techniques have great potential for studying microbial interactions with plants and the ecology of target microbial populations in soil, however, the binding of fluorescent dyes to organic matter resulting in non-specific fluorescence is a common problem in soils with high organic matter contents, such as peats, or other particles with high surface charge, such as black carbon. Image analysis software is readily available and may be 'trained' to detect only those aspects of an image that meet specified criteria.

Confocal laser scanning microscopy (CLSM), combined with in situ hybridization techniques, has been applied with considerable success to visualize the structure of soil microbial communities (Bloemberg et al. 2000). CLSM works by first capturing an image that is composed only of emitted fluorescence signals from a single plane of focus. This is done using a pinhole aperture, which eliminates any signal that may be coming from portions of the field that are out of focus. A series of these optical sections is scanned at specific depths and then each section is 'stacked' using imaging software, giving rise to either a two-dimensional image that includes all planes of focus in the specimen, or a computer generated three-dimensional image. This approach gives us unprecedented resolution in viewing environmental specimens, allowing us to better differentiate organisms from particulate matter as well as providing insight into the three dimensional spatial relationships of microbial communities within their environment.

17.4.3 Green Fluorescent Protein (GFP) and Other Marker Gene Technologies

Introduced marker genes, such as *lux*AB (luminescence), *lac*Z (β galactosidase) and *xyl*E (catechol 2,3-dioxygenase) are used frequently in soil microbial ecology studies. One marker gene that has attracted a lot of attention in rhizosphere studies is *gfp*, which encodes the green fluorescent protein (GFP). Green fluorescent protein is a unique bioluminescent genetic marker that can be used to identify, track, and count specific organisms into which the gene has been cloned that have been reintroduced into the environment (Chalfie et al. 1994). The *gfp* gene was discovered in and is derived from the bioluminescent jelly fish, *Aequorea victoria* (Prasher et al. 1992). Once cloned into the organism of interest, GFP methods require no exogenous substrates, complex media or expensive equipment to monitor and, hence, are favored over many fluorescence methods for environmental applications (Errampalli et al. 1999). GFP-marked cells can be identified using a standard fluorescence microscope fitted with excitation and emission filters of the appropriate wavelengths. One reason for such keen interest in GFP is that there is no background GFP activity in plants or the bacteria and fungi that interact with them, thereby making *gfp* an excellent target gene that can be introduced into selected bacterial or fungal strains and used to study plant-microbe interactions (Errampalli et al. 1999). Basically, *gfp* is transformed into either the chromosome or a plasmid in a bacterial strain, where it is subsequently replicated. Various gene constructs have been made that differ in the type of promoter or terminator used and some contain repressor genes such as *lac*I for control of *gfp* expression. Once key populations in a sample are known and isolates obtained, they can be subsequently marked with *gfp* or other genes producing detectable products in order to track them and assess their functions and interactions in soil and the rhizosphere. In addition to GFP, red-shifted and yellow-shifted variants have been described. Development of *gfp* mutants with a series of different excitation and emission wavelengths makes it possible to identify multiple bacterial populations simultaneously (e.g., Bloemberg et al. 2000). The *gfp* gene has been introduced into *Sinorhizobium meliloti*, *Pseudomonas putida* and *Pseudomonas* sp., among other common soil bacteria and used widely in soil ecology studies. Marked strains can be visualized in infection threads, root nodules, colonized roots and even inside digestive vacuoles of protozoa. If the *gfp* gene is cloned along with specific promoters, such as *mel*A (α galactosidase) (Bringhurst et al. 2001) or *gus*A (β glucuronidase) (Xi et al. 1999), then the transformed bacteria can be used as biosensors to report back to the observer if the inducers, in these cases galactosides or glucuronides, respectively, are present and at what relative concentration in the surrounding environment.

Marker gene approaches are restricted to use in organisms that can be cultured. While considerable information can be gained about how marked microbes interact with soil colloids and other soil organisms; and can be used as biosensors for detecting environmental concentrations of various compounds, they do not yield information about the vast, unknown majority of soil microbes for which cultured representatives have yet to be obtained.

