MANUAL OF
Environmental Microbiology
THIRD EDITION

EDITOR IN CHIEF

Christon J. Hurst
Department of Biology, Xavier University, Cincinnati, Ohio
and
Facultad de Ingeniería, Universidad del Valle,
Ciudad Universitaria Meléndez, Santiago de Cali, Colombia

EDITORS

Ronald L. Crawford
Environmental Biotechnology Institute
University of Idaho
Moscow, Idaho

Jay L. Garland
Dynamac Corporation
Kennedy Space Center, Florida

David A. Lipson
Department of Biology
San Diego State University
San Diego, California

Aaron L. Mills
Department of Environmental Sciences
University of Virginia
Charlottesville, Virginia

Linda D. Stetzenbach
Department of Environmental and
Occupational Health
School of Public Health
University of Nevada, Las Vegas
Las Vegas, Nevada

ASM PRESS
Washington, D.C.

Address editorial correspondence to ASM Press,
1752 N St. NW, Washington, DC 20036-2904, USA

Send orders to ASM Press, P.O. Box 605, Herndon, VA 20172, USA
Phone: (800) 546-2416 or (703) 661-1593
Fax: (703) 661-1501
E-mail: books@asmusa.org
Online: estore.asm.org

American Society for Microbiology
1752 N St. NW
Washington, DC 20036-2904

Library of Congress Cataloging-in-Publication Data

Manual of environmental microbiology / editor in chief, Christon J. Hurst ;
editors, Ronald L. Crawford ... [et al.].—3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-1-55581-379-6 (hardcover)
ISBN-10: 1-55581-379-8 (hardcover)
1. Microbial ecology—Laboratory manuals. 2. Sanitary microbiology—Laboratory manuals. I. Hurst, Christon J. II. Crawford, Ronald L., 1947–
[DNLM: 1. Environmental Microbiology. QW 55 M294 2007]

QR100.M36 2007
579'.17—dc22

2006034637

10 9 8 7 6 5 4 3 2 1

Molecular Approaches for the Measurement of Density, Diversity, and Phylogeny

WEN-TSO LIU AND DAVID A. STAHL

12

This chapter considers the use of molecular methods for direct measures of abundance, diversity and phylogeny of environmental populations of microorganisms. These molecular methods are mainly based on direct nucleic acid sequence recovery, genomic DNA hybridization, and nucleic acid fingerprinting. Although these methods in part cannot be separated from fundamental questions (e.g., species concept and molecular systematics) in ecology and microbial systematics, full coverage of these interrelated topics is beyond the scope of this chapter. Our primary goal is to provide a reasonably complete accounting of available technology and associated methodological biases. This emphasis also results in the exclusion of certain molecular techniques from detailed discussion, since they have not been productively applied to the description of natural systems.

NUCLEIC ACID SEQUENCE MEASURES OF COMMUNITY STRUCTURE

There are three basic formats now used to recover DNA sequence information isolated from either pure culture or environmental samples: DNA probe hybridization, restriction enzyme digestion, and chain termination sequencing of cloned (or PCR-amplified) DNA templates. The former two methods are used to identify relatively short sequence elements. For example, restriction enzymes commonly recognize four to eight nucleotide sequence elements. DNA probes, with the capacity for single-nucleotide-mismatch discrimination, are usually around 20 nucleotides in length. Longer DNA probes are commonly used to identify homologous targets but do not provide defined sequence information. However, these approaches as well as community fingerprinting techniques (see below) all have some basic limitations.

Limitations on Recovery of Nucleic Acids from Environmental Samples

The efficiency and representativeness of nucleic acid recovery (both RNA and DNA) from environmental samples are fundamental concerns in all community studies using molecular approaches. These issues have been thoroughly discussed by a number of investigators (1, 3, 47, 115, 150, 160). For DNA-based analyses, possible biases associated with DNA recovery, PCR amplification, and DNA cloning are well recognized (160, 172). Experience in our laboratories has shown that, even when one uses nucleic acids extracted from pure cultures, PCRs are sometimes inconsistent. Amplification of rRNA gene sequences by using general primers has been shown in some cases to exclude important environmental populations (6, 160, 172). The issues of PCR-generated sequence hybrids (chimeras) and the extent of sequence variation between rRNA operons of individual organisms remain to be fully evaluated but can be partially improved through a PCR conditioning step (57). Thus, the proportional recovery of specific sequences cannot be equated with abundance.

For rRNA recovery, two aspects that are distinguished by the analytical approach are reviewed here. The first is the efficiency of extraction: what fraction of total nucleic acid is recovered from the environmental matrix? For example, although the breakage technique may disrupt all microorganisms present, recovery might be reduced by degradation or adsorption of nucleic acids to matrix material (e.g., clays). The second consideration is representative rRNA recovery. Does the fractional recovery correspond to the environmental abundance of the corresponding nucleic acids present in the environment? For example, a population resistant to breakage would be fractionally underrepresented, or conversely, an exceptionally easy-to-break microorganism would be overrepresented. The use of a universal hybridization probe to evaluate relative breakage efficiency of different groups is discussed below. Even though hybridization to extracted nucleic acids should provide more direct information on abundance, the meaning of abundance as defined in molecular terms is very different from established microbiological criteria, based on CFUs or direct cell counts. Thus, comparison of different measures of total biomass should be very informative. For example, phospholipid and respiratory quinone analyses might provide a relatively independent quantification of biomass that could be related to nucleic acid-based measures (43, 46, 68).

Another important consideration is the suitability of a method for identifying and quantifying environmental nucleic acid sequences. Methods suitable for identification may not be well suited to quantification. For example, DNA restriction fragment length polymorphism (RFLP) cannot be easily interpreted when applied to environmental systems of unknown complexity and population composition. Techniques that rely upon the recovery of relatively

intact DNA (e.g., for PCR amplification, cloning, or restriction digestion) generally must use less disruptive extraction techniques than do methods that analyze RNA. This is because DNA is much more sensitive to mechanical shearing than is RNA by the most disruptive of extraction methods, for example, mechanically reciprocal shaking with zirconium beads (142, 144). These are some aspects of nucleic acid recovery that must be more fully resolved before the application of any nucleic acid technology can be routinely applied to environmental studies.

Nucleic Acid Probes

Either DNA or RNA can serve as a nucleic acid probe. However, for a variety of technical reasons (e.g., ease of synthesis and stability), most studies have used DNA probes. There are two basic categories of DNA probes, functional and group specific. Functional probes targeting genes encoding specific enzymes are used to evaluate specific chemical transformations or potential activity of environmental populations (44, 67, 167). Although they will not be specifically addressed in this chapter, functional probes provide an essential connection between the different measures of environmental diversity (phylogeny) and ecology. In this regard, the use of phylogenetic groups as a measure of environmental diversity ultimately must include better understanding of unifying phenotypic characteristics of circumscribed groups. If certain traits are found to reflect membership within a group, these traits should serve to better relate community structure and function. Some examples of traits, and corresponding genes, conserved within phylogenetically defined groups include genes for nitrogen fixation (44), Ni-Fe hydrogenase (167), cellulases within some cellulolytic assemblages (67), and dissimilatory (bi)sulfite reductase (179). Although the emphasis of this chapter is molecular, it is the phenotypic reflection of molecular diversity that must remain the central consideration in environmental microbiology.

Group-specific (phylogenetic or taxonomic) probes generally target conserved biopolymers that can be used to infer phylogenetic relationships among the host organisms (5, 142). rRNA is the most widely used target molecule, and probes can be designed to target phylogenetic groups of differing evolutionary diversity, in other words, to provide explicit measures of community structure at different levels of resolution. Three types of rRNA are common to the ribosomes of prokaryotes and eukaryotes: the 16S (18S for eukaryotes), 23S, and 5S rRNAs. Since the larger rRNAs of many eukaryotes and some prokaryotes differ significantly in size, the terms small-subunit (SSU) rRNA and large-subunit (LSU) rRNA are used to refer to the 16S (18S) and 23S rRNAs, respectively, in this contribution. For a variety of technical and practical considerations (size, information content, and ease of sequencing), the SSU rRNA has become the standard measure for defining phylogenetic affiliation. Well over 160,000 SSU rRNA sequences are now available in the Ribosomal Database Project database (http://www.cme.msu.edu/RDP) (19) along with other databases for LSU and 5S rRNA sequences (24, 146, 157, 158), and the number of sequences is rapidly increasing. They have provided the most encompassing of available molecular frameworks to explore natural microbial diversity and phylogeny (50, 110, 140, 166).

There are two basic formats for using phylogenetic probes to study the environmental distribution of microorganisms. They are hybridization to total rRNA extracted from the environment (114, 116-118, 144) and hybridization to whole cells for subsequent microscopic visualization and enumeration. Nevertheless, a full presentation or review of these two methods and the rRNA-targeted probes used would require extensive referencing of recent applications to environmental, diagnostic, and determinative research. In this regard, the reader is directed to recent reviews (4, 5, 73, 142, 166) and references therein for a more complete description, and we apologize for any exceptional omissions. The following review is intended to provide an overview of key considerations on the design and optimization of rRNA probes in the quantitative analysis of natural systems.

In addition, the use of total genomic DNA probes is another category of group-specific probes for species-level identification (e.g., reverse sample genome probing [RSGP]) and is discussed below in "Genomic DNA Hybridization Measures of Community Structure."

Phylogenetic Probe Design

The essential attribute of the rRNAs with regard to probe design is the regional conservation of nucleotide sequence. Although overall conserved in sequence, these biopolymers exhibit great variation in regional sequence conservation. Some nucleotide positions and locales have remained virtually unchanged since the divergence of all existing life (universal sequences), whereas other regions vary so quickly that they can be used to differentiate among species of bacteria. In addition, the generally high copy number of rRNA per cell lends greater sensitivity to direct detection that uses hybridization techniques.

An extensive set of probes (~15 to 25-mers) has been designed to complement the rRNAs (5, 142), and among them a large fraction has been collected in two different databases (http://www.cme.msu.edu/OPD/ and http://www.probebase.net) (2, 73). Probes vary according to the region of the molecule selected as the hybridization target. The target region defines the diversity of organisms encompassed by a probe. Species-specific probes usually complement the most variable regions. More-general probes, identifying phylogenetic groups of rank greater than species, target more conserved regions of the molecule (142, 144, 166). The publications describing the development of SSU rRNA probes for clinical or environmental applications are much too numerous to address here. Thus, we will not discuss the more specialized applications but will limit discussion to the more general class of probes. Since these probes are designed to encompass larger phylogenetic groupings, they can be used to provide a phylogenetic overview to community structure. Figure 1 presents a very general overview of the character of probe design within a phylogenetic framework. It is an incomplete representation of general probes, and the reader is directed to an online resource (probeBase) for the most up-to-date and complete listing of probes (http://www.microbial-ecology.net/probebase/).

Probe Characterization: Phylogenetic Nesting of Probes

An essential aspect of probe development is the demonstration of target group specificity. To some extent, this can be demonstrated empirically by using a collection of target and nontarget group nucleic acids or fixed reference cells for studies that use whole-cell hybridization (5). Our laboratories and others have routinely used a panel of diverse rRNAs ("phylogrid") to characterize new probes (25, 66, 118). Prior to an evaluation by using the reference panel, it is essential that the temperature of dissociation (T_d) of the

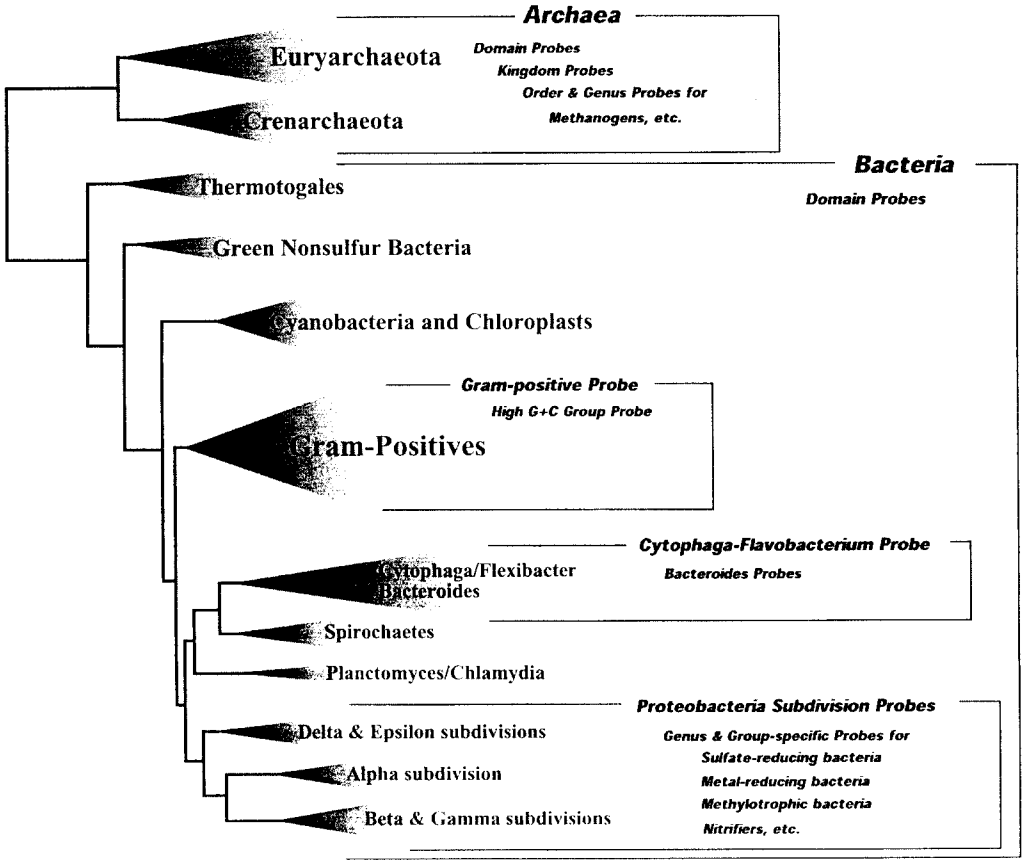

FIGURE 1 Partial listing of group-specific probes in relationship to prokaryote phylogeny. The phylogeny was adapted by Stahl (139) using the maximum likelihood analysis of Olsen, Woese, and Overbeek (107) to provide the relative branching order of the major prokaryotic lineages. The probes for the indicated target groups have been described: archaeal domain (38, 118, 143), bacterial domain (20, 38, 142), archaeal kingdoms (15), gram-positive organisms (S. Toze and D. Stahl, unpublished data), low-G+C gram-positive organisms (88), high-G+C gram-positive organisms (122), *Cytophaga-Flavobacterium* and *Bacteroides* (81), proteobacterial subdivisions (82), sulfate-reducing bacteria (25), methylotrophic bacteria (14, 154), and nitrifying bacteria (89, 165). The reader is referred to these references for more complete descriptions of probe design and characterization.

probe/target complex be experimentally determined (5, 142). It is generally insufficient to use one of the available formulas to predict T_d (142). Also, the temperature interval over which probe dissociates from the target RNA varies considerably for probes having comparable T_ds. Knowledge of the temperature range over which dissociation occurs is essential for adjusting hybridization conditions as needed to discriminate between closely related nontarget species rRNA. Thus, the initial T_d characterization should include closely related nontarget species rRNA directly extracted from pure cultures or synthesized by in vitro transcription of SSU rRNA clones (22a, 89, 113). An independent T_d evaluation must be used to characterize probes used for whole-cell hybridization, since there may be a significant difference between the T_d values and transition temperature ranges determined for the same probe by using different formats such as membrane examination, whole-cell hybridization, and DNA microarray.

There is a limitation to an empirical characterization of probe specificity. If we have only a limited appreciation of microbial diversity, it is impossible to construct a reference panel to unequivocally demonstrate specificity. However, there are a couple of additional methods to further evaluate the situation. The first is to use multiple probes, each having the same target group specificity, to quantify a single target population. For example, two probes for the domain *Archaea* (118) were used by DeLong and coworkers to independently confirm estimates of high archaeal abundance in Antarctic waters (22). Both probes hybridized to the same fraction of SSU rRNA extracted from these waters. The second approach to probe validation takes advantage of the phylogeny. As already noted, it is generally possible to construct probes for phylogenetic groups of various evolutionary depths. These probes are of a hierarchical specificity and have been described as being "nested." The use of a nested set of probes to characterize environmental diversity provides yet another consistency check. If the more specific probes fully represent the larger phylogenetic group, then the sum of the specific probe hybridization values should equal that obtained by using the more general probe. For example, the sum quantification obtained by using a complete set of species-specific probes should equal that of the corresponding genus-level probe (114, 116-118, 144). This approach was used to identify a novel lineage ("species") of

cellulolytic bacteria in the equine cecum (66) and was frequently used to evaluate consistency between domain-probe summation and total population abundance determined by using a universal probe (116).

General caveats relating to the use of phylogenetic probes include the following: the probes are tools subject to refinement through experimentation. Only through general application and combination use with other methods of community analysis will they be fully evaluated or, as necessary, refined. Also, the resolution of the SSU rRNA probes is approximately at the level of species. Questions relating to the abundance and distribution of subspecies and strains will require the combination of different approaches and methods (e.g., fluorescent antibody techniques). Another concern relates to the extraction of RNA from environmental samples. Although resistant to mechanical breakage, RNA is more subject to degradation during and following extraction, generally as a consequence of endogenous nucleases or nuclease contamination. One consequence of partial degradation of sample is variable destruction of different probe target sites. For example, one of the regions used as a target site for hybridization to a "universal" probe is very sensitive to degradation (120). Thus, for methods of quantification that use extracted rRNA, it is essential that sample integrity be evaluated. This is most conveniently accomplished by using acrylamide gel electrophoresis to demonstrate recovery of high-molecular-weight species (1, 3, 128). Furthermore, the probe target position within the rRNA significantly affects the hybridization efficiency and specificity when used for hybridization to whole cells, e.g., fluorescence in situ hybridization (FISH) (36). Based on a systematic analysis that used 171 nucleotide probes (mostly 18-mers) to encompass the SSU rRNA of *Escherichia coli*, the hybridization signal (fluorescent intensity) obtained from whole-cell hybridization was shown to vary as much as 75-fold among probes. Analysis of signal intensity relative to target position suggested that hybridization intensity was highly related to higher-order structure within the rRNA and could be classified into six different "brightness" regions (36). This information is very useful for probe design with commercial packages or free software available on the Internet (e.g., ARB) (76) and for improving the whole-cell hybridization signal (35). For example, by use of unlabeled oligonucleotides (helper probes) adjacent to the target region of a fluorescently labeled probe in whole-cell hybridization, the hybridization efficiency could be improved significantly with an enhancement of signal intensities varying from 4- to 25-fold (35).

Restriction Enzymes

The most common format for using restriction enzymes to define environmental diversity is to combine digestion and fractionation of DNA extracted from an environmental sample with hybridization by using nucleic acid probes complementary to conserved gene sequences common to all or many of the organisms present in the sample. The general format was first described by Southern and is often termed "Southern blotting" (137). The rRNAs are optimal targets for hybridization (ribotypes) (41), but other conserved elements also have been used for environmental studies (e.g., nitrate reductase and formyltetrahydrofolate synthetase) (72, 135). In application, restricted DNA is separated by size on an agarose gel and transferred to a membrane support for hybridization with radiolabeled probe or a label appropriate for use with one of a variety of nonradioactive detection formats (e.g., digoxigenin) (142). The resulting population of different-sized DNA fragments hybridizing to the probe is then used to infer relationships between individual isolates or resolve different environmental populations. The separation of genes derived from different populations requires that they differ in sequence at the sites of DNA restriction or differ in length of DNA flanked by common restriction sites. For this reason, more than one restriction enzyme is generally used for restriction enzyme analysis and the resulting size distribution patterns (banding patterns) are compared.

RFLP analysis has been used to characterize extracted total DNA (164) or specific PCR-amplified DNA (92). Sequence diversity is evaluated by digesting the native or amplfied DNA with a restriction endonuclease(s) followed by size fractionation by electrophoresis on an agarose or a polyacrylamide gel. The separation of genes derived from different populations requires that they differ in sequence at the sites of DNA restriction or differ in length of DNA flanked by common restriction sites. For this reason, more than one restriction enzyme is generally used for RFLP analyses.

The primary use of RFLP analysis has been to evaluate sequence variation among rRNA gene sequences. This general method is sometimes termed "amplified rRNA restriction analysis" and is commonly used to estimate diversity among different microbial isolates (60) or rRNA clones recovered from the environment (92). It requires the use of PCR primers complementary to conserved regions of the SSU or LSU rRNA genes or the 16S-23S intergenic regions (87) and high-resolution agarose gels to effectively separate fragments with length down to approximately 50 bp. Computer-simulated RFLPs on 106 bacterial SSU rRNA sequences from representative bacterial phyla indicate that combined RFLPs of at least three separate digestions with tetrameric restriction enzymes (i.e., four-base recognition site) are required to resolve those bacterial populations to the genus level (median sequence identity $< 96.1\%$) (93). The efficacy of detecting and differentiating bacterial rRNA genes is dependent upon selection of appropriate restriction enzymes, since the RFLP patterns are not due to true restriction site polymorphism but result primarily from insertions and deletions in the SSU rRNA sequence (13). A major limitation of this approach is that it is very time consuming and cannot be used to screen a large set of isolates or clones in a cost-effective manner.

RFLP has also been used for the analysis of environmental DNA. For example, restriction digestion of rRNA gene sequences recovered by PCR amplification can produce a banding pattern (i.e., band numbers and intensity) that serves as a fingerprint of the community. The fingerprint provides a quick assessment of genotypic changes that may result from temporal or spatial changes within a habitat (86). RFLP can be combined with hybridization by using group or functional probes to better define microbial diversity in an environmental sample. However, this approach is of limited use for demonstrating the presence of a specific phylogenetic group or for estimating species richness and evenness, since each SSU rRNA gene amplicon can contribute to multiple restriction fragments (93).

To further apply restriction enzyme for quantitative analysis of microbial community structure, we highlight a new method that utilizes the endonuclease property of RNase H to cleave the RNA strand of RNA-DNA hybrid duplexes at a specified site into two fragments (156). The first step of this method is to direct the cleavage site of the RNase H on RNA molecules by hybridizing the target RNA with an olignucleotide probe or "scissor" probe. After the formation of DNA (i.e., scissor probe)-RNA hybrids,

RNase H is added to cleave the RNA strand at the probe-binding site, followed by quantification of the cleaved RNA and intact RNA by using capillary electrophoresis systems. By varying the hybridization conditions (e.g., formamide), the digestion reaction, and probe specificity (i.e., perfectly matched versus mismatched), curves for cleavage efficiency of the RNA strand of perfectly matched and mismatch probe-RNA duplexes with respect to formamide concentrations can be produced. By selecting an optimal formamide concentration, RNase H is shown to cleave the RNA strand of perfectly matched duplexes but not mismatched duplexes. By use of rRNA-based scissor probes at different levels of phylogenetic specificity (e.g., domain, group, and species), this method was demonstrated to successfully quantify bacterial and archaeal fractions in model and complex communities and to detect a specific microbial population (≥1 or 2% of total rRNA) within a microbial ecosystem. The sensitivity of detection and quantification of microbial populations is reported to be dependent on (i) the use of high-quality intact rRNA extracted from the environments studied, (ii) the mismatched position of the nontarget rRNA molecules, and (iii) the resolution of electrophoresis systems for small cleaved rRNA fragments (156).

Direct Sequence Analyses of SSU rRNA

By far, the sequence information most commonly extracted from natural systems is that for the rRNAs. There are three basic methods to recover rRNA sequence information from nucleic acid extracted from environmental samples (see reference 124 for a detailed description). They are (i) shotgun cloning, screening, and sequencing (106, 110, 129); (ii) cDNA cloning and sequencing of rRNA (168); and (iii) PCR amplification, cloning, and sequencing (6, 166). The PCR-based methods can use either DNA or RNA as template, the latter requiring the use of reverse transcriptase to generate cDNA from rRNA. These sequencing approaches all require the screening and analysis of large clone collections. The screening of a shotgun library derived from total environmental DNA is the more difficult, since only a small fraction of the clones contain part or all of the rRNA gene, occurring at about 0.125 to 0.3% of the clones (129). They can be identified by hybridization, for example by using total rRNA derived from the environmental sample as probe. Alternatively, DNA probes targeting highly conserved regions of the molecule (phylogenetic probes) may be used. This first screening step is generally not necessary for PCR-based recovery methods, since the majority of clones will contain rRNA sequences. The second phase of the analysis is the elimination of redundant clones in order to avoid expense and time associated with unnecessary sequence determinations. A variety of approaches have been used to identify redundant clones, using or combining the following strategies: complete or single-nucleotide sequencing of a small variable region, species- or group-specific phylogenetic probe screening, restriction analysis, and denaturing gradient gel electrophoresis (DGGE) (97, 106, 166). A more recently described method for rapid sequence analysis of rRNA sequence types in environmental samples employs a series of enzymatic reactions to amplify and ligate short sequences (sequence tags) from a variable region of the rRNA gene. The ligated products, each containing as many as 20 sequence tags, are then cloned and sequenced. This approach, termed serial analysis of ribosomal sequence tags, provides a much higher-throughput survey of sequence diversity but has the associated limitation of recovering much less sequence information for each rRNA variant (59, 100).

The resolution of the different screening techniques must be balanced against time and expense associated with each strategy. This determination can be made only in consideration of community diversity (number and frequency of different rRNA genes) and research objectives. The general approach as developed for community-level analyses based on rRNA sequence content could be applied to any biopolymer of appropriate conservation and community representation.

GENOMIC DNA HYBRIDIZATION MEASURES OF COMMUNITY STRUCTURE

Genomic DNA hybridization is mainly used in bacterial systematics to determine the degree of genetic similarity between genomic DNA sequences and thus provides limited information of specific sequence content. The extent to which these methods can be used to identify individual populations or to estimate genomic diversity (total number of different genomes) in a microbial community varies with method, genomic diversity, and the aforementioned questions of microbial species definition. Two methods are introduced.

