Kukita, Y., Higasa, K., Baba, S., Nakamura, M., Manago, S., Suzuki, A., Tahira, T., and Hayashi, K. (2002). A single-strand conformation polymorphism method for the large-scale analysis of mutations/polymorphisms using capillary array electrophoresis. *Electrophoresis* **23**, 2259–2266.

Landegren, U., Nilsson, M., and Kwok, P. Y. (1998). Reading bits of genetic information: Methods for single-nucleotide polymorphism analysis. *Genome Res.* **8**, 769–776.

Larsen, L. A., Christiansen, M., Vuust, J., and Andersen, P. S. (1999). High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: Robust multiplex analysis and pattern-based identification of allelic variants. *Human Mutat.* **13**, 318–327.

Larsen, L. A., Christiansen, M., Vuust, J., and Andersen, P. S. (2000). High throughput mutation screening by automated capillary electrophoresis. *Comb. Chem. High Throughput Screen* **3**, 393–409.

Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879.

Ren, J. C., and Ueland, P. M. (1999). Temperature and pH effects on single-strand conformation polymorphism analysis by capillary electrophoresis. *Human Mutat.* **13**, 458–463.

Schneider, K., Weisshaar, B., Borchardt, D. C., and Salamini, F. (2001). SNP frequency and allelic haplotype structure of beta vulgaris expressed genes. *Mol. Breeding* **8**, 63–74.

[16] Ribosomal RNA Probes and Microarrays: Their Potential Use in Assessing Microbial Biodiversity

By KATJA METFIES and LINDA MEDLIN

Abstract

The awareness that global biological diversity is affected by numerous, mostly human-made threats has made biodiversity assessment an important scientific issue for decades. Biodiversity includes different levels of complexity, such as community diversity, habitat diversity, genetic diversity, and species diversity. The application of molecular methods to answer ecological questions permits issues of biodiversity to be addressed at all levels. Microorganisms dominate global biological diversity in terms of their species numbers. However, their small size and limited morphological features make it challenging to obtain a comprehensive view of their biodiversity. The application of ribosomal RNA (rRNA) probes contributes significantly to the assessment of biodiversity at the molecular level. DNA microarrays offer a great potential to facilitate the application of molecular probes and other DNA analytical methods to answer ecological and biodiversity questions. We provide an introduction into the application of rRNA probes and DNA microarrays for the assessment of microbial biodiversity, as well as protocols for the implementation of DNA microarrays.

Copyright 2005, Elsevier Inc.
All rights reserved.

METHODS IN ENZYMOLOGY, VOL. 395
0076-6879/05 $35.00

Introduction

Biodiversity is a popular and widely discussed topic. It is generally accepted that the global biological diversity is under threat. Despite or because of the broad public attention biodiversity has received, the term is often used too simply. Most people associate it with total numbers of different species, but biodiversity is more than that: It also comprises habitat diversity, genetic diversity, and community diversity (Harper and Hawksworth, 1994; Purvis and Hector, 2000). Although biodiversity has been the focus of popular and scientific interest for the past two decades, scientists can only guess at global biodiversity if just species numbers are considered. These estimates range from 10 million to 30 million species. In general, there is a consensus in total numbers at about 13–14 million species on earth (Cruz, 1996; Eldrege, 1998; Mann and Plummer, 1996; Myers, 1998). Even though people want to explore other planets and outer space, our knowledge about the species on earth is far from being complete (Fig. 1). The number of scientifically described species is as few as 13%, or 1.7 million species (Myers, 1998). Among the described species, insects are the largest group, but it is estimated that the microbes are numerically the most dominant organisms on the planet. Nevertheless, most microbes are still unknown. These microscopic aerobic and anaerobic cell factories are the essential catalysts for all chemical reactions within biogeochemical cycles. Macroscopic life and planetary habitability completely depend on the transformations mediated by complex microbial communities. Considering the vital role of microorganismal ecology, it is believed that the microbial diversity is an important prerequisite for ecosystem stabilization. This is of particular interest for the marine environment, where protists and other microorganisms dominate the biodiversity and represent the basis of the food chain. However, it appears to be very difficult to study the

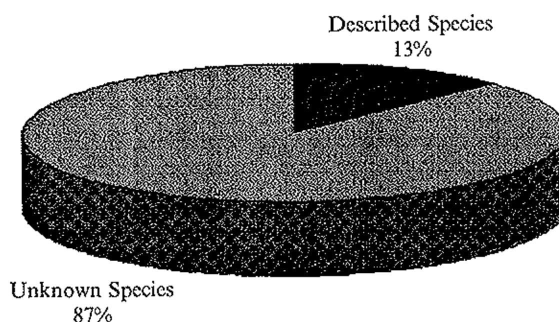

Fig. 1. Current status of the knowledge of global biological diversity in terms of species numbers (Myers, 1998).

biodiversity of microbial organisms by conventional methods. Knowing what "kinds" of organisms exist within a microbial population and how the community structure changes in response to environmental shifts challenges even the most advanced genetic technology and evolutionary theory. One of the fundamental tenets of biodiversity assessment regarding species composition is a reliable taxonomy that provides a hierarchical system for the identification and classification of diverse biological species. However, because of their small size and the presence of only limited or in some cases uninformative morphological features, microbial organisms are taxonomically very challenging. During the first half of the past century, microbial taxonomy based on morphological and physiological characteristics was controversial (Palleroni, 2003). With the introduction of molecular methods into microbial biology, methods became available that address questions related to microbial biodiversity with respect to compositional, structural, or functional aspects.

Assessment of Microbial Species Diversity with Ribosomal RNA

It was first demonstrated in the 1960s that ribosomal genes (ribosomal DNA [rDNA]) and their gene products (ribosomal RNA [rRNA]) could be used for a taxonomic classification of microbial species (Doi and Igarashi, 1965; Dubnau *et al.*, 1965; Pace and Campbell, 1971a,b). This was a big step forward in the study of microbial species diversity. Since then, the comparative analysis of rRNA sequences has become an indispensable method to gain new insights into the phylogeny and diversity of microbial organisms (Díez *et al.*, 2001; John *et al.*, 2003; Moon-van der Staay *et al.*, 2001). Homologous gene sequences possess far more phylogenetic information than phenotypic features. In comparison to other genes, the genes coding for the rRNA are particularly well suited for phylogenetic analysis, because they are universal—found in all cellular organisms; they are of relatively large size; and they contain both highly conserved and variable regions with no evidence for lateral gene transfer (Woese, 1987). The evolutionary information of the rRNA can be uncovered by comparative sequence analysis. Numerous publications have shown the power of the analysis of ribosomal sequences to identify both prokaryotic and eukaryotic microorganisms (Amann *et al.*, 1990; Groben *et al.*, 2004; Simon *et al.*, 2000). Direct cloning and sequencing of the small subunit rDNA (18S rDNA) from natural samples has, for example, permitted a broader view of the structure and composition of picoplankton communities (Giovannoni *et al.*, 1990; Giuliano *et al.*, 1999; López-Garcia *et al.*, 2001). Because of their small size, picoplankton samples are particularly difficult to analyze by conventional methods (Díez *et al.*, 2001; Moon-van der Staay

et al., 2001). The analysis of the 18S rDNA circumvents the selective step of laboratory cultivation (Giovannoni *et al.*, 1990).