17.4.4 Microarrays

Microarrays represents an exciting new development in microbial community analysis. Nucleic acid hybridization is the principle on which the technique is based. The main difference between past protocols and microarrays is that the oligonucleotide probes, rather than the extracted DNA or RNA targets, are immobilized on a solid surface in a miniaturized matrix. Thus, thousands of probes can be tested for hybridization with sample DNA or RNA simultaneously. In contrast to other hybridization techniques, the sample nucleic acids to be probed are fluorescently-labeled, rather than the probes themselves. After the labeled sample nucleic acids are hybridized to the probes contained on the microarray, positive signals are detected by use of CSLM or other laser microarray scanning device. A fully-developed DNA microarray could include a set of probes encompassing virtually all known natural microbial groupings and thereby serve to simultaneously monitor the population structure at multiple levels of resolution (see Guschin et al. 1997; Ekins and Chu 1999; Wu et al. 2001; Zhou and Thompson 2002; Zhou 2003). Such an array would potentially allow for an enormous increase in sample throughput. A major drawback of microarrays for use in soil ecology studies currently is their need for a high copy number of target DNA/RNA to obtain a signal that is detectable with current technologies. Targets in concentrations less than 10^3–10^4 are difficult to detect using this approach. Non-specific binding of target nucleic acids to the probes is also a serious issue that needs to be overcome (Zhou and Thompson 2002).

There are three basic types of arrays used in soil ecology: (i) community genome arrays (CGA), used to compare the genomes of specific groups of organisms; (ii) functional gene arrays (FGA), used to detect the presence of genes of known function in microbial populations in prepared soil samples and more recently used to detect gene expression; and (iii) phylogenetic oligonucleotide arrays (POA), used to characterize the relative diversity of organisms in a sample through use of rRNA sequence-based probes. The details of microarray construction and types of arrays can be found in Ekins and Chu (1999) and ecological applications are reviewed in Zhou (2003).

17.5 PCR-based Methods

PCR involves separating a double-stranded DNA template into two strands (denaturation), hybridizing (annealing) oligonucleotide primers (short strands of nucleotides of a known sequence) to the template DNA and then elongating the primer-template hybrid by a DNA polymerase enzyme. The potential target genes for PCR are many and varied, limited only by available sequence information. The primers most frequently used for soil ecological studies are designed to target specific DNA fragments, such as 16 S or 18 S rRNA genes, functional genes, repetitive sequences, e.g., REP (Repetitive Extragenic Palindromic) sequences, or arbitrary primers, e.g., randomly amplified polymorphic DNA (RAPD). The discovery of thermal-stable DNA polymerases from organisms such as *Thermus aquaticus*

(*Taq* polymerase) has made PCR a standard protocol in laboratories around the world (Mullis and Faloona 1987; Saiki et al. 1998).

By far the more common targets for characterizing microbial communities are the rRNA genes because of their importance in establishing phylogenetic and taxonomic relationships (Woese et al. 1990). These are the small subunit (SSU) rRNA genes, 16 S in Bacteria and Archaea or 18 S in Eucarya; the large subunit (LSU) rRNA genes, 23 S in Bacteria and Archaea or 28 S in Eucarya; or the internal transcribed spacer (ITS) regions, sequences that lie between the SSU and LSU genes. Other defined targets are genes that code for ecologically significant functions, such as genes involved in nitrogen fixation, e.g., *nif*H (Chelius and Lepo 1999; Rösch et al. 2002); *amo*A which codes for ammonium monooxygenase, a key enzyme in nitrification reactions (Rotthauwe et al. 1997); and *nir*S and *nir*K which codes for nitrite reductase, a key enzyme in denitrification reactions (Rösch et al. 2002; Henry et al. 2004).

In any study where PCR is used, sources of bias must be considered (Wintzingerode et al. 1997). The main sources of bias in amplifying soil community DNA are: (i) the use of very small sample sizes (typically only 500 mg of soil); (ii) preferential amplification of some DNA templates over others; and, (iii) for amplification of the rRNA genes, the fact that many bacteria contain multiple copies of these operons (e.g., *Bacillus* and *Clostridium* species contain 15 copies), hence sequences from such species will be over-represented among the amplification products. In addition, chimeras, composed of double-stranded DNA where each strand was derived from a different organism rather than a single organism, may be generated. Acknowledged biases associated with PCR are generally why diversity indices calculated from the results of PCR-based experiments may not be very robust.

17.5.1 DNA Fingerprinting

PCR fingerprinting is used to distinguish differences in the genetic makeup of microbial populations from different samples and can be accomplished by several different methods. The advantages of these techniques are that they are rapid and inexpensive and thus enable high sample throughput and can be used to target sequences that are phylogenetically or functionally significant (Fjellbirkeland et al. 2001). Depending on the primers chosen, PCR fingerprints can be used to distinguish between isolates at the strain level or to characterize target microbes at the community level. The more common PCR fingerprinting techniques in use today for characterizing soil microbial community composition are denaturing gradient gel electrophoresis (DGGE) or temperature gradient gel electrophoresis (TGGE) (Muyzer and Smalla, 1998) and terminal restriction fragment length polymorphism (T-RFLP) analysis (Liu et al. 1997; Marsh 2005; Thies 2007b). Both techniques can be used to separate PCR products that are initially the same length by employing additional methods to separate the amplicons into a greater

number of bands or operational taxonomic units (OTUs) that are then used for community comparisons.