RSGP

RSGP is developed by assuming that the entire genome of a microorganism can be used as a specific probe for its detection in the environment (for detailed review, see reference 40). Whole-genome probes have been used to detect *Mycobacterium*, *Mycoplasma*, *Chlamydia*, *Bacteroides*, *Pseudomonas*, *Sphingomonas*, and *Campylobacter* species (7, 28, 29, 49, 91, 134, 169). RSGP reverses the usual relationship of sample DNA and probe. The genomic DNA from different reference organisms is denatured and immobilized on a membrane support, the reference panel. DNA extracted from the environment, containing an unknown diversity of organisms, is randomly labeled and is hybridized to the reference panel. Under conditions of high stringency, whole-genome probes hybridize only to identical or closely related genotypes at the level of species or subspecies. For example, when a reference panel of DNA from different sulfate-reducing bacteria was hybridized with randomly labeled genomic DNA from any species represented on the panel, only self-hybridization or hybridization with nearly identical isolates was observed (162, 163).

The following overview addresses only key technical considerations (161–163). The method requires the antecedent isolation of reference organisms from the environment. Chromosomal DNAs are then isolated from the different strains, measured amounts are applied to membranes, and cross-hybridization among strains is evaluated by using "stringent" hybridization conditions that are empirically defined by the researcher. Strongly cross-hybridizing DNA preparations could be combined and treated as the same standard (either a single species or a set of closely related species). Different reference standards are bacterial genomes that generally show undetectable cross-hybridization signals under stringent hybridization but may have an arbitrary DNA/DNA hybridization homology below or above 70% as used for the "species" definition. The standards are then used to prepare a master filter, using bacteriophage lambda as an internal control. The amount of genomic DNA applied to the membrane varies with analytical need. For example, 20 ng is needed for analysis of DNAs obtained from pure or enrichment cultures, whereas

200 ng was appropriate for analysis of total community DNAs reported in one study. A reference concentration series of bacteriophage DNA is applied on the same membrane (e.g., 10, 20, 50, or 100 ng). Sample DNA (ca. 100 ng) and lambda DNA (ca. 200 pg) are combined, boiled, and placed on ice. A probe is prepared by random hexamer labeling by using [α-^{32}P]dCTP and Klenow polymerase. Following denaturation, the probe is hybridized to a master filter under stringent conditions. Following washing of the filter under defined conditions, bound probe is quantified (e.g., by using autoradiography), and the fraction of community DNA composed of individual component genomes (represented by the individual reference DNAs on the master membrane) is calculated from the hybridization to individual DNA standards relative to the lambda reference series (161). This practice assumes that the lambda DNA and environmentally derived DNA present in the probe mixture are labeled to the same specific activity and hybridize with comparable efficiency. One possible concern is that impurities associated with the environmental DNA may influence both relative labeling and extent of hybridization.

The technique was first applied to the analysis of microbial populations associated with oil fields (161–163). RSGP revealed a significant difference between planktonic and biofilm-associated populations in oil recovery systems. Planktonic populations were more diverse and dominated by organochemotrophs. In contrast, biofilm populations were typically dominated by one to three populations of sulfate-reducing bacteria from the *Desulfovibrionaceae*, with much lower representation by organochemotrophs. More recently, this technique was used to evaluate the impact of pollutants (e.g., benzene and toluene) on soil populations (49, 134). In these studies, a master filter was first prepared by spotting genomic DNAs of bacteria isolated from the contaminated soil. Cross-hybridization among these bacteria was first determined to normalize the environmental hybridization patterns. This study documented a clear impact of the pollutant on community structure and metabolic state of the system.

The RSGP technique is advantageous in that, once an appropriate microbial survey of the target environment has been completed, master filters can be prepared rapidly and economically in large numbers. These filters can be stored indefinitely for immediate use when new sample DNAs become available for analysis. Further, these master filters can incorporate newly isolated bacterial DNA by spotting on side strips that are hybridized with the sample DNA probe together with the master filter. Consequently, routine screening of sample DNAs against a large number of different standards is feasible when a large set of sample DNAs is simultaneously prepared.

A disadvantage of the technique is that, although the actual assay does not involve culturing, the microbial community is described only in terms of its culturable component. It generally accounts for a very small portion of the bacterial populations in a given environment. The calculated fractions can also be subject to systematic errors even though RSGP has good precision. For example, the calculated fractions are sensitive to label allocation to sample and internal standard DNA (i.e., differences in specific activity). Furthermore, since detection sensitivity of RSGP is defined by the extent of cross-hybridization of the standard DNAs, low-abundance populations that hybridize below this experimentally defined threshold may not be detected in the presence of related organisms.

DNA Reassociation

The measurement of DNA reassociation kinetics was initially used to evaluate genomic sequence complexity, revealing repetitive DNA elements in the genomes of higher eukaryotes (11). More recently it has been used to assess the diversity of natural microbial communities (152, 153). Community-level DNA sequence complexity, as inferred from the rate of DNA reassociation, is related to population complexity. This measurement of complexity is a function of the concentration of complementary strands. Under defined conditions, strand reassociation follows second-order kinetics. Thus, the rate of reassociation is proportional to the square of the nucleotide concentration of homologous DNA strands. At a given concentration of total DNA (molar concentration of nucleotides in single-stranded DNA [ssDNA]), increasing genomic complexity (larger genomes, larger numbers of genomes, and fewer repeat elements per genome) results in a reduced concentration of complementary strands and a correspondingly reduced rate of reassociation. For example, as microbial community diversity (heterogeneity) increases (e.g., greater number of unique genomes), the rate of reassociation of DNA extracted from the community decreases for the same concentration of total DNA.

Experimentally, DNA reassociation is measured over time, and the fraction of reassociated DNA (C/C_0) is expressed as a function of C_0t, where C_0 is the initial molar concentration of nucleotides in ssDNA and where t is the time in seconds (Fig. 2). The plot of this relationship is referred to as a C_0t curve. The reaction rate constant can be expressed as $1/C_0t_{1/2}$, where $t_{1/2}$ is the time required for 50% reassociation. Under defined conditions, with temperature and monovalent ion concentration most important, $C_0t_{1/2}$ is proportional to the complexity (e.g., number of unique genomes) of the DNA. The practical and theoretical considerations of DNA reassociation are well developed (11, 170), and the reader is referred to these readings for a more complete theoretical and practical treatment.

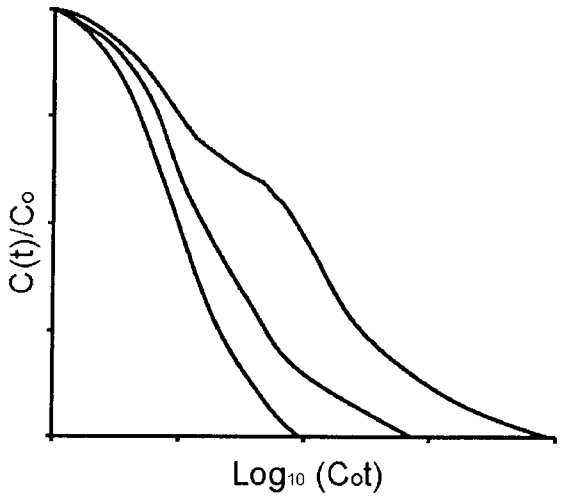

FIGURE 2 Typical C_0t curves obtained by reassociation of nucleic acids. The method can be used to estimate genome size or number by comparison to a reference genome represented by the left curve. The left and middle curves display pseudo-first-order kinetics of unique sequence DNA of different complexities, while the right curve displays pseudo-second-order kinetics.

In the application of community diversity analysis, the interpretation of DNA reassociation kinetics is made in the context of information theory, as has been developed for other diversity indices and briefly discussed above (152). It is a measure of the total amount of information in a system (richness and number of unique genomes) and the distribution of that information (evenness and abundance of individual genomes). Torsvik has expressed diversity as the number of "standard" genomes with no homology (151, 152). The results are significantly notable in that they suggest far greater diversity than anticipated. In an initial study of a soil sample taken from a beech forest (Seim, Norway), reassociation kinetics suggested the presence of approximately 4,000 genomes (grams of soil [wet wt])$^{-1}$ (152). Another study examined the effect of heavy metal contamination on the bacterial communities of soils in Braunschweig, Germany, with different levels of sludge amendments (125). Approximately 16,000, 6,400, and 2,000 bacterial genomes were estimated in the noncontaminated soil, low-metal amendments, and high-metal amendments, respectively. The number of bacterial genomes in noncontaminated soils was fourfold higher than reported previously (152). The increase in bacterial genome number in soil is mainly attributed to the use of a longer time period at a constant temperature to measure the C_0t curve. Thus, a higher $C_0t_{1/2}$ value can be obtained by taking into account not only the dominating DNA types but also the less dominating ones, as the reaction kinetics is assumed to follow pseudo-second-order reaction (Fig. 2). This study further indicated a significant reduction of genomic diversity (ca. up to eightfold) due to the metal amendments.

Although DNA reassociation provides a generally useful measure of genomic diversity, a variety of parameters must be considered, and most have yet to be systematically evaluated. One concern in the interpretation of DNA reassociation estimates is a reduction in the rate of reassociation resulting from impurities in the DNA sample. For example, Torsvik showed that the rate of reassociation increased with repeated purification of the sample DNA. This phenomenon can be evaluated in part by the addition of exogenous DNA to serve as an internal control (152). However, it is also important to more fully evaluate changes in reassociation kinetics that might result from the use of different extraction and purification techniques (150). Another consideration for any DNA-based analysis is the source of the DNA. The persistence of "inactive" DNA, either in the environment or entrained within dead or moribund cells, is essentially unanswered and is a concern in interpreting any data obtained solely from DNA. Recently, Gans and coworkers have indicated that the information theory used can significantly affect the estimation of genetic diversity using DNA reassociation by 100-fold (36a). By using computational analysis that enables direct comparison of different abundance models, this study reevaluated the results reported by Sandaa et al. (125). Approximately, 8.3×10^5, 6.4×10^4 and 7.9×10^3 genomes (g of soil [wet wt])$^{-1}$ were estimated for the noncontaminated soil, low-metal amendments, and high-metal amendments, respectively. This estimation is significantly larger than the original values, by a factor of 4 to 500. This improved computation analysis supports the use of DNA reassociation kinetics to study important questions related to the genomic diversity in different microbial ecosystems, for example, due to environmental perturbation and mapping of diversity geographically. This example further highlights the need for technical improvement.

NUCLEIC ACID FINGERPRINTS OF COMMUNITY STRUCTURE

The term "molecular fingerprint," otherwise rendered as "community fingerprint" or "phylogenetic fingerprint," is reserved for methods of analysis that generate a pattern-based characterization of community structure, most commonly represented by a banding pattern of nucleic acid fragments resolved by gel electrophoresis. Commonly used molecular fingerprint methods as summarized in Table 1 can be generalized based on the theory of pattern classification into three categories: (i) RFLP of total DNA or PCR products, (ii) fragment length heterogeneity of rRNA or PCR amplicons, and (iii) conformation or melting behavior of DNA products. Although DNA microarrays generate a type of fingerprint, the pattern of hybridization is mapped to sequence information and we treat this format separately. In principle, any gene can be explored by these fingerprint methods, but SSU and LSU rRNA genes that exhibit the trait of a molecular chronometer as aforementioned are presently the most commonly used ones. Other candidate markers include genes encoding the protein elongation factor, heat shock proteins, glutamine synthetase, ATPases, and topoisomerases, and the present databases for these genes are gradually improved.

Molecular fingerprinting methods can be used for rapid surveys using genes that provide for either phylogenetic or functional assessment of populations present in an environmental sample. Species richness and species evenness are two of those important ecological elements that can be inferred by using these methods: the former is usually estimated from the total number of different observable genetic units. The total number is determined by physical separation (e.g., electrophoretic separation of PCR amplicons by length, sequence, or conformational variation). Depending on the method and genetic marker used, each unit is often defined as a population, an operational taxonomic unit, a ribotype, a phylotype, or a genotype. However, most of these units provide little or no direct information of specific microbial population identity prior to further analysis, for example, by band purification and sequencing (68, 97). Furthermore, the extracted microbial ecology information is subject to general pitfalls associated with the use of molecular fingerprint methods (e.g., variability in cell lysis, DNA/RNA extraction, interoperon difference of SSU rRNA genes, SSU rRNA gene copy number, and PCR amplification) (see references 160 and 172). The following discussion of methods will address the theory of classification and the approaches used in the characterization of microbial communities in various environmental systems. Due to the large number of methods developed, only those commonly used ones are described in detail. Table 1 provides general advantages and disadvantages associated with individual methods.

T-RFLP

Terminal RFLP (T-RFLP) of PCR-amplified DNAs is a refined fingerprinting technique based on RFLP. The general steps include PCR amplification of a conserved target sequence (most commonly a region of the SSU rRNA gene) followed by restriction enzyme digestion and gel fractionation of resulting fragments. However, one of the two PCR primers is fluorescently labeled at the 5′ end. This labeling results in PCR amplification products that are tagged with a fluorescent dye at only one terminus. Following restriction enzyme digestion (usually with a

TABLE 1 Molecular fingerprinting methods for microbial community identification, monitoring, and tracking

Method(s)[a]	Procedure	Commonly used genetic markers	Diversity resolution	Advantages	Disadvantages	Reference(s)
T-RFLP	PCR-amplified DNA, restriction enzyme digestion, and terminal fragment size separation by a DNA autosequencer	rRNA, rRNA gene, conserved functional genes	By restriction site, length	Good gel-to-gel comparison; possible prediction of phylogenetic affiliation based on terminal fragment size and restriction enzyme(s) used	Only semiquantitative; limited resolution using unspecific primers; overestimate of diversity (>1 species from one terminal fragment)	69
PCR-ALH/ ARISA	PCR-amplification, size separation by a polyacrylamide gel or a DNA autosequencer	rRNA, rRNA gene, 16S-23S intergenic spacer	By length	Good gel-to-gel comparison; possible sequence identification of interested fragments	Only semiquantitative; limited resolution using unspecific primers; overestimate of diversity (>1 species from one fragment)	121, 145
SSCP	PCR-amplified DNA, heat denaturation, polyacrylamide gel electrophoresis, possibly by a DNA autosequencer	rRNA, rRNA gene, conserved functional genes	1-bp difference	Good gel-to-gel comparison	Only semiquantitative; limited resolution using unspecific primers; no sequence information; overestimate of diversity (multiple conformed-ssDNA bands from one dsDNA fragment)	61, 178
DGGE/ TGGE	PCR-amplified DNA, electrophoresis in a polyacrylamide gel with an increasing gradient of denaturants or temperature	rRNA, rRNA gene, conserved functional genes	1-bp difference	Inexpensive equipment setup; possible sequence identification using sequencing or hybridization	Poor reproducibility; limited resolution using unspecific primers; not quantitative; cumbersome gel preparation and staining; overestimate of diversity (formation of chimerical DNA, >1 band from one species)	96
Clone libraries	PCR-amplified DNA, clone screening, and sequencing with a DNA autosequencer	rRNA, rRNA gene, conserved functional genes	By nucleotide sequences	Fine-scale community structure analysis; identification of new genes	Expensive and labor intensive; only for dominant community members	22, 110
DNA microchip	Massively parallel hybridization among labeled targets and stationary oligonucleotide probes	RNA, rRNA gene, conserved functional genes	1-bp mismatch	High throughput; good chip-to-chip comparison; standardized procedure	Expensive for chip production and equipment setup; uncertainty with signal specificity (i.e., false-negative and false-positive signals); only for known target sequences	42, 71, 173, 174

[a] TGGE, temperature gradient gel electrophoresis.

tetrameric restriction enzyme), the restricted products are resolved by using an automated DNA sequencer equipped with a laser-induced fluorescence detector. Only the fluorescently tagged terminal fragments are detected and quantified. This is an extremely sensitive technique, since as little as 100 amol of a fluorescent tag (e.g., 6-carboxy-fluorescein) can be detected by using an ABI DNA autosequencer. Fragment lengths can be assigned by comparison to a set of size markers labeled with a different fluorescent dye and incorporated in each sample prior to electrophoresis (69). However, sizing accuracy is reported to be affected by the size calling method employed (109), the fragment length, and the type (i.e., gel-slab versus capillary) of electrophoresis systems used (99). As a whole, each T-RFLP pattern can be regarded as a fingerprint of population structure that provides information of population diversity (each terminal fragment and associated restriction site sequence define different "ribotypes") and a semiquantitative estimate of relative abundance (peak area). This information can be used for rapid comparative analysis of microbial and functional gene diversity in various microbial ecosystems.

Genes coding for the rRNA sequences are the most commonly used markers in microbial community analysis by T-RFLP. Normally, primers specific for the domains *Bacteria* (69), *Eucarya* (83), and *Archaea* (159) were used in the PCR to analyze the overall microbial diversity. Prior to the experiment, the proper combination of primer and restriction enzyme that gives a better resolution could be systematically evaluated in silico by computer simulation programs such as PatScan (http://www.unix.mcs.anl.gov/compbio/PatScan/HTML/patscan.html) (27) and TAP T-RFLP (http://35.8.164.52/html/TAP-trflp.html) (84). For example, computer-analyzed T-RFLP for 686 amplifiable sequences (8–927, *E. coli* numbering) from 1,102 complete SSU rRNA sequences of the Ribosomal Database Project indicated that those sequences could be classified into 233 different terminal restriction fragments (i.e., ribotypes) (69). Apparently, because of the high conservation of the restriction site positions with the SSU bacterial rRNA sequences, one terminal fragment may actually include one or more than one different or closely related sequences. Thus, T-RFLPs could underestimate the levels of microbial diversity by a factor of three or four, could resolve the phylogenetic population only at the level of higher-order groups, and could not correctly reflect the phylogenetic position based on the terminal restriction fragment lengths without prior knowledge of the bacterial composition (69, 84). To further increase the sensitivity of T-RFLP analysis, a hierarchical approach by employing phylum- and group-level (or below) primers is applied, for example, for the studies of sulfate-reducing bacteria (171), the *Acidobacteria* (58), the *Actinobacteria* (70), and the *Planctomycetes* (23).

A growing number of studies have demonstrated T-RFLP as a sensitive, reproducible, and robust method for the comparative analysis of microbial diversity in environmental samples. On average, 30 to 50 predominant terminal restricted fragments (i.e., different operational taxonomic units, or ribotypes) are observed within a microbial ecosystem. The least predominant account for less than 0.2% of the total amplified community rRNA (18, 54, 69). This resolution is equivalent to the screening of approximately 500 rRNA clones and was suggested to be slightly more sensitive than another popular fingerprint method, DGGE (see below), based on a comparative study of the same marine bacterioplankton community (90). Furthermore, the digitized molecular fingerprints derived from individual samples provide for a rapid and analytically based comparison of community similarity by the combined use of image analysis software (e.g., GelCompar by Applied Math, Inc.) and cluster analysis (69) or principal-component analysis (18).

T-RFLP provides an advantage over most molecular fingerprinting methods in that phylogenetic inference of predominant terminal fragment lengths within a community can be made once an appropriate database for the given sample is constructed (17, 80, 159). For example, the study by van der Maarel et al. (159) demonstrated that the observed archaeal terminal fragment lengths found in digestive tracts of marine fish could be assigned to specific marine archaeal sequences recovered from the same sample. The same approach was used in other studies (17, 80) to identify the predominant terminal fragment lengths found in archaeal or bacterial populations under changed environmental conditions. Furthermore, there are several software programs developed to infer the possible phylogenetic affiliation of predominant terminal restriction fragments within a sample. These include TAP T-RFLP (http://35.8.164.52/html/TAP-trflp.html), the T-RFLP Phylogenetic Assignment Tool (PAT) (http://trflp.limnology.wisc.edu/index.jsp), Microbial Community Analysis (MiCA) (http://mica.ibest.uidaho.edu/), and T-RFLP fragment sorter version 4.0 (http://www.oardc.ohio-state.edu/trflpfragsort/default.htm). However, the inference of phylogenetic positions based on terminal fragment length should be carefully intepretated. It is possible that the predicted fragment length can be different from observed ones due to possible variations occurred during sequencing and fragment sizing. Further, the SSU rRNA database still cannot present the total microbial diversity in microbial environments well.

In addition to community structure information, T-RFLP provides a basis for evaluating relative activity of individual populations. One of the most general measures of cellular activity is ribosome content, and the ratio of rRNA to rRNA genes generally increases with increasing growth rate (activity). Thus, comparison of T-RFLP fingerprints derived from both the rRNA gene and the rRNA of an environmental sample serves to confirm that microbial populations detected at the rRNA gene level are metabolically active and also provides some information of relative activity (80, 90). T-RFLP fingerprinting could be further used to correlate community shifts and activities with functional change by using a self-learning neural network analysis (26), community-level physiological profiles (BIOLOG) (54), and a stable-isotope probing technique (79). Since the use of an autosequencer allows for simultaneous analysis of at least 96 different samples, T-RFLP can be used as a high-throughput method to rapidly monitor community changes associated with temporal and spatial variation in the environment or occurring in response to environmental perturbations.

T-RFLP has also been extensively applied to examine variation among different functional genes in different environmental settings. The first study by Bruce (12) showed that T-RFLP could be used to rapidly characterize the genetic heterogeneity of a mercury resistance gene among different soil types (i.e., polluted and pristine) and within different areas of the same soil. Comparison of these environmentally derived T-RFLP patterns to the current sequence database suggested the presence of novel *mer* genes. Another two studies successfully used T-RFLP to reveal the diversity and expression of nitrogen fixation genes (*nif*H and *anf*) in different termite guts that exhibited different levels of nitrogen fixation or in a termite species

under different feeding modes (102, 104). The other functional genes used include monooxygenase (*amoA*) (48), ammonia monooxygenase (*amoA*) gene, nitrous oxide reductase (*nosZ*) gene (123), and methyl-coenzyme reductase (*mcrA*) (77).

As for all molecular methods dependent upon DNA isolation and PCR amplification, the results of a T-RFLP characterization may not fully reflect the native population structure. It is reported that highly degenerated primers can cause a biased representation on the abundance of individual targets (78). Possible "pseudo" terminal restriction fragments can also be produced due to single-stranded amplicons in the T-RFLP but can be eliminated by digesting these pseudofragments with single-strand-specific mung bean nuclease prior to T-RFLP analysis (30). Osborn et al. (109) also demonstrated that some of the rRNA-related biases can be minimized if the key experimental parameters (e.g., sample replication and handling, PCR DNA template concentration, PCR cycle time, *Taq* polymerase types, and the minimum restriction enzyme digestion time) are systematically evaluated.

PCR-ALH and Automated Ribosomal Intergenic Spacer Analysis (ARISA)

The estimation of microbial diversity by PCR-amplicon length heterogeneity (ALH) is based on the capacity to resolve the length and length heterogeneity of the PCR amplicons from the variable regions encoding the SSU rRNA (121, 145). To do so, domain-specific PCR primers are initially used to amplify DNA fragments flanking the variable region(s). The resulting amplication products are then separated by size on polyacrylamide gels and are viewed by using a sensitive staining method, e.g., silver staining. Like T-RFLP, fluorescently tagged primers and automated analysis have been used to increase the sensitivity and reproducibility for detecting small amounts of DNA and to separate fragments differing by only 1 or 2 bp (34, 145).

The resolution of PCR-ALH is dependent on the variation within the coding regions. Among 366 marine bacterial SSU rRNA sequences examined, two variable regions corresponding to *E. coli* numbering 8–355 and 8–536 have length variations of 312 to 363 bp and 472 to 574 bp and could be used to classify microbial populations for up to 52 and 103 different categories, respectively (145). Since amplicon length generally corresponds to phylogenetic relationship, this classification has a natural order. To further infer the phylogenetic position of the organisms represented by different length amplicons, the fragments can be cloned and sequenced. Alternatively, PCR-ALH can be used to analyze a microbial community for which complete or nearly complete sequences of the target genes are available (e.g., via selective amplification, cloning, and sequencing of SSU rRNA genes). However, since most fragment-length categories contain more than one different but closely related sequences, the resolution of the different natural groupings is somewhat compromised. This approach is further limited by the precision of fragment sizing when 1-bp differentiation between two fragments is needed for discrimination or when long fragments are compared (145).

Similar to PCR-ALH, ARISA was developed for estimating microbial diversity (9, 34). It takes advantage of a range of length variation in the 16S-23S rRNA gene intergenic spacer (usually encoding tRNAs) wider than that of the coding regions for the rRNA molecules (34, 37). A recent study indicated that the measurement of microbial diversity in terms of the number of peaks, the range of peak size, and the reproducibility of ARISA fingerprints can be significantly affected by the use of different forward and reverse primers (16). This issue was likely related to the poor design of primer sets based on the limited numbers of rRNA sequences in the database previously and could be improved by using a new set of forward and reserve primers encompassing most of the 16S and 23S rRNA sequences in the database with similar anneal temperatures (16). It should be further noted that (i) 16S-23S spacer length variation does not necessarily correspond to different populations, since individual organisms generally contain multiple operons that may encode spacers of different length and (ii) intergenic DNA fragments of the same length can encompass more than one species of microorganism. Thus, these factors limit the use of ARISA for direct inference of microbial diversity. Overall, PCR-ALH and ARISA remain useful methods for comparative analysis of natural microbial populations along different temporal and spatial gradients (9, 34, 145).

SSCP

Single-strand conformation polymorphism (SSCP) is a technique that separates ssDNA fragments differing in strand length and conformation (conformers) by gel electrophoresis. SSCP has been most frequently used in conjunction with PCR to generate the target DNA fragments. SSCP-PCR was initially used to detect minor sequence variants caused by point mutations in human alleles (108) and only more recently has been used to analyze microbial community structure (61, 130, 178). Immediately before the SSCP analysis, the PCR-amplified rRNA is denatured at a high temperature (i.e., 95°C for 2 or 3 min), usually in the presence of denaturants (e.g., formamide). This process is followed by rapid cooling on ice to "trap" different structural forms. The folding of an ssDNA into different stable structures is highly related to its sequence (i.e., nucleotide sequences and composition). Thus, different sequences contribute to different structures that can be resolved on the basis of various migration rates during electrophoresis on a nondenaturing polyacrylamide gel (usually run at a constant ambient temperature). Three major structural types observed are "self-folded" ssDNA, heteroduplex double-stranded DNA (dsDNA), and homoduplex dsDNA (formed as a result of the high rate of DNA reassociation) (61). Differentiation between ssDNA and dsDNA is based upon relative migration distance (ssDNA, heteroduplex dsDNA, and homoduplex dsDNA migrating in increasing order) and by variation in silver stain color. As described above for T-RFLP analysis, the use of fluorescent-dye-labeled PCR primers in combination with analysis on an automated DNA sequencer provides for the greatest reproducibility and sensitivity (112, 178).