Ribosomal RNA–Targeted Molecular Probes

The continually growing number of available algal 18S rDNA sequences (e.g., in the Ribosomal Database Project [RDP]) (Maidak *et al.*, 2001) makes it possible to design hierarchical sets of probes that specifically target the 18S rDNA from higher taxonomic levels down to the species level (Groben *et al.*, 2004; Guillou *et al.*, 1999; Lange *et al.*, 1996). This approach makes it possible to quantify representatives from very broad to circumscribed phylogenetic groups in a systematic fashion. In general, the principle of molecular probes identifies an organism by targeting any specific sequence in the genome of the organism. However, the binding sites of phylogenetically derived probes are mainly located within the polymorphic regions of the rRNA gene.

These characters make it possible to detect organisms specifically by the binding to a homologous site within the ribosomal gene sequence of the target species. The design of molecular probes involves the identification of unique sites within the ribosomal sequence of a taxon of interest. Molecular probes usually have a length of approximately 15–25 bases and at least one mismatch to the same region in all other known sequences. A useful tool for the design of molecular probes is the ARB software package (Ludwig *et al.*, 2004). It comprises tools for the handling of sequence information and any other type of additional data linked to the respective sequence. The database that is produced with the ARB software package can be structured according to phylogeny, thus providing the possibility to develop a hierarchical set of specific molecular probes at different taxonomic levels with the help of a probe design function in the program. The ARB software package is available as freeware from the Department of Microbiology, Technical University, Munich (http://www.arb-home.de). The program initially ran on a UNIX platform, but Linux versions are now available.

Applications of Molecular Probes

Molecular probes can be applied for the quantitative analysis of microbial communities with detection by flow cytometry, fluorescence *in situ* hybridization (FISH) (Lim *et al.*, 1999; Miller and Scholin, 1998) or other methods that take advantage of the hybridization principle (Fig. 2). With respect to the application of molecular probes for FISH, it is advantageous if the probes are specific for rRNA because rRNA is a highly abundant molecule in proliferating cells. The efficiency and practicality of molecular probe techniques have been successfully proven in numerous examples from

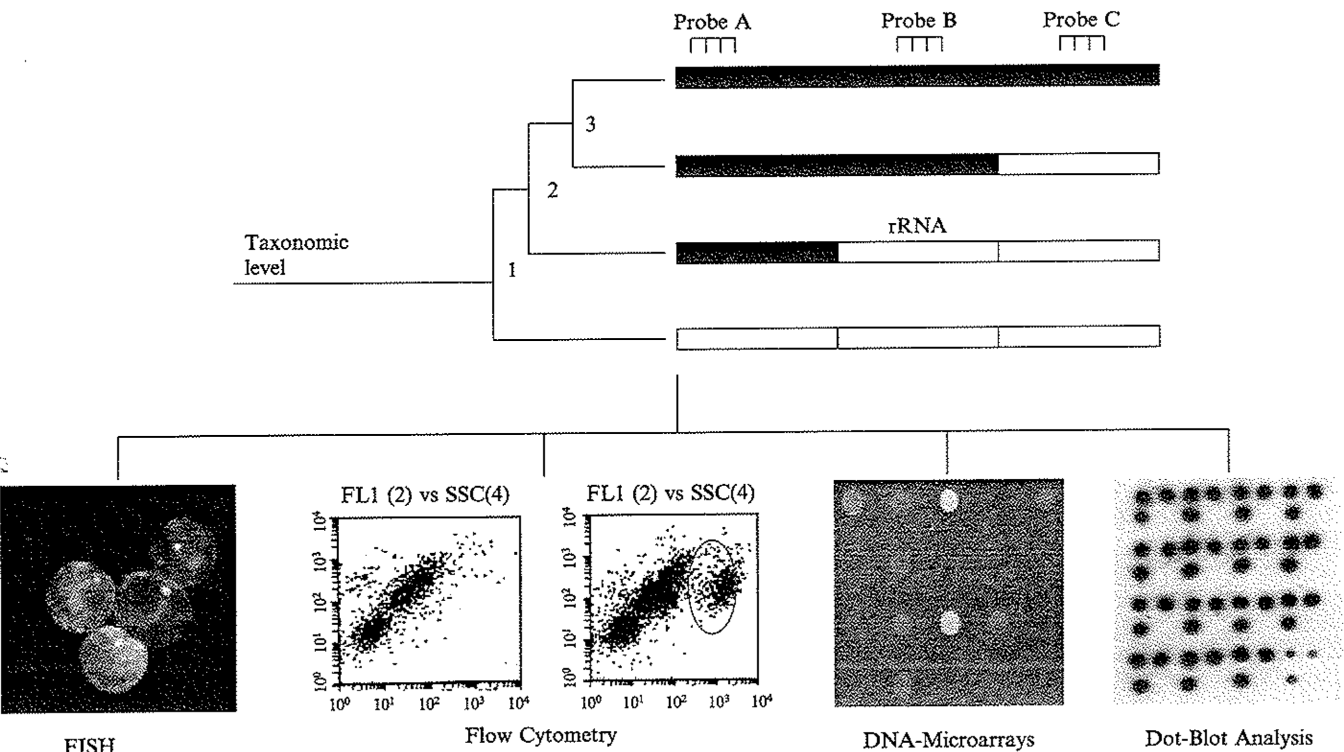

FIG. 2. Development and applications of molecular probes targeting ribosomal RNA (rRNA) sequences. Sequence databases are screened for short stretches of sequences within the rRNA that are shared among the organisms belonging to the same taxon. Hierarchical sets of probes target the rRNA at different taxonomic levels. The probes can be detected by various means.

both prokaryotic and eukaryotic target organisms (Biegala *et al.*, 2003; Sekar *et al.*, 2003). Nevertheless, these well-established approaches have the major disadvantage that they can be used to identify only one or a few organisms at a time and are not yet automated (DeLong *et al.*, 1989). This makes it very time consuming to get a broader view of a microbial sample.