DGGE and TGGE are identical in principle. Both techniques impose a parallel gradient of denaturing conditions along a polyacrylamide gel. Double-stranded DNA PCR amplicons are loaded in wells at the top of the gel and, as the DNA migrates, the denaturing conditions of the gel gradually increase. In DGGE, the denaturants are typically urea and formamide; in TGGE it is temperature. Because native double-stranded DNA is a compact structure, it migrates faster than partially denatured DNA. The sequence of a fragment determines the point in the gradient gel at which denaturation will start to retard mobility. Sequence affects duplex stability by both percentage G+C content and neighboring nucleotide interactions (e.g., GGA is more stable than GAG). The resulting gel yields a ladder of bands in each lane characteristic of the DNA extracted and amplified from the original sample. There is not a direct correspondence between bands in the DGGE gel and organism diversity, however. Sequences amplified from the DNA of different organisms may have similar melting properties in the presence of the denaturant and thus occupy the same band in the denaturing gel. DNA fragments cloned from different bands may yield as many different sequences as clones analyzed. Since there is not a one-to-one correspondence between bands and taxa, the bands are referred to as OTUs. The OTUs form the basis of similarity and multivariate analyses of data derived from various soil communities.

While the power of DGGE and TGGE to detect PCR amplicon diversity within a single gel is high, the resolving power of these and other gel-based analyses, is limited by the number of bands capable of 'fitting' and being counted as individual bands on a single gel. In practice, no more than 80–100 distinct sequence types may be resolved despite the potential for single base-pair sensitivity. An important advantage that DGGE analysis has over T-RFLP (see below) is that PCR amplicons of interest that are resolved on a DGGE gel can be excised from the gel, re-amplified, cloned and sequenced, thereby obtaining taxonomic and/or phylogenetic information about amplifiable members of the soil community. For phylogenetic assignment of cloned sequences, variable regions within the SSU rRNA genes are amplified. An important disadvantage of the gradient gel approach is that the amplicon size must be restricted to under 600 base pairs in length to optimize separation within the gel matrix. Therefore, full length rRNA gene sequences cannot be recovered using these methods. DGGE and TGGE are now being applied frequently in soil microbial ecology to compare the structures of complex microbial communities and to study their dynamics. The basic method and applications were recently reviewed by Nakatsu (2007).

T-RFLP analysis, as in DGGE analysis, begins with amplifying soil community DNA using targeted primers, but with the key differences that one or both primers are labeled with a fluorochrome(s) and that resulting amplicons are hydrolyzed with restriction enzymes to create DNA fragments of varying size that are labeled with the fluorochrome at either the 5´ or 3´ end. These terminal fragments are then sized against a standard molecular size marker using automated DNA sequencing techniques. The resulting electropherogram (peaks representing the sizes of the

terminal restriction fragments, TRFs) is used as a DNA fingerprint characteristic of the soil community sampled. Resulting TRF sizes are analogous to bands on a DGGE gel and are also referred to as OTUs, since any one terminal fragment size is not restricted to any taxonomic group per se (Marsh 2005). TRF profiles are compared subsequently between samples by use of similarity matrices and multivariate statistics.

With new capillary sequencers, up to 384 samples can be analyzed in a single run. T-RFLP also has a higher resolving power than DGGE, with often twice as many OTUs determined per sample (Jones and Thies 2007), making T-RFLP the preferred choice for a high throughput method to initially screen for differences between communities. Devare et al. (2004) applied the T-RFLP technique to compare rhizosphere bacterial communities colonizing transgenic and non-transgenic corn and found that communities clustered by sampling time and year, but not by corn hybrid. Other studies have used T-RFLP to evaluate the effects of soil management on fungal community composition (Edel-Hermann et al. 2004) and the effects of solarization and crop rotation on bacterial communities (Culman et al. 2006), among many other applications. Artursson et al. (2005) combined bromodeoxyuridine immunocapture with T-RFLP to examine the effects of mycorrhizal inoculation and plant species on the active soil bacterial metagenome. T-RFLP need not be restricted to studying the 16 S rRNA gene. This technique can be used as a quick screen for any gene for which specific primers can be devised to examine differences between communities in environmental samples, such as *nif*H to compare populations of nitrogen-fixing bacteria or *amo*A to study ammonia oxidizing bacterial populations in soil. The main drawback of the use of this approach is the inability to further characterize TRFs or obtain sequence information as the sample is lost shortly after it is sized. However, once profiles are compared, the original PCR products from samples of interest can be used for cloning and sequencing experiments as described above. Alternatively, a gel-based approach called the 'physical capture method' of T-RFLP analysis can be employed when recovery of sequence information is desired (Blackwood and Buyer 2007). The technique is not as discriminating as the capillary approaches, but this level of resolution may not be needed for some applications or may not be possible in some research settings.