Domain-specific primers for the SSU rRNA gene (61, 130, 178) have been used for analysis of microbial population structure. Criteria for selecting a good primer pair include (i) adequate fragment size (~100 to ~400 bp), (ii) low homology within the target sequence region, and (iii) high fragment length heterogeneity. Computer-assisted analysis of 1,262 available SSU rRNA bacterial sequences showed that, within the variable V3 region (*E. coli* numbering of 330 to 533), there are 34 different fragment length variants distributed among 19 different bacterial phyla or groups, and 1 to 22 different variants within individual bacterial groups (61). An alternative target region flanking the V3 to V5 region (330 to 926 [*E. coli* numbering]) of bacterial SSU rRNA sequences has also been used (130). For the

domain *Archaea*, primers are usually complementary to the V3 region (178). Individual bands can be retrieved and sequenced, but due to their relatively short sequence length, only approximate phylogenetic position can be inferred.

The application of SSCP as a fingerprint of microbial community structure can be limited by the uncertainty of the banding pattern obtained. For example, multiple bands can be generated from the same bacterium, e.g., via heteroduplex formation among closely related sequences. Alternatively, different ssDNAs can have a similar migration distance on the gel. One solution for eliminating multiple bands originating from a single bacterium was proposed (130). Briefly, one of the two PCR primers was labeled with a phosphate group at the 5′ terminus. Thus, the amplified DNA strands having a 5′ phosphate could be removed by digestion with λ exonuclease, leading to one conformed-ssDNA band for each microorganism (130). A similar approach is to use biotinylated primers in the PCR procedure and to separate biotinylated DNA and nonbiotinylated DNA with magnetic beads prior to SSCP analysis (131). However, both approaches require additional steps and a significant investment of time. As for the other fingerprinting methods described, SSCP has been used to characterize microbial population structure in relation to environmental conditions (61, 130, 178).

DGGE and TGGE

Methods of DGGE and temperature gradient gel electrophoresis (TGGE) are based upon the analytical separation of DNA fragments of identical or nearly identical length but of various sequence compositions. The separation is determined by the change of electrophoretic mobility of DNA fragments migrating in a gel containing a linearly increasing gradient of DNA denaturants (urea and formamide) or temperature. The change in fragment mobility is associated with partial melting of the dsDNA in discrete regions, the so-called melting domains. These methods were first developed to detect single-base changes in genes for diagnosis of human genetic diseases and in genetic linkage studies (98). More recently DGGE and TGGE have been extended to resolve environmental populations of microorganisms by separating PCR amplification products generated by using primers targeting conserved genes (97). PCR primers designed for the amplification and cloning of SSU rRNA genes were first used to demonstrate the technique as applied to environmental microbiology (96). Subsequent studies have applied the technique to characterize the diversity of genes encoding different metabolic functions (39, 95, 97, 148).

DGGE and TGGE analysis of PCR-amplified SSU rDNA fragments provides a rapid method for the characterization and monitoring of community population structure and dynamics (for a more detailed review see references 95, 97, and 127), with consideration of some of the caveats previously discussed and below. The initial study by Muyzer et al. demonstrated the presence of several distinguishable bands (between 5 and 10) in the gel separation pattern, which were most likely derived from the predominant species within those communities characterized (96). By using DGGE or TGGE, extensive studies have examined various microbial ecosystems such as microbial mats, deep-sea hydrothermal vent samples, a stratified marine water column, rhizosphere, soils, and activated sludge (95, 97, 127). And as also discussed in relation to T-RFLP analysis, DGGE can be used to evaluate relative activity of individual populations based on changing rRNA/rRNA gene ratio. The method has been used to evaluate this ratio among different natural populations by comparing the pattern and intensity of bands derived from using either rRNA genes or rRNA (using reverse transcriptase to generate cDNA) as templates (147). A similar approach was used to evaluate the expression of a Ni-Fe hydrogenase conserved among natural populations of *Desulfovibrio* spp. (167).

More-specific information of population composition can be obtained by secondary analysis on the predominant DGGE or TGGE banding pattern by using sequencing or hybridization. Sequences of individual bands (fragments) are determined following their extraction from the gel, a second round of PCR amplification, and sequencing (direct or after cloning). This sequence information can also be used to design specific oligonucleotide probes for membrane hybridization or FISH analyses. For example, this approach was used to verify the predominance of microbial populations identified by DGGE with FISH (101). Also, group- and species-specific DNA hybridization probes have been used to identify specific populations within the pattern of resolved bands following transfer of the DNA to nylon membranes (96). Group-specific PCR primers have been used to restrict population analysis to specific microbial groups (45, 56, 103). Alternatively, microbial diversity of a specific functional or phylogenetic group within a microbial ecosystem can be achieved by using PCR with group- or species-specific DGGE primers (103) or combining DGGE with nested PCR (21).

The caveats of the DGGE and TGGE methods include the following: beyond the usual concerns of representative DNA extraction, the questions of representative PCR amplification of individual populations within the target collection and formation of amplification artifacts (e.g., chimerical sequences, heteroduplexes, and polymerase error) between populations remain partly unanswered (65, 160, 172). The reader is referred to studies that have directly addressed artifacts associated with this method (51, 138). The separation of the many fragments amplified from a highly diverse bacterial community is not possible by using available technology, although resolution may be improved by using a narrower range of denaturants or two-dimensional electrophoresis (33). The phylogenetic information obtained from sequencing individual bands is limited, because only fragments up to approximately 500 bp can be well separated. Another concern associated with the technique as well as with other aforementioned molecular fingerprinting methods is the a priori assignment of individual bands to individual populations. As discussed in the introductory considerations of diversity assessments, there can be significant sequence variation among rRNA operons of an individual organism, and so individual organisms could potentially contribute to multiple bands (103).

DNA Microarrays

A DNA microarray (DNA microchip or DNA chip) is an orderly, high-density arrangement of hundreds (or thousands) of individual cDNA probes (~500 to 5,000 bp in length) or oligonucleotides (15- to 70-mers) bound directly or indirectly to a solid surface (e.g., membrane, silicon wafer, or glass) (132). Unlike the membrane hybridization format, the chip is a high-density format that allows for simultaneous hybridization of a labeled DNA/RNA target to a large set of probes, thus providing high throughputs. In recent years, the oligonucleotide-based microarray has been receiving increased attention from researchers due to its advantages over its cDNA counterpart. These include

simple methodology to obtain probes, good quality control of probe manufacturing, options to select high-specificity probes to prevent nonspecific hybridization, and the potential to detect alternative spliced variants of genes. Detailed information on the types and manufacturing of DNA microchips can be found in recent articles (31, 85, 132, 141).

DNA microarray technology has emerged as a popular tool in studies of environmental microbiology. So far, different types of microarrays as classified by the type of markers used have been developed for community structure analysis, community functions, species relatedness, and community gene expression. These include phylogenetic oligonucleotide arrays (POAs) (42, 74, 173), functional genomic arrays (176), community genome array (175), whole-genome open-reading-frame array (94), and gene expression array (52). However, due to the space limitation, we will focus discussion on the use of POA for microbial community structure analysis, its present limitations and challenges, and the possible solutions to these challenges.

In general, POA uses rRNA molecules as the phylogenetic marker. Within the natural framework provided by rRNA sequence variation, a fully developed POA could include a set of probes to encompass virtually all natural microbial grouping (phylotypes) and thereby serve to simultaneously monitor the population structure at multiple levels of resolution (e.g., at the approximate taxonomic ranks of domain, kingdom, order, genus, and species) (71, 173). This situation provides a basis to adjust monitoring strategies to the resolution required and, as previously discussed, serves as a mechanism to validate hybridization signal output by requiring consistency between quantifications at different taxonomic ranks (5, 139). With this hierarchical probe design strategy, POAs have been developed and applied for the study of microbial community analysis and microbial identification, for example, for a small subset of nitrifying populations (42, 53) and for analyzing oral microbiota (136), *Bacillus anthracis* and its related species (71), all recognized lineages of sulfate-reducing prokaryotes in both natural and clinical environments (74), and the beta-proteobacterial order "Rhodocyclales" (75).

These initial studies have demonstrated the potential use of POAs for rapid and comprehensive characterization of community structure but have also identified additional areas of needed technology development (for a more detailed review see references 8, 64, and 176). For example, the high cost of special equipment and a large number of oligonucleotide probes can limit the use of DNA microchip technology in its present stage of development. Design and optimization of probes remain something of an art form because the rules for predicting duplex stability are not yet fully developed (126, 149). Further, the efficiency in achieving good discrimination between targets and nontargets, eventually at a one-mismatch resolution, is still not satisfactory. This failing is primarily related to the use of the same hybridization and washing conditions to achieve target discrimination among a large set of oligonucleotide probes that differ in sequences and thermal stability. In other words, for any two given probes, the optimal conditions required for achieving perfect-match discrimination are not always the same. Furthermore, the efficiency of mismatch discrimination is complicated by the secondary structures formed within ssDNA or single-stranded RNA targets (36, 111). As a result, the occurrence of false-positive and false-negative signals can be significant (63, 74), and consequently they reduce the detection specificity of DNA microarrays.

Likewise, quantitation by using DNA microarray remains very challenging. One reason is the inherently high variation associated with array fabrication, target labeling, hybridization, and image processing (177). For example, when targets used in hybridizations are insufficiently fragmented or are not fragmented at all, it is possible for different immobilized probes (of various taxonomic ranks or targeting different regions on the rRNA sequence) to compete for the limited number of targets, resulting in incorrect estimations that deviate significantly from the actual population numbers. Differences in the accessibility of different targets to the immobilized probes can further lead to erroneous results (36).

To mitigate some of these problems, solutions have been demonstrated. One indirect solution to confirm the hybridization specificity from DNA microarray studies is to combine this technique with other well-established molecular methods. For example, Koizumi et al. (55) applied PCR-DGGE and RNA-DNA membrane hybridization in parallel with POAs to confirm the presence of specific sulfate-reducing organisms in anaerobic aromatic-degrading communities. Loy et al. (74) validated the presence of sulfate-reducing organisms, as suggested by positive signals from multiple probes of a sulfate-reducing prokaryote-specific POA, with specific PCR amplification and cloning and sequencing of 16S rRNA genes and genes encoding dissimilatory (bi)sulfate reductase. However, these DNA-microarray-independent techniques further contribute to the time required for community analysis. Alternatively, to minimize the occurrence of false-positive signals observed in the DNA microarray, a nonequilibrium dissociation curve approach, whereby the dissociation process of all positive probe-target duplexes from low to high temperature is performed and analyzed simultaneously within a short time, is proposed (87) and demonstrated (32, 63, 71, 155). By using this approach, a false-positive signal could be identified by comparing the observed dissociation curve and the observed T_d, at which 50% of the probe-target duplex is dissociated, with those of the perfectly matched duplexes. However, further efforts are still needed to resolve issues related to the sensitivity and specificity of the DNA microarray. Eventually, the DNA microarray with its massively parallel hybridization capacity should emerge as a powerful tool for microbial ecology studies.

In summary, the development and application of molecular fingerprinting are rapidly changing our understanding of microbial communities in a wide range of microbial ecosystems. All methods provide a rapid and robust means to monitor population changes. However, as yet no single method can resolve all microbial populations in a typical habitat. We emphasize that these methods are all part of a growing molecular toolbox. Most successful applications will generally come from studies that combine them with other methods. For example, some studies have combined chemotaxonomic methods (e.g., respiratory quinone and cellular fatty acid profiling) with molecular fingerprinting methods (68, 133). Others have compared molecular fingerprints with metabolic activity measurements (e.g., Biolog, enzyme activity, microelectrode, and microautoradiography) to better resolve structure-function relationships in a community (10, 54, 62, 105). Finally, we also emphasize that this chapter's focus on molecular techniques does not reduce the need for culture-based methods to fully define the character of populations rendered to sequences and hybridization signals by molecular methods.

Research originating in our laboratories and discussed in this chapter was supported by grants from NSF (Systematics DEB 0213186,

*Ecosystems Studies DEB 9904047, Environ. Eng. BES 0332118), NIH/NIDCR (U01 DE 14955), ONR, DARPA, NASA (ROSS and NAI), DOE (GTL Contract DE-AC03-76F00098), and A*STAR-Singapore.*

REFERENCES

1. **Alm, E., and D. A. Stahl.** 2000. Critical factors influencing the recovery and integrity of rRNA extracted from environmental samples: use of an optimized protocol to measure depth-related biomass distribution in freshwater sediments. *J. Microbiol. Methods* **40:**153–162.
2. **Alm, E. W., D. B. Oerther, N. Larsen, D. A. Stahl, and L. Raskin.** 1996. The oligonucleotide probe database. *Appl. Environ. Microbiol.* **62:**3557–3559.
3. **Alm, E. W., and D. A. Stahl.** 1996. Evaluation of key parameters for extraction of native rRNA from different environmental matrices, p. 1.1.5.1–1.1.5.29. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), *Molecular Microbial Ecology Manual.* Kluwer Academic Publishers, Dordrecht, The Netherlands.
4. **Amann, R., and W. Ludwig.** 2000. Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiol. Rev.* **24:**555–565.
5. **Amann, R. I., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59:**143–169.
6. **Amann, R. I., J. Stromley, R. Devereux, R. Key, and D. A. Stahl.** 1992. Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. *Appl. Environ. Microbiol.* **58:**614–623.
7. **Andrew, P. W., and G. J. Boulnois.** 1990. Early days in the use of DNA probes for Mycobacterium tuberculosis and Mycobacterium avium complexes, p. 179–198. *In* A. J. L. Macario and E. Conway de Macario (ed.), *Gene Probes for Bacteria.* Academic Press, Inc., San Diego, Calif.
8. **Bodrossy, L., and A. Sessitsch.** 2004. Oligonucleotide microarrays in microbial diagnostics. *Curr. Opin. Microbiol.* **7:**245–254.
9. **Borneman, J., and E. W. Triplett.** 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl. Environ. Microbiol.* **63:**2647–2653.
10. **Brinkhoff, T., C. M. Santegoeds, K. Sahm, J. Kuever, and G. Muyzer.** 1998. A polyphasic approach to study the diversity and vertical distribution of sulfur-oxidizing *Thiomicrospira* species in coastal sediments of the German Wadden Sea. *Appl. Environ. Microbiol.* **64:**4650–4657.
11. **Britten, R. J., and D. E. Kohne.** 1968. Repeated sequences in DNA. *Science* **161:**529–540.
12. **Bruce, K. D.** 1997. Analysis of *mer* gene subclasses within bacterial communities in soils and sediments resolved by fluorescent-PCR–restriction fragment length polymorphism profiling. *Appl. Environ. Microbiol.* **63:**4914–4919.
13. **Brunk, C. F., E. Avaniss-Aghajani, and C. A. Brunk.** 1996. A computer analysis of primer and probe hybridization potential with bacterial small-subunit rRNA sequences. *Appl. Environ. Microbiol.* **62:**872–879.
14. **Brusseau, G. A., E. S. Bulygina, and R. S. Hanson.** 1994. Phylogenetic analysis and development of probes for differentiating methylotrophic bacteria. *Appl. Environ. Microbiol.* **60:**626–636.
15. **Burggraf, S., T. Mayer, R. Amann, S. Schadhauser, C. R. Woese, and K. O. Stetter.** 1994. Identifying members of the domain *Archaea* with rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* **60:**3112–3119.
16. **Cardinale, M., L. Brusetti, P. Quatrini, S. Borin, A. M. Puglia, A. Rizzi, E. Zanardini, C. Sorlini, C. Corselli, and D. Daffonchio.** 2004. Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl. Environ. Microbiol.* **70:**6147–6156.
17. **Chin, K.-J., T. Lukow, and R. Conrad.** 1999. Effect of temperature on structure and function of the methanogenic archaeal community in an anoxic rice field soil. *Appl. Environ. Microbiol.* **65:**2341–2349.
18. **Clement, B. G., L. E. Kehl, K. L. DeBord, and C. L. Kitts.** 1998. Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *J. Microbiol. Methods* **31:**135–142.
19. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33:**D294–D296.
20. **Daims, H., A. Bruhl, R. Amann, K.-H. Schleifer, and M. Wagner.** 1999. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22:**434–444.
21. **Dar, S. A., J. G. Kuenen, and G. Muyzer.** 2005. Nested PCR-denaturing gradient gel electrophoresis approach to determine the diversity of sulfate-reducing bacteria in complex microbial communities. *Appl. Environ. Microbiol.* **71:**2325–2330.
22. **DeLong, E. F., K. Y. Wu, B. B. Prézelin, and R. V. M. Jovine.** 1994. High abundance of archaea in Antarctic marine picoplankton. *Nature* **371:**695–697.

22a. **de los Reyes, F. L., W. Ritter, and L. Raskin.** 1997. Group-specific small-subunit rRNA hybridization probes to characterize filamentous foaming in activated sludge systems. *Appl. Environ. Microbiol.* **63:**1107–1117.

23. **Derakshani, M., T. Lukow, and W. Liesack.** 2001. Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl. Environ. Microbiol.* **67:**623–631.
24. **De Rijk, P., J. Wuyts, Y. Van de Peer, T. Winkelmans, and R. De Wachter.** 2000. The European large subunit ribosomal RNA database. *Nucleic Acids Res.* **28:**177–178.
25. **Devereux, R., S.-H. He, C. L. Doyle, S. Orkland, D. A. Stahl, J. LeGall, and W. B. Whitman.** 1990. Diversity and orgin of *Desulfovibrio* species: phylogenetic definition of a family. *J. Bacteriol.* **172:**3609–3619.
26. **Dollhopf, S. L., S. A. Hashsham, and J. M. Tiedje.** 2001. Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal component analysis to describe community dynamics and convergence. *Microb. Ecol.* **42:**495–505.
27. **Dsouza, M., N. Larsen, and R. Overbeek.** 1997. Searching for patterns in genomic data. *Trends Genet.* **13:** 497–498.
28. **Dular, R.** 1990. Gene probe detection of human and cell culture mycoplasmas, p. 417–449. *In* A. J. L. Macario and E. Conway de Macario (ed.), *Gene Probes for Bacteria.* Academic Press, Inc., San Diego, Calif.
29. **Dutilh, B., C. Bebaer, and P. A. D. Grimont.** 1990. Detection of Chlamydia trachomatis with DNA probes, p. 46–64. *In* A. J. L. Macario and E. Conway de Macario (ed.), *Gene Probes for Bacteria.* Academic Press, Inc., San Diego, Calif.
30. **Egert, M., and M. Friedrich.** 2003. Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl. Environ. Microbiol.* **69:**2555–2562.

31. **Ekins, R., and F. W. Chu.** 1999. Microarrays: their origins and applications. *Trends Biotechnol.* **17:**217–218.
32. **El Fantroussi, S., H. Urakawa, A. E. Bernhard, J. J. Kelly, P. A. Noble, H. Smidt, G. M. Yershov, and D. A. Stahl.** 2003. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. *Appl. Environ. Microbiol.* **69:**2377–2382.
33. **Fischer, S. G., and L. S. Lerman.** 1979. Length-independent separation of DNA restriction fragments in two dimensional gel electrophoresis. *Cell* **16:**191–200.
34. **Fisher, M. M., and E. W. Triplett.** 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.* **65:**4630–4636.
35. **Fuchs, B. M., F. O. Glockner, J. Wulf, and R. Amann.** 2000. Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.* **66:** 3603–3607.
36. **Fuchs, B. M., G. Wallner, W. Beisker, I. Schwippl, W. Ludwig, and R. Amann.** 1998. Flow cytometric analysis of the in situ accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.* **64:**4973–4982.
37. **García-Martinez, J., I. Bescós, J. J. Rodríguez-Sala, and F. Rodríguez-Valera.** 2001. RISSC: a novel database for ribosomal 16S-23S RNA genes spacer regions. *Nucleic Acids Res.* **29:**178–180.
38. **Giovannoni, S. J., E. F. DeLong, G. J. Olsen, and N. R. Pace.** 1988. Phylogenetic, group-specific oligonucleotide probes for identification of single microbial cells. *J. Bacteriol.* **170:**720–726.
39. **Goregues, C. M., V. D. Michotey, and P. C. Bonin.** 2005. Molecular, biochemical, and physiological approaches for understanding the ecology of denitrification. *Microb. Ecol.* **49:**198–208.
40. **Greene, E. A., and G. Voordouw.** 2003. Analysis of environmental microbial communities by reverse sample genome probing. *J. Microbiol. Methods* **53:**211–219.
41. **Grimont, F., and P. A. D. Grimont.** 1991. DNA fingerprinting, p. 249–279. *In* E. Stackebrandt and M. Goodfellow (ed.), *Sequencing and Hybridization Techniques in Bacterial Systematics*. John Wiley & Sons, Chichester, England.
42. **Guschin, D. Y., B. K. Mobarry, D. Proudnikov, D. A. Stahl, B. E. Rittmann, and A. D. Mirzabekov.** 1997. Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl. Environ. Microbiol.* **63:**2397–2402.
43. **Hedrick, D. B., T. White, J. B. Guckert, W. J. Jewell, and D. C. White.** 1992. Microbial biomass and community structure of a phase separated methanogenic reactor determined by lipid analysis. *J. Ind. Microbiol.* **9:**193–199.
44. **Hennecke, H., K. Kaluza, B. Thony, M. Fuhrmann, W. Ludwig, and E. Stackebrandt.** 1985. Concurrent evolution of nitrogenase genes and 16S rRNA in Rhizobium species and other nitrogen fixing bacteria. *Arch. Microbiol.* **142:**342–348.
45. **Heuer, H., M. Krsek, P. Baker, K. Smalla, and E. M. Wellington.** 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl. Environ. Microbiol.* **63:**3233–3241.
46. **Hiraishi, A., K. Masamune, and H. Kitamura.** 1989. Characterization of the bacterial population structure in an anaerobic-aerobic activated sludge system on the basis of respiratory quinone profiles. *Appl. Environ. Microbiol.* **55:**897–901.
47. **Holben, W. E.** 1994. Isolation and purification of bacterial DNA from soils, p. 727–751. *In* R. W. Weaver et al. (ed.), *Methods of Soil Analysis*, part 2. *Microbiological and Biochemical Properties*. SSSA Book Series no. 5. Soil Science Society of America, Madison, Wis.
48. **Horz, H. P., J. H. Rotthauwe, T. Lukow, and W. Liesack.** 2000. Identification of major subgroups of ammonia-oxidizing bacteria in environmental samples by T-RFLP analysis of amoA PCR products. *J. Microbiol. Methods* **39:**197–204.
49. **Hubert, C., Y. Shen, and G. Voordouw.** 1999. Composition of toluene-degrading microbial communities from soil at different concentrations of toluene. *Appl. Environ. Microbiol.* **65:**3064–3070.
50. **Hugenholtz, P., C. Pitulle, K. L. Hershberger, and N. R. Pace.** 1998. Novel division-level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180:**366–376.
51. **Janse, I., J. Bok, and G. Zwart.** 2004. A simple remedy against artifactual double bands in denaturing gradient gel electrophoresis. *J. Microbiol. Methods* **57:**279–281.
52. **Johnson, M. R., C. I. Montero, S. B. Conners, K. R. Shockley, S. L. Bridger, and R. M. Kelly.** 2005. Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium Thermotoga maritima. *Mol. Microbiol.* **55:**664–674.
53. **Kelly, J. J., S. Siripong, J. McCormack, L. R. Janus, H. Urakawa, S. El Fantroussi, P. A. Noble, L. Sappelsa, B. E. Rittmann, and D. A. Stahl.** DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res.*, in press.
54. **Kerkhof, L., M. Santoro, and J. Garland.** 2000. Response of soybean rhizosphere communities to human hygiene water addition as determined by community level physiological profiling (CLPP) and terminal restriction fragment length polymorphism (TRFLP) analysis. *FEMS Microbiol. Lett.* **184:**95–101.
55. **Koizumi, Y., J. J. Kelly, T. Nakagawa, H. Urakawa, S. El-Fantroussi, S. Al-Muzaini, M. Fukui, Y. Urushigawa, and D. A. Stahl.** 2002. Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology. *Appl. Environ. Microbiol.* **68:**3215–3225.
56. **Kowalchuk, G. A., J. R. Stephen, W. De Boer, J. I. Prosser, T. M. Embley, and J. W. Woldendorp.** 1997. Analysis of ammonia-oxidizing bacteria of the β subdivision of the class *Proteobacteria* in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments. *Appl. Environ. Microbiol.* **63:**1489–1497.
57. **Kurata, S., T. Kanagawa, Y. Magariyama, K. Takatsu, K. Yamada, T. Yokomaku, and Y. Kamagata.** 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Appl. Environ. Microbiol.* **70:**7545–7549.
58. **Kuske, C., L. Ticknor, M. Miller, J. Dunbar, J. Davis, S. Barns, and J. Belnap.** 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Appl. Environ. Microbiol.* **68:**1854–1863.
59. **Kysela, D. T., C. Palacios, and M. L. Sogin.** 2005. Serial analysis of V6 ribosomal sequence tags (SARST-V6): a method for efficient, high-throughput analysis of microbial community composition. *Environ. Microbiol.* **7:**356–364.
60. **Laguerre, G., M.-R. Allard, F. Reboy, and N. Amarger.** 1994. Rapid identification of rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl. Environ. Microbiol.* **60:**56–63.
61. **Lee, D.-H., Y.-G. Zo, and S.-J. Kim.** 1996. Non-radioactive method to study genetic profiles of natural bacterial

communities by PCR–single-strand conformation polymorphism. *Appl. Environ. Microbiol.* **62:**3112–3120.