DNA Microarrays

The introduction of the concept of DNA microarrays about 10 years ago suggests a solution for the limitations of molecular probe applications (Lockhart *et al.* 1996). They present an opportunity for high-throughput analysis of hybridization-based methods by miniaturization, and they possess the potential for highly multiplexed assays. DNA microarrays allow the simultaneous analysis of almost infinite numbers of probes simultaneously on just one DNA microchip (Brown and Botsein, 1999; Lockhart *et al.*, 1996). The DNA microchip is at the heart of the technology. It contains a high-density array of oligonucleotides, polymerase chain reaction (PCR) products or cDNAs spotted in an ordered way onto the surface of the DNA microchip (Lockhart and Winzeler, 2000). There are two fundamentally different ways for the manufacturing of gene chips. In the first, oligonucleotides are synthesized directly on the chip surface (Singh-Gasson *et al.*, 1999), which enables the production of gene chips with a high spot density. Currently gene chips exist with numbers as high as 100,000/cm^2 spots. Alternatively, it is possible to print or spot nucleic acids onto the surface of a glass slide (Okamato *et al.*, 2000) via a spotting device that features a high-speed robotic arm fitted with a number of pins or a piezoelectric pipette. Several companies (e.g., Qiagen Operon [Germany] or Affymetrix [USA]) produce customized DNA microarrays. In either case, the most broadly used support of rDNA microchips is a glass slide in the format of a commonly known microscope slide. Glass is an appropriate material for the production of DNA microarrays because it has low background fluorescence and is, therefore, compatible with fluorescence labeling. This is of particular importance because target nucleic acids like cDNA, RNA, or PCR products are usually detected via fluorescence. The fluorescence can be either incorporated in the nucleic acid directly via a fluorescent dye or indirectly by some other moiety such as biotin, which permits detection with a secondary fluorescent label (Cheung *et al.*, 1999; Southern *et al.*, 1999). Subsequent to the hybridization of the target nucleic acid to the immobilized probes on the DNA microchip, the analysis of the fluorescence pattern is performed with a microarray scanner. A microarray scanner is a device that bears a laser or a polychromatic light source for the specific excitation of the fluorophore (DeRisi *et al.*, 1997).

Assessment of Microbial Species Composition with DNA Microarrays

Microarray technology provides a promising tool based on molecular probes to identify species in samples from complex environments quickly without a cultivation step (Figs. 3 and 4). Consequently, it has great potential as an application for the assessment of microbial biodiversity in terms of species composition. This is of special interest for the identification of prokaryotic and eukaryotic cells with very small sizes and few distinct morphological features (Figs. 3 and 4). In addition, because the number of taxonomists in phycology has been decreasing steadily, DNA microarrays also might be of special value for phycological studies because they represent a tool that does not require a broad taxonomic knowledge to identify cells. Consequently, a growing number of publications report the use of microarrays bearing molecular probes that target the rRNA for the identification of microbial species. They have been used successfully in combination with an amplification of the rRNA gene for the identification of phytoplankton, nitrifying bacteria, bacterial fish pathogens, and sulfate-reducing prokaryotes (Call *et al.*, 2003; Guschin *et al.*, 1997; Loy *et al.*, 2002; Metfies and Medlin, 2004).

Even though DNA microarrays appear to have great potential to facilitate the analysis of the species composition of microbial samples, it has to be considered that amplification of nucleic acids from complex environmental samples can be very susceptible to biases leading to underrepresentation of targets or problems with the amplification of nucleic acids from certain species. One possibility to circumvent these possible

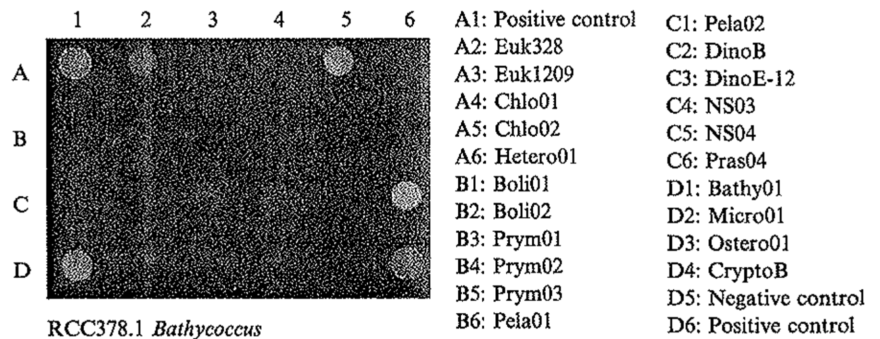

FIG. 3. Image of a scanned DNA microarray. The DNA chip contained the indicated probes. A polymerase chain reaction (PCR) fragment of the 18S ribosomal DNA (rDNA) of a species belonging to the picoplanktonic genus *Bathycoccus* was identified with a hierarchical set of probes that targeted the 18S rDNA at the level of kingdom (Euk1209), phylum (Chlo02), class (Pras04), and genus (Bathy01).

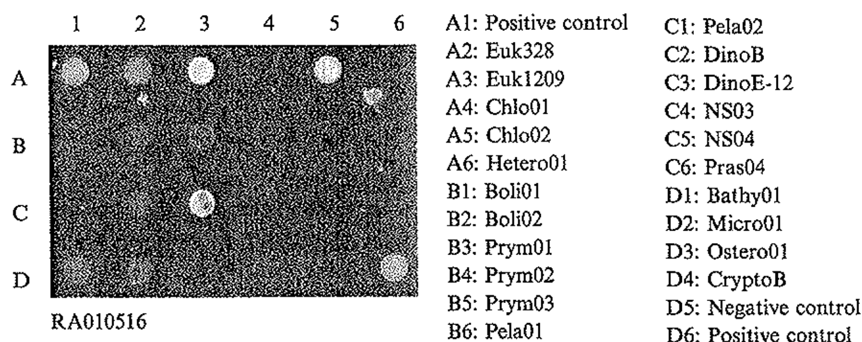

FIG. 4. Analysis of a picoplankton sample without previous cultivation. The 18S ribosomal DNA (rDNA) was amplified by polymerase chain reaction (PCR) and hybridized to a DNA chip that contained the indicated probes. The sample included Chlorophytes (Prasinophytes), Prymnesiophytes, Dinoflagellates, and Bolidophytes.

biases introduced by PCR amplification is to directly detect the rRNA from the species in the environmental sample. The use of rRNA as the target molecule for the organisms of choice could circumvent the PCR amplification because the molecule is present in high numbers in a cell. In prokaryotes, up to 80% of the RNA in a cell is rRNA (Woese, 1987). As another potential bias, it also must be considered that the specificity of the 18S rDNA-targeted probes on the DNA chip may have changed since the probes were initially designed. Probe development must always be subject to refinement. Because the known rRNA sequence database is only a small fraction of the total diversity, it is important to reevaluate the specificity of rRNA-targeted probes on a regular basis with respect to the continually increasing number of 18S rRNA sequences in public databases.