T-RFLP often yields a higher number of OTUs for use in comparative analyses than DGGE. However, all of these techniques yield numbers of OTUs that do not come close to the estimates of extant diversity in soil populations as estimated by DNA:DNA reassociation kinetics (discussed above). Hence, we are still viewing the tip of the iceberg as far as characterizing soil microbial diversity with these higher throughput DNA fingerprinting techniques.

A new DNA-based fingerprinting approach, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), can be used to separate PCR amplicons of the ITS regions first by size in a non-denaturing gel and then by melting characteristics in a second, denaturing gradient gel (Jones and Thies 2007). This approach yielded an order of magnitude higher number of OTUs than DGGE alone and three times the number of OTUs obtained by use of T-RFLP. Because the technique is gel-based, DNA spots of interest can be excised from the second dimension gel and then

cloned and sequenced. Far fewer OTUs were found in spots on the second dimension gel than were recovered from corresponding bands on the first dimension sizing gel, thus the technique allows OTUs of interest to be recovered much more easily. The disadvantage of this technique is that it is more laborious, therefore it does not lend itself to high sample throughput. Yet, its improved ability to discriminate between soil communities and retrieve sequence information make it a powerful technique for elucidating key differences in community structure between studied samples. Jones and Thies (2007) used the technique to study changes in soil bacterial community composition in relation to a naturally occurring gradient of Zn and Cd content in a soil in upstate New York.

Several additional PCR fingerprinting techniques target the ribosomal gene sequences. Ribotyping makes use of differences in the chromosomal positions or structure of rRNA genes to identify or group isolates of a particular genus or species. Ribotyping has been shown to be reproducible and hence has gained popularity for isolate fingerprinting and has found use in bacterial source tracking and other studies where the similarity of isolates obtained from different samples needs to be compared. The most frequently used ribotyping method is to identify RFLPs of rRNA genes by probing a Southern transfer of genomic DNA that has been hydrolyzed with an endonuclease. In amplified ribosomal DNA restriction analysis (ARDRA), rRNA gene sequences are amplified. In automated ribosomal intergenic spacer analysis (ARISA), the ITS region is amplified. PCR amplicons resulting from use of both methods are hydrolyzed subsequently with restriction enzymes and the resulting variations in restriction fragment sizes are analyzed on a gel. Chelius and Lepo (1999) used RFLPs of PCR amplified *nif*H sequences to study the diversity of nitrogen-fixing bacteria in the rhizosphere of wetland plant communities. In these applications, bands in the gel are again termed OTUs and similarities and differences between the fingerprints from different samples are analyzed using multivariate techniques. Use of ARISA may yield more OTUs from a given sample, but as the number of bands on the gel increases, the more difficulty one has in resolving individual bands in the analysis.

17.5.2 Quantitative and Real-Time PCR

An advance in PCR analysis that allows specific gene targets to be quantified is quantitative PCR (qPCR), also called real-time PCR. qPCR is a method that employs fluorogenic probes or dyes to quantify the number of copies of a target DNA sequence in a sample. This approach has been used successfully to quantify target genes that reflect the capacity of soil bacteria to perform given functions. Examples include the use of ammonia monooxygenase (*amo*A), nitrite reductase (*nir*S or *nir*K), and particulate methane monooxygenase (*pmo*A) genes to quantify ammonia oxidizing (Hermansson and Lindgren 2001), denitrifying (Henry et al. 2004) and methanotrophic (Kolb et al. 2003) bacteria, respectively, in soil samples. qPCR coupled with primers to specific internal transcribed spacer (ITS) or rRNA

gene sequences has also been used to quantify ectomycorrhizal (Landeweert et al. 2003) and endomycorrhizal fungi (Filion et al. 2003) as well as cyst nematodes (Madani et al. 2005) in soil.