62. **Lee, N., P. H. Nielsen, K. H. Andreasen, S. Juretschko, J. L. Nielsen, K. H. Schleifer, and M. Wagner.** 1999. Combination of fluorescent in situ hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* **65:**1289–1297.
63. **Li, E. S., J. K. Ng, J. H. Wu, and W. T. Liu.** 2004. Evaluating single-base-pair discriminating capability of planar oligonucleotide microchips using a non-equilibrium dissociation approach. *Environ. Microbiol.* **6:**1197–1202.
64. **Li, E. S. Y., and W. T. Liu.** 2004. DNA microarray technology in microbial ecology studies—principle, application and current limitations. *Microb. Environ.* **18:**175–187.
65. **Liesack, W., H. Weyland, and E. Stackebrandt.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb. Ecol.* **21:**199–209.
66. **Lin, C., B. Flesher, W. C. Capman, R. I. Amann, and D. A. Stahl.** 1994. Taxon specific hybridization probes for fiber-digesting bacteria suggest novel gut-associated Fibrobacter. *Syst. Appl. Microbiol.* **17:**418–424.
67. **Lin, C., and D. A. Stahl.** 1995. Comparative analyses reveal a highly conserved endoglucanase in the cellulolytic genus *Fibrobacter*. *J. Bacteriol.* **177:**2543–2549.
68. **Liu, W.-T., K. D. Linning, K. Nakamura, T. Mino, T. Matsuo, and L. J. Forney.** 2000. Microbial community changes in biological phosphate-removal systems on altering sludge phosphorus content. *Microbiology* **146:**1099–1107.
69. **Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **63:**4516–4522.
70. **Liu, W.-T., T. L. Marsh, and L. J. Forney.** 1998. Determination of the microbial diversity of anaerobic-aerobic activated sludge by a novel molecular biological technique. *Water Sci. Technol.* **37:**417–422.
71. **Liu, W. T., A. D. Mirzabekov, and D. A. Stahl.** 2001. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ. Microbiol.* **3:**619–629.
72. **Lovell, C. R., and Y. Hui.** 1991. Design and testing of a functional group-specific probe for the study of natural populations of acetogenic bacteria. *Appl. Environ. Microbiol.* **57:**2602–2609.
73. **Loy, A., M. Horn, and M. Wagner.** 2003. probeBase: an online resource for rRNA-targeted oligonucleotide probes. *Nucleic Acids Res.* **31:**514–516.
74. **Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K. H. Schleifer, and M. Wagner.** 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.
75. **Loy, A., C. Schulz, S. Lücker, A. Schöpfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. Wagner.** 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "*Rhodocyclales*." *Appl. Environ. Microbiol.* **71:**1373–1386.
76. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32:**1363–1371.
77. **Lueders, T., K. Chin, R. Conrad, and M. Friedrich.** 2001. Molecular analyses of methyl-coenzyme M reductase alpha-subunit (mcrA) genes in rice field soil and enrichment cultures reveal the methanogenic phenotype of a novel archaeal lineage. *Environ. Microbiol.* **3:**194–204.
78. **Lueders, T., and M. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl. Environ. Microbiol.* **69:**320–326.
79. **Lueders, T., B. Pommerenke, and M. W. Friedrich.** 2004. Stable-isotope probing of microorganisms thriving at thermodynamic limits: syntrophic propionate oxidation in flooded soil. *Appl. Environ. Microbiol.* **70:**5778–5786.
80. **Lúmann, H., I. Arth, and W. Liesack.** 2000. Spatial changes in the bacterial community structure along a vertical oxygen gradient in flooded paddy soil cores. *Appl. Environ. Microbiol.* **66:**754–762.
81. **Manz, W., R. Amann, W. Ludwig, M. Vancanneyt, and K. H. Schleifer.** 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology* **142:**1097–1106.
82. **Manz, W., R. Amann, W. Ludwig, M. Wagner, and K.-H. Schleifer.** 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Syst. Appl. Microbiol.* **15:**593–600.
83. **Marsh, T. L., W.-T. Liu, L. J. Forney, and H. Cheng.** 1998. Beginning a molecular analysis of eukaryal community in activated sludge. *Water Sci. Technol.* **37:**455–460.
84. **Marsh, T. L., P. Saxman, J. Cole, and J. Tiedje.** 2000. Terminal restriction fragment length polymorphism analysis program, a web-based tool for microbial community analysis. *Appl. Environ. Microbiol.* **66:**3616–3620.
85. **Marshall, A., and J. Hodgson.** 1998. DNA chips: an array of possibilities. *Nat. Biotechnol.* **16:**27–31.
86. **Martínez-Murcia, A. J., S. G. Acinas, and F. Rodriguez-Valera.** 1995. Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. *FEMS Microbiol. Ecol.* **17:**247–256.
87. **Massol-Deya, A. A., D. A. Odelson, R. F. Hickey, and J. M. Tiedje.** 1995. Bacterial community finger printing of amplified 16S and 16-23S ribosomal DNA gene sequences and restriction endonuclease analysis (ARDRA), p. 3.3.2.1–3.3.2.8. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), *Molecular Microbial Ecology Manual*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
88. **Meier, H., R. Amann, W. Ludwig, and K. H. Schleifer.** 1999. Specific oligonucleotide probes for in situ detection of a major group of gram-positive bacteria with low DNA G+C content. *Syst. Appl. Microbiol.* **22:**186–196.
89. **Mobarry, B. K., M. Wagner, V. Urbain, B. E. Rittmann, and D. A. Stahl.** 1996. Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. *Appl. Environ. Microbiol.* **62:**2156–2162.
90. **Moeseneder, M. M., J. M. Arrieta, G. Muyzer, C. Winter, and G. J. Herndl.** 1999. Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* **65:**3518–3525.

91. **Morotomi, M., T. Ohno, and M. Mutai.** 1988. Rapid and correct identification of intestinal *Bacteroides* spp. with chromosomal DNA probes by whole-cell dot blot hybridization. *Appl. Environ. Microbiol.* **54:**1158–1162.
92. **Moyer, C. L., F. C. Dobbs, and D. M. Karl.** 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl. Environ. Microbiol.* **60:**871–879.
93. **Moyer, C. L., J. M. Tiedje, F. C. Dobbs, and D. M. Karl.** 1996. A computer-simulated restriction fragment length polymorphism analysis of bacterial small-subunit rRNA genes: efficacy of selected tetrameric restriction enzymes for studies of microbial diversity in nature. *Appl. Environ. Microbiol.* **62:**2501–2507.
94. **Murray, A. E., D. Lies, G. Li, K. Nealson, J. Zhou, and J. M. Tiedje.** 2001. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc. Natl. Acad. Sci. USA* **98:**9853–9858.
95. **Muyzer, G.** 1999. DGGE/TGGE: a method for identifying genes from natural ecosystems. *Curr. Opin. Microbiol.* **2:**317–322.
96. **Muyzer, G., E. C. DeWall, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59:**695–700.
97. **Muyzer, G., and K. Smalla.** 2000. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Leeuwenhoek* **73:**127–141.
98. **Myers, R. M., T. Maniatis, and L. S. Lerman.** 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155:**501–527.
99. **Nachamkin, I., N. J. Panaro, M. Li, H. Ung, P. K. Yuen, L. J. Kricka, and P. Wilding.** 2001. Agilent 2100 bioanalyzer for restriction fragment length polymorphism analysis of the *Campylobacter jejuni* flagellin gene. *J. Clin. Microbiol.* **39:**754–757.
100. **Neufeld, J. D., Z. T. Yu, W. Lam, and W. W. Mohn.** 2004. Serial analysis of ribosomal sequence tags (SARST): a high-throughput method for profiling complex microbial communities. *Environ. Microbiol.* **6:**131–144.
101. **Nielsen, A. T., W.-T. Liu, C. Filipe, L. Grady, Jr., S. Molin, and D. A. Stahl.** 1999. Identification of a novel group of bacteria in sludge from a deteriorated biological phosphorus removal reactor. *Appl. Environ. Microbiol.* **65:** 1251–1258.
102. **Noda, S., M. Ohkuma, R. Usami, K. Horikoshi, and T. Kudo.** 1999. Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Appl. Environ. Microbiol.* **65:**4935–4942.
103. **Nübel, U., F. Garcia-Pichel, and G. Muyzer.** 1997. PCR primers to amplify 16S RNA genes from cyanobacteria. *Appl. Environ. Microbiol.* **63:**3327–3332.
104. **Ohkuma, M., S. Noda, and T. Kudo.** 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl. Environ. Microbiol.* **65:**4926–4934.
105. **Okabe, S., T. Itoh, H. Satoh, and Y. Watanabe.** 1999. Analyses of spatial distributions of sulfate-reducing bacteria and their activity in aerobic wastewater biofilms. *Appl. Environ. Microbiol.* **65:**5107–5116.
106. **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40:** 337–365.
107. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* **176:**1–6.
108. **Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi, and T. Sekiya.** 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86:** 2766–2770.
109. **Osborn, A. M., E. R. B. Moore, and K. N. Timmis.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ. Microbiol.* **2:**39–50.
110. **Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen.** 1986. The analysis of natural microbial populations by ribosomal RNA sequences, p. 1–55. *In* K. C. Marshall (ed.), *Advances in Microbial Ecology*. Plenum Press, New York, N.Y.
111. **Peplies, J., F. O. Glockner, and R. Amann.** 2003. Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Appl. Environ. Microbiol.* **69:**1397–1407.
112. **Pirie, F. J., D. F. York, A. A. Motala, and M. A. Omar.** 1999. Fluorescent automated single-stranded conformation (F-SSCP) analysis is able to detect a point mutation at the extreme 5′ end of a PCR product. *Clin. Biochem.* **32:**481–484.
113. **Polz, M. F., and C. M. Cavanaugh.** 1997. A simple method for quantification of uncultured microorganisms in the environment based on in vitro transcription of 16S rRNA. *Appl. Environ. Microbiol.* **63:**1028–1033.
114. **Raskin, L., W. C. Capman, M. D. Kane, B. E. Rittmann, and D. A. Stahl.** 1996. Critical evaluation of membrane supports for use in quantitative hybridizations. *Appl. Environ. Microbiol.* **62:**300–303.
115. **Raskin, L., W. C. Capman, R. Sharp, and D. A. Stahl.** 1996. Molecular ecology of gastrointestinal ecosystems, p. 243–298. *In* R. I. Mackie, B. A. White, and R. E. Isaacson (ed.), *Gastrointestinal Microbiology and Host Interactions*. Chapman & Hall, London, United Kingdom.
116. **Raskin, L., L. K. Poulsen, D. R. Noguera, B. E. Rittmann, and D. A. Stahl.** 1994. Quantification of methanogenic groups in anaerobic biological reactors using oligonucleotide probe hybridizations. *Appl. Environ. Microbiol.* **60:**1241–1248.
117. **Raskin, L., B. E. Rittmann, and D. A. Stahl.** 1996. Competition and coexistence of sulfate-reducing and methanogenic populations in anaerobic biofilms. *Appl. Environ. Microbiol.* **62:**3846–3867.
118. **Raskin, L., J. M. Stromley, B. E. Rittmann, and D. A. Stahl.** 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* **60:**1232–1240.
119. Reference deleted.
120. **Risatti, J. B., W. C. Capman, and D. A. Stahl.** 1994. Community structure of a microbial mat: the phylogenetic dimension. *Proc. Natl. Acad. Sci. USA* **91:**10173–10177.
121. **Ritchie, N. J., M. E. Schutter, R. P. Dick, and D. D. Myrold.** 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. *Appl. Environ. Microbiol.* **66:**1668–1675.
122. **Roller, C., M. Wagner, R. Amann, W. Ludwig, and K. H. Schleifer.** 1995. In situ probing of gram-positive bacteria with high DNA G+C content using 23S rRNA-targeted oligonucleotides. *Microbiology* **141:**1267.

123. **Sakano, Y., K. Pickering, P. Strom, and L. Kerkhof.** 2002. Spatial distribution of total, ammonia-oxidizing, and denitrifying bacteria in biological wastewater treatment reactors for bioregenerative life support. *Appl. Environ. Microbiol.* **68:**2285–2293.
124. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
125. **Sandaa, R. A., V. Torsvik, O. Enger, F. L. Daae, T. Castberg, and D. Hahn.** 1999. Analysis of bacterial communities in heavy metal-contaminated soils at different levels of resolution. *FEMS Microbiol. Ecol.* **30:**237–251.
126. **SantaLucia, J., Jr., and D. Hicks.** 2004. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33:**415–440.
127. **Schafer, H., and G. Muyzer.** 2001. Denaturing gradient gel electrophoresis in marine microbial ecology. *Methods Microbiol.* **30:**425–468.
128. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:**467–470.
129. **Schmidt, T. M., E. F. DeLong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173:** 4371–4378.
130. **Schwieger, F., and C. C. Tebbe.** 1998. A new approach to utilize PCR–single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl. Environ. Microbiol.* **64:**4870–4876.
131. **Selvakumar, N., B. C. Ding, and S. M. Wilson.** 1997. Separation of DNA strands facilitates detection of point mutations by PCR-SSCP. *BioTechniques* **22:**604–606.
132. **Service, R. F.** 1998. Microchip arrays put DNA on the spot. *Science* **282:**396–399.
133. **Sheffield, V. C., J. S. Beck, E. M. Stone, and R. M. Myers.** 1992. A simple and efficient method for attachment of a 40-base pair, GC-rich sequence to PCR-amplified DNA. *BioTechniques* **12:**386–387.
134. **Shen, Y., L. G. Stehmeier, and G. Voordouw.** 1998. Identification of hydrocarbon-degrading bacteria in soil by reverse sample genome probing. *Appl. Environ. Microbiol.* **64:**637–645.
135. **Smith, G. B., and J. M. Tiedje.** 1992. Isolation and characterization of a nitrate reductase gene and its use as a probe for denitrifying bacteria. *Appl. Environ. Microbiol.* **58:**376–384.
136. **Smoot, L. M., J. C. Smoot, H. Smidt, P. A. Noble, M. Könneke, Z. A. McMurry, and D. A. Stahl.** 2005. DNA microarrays as salivary diagnostic tools for characterizing the oral cavity microbial community. *Adv. Dent. Res.* **18:** 6–11.
137. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:**503–517.
138. **Speksnijder, A., G. A. Kowalchuk, S. De Jong, E. Kline, J. R. Stephen, and H. J. Laanbroek.** 2001. Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl. Environ. Microbiol.* **67:**469–472.
139. **Stahl, D. A.** 1995. Application of phylogenetically based hybridization probes to microbial ecology. *Mol. Ecol.* **4:**535–542.
140. **Stahl, D. A.** 1986. Evolution, ecology and diagnosis: unity in variety. *Bio/Technology* **4:**623– 628.
141. **Stahl, D. A.** 2004. High-throughput techniques for analyzing complex bacterial communities, *Adv. Syst. Biol.* **547:**5–17.
142. **Stahl, D. A., and R. Amann.** 1991. Development and application of nucleic acid probes in bacterial systematics, p. 205–248. *In* E. Stackebrandt and M. Goodfellow (ed.), *Sequencing and Hybridization Techniques in Bacterial Systematics*. John Wiley & Sons, Chichester, England.
143. **Stahl, D. A., R. Devereux, R. I. Amann, B. Flesher, C. Lin, and J. Stromley.** 1989. Ribosomal RNA based studies of natural microbial diversity and ecology, p. 669–673. *In* T. Hattori et al. (ed.), *Recent Advances in Microbial Ecology*. Japan Scientific Societies Press, Tokyo, Japan.
144. **Stahl, D. A., B. Flesher, H. Mansfield, and L. Montgomery.** 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Appl. Environ. Microbiol.* **54:**1079–1084.
145. **Suzuki, M., M. S. Rappé, and S. J. Giovannoni.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Appl. Environ. Microbiol.* **64:**4522–4529.
146. **Szymanski, M., M. Z. Barciszewska, J. Barciszewski, and V. A. Erdmann.** 2000. 5S ribosomal RNA database Y2K. *Nucleic Acids Res.* **28:**166–167.
147. **Teske, A., C. Wawer, G. Muyzer, and N. B. Ramsing.** 1996. Distribution of sulfate–reducing bacteria in a stratified fjord (Mariager Fjord, Denmark) as evaluated by most-probable-number counts and denaturing gradient gel electrophoresis of PCR-amplified ribosomal DNA fragments. *Appl. Environ. Microbiol.* **62:**1405–1415.
148. **Throback, I. N., K. Enwall, A. Jarvis, and S. Hallin.** 2004. Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. *FEMS Microbiol. Ecol.* **49:**401–417.
149. **Tijssen, P.** 1993. *Hybridization with Nucleic Acid Probes*. Elsevier, Amsterdam, The Netherlands.
150. **Torsvik, V., F. L. Daae, and J. Goksoyr.** 1995. Extraction, purification, and analysis of DNA from soil bacteria, p. 29–48. *In* J. T. Trevors and J. D. van Elsas (ed.), *Nucleic Acids in the Environment: Methods and Applications*. Springer-Verlag, New York, N.Y.
151. **Torsvik, V., F. L. Daae, R. A. Sandaa, and L. Ovreas.** 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* **64:** 53–62.
152. **Torsvik, V., J. Goksøyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56:**782–787.
153. **Torsvik, V., K. Salte, R. Sørheim, and J. Goksøyr.** 1990. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Appl. Environ. Microbiol.* **56:**776–781.
154. **Tsien, H. C., B. J. Bratina, K. Tsuji, and R. S. Hanson.** 1990. Use of oligodeoxynucleotide signature probes for identification of physiological groups of methylotrophic bacteria. *Appl. Environ. Microbiol.* **56:**2858–2865.
155. **Urakawa, H., P. A. Noble, S. El Fantroussi, J. J. Kelly, and D. A. Stahl.** 2002. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. *Appl. Environ. Microbiol.* **68:**235–244.
156. **Uyeno, Y., Y. Sekiguchi, A. Sunaga, H. Yoshida, and Y. Kamagata.** 2004. Sequence-specific cleavage of small-subunit (SSU) rRNA with oligonucleotides and RNase H: a rapid and simple approach to SSU rRNA-based quantitative detection of microorganisms. *Appl. Environ. Microbiol.* **70:**3650–3663.
157. **Van de Peer, Y., P. De Rijk, J. Wuyts, T. Winkelmans, and R. De Wachter.** 2000. The European small subunit ribosomal RNA database. *Nucleic Acids Res.* **28:**175–176.

158. **Van de Peer, Y., I. Van de Broeck, P. De Rijk, and R. De Wachter.** 1994. Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* **22:**3488–3494.
159. **van der Maarel, M. J. E. C., R. R. E. Artz, R. Haanstra, and L. J. Forney.** 1998. Association of marine archaea with the digestive tracts of two marine fish species. *Appl. Environ. Microbiol.* **64:**2894–2898.
160. **von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21:**213–229.
161. **Voordouw, G., Y. Shen, C. S. Harrington, A. J. Telang, T. R. Jack, and D. W. S. Westlake.** 1993. Quantitative reverse sample genome probing of microbial communities and its application to oil field production waters. *Appl. Environ. Microbiol.* **59:**4101–4114.
162. **Voordouw, G., J. K. Voordouw, T. R. Jack, J. Foght, P. M. Fedorak, and D. W. S. Westlake.** 1992. Identification of distinct communities of sulfate-reducing bacteria in oil fields by reverse sample genome probing. *Appl. Environ. Microbiol.* **58:**3542–3552.
163. **Voordouw, G., J. K. Voordouw, R. R. Karkhoff-Schweizer, P. M. Fedorak, and D. W. S. Westlake.** 1991. Reverse sample genome probing, a new technique for identification of bacteria in environmental samples by DNA hybridization, and its application to the identification of sulfate-reducing bacteria in oil field samples. *Appl. Environ. Microbiol.* **57:**3070–3078.
164. **Vos, M., R. M. Warren, C. J. van Rensburg, J. H. Hauman, J. T. Epplen, and P. D. van Helden.** 1999. Rapid genomic diversification of Helicobacter pylori. *Electrophoresis* **20:**1476–1485.
165. **Wagner, M., G. Rath, R. Amann, H.-P. Koops, and K.-H. Schleifer.** 1995. In situ identification of ammonia-oxidizing bacteria. *Syst. Appl. Microbiol.* **18:**251–264.
166. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature, p. 219–286. *In* K. C. Marshall (ed.), *Advances in Microbial Ecology*. Plenum Press, New York, N.Y.
167. **Wawer, C., and G. Muyzer.** 1995. Genetic diversity of *Desulfovibrio* spp. in environmental samples analyzed by denaturing gradient gel electrophoresis of [NiFe] hydrogenase gene fragments. *Appl. Environ. Microbiol.* **61:**2203–2210.
168. **Weller, R., and D. M. Ward.** 1989. Selective recovery of 16S rRNA sequences from natural microbial communities in the form of cDNA. *Appl. Environ. Microbiol.* **55:**1818–1822.
169. **Wetherall, B. L., and A. M. Johnson.** 1990. Nucleic acid probes for Campylobacter species, p. 255–285. *In* A. J. L. Macario and E. Conway de Macario (ed.), *Gene Probes for Bacteria*. Academic Press, Inc., San Diego, Calif.
170. **Wetmur, J. G., and N. Davidson.** 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31:**349–370.
171. **Wieland, A., M. Kuhl, L. McGowan, A. Fourcans, R. Duran, P. Caumette, T. Garcia de Oteyza, J. Grimalt, A. Sole, E. Diestra, I. Esteve, and R. Herbert.** 2003. Microbial mats on the Orkney Islands revisited: microenvironment and microbial community composition. *Microb. Ecol.* **46:**371–390.
172. **Wilson, I. G.** 1997. Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* **63:**3741–3751.
173. **Wilson, K. H., W. J. Wilson, J. L. Radosevich, T. Z. DeSantis, V. S. Viswanathan, T. A. Kuczmarski, and G. L. Andersen.** 2002. High-density microarray of small-subunit ribosomal DNA probes. *Appl. Environ. Microbiol.* **68:**2535–2541.
174. **Wu, L. Y., D. K. Thompson, G. S. Li, R. A. Hurt, J. M. Tiedje, and J. Z. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67:**5780–5790.
175. **Wu, L. Y., D. K. Thompson, X. D. Liu, M. W. Fields, C. E. Bagwell, J. M. Tiedje, and J. Z. Zhou.** 2004. Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ. Sci. Technol.* **38:**6775–6782.
176. **Zhou, J.** 2003. Microarrays for bacterial detection and microbial community analysis. *Curr. Opin. Microbiol.* **6:**288–294.
177. **Zhou, J. Z., and D. K. Thompson.** 2002. Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol.* **13:**204–207.
178. **Zumstein, E., R. Moletta, and J.-J. Godon.** 2000. Examination of two years of community dynamics in an anaerobic bioreactor using fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis. *Environ. Microbiol.* **2:**69–78.
179. **Zverlov, V., M. Klein, S. Lucker, M. W. Friedrich, J. Kellermann, D. A. Stahl, A. Loy, and M. Wagner.** 2005. Lateral gene transfer of dissimilatory (bi)sulfite reductase revisited. *J. Bacteriol.* **187:**2203–2208.

Phylogenetic and Genomic Analysis

DIRK GEVERS AND TOM COENYE

13

INTRODUCTION

Polyphasic Taxonomy and 16S rRNA Gene-Based Phylogeny

Polyphasic taxonomy aims at the integration of different kinds of data and information on microorganisms, and in principle all genotypic, phenotypic, and phylogenetic information may be incorporated in polyphasic taxonomy (113). Genotypic methods are directed toward DNA or RNA molecules, and these methods dominate modern taxonomy, because it is now generally accepted that a classification should reflect the natural relationships as encoded in the DNA. Genotypic methods include determination of the DNA base ratio (%G+C), DNA-DNA hybridization studies, sequence determination of rRNA genes, and a large variety of DNA-based typing methods. Determination of the number of moles of guanine + cytosine is one of the classical genotypic methods, and the moles percent G+C is considered to be part of the standard description of new bacterial taxa (113). The base composition of bacterial chromosomal DNA ranges from about 25 to 80%, and in general the moles percent G+C of a species shows a narrow range of about 1 to 3 mol% (104). The percent DNA-DNA hybridization is an indirect parameter of the sequence similarity between two entire genomes and is the "gold standard" for the delineation of species. Several methods are available, and these are reviewed by Johnson (67). Wayne et al. defined a prokaryotic species as an entity that included strains with approximately 70% or greater DNA-DNA relatedness (116). However, several examples have shown that, although data derived from DNA-DNA hybridizations are important, it is essential to be flexible about the boundaries of species demarcation to obtain an optimal classification system that facilitates identification (27, 111, 120).

There is a consensus that the phylogenetic relationships between prokaryotes can be deduced from sequence comparisons of conserved macromolecules. rRNA genes are among the best targets for these phylogenetic studies because they are universally present and functionally constant and have a mosaic structure of highly conserved and more variable domains (22, 121). The direct sequencing of the genes coding for small (16S) or large (23S) rRNA molecules by PCR technology has provided a phylogenetic framework that serves as a backbone for modern microbial taxonomy (45, 78), and this technique is now standard practice in taxonomic studies—nowadays, it is hard to imagine a new prokaryotic species being described without its 16S rRNA gene sequence being determined first! The availability of international online databases (some of them entirely devoted to rRNA gene sequences) (8, 28, 69, 126) facilitates the exchange of 16S rRNA gene sequences and has contributed significantly to their widespread use. However, there are some pitfalls associated with the use of 16S rRNA gene sequences. First of all, it should be clear that comparison of these sequences cannot replace DNA-DNA hybridizations for species delineation (43, 101). Although organisms that share less than 97% 16S rRNA gene sequence similarity will not show DNA-DNA binding values higher than 60%, isolates that share more than 97% 16S rRNA gene sequence similarity may or may not belong to the same species, and there is no threshold value for 16S rRNA sequence similarity for species recognition. Secondly, prokaryotes often contain multiple rRNA operons, and although interoperon variability of 16S rRNA gene sequences within a single genome are generally low and unlikely to affect phylogenetic analyses, exceptions exist (2, 14, 25). It should also be mentioned that overall little is known about intraspecific variation in 16S rRNA gene sequences, although several studies indicate that it may be higher than generally assumed (23, 51, 112). One of the reasons why the 16S rRNA has become the molecule of choice for phylogenetic analysis is that it was considered unlikely to be prone to horizontal gene transfer and/or recombination; however, more recently, several studies have highlighted that distinct segments along 16S rRNA gene sequences may have a reticulate evolutionary history (see, for example, reference 110). Finally, there are some characteristics of 16S rRNA genes that may lead to inaccurate trees, including (i) considerable differences in mutation rates between different lineages (123), (ii) overestimation of relatedness of species with similar nucleotide frequencies (117), and (iii) difficulties in aligning distantly related taxa.