Gene Expression Profiling with DNA Microarrays

The application of microarrays for the identification of organisms is a younger application of the technology. However, the assessment of gene expression profiles was one of the first applications of microarray technology and is probably the most widely distributed application (Brown and Botstein, 1999). Gene expression profiling with DNA microarrays has numerous potential applications to address questions in ecology and evolution. The success of a species in any given environment is determined by its ability to recognize and respond to environmental changes. This involves the interaction of an organism with the environment, as well as the interaction of an organism with other organisms in a habitat. A response to an environmental change usually produces significant change

in the gene expression and protein composition of an organism. The estimation of the messenger RNA (mRNA) composition of an organism is one of the keys to understanding the cellular response to environmental forces. In this respect, the widely distributed methods for gene expression profiling in biomedicine have been adapted to address ecologically related questions. Gene expression profiling generates information about gene expression on a genomewide scale for different organisms at certain phases of development and under different environmental conditions. The response of the yeast, *S. cerevisiae,* to environmental changes was one of the first applications of expression profiling and has been intensively exploited (DeRisi *et al.*, 1997; Holstege *et al.*, 1998). In *S. cerevisiae,* it could be shown that the transcriptional activators, Msn2/Msn4, are involved in the activation of almost all of the genes in the cell that were induced by environmental stresses (Causton *et al.*, 2001). With the continuously growing number of completed genomes, it is to be expected that the number of this kind of study is going to increase, and it is conceivable that it could be possible to identify specific stress-induced genes or genes related to particular biochemical processes for a variety of organisms (Fig. 5).

In terms of biochemical microbial processes, the productivity of the ocean is an important issue. The photosynthetic carbon dioxide fixation by oceanic phytoplankton is one of the most important global carbon sinks (Chisholm, 2000). The measurement of the expression of carbon fixation genes in phytoplankton can serve as an indicator for the community photosynthesis. In this respect, the species-specific determination of the gene expression of RuBisCo (ribulose-1, 5-bisphosphate carboxylase/oxygenase)

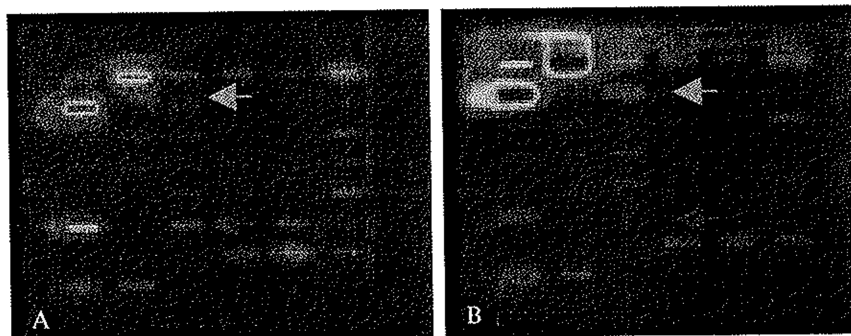

Fig. 5. Gene expression profiles of a subset of genes of the Antarctic diatom *Fragilariopsis cylindrus* at different temperatures (from Mock [2003]). (A) The diatom was grown at 2°. (B) The diatom was grown at 5°. Light arrows indicate a gene that is upregulated in panel B in comparison to panel A and dark arrows indicate a gene that is downregulated in panel B in comparison to panel A.

provided an insight into which phytoplankton groups were photosynthetically active at an oligotrophic site of the Gulf of Mexico (Wawrik *et al.*, 2002). However, understanding the response of an organism to its environment involves the elucidation of gene functions. Understanding gene functions may provide important insights to how organisms function and interact in the environment. Expression profiling could contribute to the understanding of other gene functions by the comparison of transcription profiles of unknown genes to transcription profiles of known genes. As probes for the elucidation of gene function and genes expressed specifically under certain environmental conditions or in the course of certain biochemical processes are under development, the possibility exists that a DNA microarray could be used to characterize an ecosystem in terms of the abundance and biodiversity of organisms that perform specific processes or respond to the stresses that are impacting populations in the field.

Applications of DNA Microarrays for Genotyping

Besides the determination of species numbers, biodiversity assessment may involve the estimation of genetic diversity and the investigation of population structure at the subspecies level. In the premolecular era, the study of microbial genetic diversity and population structures was often hampered by the same limitations as the identification of species, for example, the small size of the cells, a lack of considerable morphological features, and the inability to bring cells into culture (Medlin and Simon, 1998). The introduction of molecular methods was not only a quantum leap in regard to microbial phylogenetics, but it also made it possible to increase the resolution of the analysis of microbial population structure. The application of molecular marker techniques represents a means to study population structures at subspecies level unlike that available with conventional methods. Molecular markers can be locus specific and polymorphic in a studied population; therefore, they are very well suited to differentiate organisms in a given population. Molecular markers revealed, for example, the general trend that multiple isolates of a single species are often geographically related (Bakker *et al.*, 1995), but sometimes the geographic groupings uncover polyphyletic or paraphyletic taxa (Bakker *et al.*, 1995). In the past 2 decades, a number of genotyping methods based on molecular markers have been developed, and in theory most of them could be adapted to an analysis in combination with DNA microarrays. Fingerprinting methods such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP) are widely used and have been applied in the identification of genetic variation of different microorganism species

(Gomez and Gonzalez, 2001; Jackson *et al.*, 1999; Johannsson *et al.*, 1995). However, these methodologies based on the size separation of DNA fragments with gel electrophoresis are labor intensive and time consuming and are difficult to analyze statistically (Kingsley *et al.*, 2002). Moreover, they possess difficulties if bands on the gel have to be precisely correlated with genetic variants, because in some cases two bands on a gel that have the same size do not necessarily also have the same sequence.