17.5.3 Statistical Methods

The successful application of molecular techniques to population studies, particularly those based on the analysis of DNA or RNA in a gel matrix, relies heavily on the correct interpretation of the banding or spot patterns observed on electrophoretic gels. Gel images are typically digitized and band detection software is used to mark the band locations in the gel. The resulting band pattern is then exported to a statistical software package for analysis. Some analyses require that the fingerprint patterns obtained are first converted to presence/absence matrices; although average band density data are also used. The matrices generated are then compared using cluster analysis, multi-dimensional scaling, principal component analysis, redundancy analysis, canonical correspondence analysis, or additive main effects with multiplicative interaction model, among others. Each analysis will allow community comparisons, yet each has associated strengths and weaknesses. There are a number of software packages available that will enable one to compare and score PCR-fingerprints and produce similarity values for a given set of samples. Software packages, such as BioNumerics and GelCompar (Applied Maths, Kortrijk, Belgium), Canoco™ (Microcomputer Power, Ithaca, NY), PHYLIP (freeware via GenBank and the RDPII) and MatModel™ (Microcomputer Power) among others are used commonly. An advantage of using analysis programs, such as Bionumerics or GelCompar, is that fingerprints of communities generated from the use of several different markers can be combined. Generating a combined fingerprint in this way increases the robustness of similarity analyses based on PCR-fingerprints because it reduces the impact that one or two minor band differences has on the similarity matrices produced. The RDPII (Release 8.3) website provides analytical support for the analysis of T-RFLP data. The details of other analytical programs that support the analysis of data based on operational taxonomic units have also been published lately (Schloss and Handelsman 2005, 2006a,b).

The information that can be obtained from molecular characterization depends on the analysis technique. 16 S rRNA gene sequencing can aid in assigning species into genera and can be used for determining relationships between genera, but the information is frequently unable to resolve differences between closely related species. To overcome this limitation, one could use additional genetic information contained within ITS regions either by sequencing or by RFLP to further discriminate between closely related species.

To add value to the study of soil community ecology, a technique must be robust, that is, yield specific information about communities at the level of resolution required; it must be rapid and allow high throughput in order for the large number of samples needed for landscape studies to be processed with moderate effort.

17.6 Conclusions

Molecular tools are offering unparalleled opportunities to characterize Bacteria, Archaea and Eucarya in culture and directly from field soils. These tools are allowing us to ask questions at much larger geographic scales than have been possible previously. We are now able to examine such issues as how microbial populations vary across soil types and climatic zones (Fierer and Jackson 2006), in association with plant roots and between various plant species (Cardon and Gage 2006; Costa et al. 2006), and in response to soil management (e.g., Culman et al. 2006) or soil pollution (Liu et al. 1997). Molecular approaches also provide improved tools for seeking new inoculant consortia that may provide benefit in cropping systems. Genotypes that enjoy high representation in the soil population are likely to be competent saprophytes and be well adapted to site conditions. Pre-adapted strains that are also highly effective and genetically stable would then be excellent target organisms for future inoculants.

Soil has been dubbed 'The Final Frontier'. Modern molecular techniques developed to study microbial populations finally allow us access to the very large proportion of organisms that are present in the soil that we are currently unable to culture under laboratory conditions (Handelsman and Smalla 2003). They are also allowing us to begin to link identify with function (Dumont and Murrell 2005), which will lead to a better understanding of how changes in soil management practices may be altering ecosystem dynamics. Continually evolving technical developments open new horizons of research and applications that are enabling a far more complete and less biased view of microbial biodiversity in soil and the rhizosphere.

References

Amann RI, Ludwig W (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. FEMS Microbiol Rev 24:555–565

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Aneja MK, Sharma S, Munch JC, Schloter M (2004) RNA fingerprinting – a new method to screen for differences in plant litter degrading microbial communities. J Microbiol Methods 59:223–231

Artursson V, Finlay RD, Jansson JK (2005) Combined bromodeoxyuridine immunocapture and terminal-restriction fragment length polymorphism analysis highlights differences in the active soil bacterial metagenome due to *Glomus mosseae* inoculation or plant species. Environ Microbiol 7:1952–1966

Binnerup SJ, Bloem J, Hansen BM, Wolters W, Veninga M, Hansen M (2001) Ribosomal RNA content in microcolony forming soil bacteria measured by quantitative 16 S rRNA hybridization and image analysis. FEMS Microbiol Ecol 37:231–237

Blackwood CB, Buyer JS (2007) Evaluating the physical capture method of terminal restriction fragment length polymorphism for comparison of soil microbial communities. Soil Biol Biochem 39:590–599