Alternative Molecular Markers

Because of the above-mentioned limitations, alternative phylogenetic markers have been used since the early 1990s (34, 114). The choice of which additional molecule(s) to include in comparative sequence analysis is not straightforward.

Nevertheless, there seems to be agreement that these genes (i) should preferably be widely distributed, (ii) should not be transmitted horizontally frequently, (iii) should be present as a single copy, and (iv) must be long enough to contain sufficient information but short enough to allow convenient sequencing (114, 131). In addition, they should contain the "right" amount of phylogenetic information (resolution); i.e., they should neither be too conserved nor too variable. However, several studies have shown that, based on these criteria, very few (if any) genes will be found that can be applied universally, and it is most likely that the development of group-specific approaches will be necessary (24, 92, 131). Rather than presenting an exhaustive overview of studies relying on a wide range of different alternative phylogenetic markers, we would like to illustrate the use of these markers with the *recA* gene as an example.

The use of the *recA* gene in phylogenetic studies was pioneered by Lloyd and Sharp (80), Eisen (34), and Karlin et al. (70). RecA is a relatively small protein (approximately 350 amino acids) involved in homologous DNA recombination, SOS induction, and DNA-damage-induced mutagenesis, and it is capable of binding single- and double-stranded DNA, unwinding duplex DNA, and finding homologous regions (34, 70). From these early studies it was already obvious that the *recA* gene was a good candidate alternative marker: (i) it is present in most prokaryotes, (ii) some regions are conserved while others are more variable, and (iii) the extensive size and sequence conservation and the fact that *recA* is a protein-coding gene make alignments virtually unambiguous. The first large-scale study using *recA* sequences from *Bacteria* (*Proteobacteria*, *Cyanobacteria*, and gram-positive organisms) as well as from *Archaea* (34) showed that phylogenetic trees based on *recA* sequences displayed topologies and resolution similar to those of 16S-rRNA-gene-based trees. Subsequent studies have focused on various groups of bacteria.

Mycobacterial species often have very similar 16S rRNA gene sequences (e.g., *Mycobacterium kansasii* and *Mycobacterium gastri* have identical 16S rRNA gene sequences), which may hinder their accurate identification (109). Comparison of phylogenetic trees based on 16S rRNA gene and *recA* sequences revealed a general likeness in topology (11). However, unlike the 16S rRNA gene, sequence similarities of the *recA* gene are significantly lower between species (e.g., 96.2% between M. *kansasii* and M. *gastri*), allowing a more reliable species-level identification (11).

The family *Vibrionaceae* contains six genera, including the genus *Vibrio* (containing 44 species). The classification and identification of vibrios to the species level require the application of state-of-the-art genomic analyses, including amplified fragment length polymorphism fingerprinting and DNA-DNA hybridizations (108). Thompson et al. used *recA* as an alternative phylogenetic marker in the *Vibrionaceae* (106). Their data showed that there was a relatively good correlation between *recA* and 16S rRNA gene sequence data but that overall *recA* gene sequences were much more discriminatory than were 16S rRNA gene sequences. The *recA* data also showed that *Vibrio* species appear to be polyphyletic.

Species-level identification of *Burkholderia* species, and most notably of members of the *Burkholderia cepacia* complex (a group of closely related species sharing >97.7% 16S rRNA gene sequence similarity), is not straightforward (27). 16S rRNA gene sequence analysis is of limited value in the genus *Burkholderia* (especially in the *B. cepacia* complex), as its resolution is too low to allow accurate identifications. Recently a *recA*-based approach for identifying *Burkholderia* species was developed (85). With this approach, it was possible to identify all *Burkholderia* species to the species level. In particular the degree of resolution of the *recA* phylogenetic tree for members of the *B. cepacia* complex was much greater than that observed with 16S rRNA gene sequence analysis (Fig. 1). However, the presence of discrete *recA* lineages within some members of the *B. cepacia* complex adds an additional layer of complexity.

Other frequently used molecular markers include *gyrB*, *rpoB*, *rpoD*, and *hsp60* (see, for example, references 3, 19, 20, 29, 56, 75, 76, 98, and 127).

Multilocus Sequence Analysis

However, even trees based on these alternative markers have been criticized, as concern was raised that any single-gene tree cannot adequately reflect phylogenetic relationships because of the possibility of horizontal gene transfer, variable mutation rates, variable rates of recombination, and simple stochastic variation (24, 46). The ad hoc committee for the reevaluation of the species definition in bacteriology proposed that a minimum of five housekeeping genes be sequenced to achieve an adequately informative level of phylogenetic data (100). In analogy with multilocus sequence typing (a sequence-based typing method primarily used to distinguish infraspecific groups within named species), this polygenic approach has been called multilocus sequence analysis (MLSA) (46). MLSA is starting to become common practice in taxonomic studies (see, for example, references 3, 19, 20, 56, 76, 82, 83, 98, and 107), and it is anticipated that in the future it may replace DNA-DNA hybridizations (46, 100).

Classification of microorganisms by MLSA could rely on either one of two approaches. In the first approach, one may select a universal set of genes that allows for the hierarchical classification of all prokaryotes (92, 131). However, this idea may be impractical, and since we are (in general) less interested in the deeper phylogenetic relationships between genera, a more realistic approach is to look for sets of genes that can be used within all strains of a particular group (a genus or family). The reason behind this thinking is that genes that are informative within a given genus or family may not be useful or even present in other taxa. Also, genes that are conserved enough to be amplified by a common set of primers from all species may not evolve quickly enough to distinguish closely related taxa. Nevertheless, some genes may be informative in more than one group-specific set, and these more widely distributed genes could provide tools for broader comparisons. The genes used in MLSA should be ubiquitous (at least in the taxon under study) and present in a single copy, but those genes in which recombination might confer a selective advantage (e.g., virulence- and antigen-encoding genes) or closely linked genes should be avoided. Recently, a novel tool was described that automatically makes a selection of representative proteins for bacterial phylogeny (9).

The simplest approach is to concatenate the sequences of the sets of homologous genes and to use the concatenated sequences to construct a tree (see below) that can delineate clusters that can aid in the division of the genus into species. Prokaryotic identification in this scenario is a two-step process. In the first step, 16S rRNA gene sequencing is used to assign an unknown isolate to a group (genus or family). In the second step, genes and primers specific for that group are used in MLSA to assign the isolate to a species. We would like to emphasize that, at present, there are no

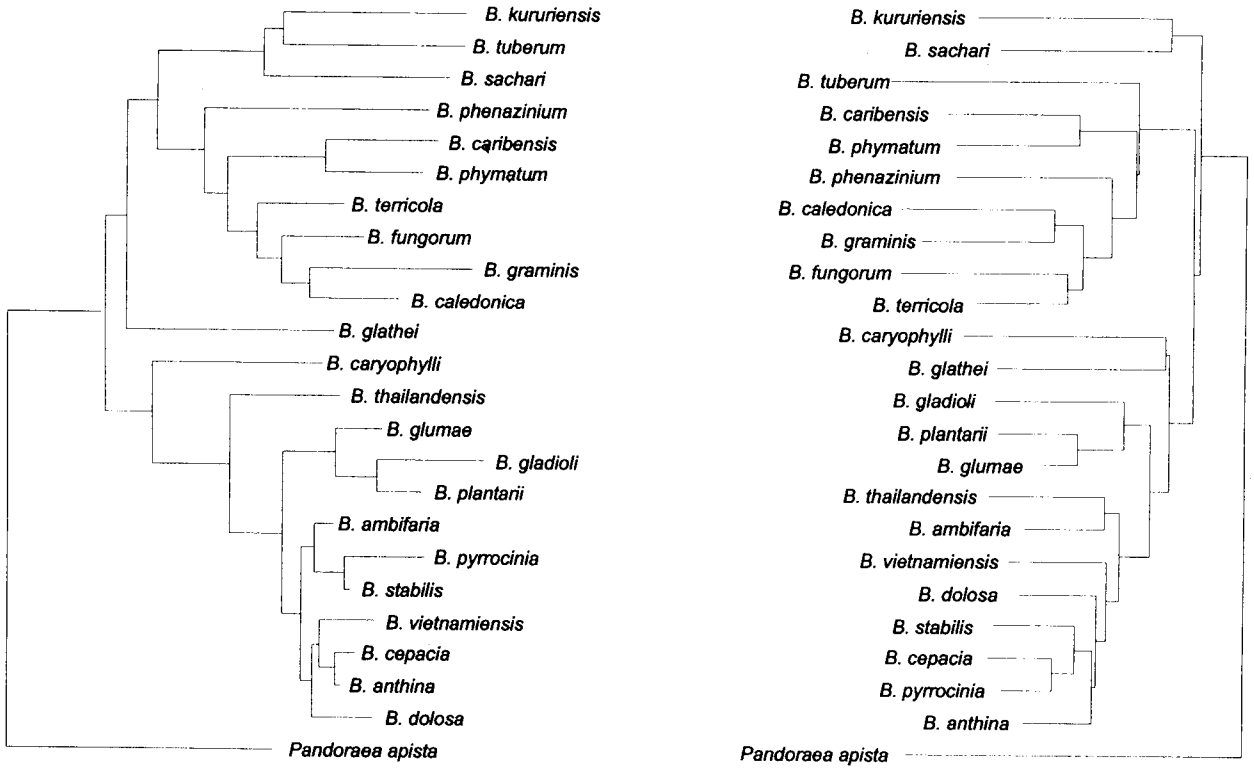

FIGURE 1 Comparison of phylogenetic trees based on 16S rRNA gene (right; scale bar indicates 1% sequence dissimilarity) and *recA* (left; scale bar indicates 5% sequence dissimilarity) sequences of a selection of *Burkholderia* species. Sequences were extracted from GenBank and were aligned by using BioNumerics 4.0 (Applied Maths, Sint-Martens-Latem, Belgium), and trees were constructed based on the NJ algorithm as implemented in BioNumerics 4.0.

guidelines to delineate species based on MLSA and that it is therefore necessary to "calibrate" novel MLSA schemes by using DNA-DNA hybridization values (76, 100). Two examples of the use of MLSA are given below.

Godoy et al. used MLSA to investigate the relationships among the closely related organisms *Burkholderia mallei*, *Burkholderia pseudomallei*, and *Burkholderia thailandensis* (47). MLSA showed the clear separation of *B. pseudomallei* isolates from those of *B. thailandensis*, a finding consistent with other data that indicate that they are different species. However, MLSA also showed that *B. mallei* is a clone within *B. pseudomallei*, raised to species status due to its ability to cause a distinctive disease (i.e., equine glanders) in horses, mules, and donkeys.

Similarly, MLSA was used to investigate the relationships among members of the *Bacillus cereus* group (including *B. cereus*, *Bacillus thuringiensis*, *Bacillus anthracis*, and *Bacillus mycoides*) (87), confirming that strains of the insect pathogen *B. thuringiensis* correspond to several distinct lineages within the phylogeny of the free-living soil bacterium *B. cereus*. *B. anthracis*, on the other hand, appears to be a specialized clone of *B. cereus* that was recognized as a single separate species because of its capacity to cause anthrax. These conclusions are confirmed by whole-genome comparisons (5, 88), in which it was shown that (i) genomes of members of this group show a high level of synteny and a high level of protein identity, (ii) very few genes are unique to one species, and (iii) much of the "specificity" in this group can be attributed to plasmid content.

PHYLOGENETIC ANALYSIS

This section aims to briefly discuss the different steps in constructing evolutionary trees, including collecting data sets of homologous sequences, generating a multiple-sequence alignment, inferring tree topology, and assessing confidence in the tree. In addition, some theoretical background is given. For further details, the reader is referred to a number of recent reviews (6, 55, 57, 119) and two recent books that highlight the topic from a more practical perspective (49, 91).

What Is a Phylogenetic Tree?

A phylogenetic tree is a tree-like diagram that depicts the evolutionary relationships between different genes or organisms. All trees have branches and nodes. External nodes are often called operational taxonomic units (OTUs) and are the terminal points where the evolutionary process has yielded the sequences under analysis. Internal nodes are the points where branches split from each other and may be called hypothetical taxonomic units to emphasize that they are the hypothetical progenitors of OTUs, also referred to as the last common ancestor of everything arising from it. Molecular phylogenetic trees are usually drawn with proportional

branch lengths; the length of the branches corresponds to the degree of sequence divergence, or amount of evolution, between the two nodes that they connect.

Phylogeneticists often speak of rooted and unrooted trees. An unrooted tree only positions the individual taxa relative to each other without indicating the direction of the evolutionary process. In a rooted tree, a root represents a common ancestor of all sequences analyzed and therefore implies the order of branching in the rest of the tree. A tree can be rooted if one or more of the OTUs form an external point of reference and are known as, or are believed to be, the most distantly related of the OTUs (i.e., outgroup routing).

Phylogenetic trees illustrate the relationship among the sequences analyzed; therefore, they are always gene trees. Whether these gene trees are species trees and can be interpreted as representative of the organismal phylogeny—that is, the topology that traces the history of the replicating cell lineages that transmit genes and genomes to successive generations—depends on the gene selected. An example of a robust hypothesis for organismal phylogenies based on a multigene approach and applied to the case of γ-proteobacteria has been published by Lerat et al. (77). What should be considered the organismal phylogeny, if any, is a topic beyond the scope of this chapter (31, 125).

DNA or Protein?

Nucleotide sequences may be coding or noncoding. For protein-coding genes, the alignment can be accomplished based on the nucleotide or the amino acid sequences. Because it is the DNA that contains all the information to create functional proteins, it is often thought that the DNA sequence should also be used in molecular phylogeny. However, there are reasons why it may be more appropriate to use protein sequences for such analyses (84). (i) In protein-coding sequences, the first and second nucleotide of each codon are less prone to the incorporation of mutations because it almost always leads to a change in amino acid. When one compares sequences that have diverged for possibly hundreds of millions of years, it is likely that the third codon position has become saturated, resulting in positions with no phylogenetic signal. This pitfall is avoided by looking at amino acid sequence or by excluding the third position from each codon in the alignment. (ii) Because DNA is composed of only four different units, two randomly chosen aligned sequences will have on average 25% identical residues if gaps are not allowed. This percentage increases even up to 50% when gaps are allowed. This situation may obscure any genuine relationship that may exist between two gene sequences. By contrast, the alignment of proteins with their 20 amino acids is less cumbersome. On average, 5% of residues in two randomly chosen and aligned sequences would be identical. Even with gaps, still only 10 to 15% of residues are identical. This situation makes protein sequences easier to align and allows the signal-to-noise ratio to improve significantly.

In general, it is strongly recommended to analyze both DNA and protein data sets. For a group of closely related species or taxa, DNA-based analysis is probably a good method because problems like differences in codon bias or saturation of the third position of codons can be avoided. In case of ambiguity in the alignment of gene sequences, it is recommended to first translate the sequences to their corresponding protein sequence and then to align and determine the position gaps in the DNA sequence according to the protein alignment.

Collecting a Set of Homologous Sequences

In an evolutionary analysis, it is absolutely required to study homologous sequences, i.e., sequences that share a common origin, as this requirement is a basic premise for phylogenetic analyses. It makes no sense at all to generate a phylogenetic tree of unrelated sequences (38). Homologs can be orthologous, paralogous, or xenologous (66). Orthologs duplicate only when their host divides, i.e., along with the rest of the genome. They are strictly vertically transmitted, so their phylogeny traces that of their host lineage. Paralogs and xenologs are members of multigene families that arise by intragenomic gene duplication and by horizontal gene transfer, respectively. Keeping this distinction in mind is important, particularly when the object of a phylogenetic reconstruction is to establish evolutionary relationships between organisms. If paralogous or xenologous genes are unknowingly considered, recovery of incorrect species relationships is likely.

Homology between sequences is an all-or-none property (i.e., two sequences are either homologous or not) and is always a hypothesis. Therefore, an indirect method is necessary to assess experimentally whether two sequences are homologous. In most cases the level of sequence similarity is the criterion used (e.g., E-value or similarity score cutoff). However, sequences might have diverged to the extent that their common origin is untraceable by a direct sequence comparison. It becomes very difficult to correctly detect homology for pairs with a pairwise sequence identity between 20 and 30%, the so-called twilight zone (90). One should remember that highly divergent sequences should be excluded in the evolutionary analysis, as they prevent a reliable alignment.

Once one has decided which phylogenetic marker will be studied (see above), the first practical step of building the data set often consists of recovery of reference sequences from public databases. One search strategy for finding a set of related sequences is a keyword search in public databases (e.g., using Entrez or SRS). Although this strategy is easy and seems more intuitive, it is far from exhaustive. Difficulty arises mostly because many data entries are incorrectly or badly annotated and thus hard to find with keyword searches. To compile a comprehensive data set, sequence similarity searching (e.g., using BLAST or FASTA) is strongly recommended. A consideration to be made in the selection of sequence sampling is that the range of organisms sampled should accurately reflect the total diversity present within the range of organisms under consideration.

A well-chosen outgroup should not be too distantly related because the sequences may have become saturated with multiple mutations, by which information may have been erased. It should also not be too closely related, to make sure it is a true outgroup. In the absence of an outgroup, the best guess is to place the root in the middle of the tree (i.e., midpoint rooting), or better, not to root it at all (6).

Creating a Good Sequence Alignment

Multiple-sequence alignments are the essential prerequisite for most phylogenetic analyses. Any phylogenetic inference based on molecular data begins by comparing the homologous residues (i.e., those that descend from a common ancestral residue) among different sequences. The best way to do this is to align sequences on top of each other, so that homologous residues from different sequences line up in the same column.

As an optimal sequence alignment by a simultaneous comparison of all N sequences is essentially impossible for four or more sequences (due to the enormous computer memory and time needed), most multiple-sequence alignments are constructed by the method known as "progressive sequence alignment" (37). That is, an alignment is built up stepwise, starting with the most similar sequences and progressively adding the more dissimilar ones. CLUSTAL is without doubt the software program most widely used to align a set of more than two sequences (17) and constructs a multiple alignment in three steps. The first step involves performing all pairwise comparisons between the sequences and generating a distance matrix representing pairwise sequence similarities. Secondly, based on this matrix, a guide tree is constructed by using the neighbor-joining (NJ) method (see below). Note that this is a "quick and dirty" tree and is unsuitable for serious phylogenetic inference. Finally, the alignment is built up progressively by a series of pairwise alignments following the branching order of the guide tree.

Progressive alignment is fast but heuristic; i.e., it does not guarantee finding the most optimal solution. The major problem with progressive sequence alignment programs is the dependence of the ultimate multiple-sequence alignment on the initial pairwise sequence alignment. Match errors during early steps in the alignment protocol are accumulated and propagated, leading to further errors in later steps. This problem is also referred to as the "once a gap always a gap" problem (37). Gaps can only be added or enlarged, never moved or removed. The latter actions would make the alignment process much slower. This predicament obviously results in errors that need manual adjustment to minimize insertion and/or deletion events and to improve the quality of the alignment, a widespread and fairly well-accepted correction (6).

Once an alignment has been created and manually checked or adjusted, it is necessary to select which positions will be used for subsequent analyses. Since an alignment makes statements about the homology of amino acids or nucleotides present at each position, it is important to include only unambiguously aligned sites. If there are gaps in the alignment, it can be difficult to say confidently that all positions are correctly aligned. Furthermore, the presence of incomplete sequences and variations in length of terminal regions of genes can mean that some alignment positions are poorly sampled, with missing data. These are preferentially positions to throw away, referred to as "stripping of the alignment" or "masking of data." It is a complicated and controversial part of phylogenetic analyses, but it is clear that in most cases masking increases the accuracy of tree reconstruction (57). Programs such as Gblocks (16) objectively assess which parts of the alignment are sufficiently conserved and useful for phylogenetic analysis.

The use of structural information can improve an alignment substantially, because secondary structural elements of functional RNAs and structural features of proteins are often more conserved than are primary sequence features (53, 90). As almost all RNA molecules form secondary structures, one can outperform the classical approaches by taking structure information into account. The software tool StructMiner efficiently detects and aligns conserved structural patterns (128). If a three-dimensional structure of the protein is available, the secondary structure can be deduced by programs such as DSSP (68) or STRIDE (44). In the absence of tertiary structures for a query set of protein sequences, information about secondary structure can be gathered only by using secondary structure prediction programs. Applications that integrate structural information into the alignment process have been published (e.g., PRALINE [96], and MASS [32]). When DNA sequences from protein-coding genes are aligned, one can also obtain an improved alignment by first aligning the inferred amino acid sequences and then performing a codon-by-codon back alignment of the corresponding DNA sequences. Software packages that automatically perform this task are available (e.g., transAlign and RevTrans) (10, 118).

In conclusion, shortcomings of most methods to keep in mind are that (i) they are not guaranteed to find the optimal alignment, (ii) the optimal alignment does not necessarily correspond to the alignment of homologous positions, and (iii) an alignment is generated also for random sequences and in regions of the proteins where variability is too high to be reliable.

Inferring Tree Topology

Reconstructing the phylogeny from a sequence alignment is not straightforward (102). There are no uniquely correct methods for inferring phylogenies, many methods exist, and it is rarely possible to verify that one has arrived at the true phylogenetic tree. There are currently four primary methods for constructing phylogenies from protein and nucleic acid sequence alignments: the distance-based NJ method and the character-based methods, including maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference.

Distance-Based Methods

Distance-based methods (e.g., NJ) are relatively simple and straightforward; starting from a sequence alignment, the differences between all pairwise combinations of OTUs (DNA or protein sequences) are converted into a distance matrix that represents an estimate derived from applying a so-called evolutionary or substitution model of the evolutionary distance between sequences (see further). These distances are than assembled into a tree. A disadvantage of distance methods is that they reduce the phylogenetic information to one number. The major advantage is that they are much less computer intensive, which is important when many taxa have to be compared.

Character-Based Methods

Character-based methods examine each column of the alignment separately (each position in the aligned sequences is a "character"). They look for the tree that best accommodates all of this information.

MP methods select the tree that explains the data observed in terms of the minimal number of possible substitutions. It is often the case that there are several trees, typically differing only slightly, that are consistent with the same number of events and that are therefore equally parsimonious. This method is becoming less popular mainly because MP generally uses a simple model of sequence substitution (all changes are equally probable). Moreover, it has been shown that, for more divergent sequences, MP is prone to recover incorrect trees (72).

ML methods seek to identify the single most probable tree on a statistical basis, given the chosen model of sequence evolution. It is a method that allows correcting for multiple mutational events at the same site and therefore is more suited to accurately reconstructing the relationships

between sequences that have been separated for a long time or are evolving rapidly. In ML, all possible mutational pathways that are compatible with the data are considered, and the tree that has the highest probability (i.e., the likelihood of the tree) of producing the observed sequences is preferred. The main advantage of ML over MP is that ML permits the inference of phylogenetic trees by using different (more complex) evolutionary models (see below). This truth of course implies that the obtained result depends on the accuracy and assumptions included in that model. Although such models are simpler than is the true process underlying sequence evolution, they seem to be relatively robust to violation of their simplifying assumptions (60). The main obstacle of ML is the computational burden; i.e., the number of different tree topologies that have to be evaluated increases enormously as a function of the number of sequences. Even with adaptations of the ML method, such as TREE-PUZZLE (95) or the fastDNAml program (4), to the most modern computer technologies, constructing trees of 40 or more sequences becomes very impractical. Bayesian estimation of phylogeny holds promise as an alternative to ML, particularly for large molecular-sequence data sets (61). Bayesian inference of phylogeny is based upon a quantity called the posterior probability distribution of trees, which is the probability of a tree conditioned on the observations. The conditioning is accomplished by using Bayes's theorem. For more comprehensive information, it is impossible to come up with a mathematical definition; therefore, we refer elsewhere (61). The advantages of this method are numerous. Firstly, like ML methods, it is based on the likelihood function, so it inherits many of the desirable statistical properties of ML. Secondly, it allows one to incorporate prior information (if available), e.g., a systematist's prior confirmed conviction about the phylogeny of the group under study. Thirdly, it has major computational advantages and therefore allows one to study large data sets and implement complex models of sequence evolution. Finally, it not only produces a tree estimate but also measures phylogenetic uncertainty for the groups on the tree (comparable to bootstrap values on ML trees; see below). These measures are more intuitive, as they represent the posterior probability that the group is true, and are thus more easily interpreted.

Evolutionary Models

In order to estimate the genetic or evolutionary distances between pairs of sequences, a mathematical model is necessary. Simply counting the observed differences between sequences does not accurately reflect the evolutionary distances between them. It fails to take into account superimposed multiple changes at individual sites (including back mutations), different mutational rates depending on nucleotides or amino acids or among sites of alignment, and different rates of mutation for different lineages and therefore makes a correction obligatory. An evolutionary model is a set of assumptions about the process of nucleotide or amino acid substitutions. They describe the different probabilities of change from one nucleotide or amino acid to another, with the aim of correcting for unseen changes along the phylogeny. A good visualization illustrating the different models is shown by Whelan et al. (119). Various models have been developed to try to estimate the true difference between sequences based on their present states, such as amino acid substitution matrices (e.g., Dayoff, Blossom, etc.) or gamma corrections (considering among-site rate variation), etc. (for a recent overview, see reference 13). The choice of an appropriate model is paramount to accurate evolutionary reconstruction. Good fit to an accurate model should provide a robust analysis. Good fit to an inappropriate model can be seriously misleading. Whereas MP implicitly assumes a model of evolution, distance and ML methods estimate parameters according to an explicit model of evolution. However, whereas distance methods estimate only a single parameter (substitutions per site) given the model, ML can estimate all the relevant parameters of the substitution model. One should note that a model is always simplified and often makes assumptions just to turn a complex problem into a computationally tractable one. But a model becomes a powerful tool when, despite its simplified assumptions, it can fit the data and make accurate predictions about the problem at hand. Programs exist to help one in selecting the best-fit model for nucleotide substitution (MODELTEST) or protein evolution (PROTTEST); both are also available as Web tools (1, 86).