The limitations of gel-based methods for genotyping can be overcome by the application of DNA microarray–based methods for genomic fingerprinting. DNA microarrays make it feasible to analyze target organisms in a sequence-specific and quantitative way, allowing molecular assessment of genetic variation among organisms in a population. With respect to genetic variation, the analysis of single nucleotide polymorphisms (SNPs) has become a focus of interest. SNPs are polymorphic positions in the genome of organisms at which nucleotides differ because of single nucleotide substitutions or single nucleotide insertions/deletions (Wang *et al.*, 1998). The analysis of SNPs with DNA microarrays could serve as a high-throughput method for the analysis of genetic variation (Lindblad-Toh *et al.*, 2000; Raitio *et al.*, 2001). However, a prerequisite of the analysis of SNPs by DNA microarrays is detailed sequence information of a number of variants of organisms from one species. Because of missing sequence information, the application of SNPs for genotyping of organisms is not very widely distributed if the variety of species is considered. Currently, genotyping by SNPs is mostly used to address questions related to the human genome. In particular, the amount of money that is required for the development of SNPs exceeds, in many cases, the budget of laboratories that are not involved in medical biotechnology or intensive genomics studies in other model organisms. With regard to microorganisms, a collection of 186 SNPs has been developed for the green alga *Chlamydomonas reinhardtii* (Vysotskaia *et al.*, 2001).

The application of DNA microarrays for genotyping is not restricted to the analysis of SNPs. In fact, there are examples in the literature where DNA microarrays facilitate the assessment of genetic variation. For example, DNA microarrays have been applied for the analysis of microsatellites (Radtkey *et al.*, 2000). Microsatellites are specific DNA sequences that consist of tandemly repeated monomers of 1–5 bp and are highly variable among individuals in a population. Hence, they possess the possibility to make inferences about a number of questions related to the assessment of biodiversity, including population structure, the level of genetic drift, sudden changes in populations size, effects of population fragmentation and the interaction with different populations, and the identification of new and incipient populations. Microsatellites have been

identified for a wide variety of eukaryotic genomes (Toth *et al.*, 2000). With respect to the assessment of microbial diversity, microsatellites have been identified for phytoplankton (Kang and Fawley, 1997). As with other fingerprinting techniques, the analysis of microsatellites is usually discriminated on gels, which are time consuming to prepare and pose problems for a reliable quantitative analysis. The application of DNA microarrays to the analysis of microsatellites would facilitate the assay and improve the accuracy of the analysis. For the analysis of microsatellites, a DNA chip would contain probes that target sequences of microsatellite loci that have been identified previously. Thus, the development of microsatellites is also a cost-intensive issue but may be worth the time investment for large-scale screenings.

To circumvent the need for sequence information before genotyping, DNA microarrays have been developed that reproduce a broadly applicable fingerprinting technique. They contain nanomers with random sequences. Such a DNA microarray has been used to display differences between closely related bacterial organisms belonging to the genus *Xanthomonas* and different strains of *Escherichia coli* (Kingsley *et al.*, 2002). DNA chips that contain nanomers can be used to generate a fingerprint of any microorganism, because on average nanomers occur only once every 131 kbp in double stranded DNA (Kingsley *et al.*, 2002). DNA microarrays with random oligonucleotides possess strong potential for fingerprinting microorganisms in different ways. Thousands of probes could be immobilized on one DNA chip, making it possible to generate fingerprints of microorganisms by using all possible nanomers on that single DNA chip. This would greatly increase the resolution of the fingerprinting method for the study of genetic variation in populations. In comparison to gel-based fingerprinting, all DNA chip–based fingerprinting methods have the major advantage that the DNA chip offers the possibility for reliable quantitative analysis of the different fingerprinting patterns. Genetic differences among isolates and closely related species can also be identified with microarrays that contained probes targeting the whole genome of organisms. In this case, polymorphisms and insertions can be detected as a reduction or elevation of a hybridization signal if two different genomic DNA samples are used for hybridization (Gibson, 2002). The comparison of whole genomes of different strains of various microbes indicates that polymorphism for gene content is not uncommon (Riley and Serres, 2000), suggesting genomic adaptations to certain habitats.

Protocol DNA Microarray Analysis

If it comes to the application of microarrays in practice, any DNA microarray analysis involves simplified four steps until, for example, target species in a sample or differentially displayed genes in an expression

analysis can be identified. Subsequent to the experimental design, the first step of a microarray analysis is the preparation of sufficient target nucleic acid. This step differs slightly between the different approaches and is discussed in detail separately for species identification or gene expression profiling. However, the principles of the following steps are shared among all potential applications of DNA microarray analysis in biodiversity assessment. The steps are labeling of the nucleic acids, hybridization, and analysis of the hybridization pattern on the DNA microarray.

Preparation of Target Nucleic Acid for Species Identification

In terms of species composition, environmental samples can be analyzed with a DNA microarray containing probes that specifically target the small subunit of the rRNA of a diverse set of microbial taxa at different taxonomic levels (Fig. 2). In the case of prokaryotes, such a chip would contain probes that specifically target the 16S rDNA (Loy *et al.*, 2002). However, a chip dedicated to the identification of eukaryotes would contain specific probes that bind to specific motives in the 18S rDNA (Metfies and Medlin, 2004). The analysis could be either carried out with rRNA isolated from the samples or with DNA fragments that were amplified from genomic DNA of the sample. The isolation of high-quality genomic DNA or rRNA from environmental samples is a prerequisite for a successful microarray analysis. This issue is discussed in detail in Chapter 2 of this volume (Valentin *et al.*).

Preparation of Target Nucleic Acid for Gene Expression Profiling

The focus of gene expression profiling is the identification of differentially expressed genes of the same type of cells grown under different environmental conditions on a genomewide level. The target nucleic acid of gene expression profiling is mRNA, because the response of an organism to environmental changes goes along with broad changes of the gene expression, which is reflected by the mRNA composition. It was mentioned before that a successful microarray analysis requires high-quality target nucleic acids. This is of particular importance for gene expression profiling. mRNA is very susceptible to degradation. Partial degradation of the RNA may result in a wrong estimation of the relative proportions of mRNAs hybridized to DNA microarrays.