Bloemberg GV, Wijfjes AHM, Lamers GEM, Stuurman N, Lugtenberg BJJ (2000) Simultaneous imaging of *Pseudomonas fluorescens* WCS365 populations expressing three different

autofluorescent proteins in the rhizosphere: new perspectives for studying microbial communities. Mol Plant-Microbe Interact 13:1170–1176

Bringhurst RM, Cardon ZG, Gage DJ (2001) Galactosides in the rhizosphere: utilization by *Sinorhizobium meliloti* and development of a biosensor. Proc Natl Acad Sci USA 98:4540–4545

Bruns MA, Buckley DH (2002) Isolation and purification of microbial community nucleic acids from environmental samples. In: Hurst CJ, Crawford RL, Knudsen GR, McInerney MJ, Stetzenbach LD (eds) Manual of environmental microbiology. American Society for Microbiology Press, Washington, DC

Cardon ZG, Gage DJ (2006) Resource exchange in the rhizosphere: molecular tools and the microbial perspective. Ann Rev Ecol Evol Syst 37:459–488

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene-expression. Science 263:802–805

Chelius MK, Lepo JE (1999) Restriction fragment length polymorphism analysis of PCR-amplified *nif*H sequences from wetland plant rhizosphere communities. Environ Tech 20:883–889

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 33:D294–D296 doi:10.1093/nar/gki038

Costa R, Gotz M, Mrotzek N, Lottmann J, Berg G, Smalla K (2006) Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. FEMS Microbiol Ecol 56:236–249

Courtois S, Frostegard A, Goransson P, Depret G, Jeannin P, Simonet P (2001) Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. Environ Microbiol 3:431–439

Culman SW, Duxbury JM, Lauren JG, Thies JE (2006) Microbial community response to soil solarization in Nepal's rice-wheat cropping system. Soil Biol Biochem 38:3359–3371

Devare MH, Jones CM, Thies JE (2004) Effect of Cry3Bb transgenic corn and tefluthrin on the soil microbial community: biomass, activity, and diversity. J Environ Qual 33:837–843

Dumont MG, Murrell JC (2005) Stable isotope probing – linking microbial identity to function. Nat Rev Microbiol 3:499–504

Edel-Hermann W, Dreumont C, Perez-Piqueres A, Steinberg C (2004) Terminal restriction fragment length polymorphism analysis of ribosomal RNA genes to assess changes in fungal community structure in soils. FEMS Microbiol Ecol 47:397–404

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Rohwer F (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7:57

Ekins R, Chu FW (1999) Microarrays: their origins and applications. Trends Biotechnol 17:217–218

Errampalli D, Leung K, Cassidy MB, Kostrzynska M, Blears M, Lee H, Trevors JT (1999) Applications of the green fluorescent protein as a molecular marker in environmental microorganisms. J Microbiol Methods 35:187–199

Felske A, Engelen B, Nübel U, Backhaus H (1999) Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. Appl Environ Microbiol 62:4162–4167

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci 103:626–631

Filion Mm, St-Arnaud M, Jabaji-Hare SH (2003) Direct quantification of fungal DNA from soil substrate using real-time PCR. J Microbiol Methods 53:67–76

Fjellbirkeland A, Torsvik V, Ovreas L (2001) Methanotrophic diversity in an agricultural soil as evaluated by denaturing gradient gel electrophoresis profiles of *pmo*A, *mxa*F and 16S rDNA sequences. Antonie Van Leeuwenhoek 79:209–217

Fritze H, Pennanen T, Kitunen V (1998) Characterization of dissolved organic carbon from burned humus and its effects on microbial activity and community structure. Soil Biol Biochem 30:687–693

Frostegard A, Baath E, Tunlid A (1993) Shifts in the structure of soil microbial communities in limed forest as revealed by phospholipid fatty acid analysis. Soil Biol Biochem 25:723–730

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

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl Environ Microbiol 66:5488–5491

Griffiths RI, Manefield M, Ostle N, McNamara N, O'Donnell AG, Bailey MJ, Whiteley AS (2004) (CO_2)-^{13}C pulse labelling of plants in tandem with stable isotope probing: methodological considerations for examining microbial function in the rhizosphere. J Microbiol Methods 58:119–129

Guschin DY, Mobarry BK, Proudnikow D, Stahl DA, Rittmann BE, Mirzabekov AD (1997) Oligonucleotide microchips and genosensors for determinative and environmental studies in microbiology. Appl Environ Microbiol 63:2397–2402