Bootstrap Analysis

How well can one trust the tree that one has constructed? A popular way of assessing the robustness of the topology of a tree is nonparametric bootstrapping (35). The bootstrap analysis tests whether the whole data set is supporting the tree or if the tree is just a marginal winner among many nearly equal alternatives. In practice, such analysis goes as follows: from the original alignment, columns in the sequence alignment are chosen at random, until a new alignment is constructed with the same size as the original one. Some characters will not be included at all in a given bootstrap replication, while others might be included more than once (i.e., the "sampling with replacement" principle). For each artificial data set, a tree is constructed and compared with the tree based on the original alignment. The number of times that a cluster, as defined in the original tree, is also found in the bootstrap tree, is recorded, and the resulting bootstrap values are superimposed on the original tree.

One difficulty with this analysis is the precise interpretation of what bootstrap values represent (99); higher is clearly better, but what is a reasonable cutoff? Some have concluded that values of 70% or higher are likely to indicate reliable groupings (54). But this cutoff is not generally accepted. More discussion of the interpretation of bootstrap values has been published elsewhere (36).

The major advantage of the bootstrap technique is that it can be applied to basically all tree construction methods, although one has to keep in mind that applying the bootstrap method multiplies the computer time needed by the number of bootstrap samples requested. This drawback is not a concern when a fast analysis (like NJ or parsimony) is employed, but it can be an obstacle when ML is used. A Bayesian analysis takes as long as an ML analysis, but as implemented in the software package MrBayes, it does not have the same drawback because bootstrapping is not necessary. Instead, it provides Bayesian posterior probabilities as indicators of branch support (55).

General Guidelines

There are three main reasons why phylogenies may be incorrect. First, a random error occurs when the informative sites (or data points) are limited, resulting in any tree that could be generated. We can cope with this problem by applying enough data and use bootstrap values as an indication of the extent of the random error. Having many data is not enough, as a second problem is caused by bias. An estimation (tree) is biased if the data set is not representative of

the underlying distribution. For example, two sequences can be clustered together just because they both share an unusually high G+C content. And thirdly, there is a systematic error, i.e., tackling the problem in the wrong way by using an incorrect model (modeling error). Methods can be misleading no matter how many data one has. Therefore, no guarantee exists that one can produce the one tree with the correct topology. In order to be aware of the reliability of the topology of the resulting tree, one or all of the following should be done:

1. Apply different tree-building methods to the data set.
2. Vary the parameters used by the different programs.
3. Apply different evolutionary models for matrix construction.
4. Add or remove one or more OTUs to see the influence on tree topology.
5. Include an outgroup that may serve as a root for the tree.
6. Apply bootstrap analysis to the data set.

A tree should be considered robust and thus reliable only when widely different methods infer similar or identical tree topologies and when such topologies are supported by good bootstrap values.

Phylogenies from Multiple Genes

Until recently, phylogenetic analyses have been routinely based on homologous sequences of a single phylogenetic marker, i.e., the 16S rRNA gene among bacteria (122). Given the vast number of genome sequences now available, it is possible to compute trees from whole genomes (see "Use of Whole Genome Sequences for Phylogeny" below). However, researchers will not have access to full genome data from thousands of species in the immediate future. Given these constraints on data availability, we focus here on the use of multiple genes for constructing phylogenies, i.e., MLSA (see "Multilocus Sequence Analysis" above). Two main reasons for using multiple loci follow: (i) sequence-based approaches to organism phylogeny require loci that evolve more rapidly than do rRNA genes, in order to increase phylogenetic accuracy (89), and (ii) multiple genes provide a buffer against the distorting phylogenetic signals at a single locus, such as effects of recombination, gene conversion, and horizontal gene transfer (77). A software package (VisRD) allows a graphical inspection of the phylogenetic content of a sequence alignment to detect recombination and recombination breakpoints (42).

The main question that this section addresses is how to infer a phylogenetic tree from multiple genes. There are two fundamentally different ways. A simple approach is to concatenate multiple genes head to tail to form a super-gene alignment, assuming that all positions are independent and identically distributed, and construct a tree. Although there are definitely arguments in favor of this approach (30), one cannot ignore the specific evolutionary features for each of the genes (94). Estimation of a phylogeny always assumes a model of evolution, but because different genes likely have different evolutionary constraints and/or pressures, the parameters (including tree topology, branch lengths, rate heterogeneity among sites, and substitution probabilities) may change from gene to gene. For example, the substitution model may vary from gene to gene because of differences in G+C content. At the extreme, different genes may even support different tree topologies because of horizontal gene transfer. Therefore, it might be appropriate to describe the evolution of each gene by its own set of parameters. This method is a second approach, in which phylogenies are inferred separately for each gene and the resulting gene trees are used to generate a consensus phylogeny. This strategy gives a more conservative and a safer estimate of evolution, because it produces only high resolution in the branching pattern when there is at least a majority consensus among the different genes.

A test (the likelihood ratio test) to examine whether different genes support congruent trees and whether we can ignore gene-specific effects via concatenation is available (115). When faced with different histories, there are at least three obvious ways in which one might represent phylogeny: (i) as the collection of individual histories, (ii) as a tree representing the single dominant pattern among data sets, or (iii) as a network. A consensus network attempts to represent all phylogenetic signals present in the given set of gene trees, simultaneously, up to a given level of complexity. In the resulting network, regions of the evolutionary history that are undisputed within the set of gene trees appear tree-like, whereas regions containing conflicts are shown as a box of parallel edges (a split), the "dimensionality" of which reflects the number of conflicting signals. A comprehensive software package for analyzing and visualizing a multiple-gene data set is SplitsTree (62).

USE OF WHOLE GENOME SEQUENCES FOR PHYLOGENY

Genome Tree Approach

Since the publication of the complete genome sequence of *Haemophilus influenzae* Rd (41), over 350 completely sequenced microbial genomes have been published, and many more are under way (for an overview, see http://www.genomesonline.org/). Although there has been a general tendency to focus on organisms with particularly interesting properties (most notably human pathogens), by now the available complete genome sequences give a more or less adequate picture of the genomic diversity observed in culturable prokaryotes. A good starting point for browsing through completed genomes is the CBS Genome Atlas Database (50), which is available from http://www.cbs.dtu.dk/services/GenomeAtlas/. While the availability of an increasing number of completely sequenced genomes has significantly facilitated the search for alternative molecular markers, there are other ways in which these sequences can be used to deduce phylogenetic relationships between taxa. An overview of these different approaches is given elsewhere (24). Below, we focus on one of the most promising novel approaches to taxonomy, based on gene content.

When the gene contents of organisms are compared, the simplest approach is to consider genomes "bags of genes," and then to compare the contents of different "bags" (65). The identification of orthologous genes (orthologs are homologous sequences in different species that arose from a common ancestral gene during speciation) is pivotal in this approach and largely depends on the definition of orthology (125). Most studies use a minimal definition, terming putative orthologs as genes that have the highest level of significant pairwise identity when the genes are compared between genomes (i.e., they are identified as those homologous genes that show the largest identity of several possibilities above a certain threshold) (7, 26, 65). From the first analyses performed on a limited number of genomes, a few general trends emerged (65, 97). It was observed that (i) large genomes have many genes in common, (ii) the number of

genes that two genomes have in common depends on their evolutionary distance, (iii) the fraction of shared orthologous genes decreases rapidly in evolution, faster than does the identity between the shared orthologs, and (iv) the evolution of gene content shows non-tree-like aspects, as phylogenetically closely related species do not necessarily share orthologous genes that either of them shares with a phylogenetically more distant third species. The observation that large genomes have many genes in common (irrespective of phylogenetic distance) and the observations that smaller genomes (e.g., from strictly parasitic organisms) are dominated by essential genes and share a greater fraction of their genomes with other species (7, 129) suggest that it is useful to normalize or weigh the data (i.e., to correct for differences in genome size) before further analysis. Interestingly, Snel et al. also observed that trees based on gene content do not correlate well with phenotype-based trees (97). This finding is somewhat unexpected, as it has always been assumed that differences in observed phenotypes are the result of differences in gene content (see, for example, reference 64).

Several slightly different approaches to compare gene contents are used as well. The presence and absence of families of protein-encoding genes in sequenced genomes have been used to reconstruct the relationships between a number of organisms (40, 58). In this approach, proteins are grouped together in families if their pairwise similarity is greater than a preset value, thereby eliminating the need for the identification of putative orthologs in each genome and the need for specific alignments. Tekaia et al. constructed genome trees based on whole-proteome comparisons, using hierarchical classification of genomes (taking into account genome size, levels of ancestral gene redundancy due to duplications, and net gene gain or loss) (105). Protein folds are protein families that share the same basic molecular shape but not necessarily sequence similarity. The presence or absence of these features was used by Wolf et al. and by Lin and Gerstein to build genome trees (79, 124). Protein folds are considered by some to be ideal characteristics for building phylogenetic trees, as they represent fundamental molecular units used by organisms. Yang et al. focused on the fold superfamily level (129). The advantage of using the fold superfamily level instead of the fold level is that it offers a higher level of certainty that the members of each group share common ancestry.

In the concept of "extended gene content" (introduced by Gu and Zhang [48]), the status of a gene family in a given genome is recorded as absent, present as a single copy, or present as a duplicate (instead of merely being recorded as absent or present). Also, "hybrid" measures have been defined, in which both gene content and sequence conservation are expressed. "Genome blast distance phylogeny," an approach proposed by Henz et al., starts with an all-against-all pairwise comparison of genomes (52). Subsequently a distance matrix is calculated from the resulting high-scoring pairs. Kunin et al. derived a new composite measure (called "genome conservation") from the sum of alignment scores between all proteins for every pair of organisms (74).

Methods relying on shared gene content for reconstructing phylogenies have been criticized because of the tendency of gene content convergence (due to horizontal gene transfer and gene loss). Nevertheless, trees based on gene content generally correspond well to trees based on 16S rRNA gene sequences (although it should be noted that some discrepancies observed in various studies remain at present unexplained). This result indicates that, despite horizontal gene transfer, gene duplications, and gene loss, there is a strong phylogenetic signal in gene content. However, to reduce the possible impact of these genetic processes on trees based on gene content, several methods have been developed to filter out the "noise" that is associated with them. The reader can consult references 15, 21, and 33 for further details.

There are several approaches to transform the fraction of shared genes (shared protein families and protein folds, etc.) into a genome distance matrix, subsequently construct a tree or a network from these matrices, and perform statistical tests (24, 48, 63, 73, 93, 132). While a detailed discussion of these methods is outside the scope of the text, it is worth mentioning that several studies indicate that ML and MP approaches outperform distance methods for constructing trees based on gene content data (63, 132).

Comparative Microbial Genomics with DNA Microarrays

Differences and/or similarities between microorganisms can also be studied by using DNA microarrays. These methods have the advantage of not requiring the availability of a whole-genome sequence of all organisms being studied. Hybridization of DNA to whole-genome microarrays has been used to study the genetic diversity of a wide range of bacteria (see references 39 and 130 for recent reviews). The microarray technology has also been used for a number of specialized applications in taxonomy and identification. Cho and Tiedje proposed a new approach to identify bacteria based on genomic DNA-DNA similarity by employing microarray technology (18). This method (so far evaluated only with four *Pseudomonas* species) does not require laborious cross-hybridizations, and the resulting hybridization profiles can be used in statistical procedures to identify test strains and can be stored in a database.

Microarrays for the identification of specific bacteria were also developed. An overview is presented in reference 24.

Microarrays have also been developed to study specific bacterial populations and consortia, including sulfate-reducing bacteria (81), toluene- and ethylbenzene-degrading consortia (71), and methanotrophs (12, 103).

REFERENCES

1. **Abascal, F., R. Zardoya, and D. Posada.** 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**:2104–2105.
2. **Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj, and M. F. Polz.** 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* **186**:2629–2635.
3. **Adekambi, T., and M. Drancourt.** 2004. Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp*65, *sod*A, *rec*A and *rpo*B gene sequencing. *Int. J. Syst. Evol. Microbiol.* **54**:2095–2105.
4. **Alsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10**:41–48.
5. **Anderson, I., A. Sorokin, V. Kapatral, G. Reznik, A. Bhattacharya, N. Mikhailova, H. Burd, V. Joukov, D. Kaznadzey, T. Walunas, M. d'Souza, N. Larsen, G. Pusch, K. Liolios, Y. Grechkin, A. Lapidus, E. Goltsman, L. Chu, M. Fonstein, S. D. Ehrlich, R. Overbeek, N. Kyrpides,**

and N. Ivanova. 2005. Comparative genome analysis of Bacillus cereus group genomes with Bacillus subtilis. *FEMS Microbiol. Lett.* **250:**175–184.
6. **Baldauf, S. L.** 2003. Phylogeny for the faint of heart: a tutorial. *Trends Genet.* **19:**345–351.
7. **Bansal, A. K., and T. E. Meyer.** 2002. Evolutionary analysis by whole genome comparisons. *J. Bacteriol.* **184:** 2260–2272.
8. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler.** 2005. GenBank. *Nucleic Acids Res.* **33:**D34–D38.
9. **Bern, M., and D. Goldberg.** 2005. Automatic selection of representative proteins for bacterial phylogeny. *BMC Evol. Biol.* **5:**34.
10. **Bininda-Emonds, O. R.** 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* **6:**156.
11. **Blackwood, K. S., C. He, J. Gunton, C. Y. Turenne, J. Wolfe, and A. M. Kabani.** 2000. Evaluation of *recA* sequences for identification of *Mycobacterium* species. *J. Clin. Microbiol.* **38:**2846–2852.
12. **Bodrossy, L., N. Stralis-Pavese, J. C. Murrell, S. Radajewski, A. Weilharter, and A. Sessitsch.** 2003. Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ. Microbiol.* **5:**566–582.
13. **Bos, D. H., and D. Posada.** 2005. Using models of nucleotide evolution to build phylogenetic trees. *Dev. Comp. Immunol.* **29:**211–227.
14. **Boucher, Y., C. J. Douady, A. K. Sharma, M. Kamekura, and W. F. Doolittle.** 2004. Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* **186:**3980–3990.
15. **Brown, J. R., C. J. Douady, M. J. Italia, W. E. Marshall, and M. J. Stanhope.** 2001. Universal trees based on large combined protein sequence data sets. *Nat. Genet.* **28:**281–285.
16. **Castresana, J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17:**540–552.
17. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson.** 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31:**3497–3500.
18. **Cho, J. C., and J. M. Tiedje.** 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* **67:**3677–3682.
19. **Christensen, H., P. Kuhnert, J. E. Olsen, and M. Bisgaard.** 2004. Comparative phylogenies of the housekeeping genes *atp*D, *inf*B and *rpo*B and the 16S rRNA gene within the *Pasteurellaceae*. *Int. J. Syst. Evol. Microbiol.* **54:** 1601–1609.
20. **Cladera, A. M., A. Bennasar, M. Barcelo, J. Lalucat, and E. Garcia-Valdes.** 2004. Comparative genetic diversity of *Pseudomonas stutzeri* genomovars, clonal structure, and phylogeny of the species. *J. Bacteriol.* **186:**5239–5248.
21. **Clarke, G. D., R. G. Beiko, M. A. Ragan, and R. L. Charlebois.** 2002. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J. Bacteriol.* **184:**2072–2080.
22. **Clarridge, J. E., III.** 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* **17:**840–862.
23. **Clayton, R. A., G. Sutton, P. S. Hinkle, Jr., C. Bult, and C. Fields.** 1995. Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int. J. Syst. Bacteriol.* **45:**595–599.
24. **Coenye, T., D. Gevers, Y. Van de Peer, P. Vandamme, and J. Swings.** 2005. Towards a prokaryotic genomic taxonomy. *FEMS Microbiol. Rev.* **29:**147–167.
25. **Coenye, T., and P. Vandamme.** 2003. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* **228:**45–49.
26. **Coenye, T., and P. Vandamme.** 2003. Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case. *Microbiology* **149:**3507–3517.
27. **Coenye, T., P. Vandamme, J. R. Govan, and J. J. LiPuma.** 2001. Taxonomy and identification of the *Burkholderia cepacia* complex. *J. Clin. Microbiol.* **39:**3427–3436.
28. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33:**D294–D296.
29. **Dauga, C.** 2002. Evolution of the *gyr*B gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. *Int. J. Syst. Evol. Microbiol.* **52:**531–547.
30. **de Queiroz, A., M. J. Donoghue, and J. Kim.** 1995. Separate versus combined analysis of phylogenetic evidence. *Ann. Rev. Ecol. Syst.* **26:**657–681.
31. **Driskell, A. C., C. Ane, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson.** 2004. Prospects for building the tree of life from large sequence databases. *Science* **306:**1172–1174.
32. **Dror, O., H. Benyamini, R. Nussinov, and H. Wolfson.** 2003. MASS: multiple structural alignment by secondary structures. *Bioinformatics* **19** (Suppl. 1)**:**i95–i104.
33. **Dutilh, B. E., M. A. Huynen, W. J. Bruno, and B. Snel.** 2004. The consistent phylogenetic signal in genome trees revealed by reducing the impact of noise. *J. Mol. Evol.* **58:**527–539.
34. **Eisen, J. A.** 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* **41:**1105–1123.
35. **Felsenstein, J.** 1985. Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* **39:**783–791.
36. **Felsenstein, J., and H. Kishino.** 1993. Is there something wrong with the bootstrap on phylogenies—a reply. *Syst. Biol.* **42:**193–200.
37. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25:**351–360.
38. **Fitch, W. M.** 2000. Homology: a personal view on some of the problems. *Trends Genet.* **16:**227–231.
39. **Fitzgerald, J. R., and J. M. Musser.** 2001. Evolutionary genomics of pathogenic bacteria. *Trends Microbiol.* **9:**547–553.
40. **Fitz-Gibbon, S. T., and C. H. House.** 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* **27:**4218–4222.
41. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:**496–512.
42. **Forslund, K., D. H. Huson, and V. Moulton.** 2004. VisRD—visual recombination detection. *Bioinformatics* **20:**3654–3655.
43. **Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk, Jr.** 1992. How close is close: 16S rRNA sequence identity may not

be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* **42:**166–170.
44. **Frishman, D., and P. Argos.** 1995. Knowledge-based protein secondary structure assignment. *Proteins* **23:**566–579.
45. **Garrity, G. M., and J. G. Holt.** 2001. The road map to the Manual, p. 119–166. *In* D. R. Boone and R. W. Castenholz (ed.), *Bergey's Manual of Systematic Bacteriology*. Springer-Verlag, New York, N.Y.
46. **Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson, and J. Swings.** 2005. Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* **3:**733–739.
47. **Godoy, D., G. Randle, A. J. Simpson, D. M. Aanensen, T. L. Pitt, R. Kinoshita, and B. G. Spratt.** 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderfia mallei*. *J. Clin. Microbiol.* **41:**2068–2079.
48. **Gu, X., and H. Zhang.** 2004. Genome phylogenetic analysis based on extended gene contents. *Mol. Biol. Evol.* **21:**1401–1408.
49. **Hall, B. G.** 2001. *Phylogenetic Trees Made Easy: a How-To Manual for Molecular Biologists*. Sinauer Associates, Inc., Sunderland, Mass.
50. **Hallin, P. F., and D. W. Ussery.** 2004. CBS Genome Atlas Database: a dynamic storage for bioinformatic results and sequence data. *Bioinformatics* **20:**3682–3686.
51. **Harrington, C. S., and S. L. W. On.** 1999. Extensive 16S rRNA gene sequence diversity in *Campylobacter hyointestinalis* strains: taxonomic and applied implications. *Int. J. Syst. Bacteriol.* **49:**1171–1175.
52. **Henz, S. R., D. H. Huson, A. F. Auch, K. Nieselt-Struwe, and S. C. Schuster.** 2004. Whole-genome prokaryotic phylogeny. *Bioinformatics* **21:**2329–2335.
53. **Heringa, J.** 1999. Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. *Comput. Chem.* **23:**341–364.
54. **Hillis, D. M., M. W. Allard, and M. M. Miyamoto.** 1993. Analysis of DNA sequence data: phylogenetic inference. *Methods Enzymol.* **224:**456–487.
55. **Holder, M., and P. O. Lewis.** 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* **4:**275–284.
56. **Holmes, D. E., K. P. Nevin, and D. R. Lovley.** 2004. Comparison of 16S rRNA, *nifD*, *recA*, *gyrB*, *rpoB* and *fusA* genes within the family *Geobacteraceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **54:**1591–1599.
57. **Horner, D. S., and G. Pesole.** 2004. Phylogenetic analyses: a brief introduction to methods and their application. *Expert Rev. Mol. Diagn.* **4:**339–350.
58. **House, C. H., and S. T. Fitz-Gibbon.** 2002. Using homolog groups to create a whole-genomic tree of free-living organisms: an update. *J. Mol. Evol.* **54:**539–547.
59. Reference deleted.
60. **Huelsenbeck, J. P.** 1995. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* **12:**843–849.
61. **Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback.** 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294:**2310–2314.
62. **Huson, D. H.** 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14:**68–73.
63. **Huson, D. H., and M. Steel.** 2004. Phylogenetic trees based on gene content. *Bioinformatics* **20:**2044–2049.
64. **Huynen, M., T. Dandekar, and P. Bork.** 1998. Differential genome analysis applied to the species-specific features of Helicobacter pylori. *FEBS Lett.* **426:**1–5.
65. **Huynen, M. A., and P. Bork.** 1998. Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* **95:**5849–5856.
66. **Jensen, R. A.** 2001. Orthologs and paralogs—we need to get it right. *Genome Biol.* **2:**interactions1002.1–interactions1002.3.
67. **Johnson, J. L.** 1991. DNA reassociation experiments, p. 21–40. *In* E. Stackebrandt and M. J. Goodfellow (ed.), *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons Ltd., Chichester, United Kingdom.
68. **Kabsch, W., and C. Sander.** 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:**2577–2637.
69. **Kanz, C., P. Aldebert, N. Althorpe, W. Baker, A. Baldwin, K. Bates, P. Browne, A. van den Broek, M. Castro, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, J. Gamble, F. G. Diez, N. Harte, T. Kulikova, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, F. Nardone, V. Silventoinen, S. Sobhany, P. Stoehr, M. A. Tuli, K. Tzouvara, R. Vaughan, D. Wu, W. Zhu, and R. Apweiler.** 2005. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **33:**D29–D33 .
70. **Karlin, S., G. M. Weinstock, and V. Brendel.** 1995. Bacterial classifications derived from RecA protein sequence comparisons. *J. Bacteriol.* **177:**6881–6893.
71. **Koizumi, Y., J. J. Kelly, T. Nakagawa, H. Urakawa, S. El-Fantroussi, S. Al-Muzaini, M. Fukui, Y. Urushigawa, and D. A. Stahl.** 2002. Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology. *Appl. Environ. Microbiol.* **68:**3215–3225.
72. **Kolaczkowski, B., and J. W. Thornton.** 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* **431:**980–984.
73. **Korbel, J. O., B. Snel, M. A. Huynen, and P. Bork.** 2002. SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* **18:**158–162.
74. **Kunin, V., D. Ahren, L. Goldovsky, P. Janssen, and C. A. Ouzounis.** 2005. Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic Acids Res.* **33:**616–621.
75. **Kwok, A. Y., and A. W. Chow.** 2003. Phylogenetic study of *Staphylococcus* and *Macrococcus* species based on partial *hsp60* gene sequences. *Int. J. Syst. Evol. Microbiol.* **53:**87–92.
76. **La Scola, B., Z. Zeaiter, A. Khamis, and D. Raoult.** 2003. Gene-sequence-based criteria for species definition in bacteriology: the *Bartonella* paradigm. *Trends Microbiol.* **11:**318–321.
77. **Lerat, E., V. Daubin, and N. A. Moran.** 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-Proteobacteria. *PLoS Biol.* **1:**E19.
78. **Lilburn, T. G., and G. M. Garrity.** 2004. Exploring prokaryotic taxonomy. *Int. J. Syst. Evol. Microbiol.* **54:**7–13.
79. **Lin, J., and M. Gerstein.** 2000. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. *Genome Res.* **10:**808–818.
80. **Lloyd, A. T., and P. M. Sharp.** 1993. Evolution of the *recA* gene and the molecular phylogeny of bacteria. *J. Mol. Evol.* **37:**399–407.
81. **Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K. H. Schleifer, and M. Wagner.** 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.
82. **Naser, S., F. L. Thompson, B. Hoste, D. Gevers, K. Vandemeulebroecke, I. Cleenwerck, C. C. Thompson,**

M. Vancanneyt, and J. Swings. 2005. Phylogeny and identification of enterococci by *atp*A gene sequence analysis. *J. Clin. Microbiol.* **43:**2224–2230.