Labeling of Nucleic Acids

Subsequent to the isolation of RNA or DNA, the next step of the microarray analysis is the generation of labeled target nucleic acid. A number of different strategies can be used to label nucleic acids. First,

nucleic acids can be directly labeled randomly without an amplification step. For DNA, Nick translation or the application of the Klenow fragment for random-sequence oligonucleotide-primed synthesis is the classic way to directly label nucleic acids (Ausubel *et al.*, 1998). Alternatively, there are easy-to-handle labeling kits like the BrightStar Psoralen-Biotin labeling kit (Ambion, Austin, TX) or the BiotinULS labeling kit (Fermentas, Germany), which make it possible to chemically incorporate fluorophores or a biotin moiety into both DNA and RNA. The principle of these direct labeling kits is that the fluorophore or the biotin is coupled to a compound that intercalates or binds covalently to nucleotides within the target nucleic acid. These labeling methods are quick and simple and make it possible to incorporate more than one label to the nucleic acid. However, a drawback to these direct labeling methods is that they do not select for degraded RNAs. Partially degraded RNAs are shorter than their cognate display elements. Shorter RNAs lead to a reduction in signal intensity. This displays, consequently, differences in fluorescent signal intensities between two different RNA samples that reflect differences in the quality of the two RNA preparations rather than differences in numbers of molecules. The application of oligo (dT)-primer for the synthesis of labeled cDNA can help to circumvent this problem in gene expression profiling, where mRNA is the target nucleic acid. However, if rRNA is the target nucleic acid, synthesizing fluorescently labeled PCR fragments of the ribosomal gene can minimize this problem. This could be done by the incorporation of nucleotides that carry a fluorophore or a biotin moiety in the course of the PCR. Alternatively, labeled primers could be used for the amplification. As an example, biotinylated target DNA for the analysis of eukaryotic microbial species diversity can be generated with a set of primers that amplify about an 1800 bp PCR fragment of the 18S rDNA. The primer set encounters the biotinylated primer 1F-biotin (5′-AAC CTG GTT GAT CCT GCC AGT-3′) and the unlabeled primer 1528R (5′-TGA TCC TTC TGC AGG TTC ACC TAC-3′) (Medlin *et al.*, 1988).

Hybridization

One of the goals of a successful hybridization is to minimize cross-hybridization between arrayed elements on the chip and nontarget nucleic acids. It is of significant importance to achieve high signal-to-noise ratios and to ensure that the signal intensity on the chip is proportional to the amount of nucleic acid bound to the chip. These goals can be achieved by optimizing the hybridization conditions in respect to a number of

different parameters like ionic and buffering conditions, hybridization volume, or time of prehybridization and hybridization. A hybridization buffer typically contains blocking reagents such as bovine serum albumin (BSA) or a detergent like sodium dodecyl sulfate (SDS) to minimize background noise. However, the important step to minimize background noise is the prehybridization. Typically, this involves an incubation of the DNA chips for a minimum of 30 min before the actual hybridization at hybridization temperature in hybridization buffer lacking target nucleic acid. During hybridization of the nucleic acid, the annealing to the arrayed elements on the DNA chip is promoted by a high ionic strength of the hybridization buffer. The concentration of the nucleic acid in the hybridization mixture is important for the outcome of the microarray experiment. If the concentration is too low, annealing will not proceed quick enough and it will be difficult to detect a sufficient signal for analysis. As a rule of thumb, the hybridization mixture should consist of the nucleic acid at a concentration of about 30 ng/μl in hybridization buffer (1 *M* NaCl/10 m*M* Tris; pH 8/0.005% Triton X-100/1 mg/ml BSA/0.1 μg/μl Herring Sperm DNA). If these amounts are not available, it is recommended to amplify the nucleic acid (Bowtell and Sambrook, 2003). The volume of the hybridization mixture should not exceed about 35–50 μl. During hybridization, the reaction is protected from evaporation by a coverslip, which is floating on the DNA chip and dislocated from the hybridization area, if the hybridization volume is too high. Before the application to the microarray, the hybridization mixture has to be denatured for 5 min at 94°. This step is important to dissolve secondary structures that might have been formed by the target nucleic acid. The application of the hybridization mixture is a crucial step in a microarray experiment. It is recommended to perform this step in a clean, dust-free environment, because dust particles on the chip can interfere with the hybridization signals. The hybridization mixture should be applied evenly onto the microarray and a coverslip should be deposited over the solution right after the application of the hybridization solution to prevent it from evaporating. An incubation of the DNA chip in a humid environment during hybridization additionally counteracts evaporation that occurs at the edges of the coverslip. Humid environments can be generated, for example, by filling the lower part of an empty pipette box with 5–10 ml of water. In this arrangement, the DNA chip rests on the perforated surface of the pipette box. Hybridization should be carried out for 12–16 h at hybridization temperature. The hybridization temperature is another important factor that strongly affects the hybridization efficiency in a microarray experiment. However, it depends on the nature of the hybridization buffer and the melting temperatures of the arrayed elements

on the chip. If molecular probes are spotted on the chip, the hybridization temperature should be selected and optimized in respect to the average melting temperature of the probes. In respect to hybridization buffers, optimal hybridization temperatures that contain 50% formamide are on average about 20° lower than optimal hybridization temperatures of aqueous hybridization buffers. If aqueous hybridization buffers are used, it is recommended to optimize the hybridization temperature in a range between 55° and 65°. Subsequent to the hybridization, cross-hybridized nucleic acids are washed off the chip by the application of washing buffers that have lower ionic strength than the hybridization buffer. Typically, washing buffers contain between 2× SSC and 0.01× SSC. Best results are achieved if a sequence of different washings with gradually decreasing SSC concentrations is performed. The first buffer after the hybridization can contain additionally SDS as a detergent at concentrations between 0.01% and 0.05%. However, the buffer used for the last washing step should not contain SDS. The presence of SDS in the buffer leads to strong background intensities and consequently a decrease of the signal-to-noise ratio.

Acquisition of Hybridization Signals. Before the measurement of signal intensities, hybridized biotinylated target nucleic acids that are labeled with biotin have to be visualized by staining the DNA microchips with streptavidin–fluorophore conjugates (e.g., streptavidin-Cy5 or streptavidin-Cy3 [Amersham Biosciences, Germany]). Dissolved in hybridization buffer, the streptavidin–fluorophore complex binds covalently to the biotin label on the nucleic acids. The staining works well if the DNA chips are incubated at room temperature for about 30 min. The concentration of the dye should be optimized in a range of 50–400 ng/ml. To prevent evaporation during the staining procedure, a coverslip is deposited onto the microarray. For the same reasons discussed earlier for the volume of the hybridization solution, the volume of the staining solution should not exceed 35–50 μl. Fluorescence images of the hybridized microarray are taken with a microarray reader, such as the Genepix 4000B-Scanner (Axon Instruments, Inc., USA). This device is a laser-based system that uses two lasers as excitation sources for the fluorescent dyes. Alternatively, some microarray readers use a charged-coupled device (CCD) camera. These systems use a beam of white light that is directed through optical filters to gain the adequate excitation wavelength. In contrast to the laser-based systems that are restricted to certain excitation wavelengths, the CCD-based systems are more flexible in this respect. The optical filters in the systems are usually relatively easy to replace and allow choosing between wider ranges of excitation wavelengths.