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Handelsman J (2005) Metagenomics: application of genomics to uncultured microorganisms (vol 68, p 669, 2004). Microbiol Mol Biol Rev 69:195

Handelsman J, Smalla K (2003) Conversations with the silent majority. Curr Opin Microbiol 6:271–273

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–R249

Henry S, Baudoin E, Lopez-Gutierrez JC, Martin-Laurent F, Baumann A, Philippot L (2004) Quantification of denitrifying bacteria in soils by *nir*K gene targeted real-time PCR. J Microbiol Methods 59:327–335

Hermansson A, Lindgren PE (2001) Quantification of ammonia-oxidizing bacteria in arable soil by real-time PCR. Appl Environ Microbiol 67:972–976

Hurt RA, Qiu XY, Wu LY, Roh Y, Palumbo AV, Tiedje JM, Zhou JH (2001) Simultaneous recovery of RNA and DNA from soils and sediments. Appl Environ Microbiol 67:4495–4503

Jiao JY, Wang HX, Zeng Y, Shen YM (2006) Enrichment for microbes living in association with plant tissues. J Appl Microbiol 100:830–837

Jones CM, Thies JE (2007) Soil microbial community analysis using two-dimensional polyacrylamide gel electrophoresis of the bacterial ribosomal internal transcribed spacer regions. J Microbiol Methods 69:256–267

Kolb S, Knief C, Stubner S, Conrad R (2003) Quantitative detection of methanotrophs in soil by novel *pmo*A-targeted real-time PCR assays. Appl Environ Microbiol 69:2423–2429

Landeweert R, Veenman C, Kuyper TW, Fritze H, Wernars K, Smit E (2003) Quantification of ectomycorrhizal mycelium in soil by real-time PCR compared to conventional quantification techniques. FEMS Microbiol Ecol 45:283–292

Lane DJ (1991) 16 S/23 S rRNA sequencing. In: Nucleic acid techniques in bacterial systematics. Wiley, Chichester, pp 115–175

Leake JR, Ostle NJ, Rangel-Castro JI, Johnson D (2006) Carbon fluxes from plants through soil organisms determined by field (CO_2)-^{13}C pulse-labelling in an upland grassland. Appl Soil Ecol 33:152–175

Lee N, Halkjaer N, 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

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC, Schleper C (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806–809

Li Y, Dick WA, Tuovinen OH (2004) Fluorescence microscopy for visualization of soil microorganisms – a review. Biol Fert Soils 39:301–311

Liu WT, Stahl DA (2002) Molecular approaches for the measurement of density, diversity and phylogeny. In: Hurst CJ, Crawford RL, Knudsen GR, McInerney MJ, Stetzenbach LD (eds) Manual of environmental microbiology. American Society for Microbiology Press, Washington, DC

Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl Environ Microbiol 63:4516–4522

Lu YH, Abraham WR, Conrad R (2007) Spatial variation of active microbiota in the rice rhizosphere revealed by in situ stable isotope probing of phospholipid fatty acids. Environ Microbiol 9:474–481

Madani M, Subbotin SA, Moens M (2005) Quantitative detection of the potato cyst nematode, *Globodera pallida*, and the beet cyst nematode, *Heterodera schachtii*, using real-time PCR with SYBR green I dye. Mol Cell Probes 19:81–86

Manefield M, Whiteley AS, Griffiths RI, Bailey MJ (2002a) RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. Appl Environ Microbiol 68:5367–5373

Manefield M, Whiteley AS, Ostle N, Ineson P. Bailey MJ (2002b) Technical considerations for RNA-based stable isotope probing: an approach to associating microbial diversity with microbial community function. Rapid Comm Mass Spec 16:2179–2183

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu PG, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Marsh TL (2005) Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. Methods Enzymol 397:308–329

McDonald IR, Radajewski S, Murrell JC (2005) Stable isotope probing of nucleic acids in methanotrophs and methylotrophs: a review. Org Geochem 36:779–787

Mullis KB, Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. Methods Enzymol 155:335–350

Muyzer G, Smalla K (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie Van Leeuwenhoek 73:127–141

Nakatsu CH (2007) Soil microbial community analysis using denaturing gradient gel electrophoresis. Soil Sci Soc Am J 71:562–571

NCBI News (2006/2007) National Center for Biotechnology Information 15(2):2

Neufeld JD, Mohn WW (2005) Unexpectedly high bacterial diversity in arctic tundra relative to boreal forest soils, revealed by serial analysis of ribosomal sequence tags. Appl Environ Microbiol 71:5710–5718