83. **Naser, S. M., F. L. Thompson, B. Hoste, D. Gevers, P. Dawyndt, M. Vancanneyt, and J. Swings.** 2005. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpo*A and *phe*S genes. *Microbiology* **151:**2141–2150.
84. **Opperdoes, F. R.** 2003. Phylogenetic analysis using protein sequences, p. 207–235. *In* M. Salemi and A. M. Vandamme (ed.), *The Phylogenetic Handbook: a Practical Approach to DNA and Protein Phylogeny*. Cambridge University Press, Cambridge, United Kingdom.
85. **Payne, G. W., P. Vandamme, S. H. Morgan, J. J. Lipuma, T. Coenye, A. J. Weightman, T. H. Jones, and E. Mahenthiralingam.** 2005. Development of a *rec*A gene-based identification approach for the entire *Burkholderia* genus. *Appl. Environ. Microbiol.* **71:**3917–3927.
86. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:**817–818.
87. **Priest, F. G., M. Barker, L. W. Baillie, E. C. Holmes, and M. C. Maiden.** 2004. Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.* **186:**7959–7970.
88. **Rasko, D. A., M. R. Altherr, C. S. Han, and J. Ravel.** 2005. Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev.* **29:**303–329.
89. **Rokas, A., and S. B. Carroll.** 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol. Biol. Evol.* **22:**1337–1344.
90. **Rost, B.** 1999. Twilight zone of protein sequence alignments. *Protein Eng.* **12:**85–94.
91. **Salemi, M., and A. M. Vandamme (ed.).** 2003. *The Phylogenetic Handbook: a Practical Approach to DNA and Protein Phylogeny*. Cambridge University Press, Cambridge, United Kingdom.
92. **Santos, S. R., and H. Ochman.** 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ. Microbiol.* **6:**754–759.
93. **Sawa, G., J. Dicks, and I. N. Roberts.** 2003. Current approaches to whole genome phylogenetic analysis. *Brief Bioinform.* **4:**63–74.
94. **Schierup, M. H., and J. Hein.** 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156:**879–891.
95. **Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler.** 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18:**502–504.
96. **Simossis, V. A., and J. Heringa.** 2005. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33:**W289–W294.
97. **Snel, B., P. Bork, and M. A. Huynen.** 1999. Genome phylogeny based on gene content. *Nat. Genet.* **21:**108–110.
98. **Soler, L., M. A. Yanez, M. R. Chacon, M. G. Aguilera-Arreola, V. Catalan, M. J. Figueras, and A. J. Martinez-Murcia.** 2004. Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *Int. J. Syst. Evol. Microbiol.* **54:**1511–1519.
99. **Soltis, P. S., and D. E. Soltis.** 2003. Applying the bootstrap in phylogeny reconstruction. *Stat. Sci.* **18:**256–267.
100. **Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. Grimont, P. Kampfer, M. C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward, and W. B. Whitman.** 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **52:**1043–1047.
101. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44:**846–849.
102. **Stevens, J. R., and C. J. Schofield.** 2003. Phylogenetics and sequence analysis—some problems for the unwary. *Trends Parasitol.* **19:**582–588.
103. **Stralis-Pavese, N., A. Sessitsch, A. Weilharter, T. Reichenauer, J. Riesing, J. Csontos, J. C. Murrell, and L. Bodrossy.** 2004. Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ. Microbiol.* **6:**347–363.
104. **Tanaoka, J.** 1994. Determination of the DNA base composition, p. 463–470. *In* M. Goodfellow and A. G. J. O'Donnell (ed.), *Chemical Methods in Prokaryotic Systematics*. John Wiley & Sons Ltd., Chichester, United Kingdom.
105. **Tekaia, F., A. Lazcano, and B. Dujon.** 1999. The genomic tree as revealed from whole proteome comparisons. *Genome Res.* **9:**550–557.
106. **Thompson, C. C., F. L. Thompson, K. Vandemeulebroecke, B. Hoste, P. Dawyndt, and J. Swings.** 2004. Use of *rec*A as an alternative phylogenetic marker in the family *Vibrionaceae*. *Int. J. Syst. Evol. Microbiol.* **54:**919–924.
107. **Thompson, F. L., D. Gevers, C. C. Thompson, P. Dawyndt, S. Naser, B. Hoste, C. B. Munn, and J. Swings.** 2005. Phylogeny and molecular identification of vibrios based on multilocus sequence analysis. *Appl. Environ. Microbiol.* **71:**5107–5115.
108. **Thompson, F. L., T. Iida, and J. Swings.** 2004. Biodiversity of vibrios. *Microbiol. Mol. Biol. Rev.* **68:**403–431.
109. **Tortoli, E.** 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin. Microbiol. Rev.* **16:**319–354.
110. **van Berkum, P., Z. Terefework, L. Paulin, S. Suomalainen, K. Lindström, and B. D. Eardly.** 2003. Discordant phylogenies within the *rrn* loci of rhizobia. *J. Bacteriol.* **185:**2988–2998.
111. **Vancanneyt, M., P. Vandamme, and K. Kersters.** 1995. Differentiation of *Bordetella pertussis*, *B. parapertussis*, and *B. bronchiseptica* by whole-cell protein electrophoresis and fatty-acid analysis. *Int. J. Syst. Bacteriol.* **45:**843–847.
112. **Vandamme, P., C. S. Harrington, K. Jalava, and S. L. On.** 2000. Misidentifying helicobacters: the *Helicobacter cinaedi* example. *J. Clin. Microbiol.* **38:**2261–2266.
113. **Vandamme, P., B. Pot, M. Gillis, P. Devos, K. Kersters, and J. Swings.** 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60:**407–438.
114. **Vandamme, P. A. R.** 2003. Taxonomy and classification of bacteria, p. 271–285. *In* P. R. Murray, E. J. Baron, J. H. Jorgensen, M. A. Pfaller, and R. H. Yolken (ed.), *Manual of Clinical Microbiology*, 8th ed. ASM Press, Washington, D.C.
115. **Waddell, P. J., H. Kishino, and R. Ota.** 2000. Rapid evaluation of the phylogenetic congruence of sequence data using likelihood ratio tests. *Mol. Biol. Evol.* **17:**1988–1992.
116. **Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Truper.** 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37:**463–464.
117. **Weisburg, W. G., S. J. Giovannoni, and C. R. Woese.** 1989. The *Deinococcus-Thermus* phylum and the effect of

rRNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* **11:**128–134.

118. **Wernersson, R., and A. G. Pedersen.** 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31:**3537–3539.
119. **Whelan, S., P. Lio, and N. Goldman.** 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* **17:**262–272.
120. **Willems, A., E. Falsen, B. Pot, E. Jantzen, B. Hoste, P. Vandamme, M. Gillis, K. Kersters, and J. De Ley.** 1990. *Acidovorax*, a new genus for *Pseudomonas facilis*, *Pseudomonas delafieldii*, E. Falsen (EF) group 13, EF group 16, and several clinical isolates, with the species *Acidovorax facilis* comb. nov., *Acidovorax delafieldii* comb. nov., and *Acidovorax temperans* sp. nov. *Int. J. Syst. Bacteriol.* **40:**384–398.
121. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51:**221–271.
122. **Woese, C. R.** 1991. The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria, p. 1–24. *In* R. K. Selander, A. G. Clark, and T. S. Whittam (ed.), *Evolution at the Molecular Level*. Sinauer Associates, Inc., Sunderland, Mass.
123. **Woese, C. R., E. Stackebrandt, and W. Ludwig.** 1985. What are mycoplasmas—the relationship of tempo and mode in bacterial evolution. *J. Mol. Evol.* **21:**305–316.
124. **Wolf, Y. I., S. E. Brenner, P. A. Bash, and E. V. Koonin.** 1999. Distribution of protein folds in the three superkingdoms of life. *Genome Res.* **9:**17–26.
125. **Wolf, Y. I., I. B. Rogozin, N. V. Grishin, and E. V. Koonin.** 2002. Genome trees and the tree of life. *Trends Genet.* **18:**472–479.
126. **Wuyts, J., G. Perriere, and Y. Van De Peer.** 2004. The European ribosomal RNA database. *Nucleic Acids Res.* **32:**D101–D103.
127. **Yamamoto, S., H. Kasai, D. L. Arnold, R. W. Jackson, A. Vivian, and S. Harayama.** 2000. Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyr*B and *rpo*D genes. *Microbiology* **146:**2385–2394.
128. **Yang, Q., and M. Blanchette.** 2004. StructMiner: a tool for alignment and detection of conserved secondary structure. *Genome Inform. Ser. Workshop Genome Inform.* **15:**102–111.
129. **Yang, S., R. F. Doolittle, and P. E. Bourne.** 2005. Phylogeny determined by protein domain content. *Proc. Natl. Acad. Sci. USA* **102:**373–378.
130. **Ye, R. W., T. Wang, L. Bedzyk, and K. M. Croker.** 2001. Applications of DNA microarrays in microbial systems. *J. Microbiol. Methods* **47:**257–272.
131. **Zeigler, D. R.** 2003. Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.* **53:**1893–1900.
132. **Zhang, H., and X. Gu.** 14 November 2004, posting date. Maximum likelihood for genome phylogeny on gene content. *Stat. Appl. Genet. Mol. Biol.* **3:**no. 1, article 31. [Online.] http://www.bepress.com/sagmb/vol3/iss1/art31.

Bioreporters, Biosensors, and Microprobes

ROBERT S. BURLAGE

14

The study of microbial communities is dependent on the ability to discern individual bacterial species and microbial activities in a complex matrix. Small variations in constituents can determine whether a microorganism occupies a specific niche or whether a specific gene is active. Therefore, the tools that are used to study microbial environments must be very sensitive and the scale on which they measure might be very small indeed, perhaps the area surrounding a single microorganism. For this purpose a variety of analytical devices have been produced; they can be conveniently organized into bioreporters, microprobes, and biosensors. All of these technologies have been used in other areas of biology, notably in cell biology, but this chapter concentrates on their uses and potential uses for environmental microbiology. The danger in doing this is that a promising technology may be omitted because it is not currently used for environmental microbiology, which is certainly likely in such a fast-moving field. However, there is also an opportunity to create new tools that are relevant to specific problems in environmental microbiology.

This chapter stresses techniques that are significantly different from other molecular techniques in that they are suitable for complex environments inhabited by a diverse collection of bacteria. Other molecular assays are valuable for pure-culture work, and they certainly have great utility in environmental microbiology, but environmental microbiologists need assays that have other attributes. The assays should be nondestructive of sample material. Heterogeneity of environmental samples may be very significant, and therefore assays that examine the same sample are very valuable. The assays should be real-time procedures, since conditions can change quickly within mixed communities. In addition, the assays should be able to report conditions continuously rather than at selected assay times. Fluctuations within complex communities may be very slight yet may still be significant in understanding the ecology of the system.

Biosensors of environmental conditions and bioreporter genes of genetic activity have been developed to augment our ability to detect, identify, and quantify microorganisms in complex ecological settings and to analyze their microniches. Biosensors are devices that fuse two technologies, electronic and biological, into a unique analytical tool. Bioreporters are genes that produce a product that is easily assayed and that relates to the genetic activity of the host cell. These techniques are likely to become more generally applied as the technology matures and as commercial opportunities arise. The field has now grown so large that a comprehensive examination of the techniques is not possible here. Reference is made to relevant review articles for readers seeking a more comprehensive understanding of the technology.

BIOREPORTER TECHNOLOGY

The analysis of genetic expression is made more difficult by the lack of suitable assays for most gene products. Bioreporter genes play a surrogate role, supplying an assayable gene product when an assay for the gene product of interest is not available or very difficult to perform. Bioreporter genes have been used extensively with pure cultures to demonstrate the expression of specific genes. However, many of the most useful bioreporter genes, such as *lacZ* (encoding the β-galactosidase enzyme) and *xylE* (encoding catechol 2,3-oxygenase), are usually unsuitable for ecological studies. While any single bioreporter gene might not be present in a particular species, the presence of the gene in other members of a community is not unlikely, potentially creating a significant background problem. This can be true in a monoculture as well, such as when a *lacZ* bioreporter is used with a lac^+ strain. To circumvent this problem, researchers have used bioreporter genes with no homologs in the host strain, such as the chitobiase gene (38). This gene, derived from a deep-sea *Vibrio harveyi* strain, catalyzes a reaction that gives a colorimetric response on agar plates and therefore can be easily detected. Background is essentially absent in *Escherichia coli*, although the use of this gene in other species has been limited so far. Another interesting example is the ice nucleation gene, *inaZ*. This gene was isolated from the plant pathogen *Pseudomonas syringae* and allows ice droplet formation at higher temperatures. It is assayed by determining the temperature at which a droplet containing the strain freezes. If it freezes at a higher temperature, then the gene product is present and gene expression has occurred. This gene has been used in several experiments, often in concert with other bioreporters (1, 5, 48, 49).

Another drawback of conventional bioreporters is that performing a biochemical assay often requires the destruction of a sample of the community. Nondestructive assays allow repeated experiments to be performed on the same sample, so that changes (development) of a microbial com-

munity can be observed. Biochemical assays for conventional bioreporters require time for sampling, processing, and signal development, and these steps introduce a significant time delay between the microbial event and the analysis of the data. There is a distinct advantage in having an assay that gives real-time results.

While there are several bioreporter genes that might be used, bioreporters that make use of light for bioreporting have significant advantages. The use of either bioluminescent or fluorescent bioreporters is now an established technology, and the uses have expanded greatly over the years. They both represent nondestructive, noninvasive means of detecting gene expression. Light can be measured with great sensitivity and precision, allowing the detection of a single cell under some conditions and with appropriate light-gathering equipment (76). Bioreporters for bioluminescent and fluorescent gene products are described below.

Bioluminescence

Bioluminescence is the production of visible light by a biochemical process. Unlike most chemical reactions, which produce heat as the main by-product, these reactions also generate enough light to be detected by conventional photodetectors. This phenomenon is easily observed in fireflies (*Photinus pyralis*), although many other species are capable of producing light. The genes for these light-producing reactions have been isolated by researchers, and they are now available on cloning vectors. When genetically fused (using either a transcriptional or protein fusion) to appropriate genes from a host bacterium, these strains produce light under defined conditions. The cloned firefly luciferase gene (*luc*) has been used to observe gene expression in animals (68), plants (42, 59), and bacteria (60). Firefly luciferase is a powerful tool for genetic analysis, although it is difficult to use in microbial ecology experiments since the reaction requires the substrate luciferin, which must be added exogenously. Often these assays utilize extracts from samples, rather than whole cells, and therefore destructive sampling is required.

Several genera and species of bioluminescent bacteria are known, although *V. harveyi* and *V. fischeri* have received the most attention. These bacteria contain *lux* genes, which are responsible for bioluminescence. The *lux* operon is a complex pathway of five genes, *luxCDABE*, and efficient expression of all of these genes in the host is required for appropriate functioning of the bioreporter. Only two genes, *luxA* and *luxB*, encoding the heterodimeric luciferase enzyme, are needed for the actual bioluminescent reaction. The *luxCDE* genes have been implicated in the recycling of the required aldehyde substrate, so that a pool of substrate is continuously available (Fig. 1). The *lux* genes, comprising about 7 kbp of DNA, have been cloned and sequenced and are available for analysis of microbial consortia. Several excellent reviews describe the genetics and physiology of bacterial bioluminescence (46, 47) and the use of these fusions (11).

Advantages and Disadvantages

The advantages of bioluminescent bioreporters lie primarily in the relative ease of light measurement. Light can be measured accurately and with great sensitivity. Since light radiates in all directions from a point source, light detection can be performed in three dimensions, giving a more sophisticated analysis of an object's position in space. It can be measured quickly (in real time) and without perturbing or destroying the sample. For instance, the light detector can be introduced into the sample and left there for an extended period, or it can detect light that passes through the glass wall of a bioreactor vessel. Bacterial interactions can thus be examined in real time, for example, to study predator-prey or symbiotic relationships. There is usually no need to add any substrates or reagents for the bioluminescence assay (although the requirement for an aldehyde substrate is discussed below). In most consortia of interest to environmental microbiologists, bioluminescence is a rare trait, and therefore a background problem is unlikely as long as ambient light can be excluded from the reaction vessel.

The *lux* genes are especially useful if a qualitative analysis is sufficient, i.e., determining whether light is being produced at a given time. For these experiments, light output can be expressed as relative light units. Relative light units are acceptable when all work is performed with the same light-measuring apparatus, under identical conditions (distance from detector, temperature, composition of the vessel holding the sample), and where no comparisons are drawn to other published results. If quantification is desired, photodetectors can be calibrated (see below).

The advantages described above must be weighed against several disadvantages that are inherent to the bioluminescent bioreporters. The bioluminescent reaction requires molecular oxygen, without which the bioreporter is inoperative. Oxygen limitation results in a lower intensity of light generated, perhaps to a level that would escape detection. The physiological state of the cell can thus have a major impact on the strength of the bioluminescent signal: cells respiring at a high rate will quickly exhaust their available oxygen and leave little for the bioluminescence reaction. This is easily seen by swirling a rapidly growing culture and watching the light intensity increase as the culture medium becomes aerated. In addition, if the host strain does not have a suitable aldehyde substrate for the bioluminescent reaction, an aldehyde must be added exogenously. Few strains make sufficient aldehyde for prolonged light production, and strains utilizing *luxAB* vectors must always have an aldehyde supplement in the medium. The substrate, usually *n*-decanal at a final concentration of 0.1 to 1.0% (vol/vol), penetrates

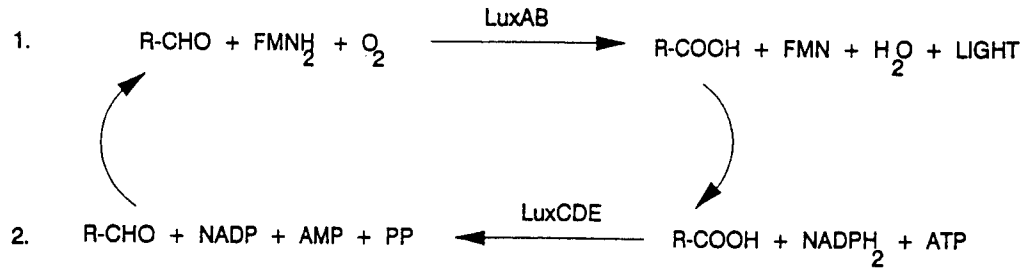

FIGURE 1 Genes and chemical intermediates involved in the bioluminescence reaction of *Vibrio fischeri*.

the cells readily, although it can be toxic at relatively low doses. Finally, the luciferase enzyme of *V. fischeri* is heat labile and is not recommended for use above 30°C. However, the luciferase from *V. harveyi* is stable at 37°C. An excellent discussion of the weaknesses of bacterial bioreporters and the means by which these techniques may be improved is available (83).

Applications

Many genetic constructions involving the *lux* genes have been prepared, most of which utilize the genes from a *Vibrio* species. Some are now available commercially. The intact *luxCDABE* cassette is available on a plasmid cloning vector, pUCD615, so that expression of bioluminescence can be placed under the control of a host promoter (70). Use of the full cassette ensures that the aldehyde substrate will be regenerated for continuous availability, although the host cell must have a suitable long-chain aldehyde (e.g., decanal) present as substrate for the luciferase reaction. The presence of such an aldehyde can be determined only empirically, and fluctuations of the aldehyde concentration are always possible. If such fluctuations are suspected, *n*-decanal can be added at the concentrations listed above to ensure an adequate supply.

The inclusion of the full cassette (i.e., the *luxA* and *luxB* luciferase and the *luxCDE* genes) obviates the need for exogenously added aldehyde, at least in strains that have a suitable aldehyde substrate to start with. The luciferase enzyme uses molecular oxygen to convert the aldehyde substrate to a carboxylic acid, with the resulting light being a by-product of the reaction. The action of the *luxCDE* genes is to recycle the carboxylic acid product of the light-producing reaction, giving a continuous supply of aldehyde. It has been observed that, at peak light emissions, the light intensity of such a strain can be boosted by the addition of aldehyde, although not greatly. This suggests that there is always an aldehyde limitation in the cell.

Alternatively, a construction utilizing only the *luxA* and *luxB* genes can be used if the aldehyde substrate is added exogenously. The *luxA* and *luxB* genes have been fused into a single open reading frame by Escher et al. (26), although the resulting *luxAB* fusion luciferase is more temperature sensitive than the native enzyme. Transposons carrying the *lux* genes are also available, such as the Tn*4431* transposon (75). This construct carries the intact *luxCDABE* cassette and a gene for antibiotic selection of transposon insertion. This construct is a valuable means of generating transposon mutants quickly. The *luxAB* genes are available on a Tn5 derivative (9) and on a mini-Tn5 derivative (23). There have been many applications of this technology in a variety of bacterial species. A *Salmonella* strain was used for genotoxicity assays (80) with the inducible *umu* gene promoter. A lambda phage was genetically engineered to carry the bioluminescence genes into *E. coli* cells in environmental settings, increasing the detection limits for this important pathogen (67). Hassler and Twiss (31) modified a *Synechococcus* strain to serve as a bioluminescent reporter of iron availability, another important environmental factor. The yeast *Saccharomyces cerevisiae* has been modified with *lux* genes to create a novel bioreporter of estrogenic compounds (71). Novel applications extend the trend of miniaturization of technology, which is also evident in the biosensors. Nivens et al. (56) combined bioreporter bacteria with a miniaturized optics detection device to make an autonomous detector. This is an important development since the environmental applications of bioreporters are dependent on inexpensive, independently monitored systems. There is little commercial appeal for a system that requires a visit from a human operator for every data point. Coupling this technology with a communications capability and the appropriate data storage and retrieval system would make it applicable to many field environments.

Light Measurement

Unlike many assays, in which a standard procedure is used to describe the results, light can be measured by a variety of means. Visualization of bacterial colonies may be sufficient for screening of clones during genetic construction, although the observer must be in a darkened room, and this method is not at all quantitative. Photographic film can be exposed to the light being emitted from colonies, although this technique is usually cumbersome. Several types of electronic equipment are suitable for the measurement of light. ATP photometers or luminometers, which are used for measurement of ATP concentrations by the luciferase assay, are common in laboratories. Liquid scintillation counters are also common. A liquid scintillation counter must be very sensitive in order to detect photons resulting from radioactive decay, and so these counters make good photodetectors for bioluminescence, although the coincidence channel should be disconnected prior to use. The coincidence channel eliminates background during its measurement of radiation, but it is a hindrance for bioluminescence work since light emanating from a single cell might not be detected by both photodetectors simultaneously. These methods are sensitive but are not designed specifically for bioluminescence work. Accordingly, there are problems in introducing representative samples to the photodetectors as well as in determining incubation conditions for the samples. That is, the samples would have to fit inside ordinary scintillation vials, which might not provide adequate aeration or mixing.

Commercial photomultipliers (e.g., those from Oriel, Stratford, Conn.) are recommended for remote sampling of light, including bioreactors and soil microcosms. These usually include flexible fiber-optic cables, which have a high efficiency of light transmittance, an important feature in measuring low levels of light. For extremely low levels of light, such as would be expected from single bacterial cells, charge-coupled devices (CCD) can be used (e.g., those from Hamamatsu, Hamamatsu City, Japan). The added sensitivity is reflected in the increased cost of this equipment, and few laboratories have access to one. A CCD can be used, however, to visualize signals that are seen through a microscope, and thus it has the potential to describe the physiological response of single cells, although integration of weak signals can delay output for several minutes. Accordingly, samples that move or drift during the integration time give a blurred image, if the image is detected at all.

The lack of standardization is a major shortcoming of bioluminescent reporter work, and its greatest impact is on the quantification of results. The output of a bioluminescent strain must be expressed in terms of specific activity to allow comparisons between laboratories. Units of light production would ideally be expressed as photons (quanta) of light per minute per milligram of total protein. However, each photodetector system has a different efficiency of light detection (for instance, different sensitivities for different wavelengths), as well as a different detector window geometry. A method to standardize photodetectors using a light-producing biochemical reaction has been described (58), and this method should be applied more generally. Calibration of a photodetector using a standard light source is possible, although the equipment is expensive and not generally available in laboratories.

Fluorescence

There are now several fluorescent proteins that can be used as bioreporters in bacterial cells, but the first successful one was green fluorescent protein (GFP). GFP addresses many of the disadvantages of the bioluminescent bioreporters, including the following: (i) the bioluminescence reaction requires oxygen and is unreliable under conditions of reduced oxygen tension (although GFP does require small amounts of oxygen for proper chromophore formation); (ii) bioluminescence requires functioning of the luciferase enzyme, which requires correct synthesis and folding of the protein; (iii) the luciferase enzyme requires a substrate, which might not be available in the cell; and (iv) the luciferase from *V. fischeri* is heat labile and is useless at 37°C. Increased stability at high temperatures would be an asset for all luciferases. GFP has the advantage that it is measured on the basis of its intrinsic properties and not on the basis of its biological activity in a certain milieu. Generally speaking, the newer versions of GFP and the new fluorescent proteins all have these attributes, and researchers typically choose a fluorescent protein based on its fluorescence characteristics (excitation and emission wavelengths).

The original GFP was a huge asset to the biological sciences, not only for microbiology but especially for cell biology. The ability to discern gene expression in a three-dimensional system, with great specificity, has been a boon for developmental biology (82). The success of GFP inspired a search for other fluorescent proteins, and now there are several commercially available proteins (Table 1). Typically these products include the purified protein, expression vectors for several systems (plant, animal, and bacterial), and perhaps antibodies that are specific for the fluorescent protein. The original GFP, often referred to as avGFP (for its origin in *Aequorea victoria*), must now compete with a variety of other fluorescent products that have been isolated from many sources, particularly deep-sea corals. Fluorescent colors are usually blue, green, yellow, and red, and different manufacturers tout the excitation and emission spectra of their products, particularly in regard to their ability to be used in tandem. Besides fluorescent characteristics, other attributes are desirable. The DsRed gene, from the deep-sea coral *Discosoma*, was originally thought to have great potential as a fluorescent marker comparable to GFP, but in most biological systems the protein folded so slowly that it had little practical value. Recent work with mutations of DsRed have yielded variants that are much improved, with a folding time reduced to less than 1 h (7), which is a fraction of the time for the wild-type protein.

The GFP gene has been cloned and sequenced, and the protein has been extensively characterized (61, 63). The protein that is synthesized from the GFP gene autocyclizes (16), producing a chromophore that is brightly fluorescent (Fig. 2). When the GFP gene is expressed in a cell (either prokaryotic or eukaryotic), it fluoresces a bright green after cyclization of the chromophore (13). The fluorescence makes the cell easy to detect with UV light (excitation, 395 nm) and conventional light-gathering equipment.

Advantages and Disadvantages

As with the measurement of bioluminescence, fluorescence can be measured accurately and with great sensitivity. Detection is dependent on the ability of the researcher to expose the GFP molecule to the excitation wavelength, and this can be performed with flexible fiber-optic cables introduced into a microbial ecosystem. Measurement is rapid, and there is no need to add any substrates or reagents. The problems of sample perturbation and destruction are therefore avoided.