Image Analysis

The fluorescent signal intensities on the DNA chip are quantified using image analysis software packages, such as the GenePix 4.0 software (Axon Instruments, Inc., Union City, CA). The first step to precisely quantify each single spot is the superimposition of a grid of individual circles onto the microarray image. In this step, the location of a spot is defined. This step is very important, because spots might be dislocated from the expected position or have different sizes due to variations during the printing process. It follows the segmentation of the image, which is the classification of pixels on the image in foreground or background. This allows fluorescence intensities to be calculated and displayed as pixel values. Each pixel value represents the level of hybridization at a certain location on the slide. The amount of hybridized nucleic acid at a certain spot of the microarray is proportional to the fluorescent signal at the spot. For later calculations, the mean signal intensity is determined for each spot. The measurement of spot intensities includes signals originating from nonspecific hybridization or fluorescence emitted by other chemicals on the DNA chip. Therefore, it is finally important to assess the background signal. If segmentation has been performed, it is possible to determine local background intensities for each spot. The values of the fluorescent intensities are typically stored as 16-bit images and can be used for data analysis.

Concluding Remarks

Microarray technology has various potential applications for biodiversity assessment at all levels. The biodiversity of natural complex microbial samples can be addressed in terms of species numbers and composition without a cultivation step by the application of microarrays that contain hierarchical sets of molecular probes like rRNA probes, which allow the quantification of representatives from very broad to circumscribed phylogenetic groups in a systematic fashion. Gene expression profiles should provide insights into how organisms interact with their environment and how they respond to environmental changes. Biodiversity in terms of genetic diversity at the subspecies level can be addressed with microarrays that target molecular markers, such as RAPDs, AFLPs, or microsatellites. A microarray that contains random oligonucleotides could be used to generate fingerprints of populations that can be easily interpreted. A polymorphism in genomic DNA content of organisms of one species can be visualized by a microarray that targets whole genomes of organisms. Usually, however, this kind of microarray is most widely used to

generate expression profiles of organisms. A protocol for the application of DNA microarrays in biodiversity assessment was described in detail in this chapter.

References

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., and Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56,** 1919–1925.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1998). "Current Protocols in Molecular Biology." John Wiley & Sons, New York.

Bakker, F. T., Olsen, J. L., and Stam, W. T. (1995). Global phylogeography in the cosmopolitan species *Cladophora vagabunda* (Chlorophyta) based on nuclear DNA ITS sequences. *Eur. J. Phycol.* **30,** 187–210.

Biegala, I. C., Not, F., Vaulot, D., and Simon, N. (2003). Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence *in situ* hybridization and flow cytometry. *Appl. Environ. Microbiol.* **69**(9), 5519–5529.

Bowtell, D., and Sambrook, J. (2003). "DNA Microarrays." Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.

Brown, P. O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21,** 33–37.

Call, D. R., Borucki, M. K., and Loge, F. J. (2003). Detection of bacterial pathogens in environmental samples using DNA microarrays. *J. Microbiol. Methods* **53**(2), 235–243.

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. (2001). Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**(2), 323–337.

Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. (1999). Making and reading microarrays. *Nat. Genet.* **21**(Suppl. 1), 15–19.

Chisholm, S. W. (2000). Stirring times in the southern ocean. *Nature* **407,** 685–689.

Cruz, M. (1996). Management options for biodiversity protection and population. *In* "Human Population, Biodiversity and Protected Areas: Science and Policy Issues" (V. Dompka, ed.), pp. 71–89. American Association for the Advancement of Science, Washington, DC.

DeLong, E. F., Wickham, G. S., and Pace, N. R. (1989). Phylogenetic stains: Ribosomal RNA-based probes for the identification of single cells. *Science* **243,** 1360–1363.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**(5338), 680–686.

Díez, B., Pedrós-Alió, C., and Massana, R. (2001). Study of genetic Diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA Gene Cloning and Sequencing. *Appl. Environ. Microbiol.* **67,** 2942–2951.

Doi, R. H., and Igarashi, R. T. (1965). Conservation of ribosomal and messenger ribonucleic acid cistrons in *Bacillus* species. *J. Bacteriol.* **90,** 384–390.

Dubnau, D., Smith, I., Morell, P., and Marmur, J. (1965). Gene conservation in Bacillus species. I. Conserved genetic and nucleic acid base sequence homologies. *Proc. Natl. Acad. Sci. USA* **54**(2), 491–498.

Eldrege, N. (1998). "Life in the Balance: Humanity and the Biodiversity Crisis." Princeton University Press, Princeton, NJ.

Gibson, G. (2002). Microarrays in ecology and evolution: A preview. *Mol. Ecol.* **11,** 17–24.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345,** 60–63.

Giuliano, L., De Domenico, M., De Domenico, E., Hofle, M. G., and Yakimov, M. M. (1999). Identification of culturable oligotrophic bacteria within naturally occurring bacterioplankton communities of the Ligurian Sea by 16S rRNA sequencing and probing. *Microb. Ecol.* **37**(2), 77–85.

Gomez, P. I., and Gonzalez, M. A. (2001). Genetic polymorphism in eight Chilean strains of the carotenogenic microalga *Dunaliella salina* Teodoresco (Chlorophyta). *Biol. Res.* **34**(1), 23–30.

Groben, R., John, U., Eller, G., Lange, M., and Medlin, L. K. (2004). Using fluorescently-labelled rRNA probes for hierarchical estimation of phytoplankton diversity. *Nova Hedwigia* **79,** 313–320.

Guillou, L., Moon-van-der-Staay, S. Y., Claustre, H., Partensky, F., and Vaulot, D. (1999). Diversity and abundance of Bolidophyceae (Heterokonta) in two oceanic regions. *Appl. Environ. Microbiol.* **65,** 4528–4536.

Guschin, D. Y., Mobarry, B. K., Proudnikov, D., Stahl, D. A., Rittmann, B. E., and Mirzabekov, A. D. (1997). Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl. Environ. Microbiol.* **63,** 2397–2402.

Harper, J. L., and Hawksworth, D. L. (1994). Biodiversity: Measurement and estimation. *Philos Trans. R. Soc. Lond. B Biol. Sci.* **345**(1311), 5–12.

Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**(5), 717–728.

Jackson, P. J., Hill, K. K., Laker, M. T., Ticknor, L. O., and Keim, P. (1999). Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. *J. Appl. Microbiol.* **87**(2), 263–269.