Neufeld JD, Yu ZT, Lam W, Mohn WW (2004) Serial analysis of ribosomal sequence tags (SARST): a high-throughput method for profiling complex microbial communities. Environ Microbiol 6:131–144

Nicol GW, Schleper C (2006) Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? Trends Microbiol 14:207–212

Ouverney CC, Fuhrman JA (1999) Combined microautoradiography–16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ. Appl Environ Microbiol 65:1746–1752

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ (1992) Primary structure of the *Aequorea-victoria* green-fluorescent protein. Gene 111:229–233

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

Rangel-Castro JI, Killham K, Ostle N, Nicol GW, Anderson IC, Scrimgeour CM, Ineson P, Meharg A, Prosser JI (2005) Stable isotope probing analysis of the influence of liming on root exudate utilization by soil microorganisms. Environ Microbiol 7:828–838

Rondon MR, August PR, Betterman AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNiel IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–2547

Roslev P, Iversen N, Henriksen K (1998) Direct fingerprinting of metabolically active bacteria in environmental samples by substrate specific radiolabelling and lipid analysis. J Microbiol Methods 31:99–111

Rotthauwe J-H, Witzel K-P, Liesack W (1997) The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl Environ Microbiol 63:4704–4712

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. Science 239:487–491

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

Schloss PD, Handelsman J (2006a) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. Appl Environ Microbiol 72:6773–6779

Schloss PD, Handelsman J (2006b) Introducing TreeClimber, a test to compare microbial community structures. Appl Environ Microbiol 72:2379–2384

Sessitsch A, Gyamfi S, Stralis-Pavese N, Weilharter A, Pfeifer U (2002) RNA isolation from soil for bacterial community and functional analysis: evaluation of different extraction and soil conservation protocols. J Microbiol Methods 51:171–179

Smalla K, Sobecky PA (2002) The prevalence and diversity of mobile genetic elements in bacterial communities of different environmental habitats: insights gained from different methodological approaches. FEMS Microbiol Ecol 42:165–175

Thies JE (2007a) Molecular methods for studying the soil biota. In: Paul EA (ed) Soil microbiology, ecology and biochemistry, 3rd edn. Academic Press. Elsevier, Oxford, pp 85–118

Thies JE (2007b) Soil microbial community analysis using terminal restriction fragment length polymorphisms. Soil Sci Soc Am J 71:579–591

Thies JE, Grossman JM (2006) The soil habitat and soil ecology. In: Uphoff N, Ball AS, Fernandes E, Herren H, Husson O, Laing M, Palm C, Pretty J, Sanchez P, Sanginga N, Thies J (eds) Biological approaches to sustainable soil systems. CRC Press, Taylor and Francis Group, Boca Raton, pp 59–78

Torsvik VL (1980) Isolation of bacterial-DNA from soil. Soil Biol Biochem 12:15–21

Torsvik V, Ovreas L (2002) Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5:240–245

Torsvik V, Goksoyr J, Daae FL (1990) High diversity in DNA of soil bacteria. Appl Environ Microbiol 56:782–787

Torsvik V, Daae FL, Sandaa RA, Ovreas L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. J Biotechnol 64:53–62

Treonis AM, Ostle NJ, Stott AW, Primrose R, Grayston SJ, Ineson P (2004) Identification of groups of metabolically-active rhizosphere microorganisms by stable isotope probing of PLFAs. Soil Biol Biochem 36:533–537

Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk HP, Schleper C (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. Environ Microbiol 7:1985–1995

Tunlid A, White DC (1992) Biochemical analysis of biomass, community structure, nutritional status and metabolic activity of microbial communities in soil. Soil Biochem 7:229–262

Ward DM (1998) A natural species concept for prokaryotes. Curr Opin Microbiol 1:271–277

Wellington EMH, 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

Wintzingerode Fv, Gobel UV, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21:213–229

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Wu LY, Thompson DK, Li GS, Hurt RA, Tiedje JM, Zhou JZ (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol 67:5780–5790

Xi C, Lambrecht M, Vanderleyden J, Michiels J (1999) Bi-functional *gfp*-and *gus*A-containing mini-Tn5 transposon derivatives for combined gene expression and bacterial localization studies. J Microbiol Methods 35:85–92

Zhou JZ (2003) Microarrays for bacterial detection and microbial community analysis. Curr Opin Microbiol 6:88–294

Zhou JZ, Thompson DK (2002) Challenges in applying microarrays to environmental studies. Curr Opin Biotech 13:204–207

Zwirglmaier K (2005) Fluorescence in situ hybridisation (FISH) – the next generation. FEMS Microbiol Lett 246:151–158