Johannsson, M. L., Molin, G., Pettersson, B., Hlen, M., and Ahrne, S. (1995). Characterization and species recognition of *Lactobacillus plantarum* strains by restriction fragment length polymorphism (RFLP) of the 16S rRNA gene. *J. Appl. Bacteriol.* **79,** 536–541.

John, U., Fensome, R. A., and Medlin, L. K. (2003). The application of a molecular clock based on molecular sequences and the fossil record to explain biogeographic distributions within the *Alexandrium tamarense* "species complex" (Dinophyceae). *Mol. Biol. Evol.* **20,** 1015–1027.

Kang, T. J., and Fawley, M. W. (1997). Variable (CA/GT)n simple sequence repeat DNA in the alga *Chlamydomonas. Plant Mol. Biol.* **35**(6), 943–948.

Kingsley, M. T., Straub, T. M., Call, D. R., Daly, D. S., Wunschel, S. C., and Chandler, D. P. (2002). Fingerprinting closely related *Xanthomonas* pathovars with random nonamer oligonucleotide microarrays. *Appl. Environ. Microbiol.* **68**(12), 6361–6370.

Lange, M., Guillou, L., Vaulot, D., Simon, N., Amann, R. I., Ludwig, W., and Medlin, L. K. (1996). Identification of the class Prymnesiophyceae and the genus *Phaeocystis* with ribosomal RNA–target nucleic acid probes detected by flow cytometry. *J. Phycol.* **32,** 858–868.

Lim, E. L., Dennett, M. R., and Caron, D. A. (1999). The ecology of *Paraphysomonas* imperforata based on studies employing oligonucleotide probe identification on coastal water samples and enrichment cultures. *Limnol. Oceanogr.* **44,** 37–51.

Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Laviolette, J. P., Ardlie, K., Reich, D. E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J. B.,

Gingeras, T., Warrington, J., Patil, N., Hudson, T. J., and Lander, E. S. (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* **24**(4), 381–386.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14,** 1675–1680.

Lockhart, D. J., and Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature* **405,** 827–836.

López-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409,** 603–607.

Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K. H., and Wagner, M. (2002). Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.* **68,** 5064–5081.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K. H. (2004). ARB: A software environment for sequence data. *Nucleic Acids Res.* **32**(4), 1363–1371.

Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T. J., Saxman, P. R., Farris, R. J., Garrity, G. M., Olson, G. J., Schmidt, T. M., and Tiedje, J. M. (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Res.* **29,** 173–174.

Mann, C. C., and Plummer, M. L. (1996). "Noah's Choice: The Future of Endangered Species." Alfred Knopf, New York.

Medlin, L., Elwood, H. J., Stickel, S., and Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71,** 491–499.

Medlin, L., and Simon, N. (1998). Phylogenetic analysis of marine phytoplankton. *In* "Molecular Approaches to the Study of the Ocean" (K. E. Cooksey, ed.), pp. 161–179. Chapman and Hall, London.

Metfies, K., and Medlin, L. (2004). DNA microchips for phytoplankton: The fluorescent wave of the future. *Nova Hedwigia* **79,** 321–327.

Miller, P. E., and Scholin, C. A. (1998). Identification and enumeration of cultured and wild *Pseudo-Nitzschia* (Bacillariophyceae) using species-specific LSU rRNA-targeted fluorescent probes and filter based whole cell hybridization. *J. Phycol.* **34,** 371–382.

Mock, T. (2003). "Photosynthesis in Antarctic sea-ice diatoms." Dissertation, University Bremen, Germany.

Moon-van der Staay, S. Y., De Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409,** 607–610.

Myers, N. (1998). Securing the fabric of life. *People Planet* **7**(4), 6–9.

Okamato, T., Suzuki, T., and Yamamoto, N. (2000). Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat. Biotechnol.* **18,** 438–441.

Pace, B., and Campbell, L. L. (1971a). Homology of ribosomal ribonucleic acid of *Desulfovibrio* species with *Desulfovibrio vulgaris*. *J. Bacteriol.* **106**(3), 717–719.

Pace, B., and Campbell, L. L. (1971b). Homology of ribosomal ribonucleic acid diverse bacterial species with *Escherichia coli* and *Bacillus stearothermophilus*. *J. Bacteriol.* **107**(2), 543–547.

Palleroni, N. J. (2003). Prokaryote taxonomy of the 20th century and the impact of studies on the genus *Pseudomonas*: A personal view. *Microbiology* **Part 1,** 1–7.

Purvis, A., and Hector, A. (2000). Getting the measure of biodiversity. *Nature* **405**(6783), 212–219.

Raitio, M., Lindroos, K., Laukkanen, M., Pastinen, T., Sistonen, P., Sajantila, A., and Syvanen, A. C. (2001). Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. *Genome Res.* **11**(3), 471–482.

Radtkey, R., Feng, L., Muralhidar, M., Duhon, M., Canter, D., DiPierro, D., Fallon, S., Tu, E., McElfresh, K., Nerenberg, M., and Sosnowski, R. (2000). Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. *Nucleic Acids Res.* **28**(7), E17.

Riley, M., and Serres, M. H. (2000). Interim report on genomics of *Escherichia coli*. *Annu. Rev. Microbiol.* **54,** 341–411.

Sekar, R., Pernthaler, A., Pernthaler, J., Warnecke, F., Posch, T., and Amann, R. (2003). An improved protocol for quantification of freshwater *Actinobacteria* by fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.* **69**(5), 2928–2935.

Simon, N., Campbell, L., Ornolfsdottir, E., Groben, R., Guillou, L., Lange, M., and Medlin, L. K. (2000). Oligonucleotide probes for the identification of three algal groups by dot blot and fluorescent whole-cell hybridization. *J. Eukaryot. Microbiol.* **47,** 76–84.

Singh-Gasson, S., Green, R. D., Yue, Y., Nelson, C., Blattner, F., Sussman, M. R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17,** 974–978.

Southern, E., Mir, K., and Shchepinov, M. (1999). Molecular interactions on microarrays. *Nature Genet.* **21**(Suppl. 1), 5–9.

Toth, G., Gaspari, Z., and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **10**(7), 967–981.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., and Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**(5366), 1077–1082.

Vysotskaia, V. S., Curtis, D. E., Voinov, A. V., Kathir, P., Silflow, C. D., and Lefebvre, P. A. (2001). Development and characterization of genome-wide single nucleotide polymorphism markers in the green alga *Chlamydomonas reinhardtii*. *Plant Physiol.* **127**(2), 386–389.

Wawrik, B., Paul, J. H., and Tabita, F. R. (2002). Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. *Appl. Environ. Microbiol.* **68**(8), 3771–3779.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51,** 221–271.