Fluorescence of GFP is very bright, and individual bacterial cells can easily be seen by epifluorescence microscopy. GFP appears to be very slow in forming the chromophore (typically taking several hours), and the speed at which it forms seems to vary with different organisms and different growth conditions, although a comprehensive analysis of this phenomenon is lacking. The protein is extremely stable and is largely unaffected by treatment with detergents, proteases, glutaraldehyde, or organic solvents. It is also very stable over a pH range of 6 to 12 and at high (65°C) temperatures. GFP may be useful in genetic analysis of ther-

TABLE 1 Commercially available fluorescent proteins

Name	Supplier	Excitation and emission wavelengths (nm)	Color	Source organism
Aequorea victoria green fluorescent protein (GFP)	Amersham	395, 509	Green	Jellyfish
Ptilosarcus green fluorescent protein	Lux Biotechnology	485, 508	Green	Sea pen
Renilla green fluorescent protein	Lux Biotechnology	485, 506	Green	Sea pansy
Ptilosarcus green fluorescent protein	Nanolight Technology	NA[a]	Green	Sea pen
Renilla green fluorescent protein	Nanolight Technology	NA	Green	Sea pansy
Anthomedusae JRed	Evrogen	584, 610	Red	Jellyfish
Pantellina TurboGFP	Evrogen	482, 502	Green	Plankton
Heteractis HcRed	Evrogen	590, 637	Red	Sea anemone
Phialidium PhiYFP	Evrogen	525, 537	Yellow	Jellyfish
Anemonia KFP-Red	Evrogen	580, 600	Red	Sea anemone
Discosoma DsRed	Clontech	556, 586	Red	Coral
Aequorea corulescens green fluorescent protein	Clontech	475, 505	Green	Jellyfish
Anthozoa AmCyan	Clontech	458, 489	Blue	Coral
Anthozoa ZsYellow	Clontech	529, 539	Yellow	Coral
Heteractis HcRed	Clontech	588, 618	Red	Coral

[a]NA, not applicable.

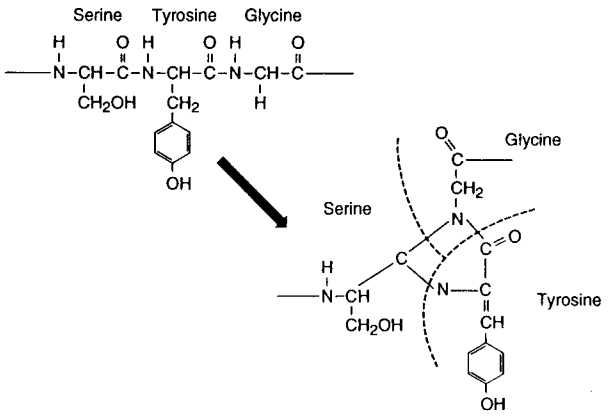

FIGURE 2 The chromophore of *Aequorea* GFP. Amino acids 65, 66, and 67 of GFP form a cyclical structure by an autocatalytic reaction. This chromophore is the source of the bright fluorescence seen with this protein. The dotted lines delineate the separate amino acids in the chromophore.

mophiles and other extremophiles. Once the protein is made, it does not degrade quickly in the cell, and therefore assays of the dynamics of gene expression, such as have been performed with the *lux* genes, are not possible with wild-type GFP. Andersen et al. (3) were able to attach a peptide tag to the carboxy terminus of GFP. This extra peptide made the mutant GFP susceptible to bacterial housekeeping proteases that recognize the carboxy ends of proteins, which resulted in a GFP with a reduced half-life, in this case ranging from 40 min to several hours, depending on the strain and the growth conditions. Because of this development, it is possible to use the bright fluorescence of GFP as an unparalleled bioreporter of real-time gene expression. Many other mutations have been introduced into GFP, producing useful variants. While wild-type GFP tends to form inclusion bodies, which limit the amount of fluorescence seen, GFP mutants have been developed that avoid this problem and which result in greatly amplified fluorescence (17).

Formation of the GFP chromophore requires molecular oxygen, although not in great amounts, and therefore is unsuitable for use under completely anaerobic conditions. Hansen et al. (30) demonstrated that GFP could be formed and detected in biofilms when dissolved oxygen was present at 0.1 ppm. Only the most oxygen-intolerant microorganisms would be unable to live under these conditions. No fluorescence is seen when cells are grown in an anaerobic environment, although once the chromophore is formed it continues to fluoresce in an anaerobic environment. The presence of GFP in bacteria does not appear to have deleterious effects on the host, although a comprehensive analysis has not been performed. Interestingly, individual molecules of GFP do not remain fluorescent at all times, but instead appear to "blink"; that is, they stop fluorescing for a few seconds and then start again (24). In any aggregate this is undetectable since the vast majority of molecules are fluorescing, but it does suggest other possible uses for GFP on a nanoscale.

Applications

Use of GFP is still a relatively new technique, and the construction of convenient cloning vectors is continuing. However, the number of applications of GFP is impressive, guaranteeing that more vectors will soon become available. These vectors typically contain the GFP gene within a polylinker region, allowing convenient manipulation of the gene for transcriptional fusions. The intact GFP gene has been inserted into a derivative of Tn*5*, and therefore random mutations with GFP are possible (12). This transposon, Tn*5GFP1*, can be introduced into a variety of gram-negative species by electroporation.

Mutations have been introduced into the GFP gene to produce fluorescent signals with altered properties. The red-shifted GFP was isolated in this manner (22). The name refers to the shift of the excitation wavelength toward the red end of the spectrum. The protein fluoresces at approximately the same wavelength (the maximum is at 505 nm instead of 510 nm) but excites at 490-nm instead of 395 nm. This shift is expected to be helpful, since the 490-nm excitation wavelength is beyond the wavelengths of excitation for cellular-protein fluorescence (due to their aromatic amino acids). A mutant GFP developed by Heim et al. (32) results in the production of a blue color instead of a green color. Multiple site mutations introduced by Anderson et al. (4) had the effect of increasing the brightness of the fluorescence. These mutants also had an altered excitation profile, allowing both mutant genes to be used in the same cell. Although they were initially tested only in eukaryotic cells, it is reasonable to expect that they will work in prokaryotes as well. Novel uses have been described for the detection of nitrates in soil systems (19) and for arsenic detection (85). In a clever adaptation of the technology, this latter group was able to detect single bioreporter cells as they responded to arsenic in a water sample, demonstrating the strength of the GFP fluorescence.

Fluorescence Measurement

Bacterial colonies expressing GFP can be easily detected on exposure to a UV light. This is easily accomplished with the UV source that is used in most molecular biology laboratories to visualize DNA in agarose gels, although an inexpensive hand-held UV light works just as well. Fluorescent bacteria can also be easily seen using epifluorescence microscopy. An appropriate filter set should be used; the filter for fluorescein detection has proved to be very useful for this purpose. A xenon or mercury lamp can be used as a source of UV excitation. For discrimination of their mutant GFPs in a fluorescence-activated cell sorter, Anderson et al. (4) used a krypton ion laser for one variant and an argon ion laser for the other. Since both proteins produced the same emission wavelength, the same filter and detector could be used.

Fluorescence spectrometry facilitates the detection of GFP fluorescence. Fluorescence spectrometers vary in sensitivity and versatility, although in general they should be able to detect GFP expression in bacteria. Quantification of bacteria in the sample is possible when a standard is examined contemporaneously. Digital imaging spectroscopy (28, 89) is an excellent means of detecting and characterizing fluorescent signals, although the expense of the system makes it unavailable to all but a few researchers.

MICROPROBES AND BIOSENSORS

Microprobes

A microprobe is a device that measures a specific physical or chemical property in a microenvironment. For instance, microprobes can be devised to test for pH, temperature, or the concentration of ionic species (72). The quality that makes these probes different from other probes is their small

size, which makes them more suitable for analysis at submillimeter resolution. The development of microprobes for the examination of microbial environments has proceeded rapidly thanks to innovative construction techniques. Microprobes have been described for ammonium (20), nitrate (36, 79), oxygen (65), denitrification (by nitrous oxide production) (15), and sulfate reduction (64). Conventional technology is used to measure the analyte; the critical development is the miniaturization of the electrode. The probe tip can be in the range of 1 to 10 μm in diameter, although 10 to 50 μm is more common. The signal is usually reported as a current on an ammeter. Delicate construction of the probe is required, as is the manipulation of the probe in three dimensions with a micromanipulator and dissecting microscope. Microprobes are especially useful to study the microecology of biofilms, including formation and activity at various depths inside the biofilm. An excellent review is available (66). The main obstacle to the general use of microprobes is that there are a limited number of commercially available microprobes (with a few notable exceptions, such as Microelectrodes [Bedford, N.H.], Unisense [Aarhus, Denmark), Abtech [Richmond, Va.], and Microprobe [Gaithersburg, Md.]), and the microprobe often must be handmade by the researcher. Techniques for construction are available, although they require skill and patience. As the usefulness of microprobes becomes better appreciated, the number of manufacturers will certainly increase.

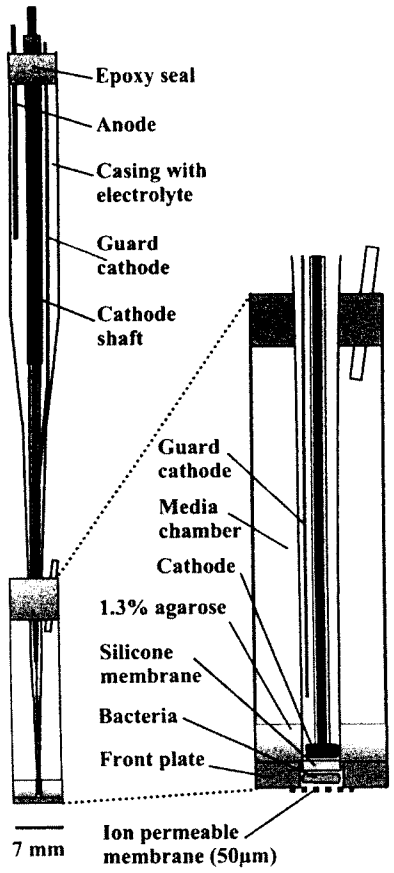

FIGURE 3 Cross-section of a typical microprobe. Reprinted from reference 55 with permission from the publisher.

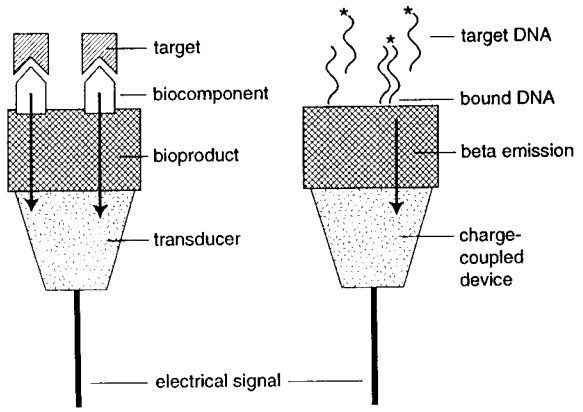

FIGURE 4 Biosensors. (Left) A generalized scheme for a biosensor. Interaction of the target analyte with the biological component results in a signal, which is transmitted to the transducer. The transducer senses the signal and converts it to an electrical signal. (Right) A DNA biosensor. The hybridization event brings the labeled DNA in contact with the transducer, a CCD camera. The CCD camera detects beta emission from ^{32}P decay and converts it to an electrical signal.

The oxygen microprobe described by Revsbech (65) provides a good example of the current microprobe technology. This microprobe has a tip that is approximately 10 μm in diameter and is sensitive to oxygen concentrations in the micromolar range. It incorporates a guard cathode that removes oxygen diffusing toward the sensor tip from the electrolyte solution, permitting stable signal acquisition from the sample. This microprobe is suitable for examination of biofilm ecology or aquatic microbiology, with the ability to discern microbial processes at the water interface. Although it is extraordinarily small for an analytical instrument, its presence is likely to disturb or influence the surrounding environment, however slightly. The prudent researcher will be attentive to possible effects from the use of these tools. Figure 3 shows a cross-section of a typical microprobe.

Biosensors

A biosensor is a type of probe in which a biological component, such as an enzyme, antibody, or nucleic acid, interacts with an analyte, which is then detected by an electronic component and translated into a measurable (electronic) signal (Fig. 4). Biosensor probes are possible because of a fusion of two technologies: microelectronics and biotechnology. Their greatest impact is in the clinical area, in which rapid test results are needed (86). However, they are also applicable to environmental analysis, and in recent years several substantial improvements have allowed them to be used more generally. A review of the application of biosensors for environmental study is available (69). There are several components to a biosensor. Different researchers may contribute to the development or optimization of a particular part or may incorporate several different technologies into one tool. For convenience, these components can be summarized as follows: the biological component (or biomolecule), the attachment method, the microfluidics, the computational component, and the electronic (sensing) component, which is also called the transducer (27).

The biological component is the molecule that interacts with the analyte of interest. A variety of substances can be used as the biological component, including nucleic acids,

proteins (particularly antibodies and enzymes), lectins (plant proteins that bind sugar moieties), and complex materials (organelles, tissue slices, and microorganisms). In each case, it is the specificity of the biological component for an analyte (or group of related analytes) that makes the biomolecule attractive for use in sensing technology. For example, a single strand of DNA hybridizes only to its complementary strand under the appropriate conditions. The conditions of the assay are very important, especially where reversibility of binding is a key factor. This is apparent in antibody-based biosensors, in which the affinity of the antibody for the antigen affects the sensitivity of the biosensor. This is particularly true for the measurement of dissociation constants of antibodies, which may vary substantially.

In most cases, isolation of the biological component is necessary to ensure that only the molecule of interest is bound or immobilized on the electronic component. In some cases this is easy, such as the isolation of DNA. It is possible to have a commercial vendor make specific oligonucleotides that are pure and that are already labeled. Antibody or enzyme extraction and purification are much more complex procedures, although crude extracts can sometimes be used. The stability of the biological component is also critical, since it is being used outside of its usual biological environment. A labile protein usually makes a poor candidate for a biosensor. Koblizek et al. (40) isolated particles from *Synechococcus elongatus* that contain the photosystem II enzymes. These particles were then trapped on an oxygen electrode with a dialysis membrane. If the enzymes are working normally, oxygen is generated and is detected by the electrode. In the presence of inhibitors of the enzymes (the authors used certain herbicides), oxygen generation is diminished and the effect becomes quantifiable. Although this device is neither field-hardy nor particularly long-lived, it is a clever example of the use of a natural product to create a reliable tool.

While it is possible for these biomolecules to be free in solution in order to function, it is more common to bind them to a surface with a known location, and therefore the attachment of the biomolecule is important. This has numerous benefits. The precise location of the biomolecules will be known. This is especially important when constructing arrays of sensors, in which a different biomolecule is attached at a discrete address on a surface and can be interrogated individually by the electronic component. And, of course, having the biomolecule tethered to a site allows other reagents to be introduced by using fluid transfer technology. The objective is to bind the biomolecule in place without disrupting its biological activity. That is, the enzyme must still catalyze a reaction, the antibody must bind its antigen, and the nucleic acid must allow hybridization with its complementary strand. A number of attachment protocols have been described, particularly for the attachment to glass or silica particles. This may involve a silane cross-bridge to which biomolecules can be attached. The need to deposit organic substances on transducers in a predictable manner has been addressed by the work of Decher et al. (21). This group demonstrated the sequential construction of layers on typical transducer surfaces, such as glass, silicon wafers, and quartz. This technique should improve biosensors by increasing the uniformity of results. Amino-derivatized oligonucleotides can be attached to glass (SiO_2) surfaces such as fiber-optic cables, glass beads, or microscope slides though covalent binding with a chemical linker. Some techniques result in the nonspecific attachment of oligonucleotides to the surface, which is an impediment to the hybridization that is required for the detection by the transducer. Losses of efficiency such as this can be avoided through careful attachment of the oligonucleotide via a modified 3′ or 5′ end of the oligomer. Graham et al. (29) used a simple procedure to attach oligomers to an evanescent-wave biosensor array. Maskos and Southern (45) described the synthesis of oligonucleotides on derivatized glass bead supports. Glycol spacers of various dimensions can be added between the oligonucleotide and the glass support, which facilitates hybridization. Other methods include immobilization within carbon paste or polymers and stabilization within hydrogels or sol-gels (90). The streptavidin-biotin interaction is often used. Ultrathin applications of biological material are usually deposited on transducers by the Langmuir-Blodgett (8) or molecular self-assembly (51) techniques.

Fortunately, attachment protocols are usually uncomplicated and very reliable. Quantification of the material on a surface, or the density of the material, is another problem entirely. A number of techniques have been developed to detect mass on a surface, including total internal reflectance fluorescence, quartz crystal vibration analysis, and optical reflectometry. These techniques often take advantage of the fluorescent properties of proteins (enzymes or antibodies) for quantification. In some cases it is possible to use an intact microbial cell as the biomolecule (77), which avoids the problem of orientation but does necessitate the creation of a suitable microenvironment.

The target of the biosensor must be brought to the biomolecule by some method. Typically this is done by suspending the target in an aqueous solution and flooding the biomolecule area with it. This is easy enough to do if the volume required is in the microliter or greater range; micropipettors do an excellent job of delivering fluids. However, the trend in biosensors has been to miniaturize, both to save on space and weight and to examine very small volumes. The sample size can therefore be very small, and waste concerns are minimal. To produce a biosensor of this size, a means of transferring fluids is needed. This is the technology of microfluidics, which is very important to the design of biosensors because fluids at very low volumes are difficult to move in a uniform manner.

The computational component is the hardware and software that interprets and reports on the signal received from the biosensor; as such, it is beyond the scope of this chapter. With some systems there is merely a digital readout, while in other systems (such as DNA microarrays) the amount of information and the need to address the data points make computer control essential.

The final component is the transducer, which is also the component that is the most unfamiliar to microbiologists. Essentially, this is a source of energy which is directed at the biomolecule and which is changed by the biomolecule in some way. This change is detected electronically and reported. Since energy is supplied to the system, it is typical to consider a type of wave that is introduced (light, electromagnetic, or sound). Generally, the transducers fall into distinct categories: electrochemical, optical, piezoelectric, and calorimetric (74).

Electrochemical transducers report changes in voltage when the current is held constant (potentiometric) or report changes in current when the voltage is held constant (amperometric). These are by far the most common electrochemical transducers, although transducers based on conductance and capacitance have also been described (74). In each case, the interaction of the analyte with the biological component causes a change in potential that is detected by the sensor.

Optical biosensors use a fiber-optic probe to receive specific wavelengths of light. The versatility of fiber-optics is due to their capacity to transmit signals that report on changes in wavelength, wave propagation time, intensity, distribution of the spectrum, or polarity of the light. In general, the signal is acquired from these devices through flexible cables which can transmit light to the biological component (such as an excitation light for fluorescence) and receive light back from the sample (such as for light generation by the sample or light absorption or reflection). Light conductance can be accomplished with great efficiency (less than 1% loss over short distances), and loss of signal is usually not a problem. Acquisition of the light signal by the detector can be a considerable problem if the light source is very weak, and the difference in light signals may be small. Light propagation over longer distances (>1 m) is usually accompanied by a loss of conductance efficiency unless laser light is used.

One of the techniques which has been more fully developed, and which now has been commercialized, is the use of evanescent-wave excitation. This is dependent on an unusual principle of physics and requires some explanation to make it relevant to the biologist. When a beam of light strikes an interface between two transparent interfaces (e.g., glass and water) the light beam is reflected off the surface and refracted through the new medium. Light can be directed through a waveguide, such as a fiber-optic cable. If the angle of incidence of the light is at the critical angle, the light is refracted at 90° from the normal, essentially following the interface of the two media. If the light enters the waveguide at an angle greater than the critical angle, it undergoes total internal reflection. This creates an evanescent wave at the interface between the two media. An evanescent wave is essentially an electromagnetic wave that decays exponentially with distance from its source. Thus, they are very weak effects and are significant only within a small distance from the waveguide. This is actually a very handy attribute, since biosensor materials attached to the waveguide are affected by the evanescent wave while other materials only a short distance away are not. Typically, the affected distance is on the order of nanometers, which is optimal for most biomolecules. An excellent review is available (81).

A related technology is surface plasmon resonance (SPR). In this technique, the surface of the waveguide is coated with a thin layer of gold (43). When light hits the gold at a certain angle, there is a decrease in the reflected energy due to the creation of an evanescent wave and its interaction with surface plasmons. Plasmons are another difficult physical principle. They are quasiparticles resulting from quantization of plasma oscillations. They can be created by reflecting a photon off a thin metal film (Fig. 5). The reflected light shows an energy loss equal to integral multiples of the plasmon energy. The decrease in reflected light is measured with a CCD camera and is related to the quantity of matter that interacts on the biomolecular side of the instrument. A particularly lucid example is presented by Mullett et al. (52) and is recommended for those new to the field. Miura et al. (50) used SPR and a tethered antibody to measure benzo[*a*]pyrene. They were able to reuse the sensor multiple times. Later models incorporated a dual-chamber system to provide a reference electrode (54). This enables faster, more accurate detection of analytes. Nelson et al. (53) constructed a DNA hybridization SPR platform and demonstrated that it can be used to detect 16S rRNA specifically. The Biacore company has successfully commercialized SPR technology. A variety of biomolecules can be immobilized to the sensor surface. Introduction of target molecules leads to binding (e.g., antibody-antigen, chelator-metal, and DNA-DNA). Signal strength is determined by a CCD camera and is correlated with binding and concentration. The Biacore X is a manual model that should be suitable for many environmental experiments.

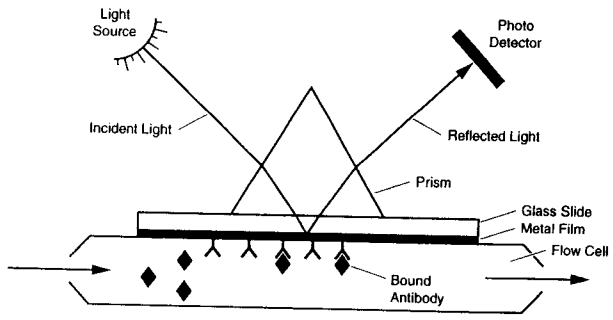

FIGURE 5 The principle of surface plasmon resonance. The sensing surface is on the opposite side of the metal film from the illuminated surface. Here an antibody-antigen-type biosensor is shown, with the sensing surface incorporated into a flow cell. The light source can be either polarized or laser light or an electron stream. The photodetector must be able to record subtle changes in light intensity.

A type of optical sensor that is related to SPR is the resonant-mirror biosensor. Instead of a metal layer as the sensing surface, a material with a high refractive index [such as titanium (IV) oxide] is used. This material is overlaid with a medium of low refractive index (such as silica), which is then connected to one side of a prism. When light enters the prism, it is totally reflected from the sensing surface. The low-index layer is typically so thin that the light may reach the high-index layer through an evanescent field, although this is dependent on the correct angle of incident light and phase matching of the resonant modes of the high-index material. When light enters the high-index medium, it propagates for a short distance before exiting through the prism. This makes the angle of the reflected light very susceptible to changes at the surface of the high-index (sensing) layer. Instead of changes in light intensity, as seen with SPR, modifications to the sensing surface are seen as phase changes and are recorded as changes in arc seconds of the reflected light. A thorough description of this technology is available (18). The resonant-mirror biosensor is quite sensitive (comparable to SPR) but is dependent on having the right materials for the sensing layer. Therefore, it is important to have a reliable attachment protocol for your molecule of interest. Just as for SPR, immunoassays have been performed with resonant-mirror biosensors (10). Complex materials can also be conjugated to the surface, and binding events can be studied in near real time (33, 34).

Piezoelectric biosensors measure changes in mass. A piezoelectric material, such as quartz crystal, oscillates at a certain frequency when a potential is applied across its surface. If the mass at the surface changes because, for example, a bound antibody complexes with a specific antigen, the frequency of oscillation will change, and this change is detectable. Lu et al. (44) used this method to detect an enzyme, glutathione *S*-transferase, at concentrations as low as 0.2 mg/ml. It is expected that sensitivity will continue to improve with other

technological developments. The surface acoustic wave biosensor depends on the propagation of an acoustic wave through a medium. The medium is attached to a piezoelectric detector. If the energy of the wave is confined near the surface, the velocity of the wave and the effects detected by the piezoelectric detector are influenced by the mass the wave encounters. Therefore, a change of mass such as antibody-antigen binding is directly measurable.

Calorimetric transducers are comparable in function to optical transducers, except that heat generation is measured instead of light. Microcalorimeters can measure very small fluctuations in temperature and are sensitive enough to detect heats of molecular interaction (such as ligand binding), substrate use by microorganisms, and responses of immune cells to antigens. The microcantilever has great potential to sensitively and selectively detect compounds of interest. The technology is based on a mechanical stress principle. The reactive end of the device resembles a tiny diving board, a thin metal (e.g., gold) platform to which biomolecules can be attached. Typically the metal platform is less than 1 μm thick, and approximately 100 μm long. When a target molecule binds to the attached biomolecule, a mechanical stress occurs in the thin metal. A laser is focused on the free end of the metal platform, where the deformation is greatest. Deviation of the reflection is determined by reflection of the laser onto a sensitive photodetector. Alternatively, the reaction can be detected as a change in resonance frequency. Although the instruments were initially rather large devices, in recent years they have been reduced to a convenient size. The number of possible applications is very great, since anything that causes a change in the metal can be measured. In addition to measuring changes in mass, this technique is very accurate at temperature detection, and thus possibly can be used to indicate exothermic and endothermic reactions. The use of microcantilevers for temperature sensing has the potential to detect changes of as little as 10^{-6}°C. Fast response times are possible (e.g., within seconds), although care must be taken to avoid temperature fluctuations due to external conditions. Dual-material metal platforms (e.g., aluminum and silicon) have also been used. Analytes that interact differently with the two materials will deform the platform in a measurable way, allowing the interaction to be detected and quantified. A biosensor of heavy metals has been described that utilizes metal-binding proteins as the biomolecule (14). The latest models are very compact and will probably be commercialized in the near future.

Applications

The essence of the biosensor is in matching the appropriate biological and electronic components to produce a relevant signal during analysis. For example, antibodies can be attached to a piezoelectric transducer so that the binding of the antigen is recorded as a change in the attached mass. Alternatively, the antibodies can be attached to an optical fiber and the antigen binding can be recorded by evanescent-wave detection (8, 52). This technique is particularly suitable for immunoassays because the evanescent waveform is operational for only a short distance from the surface of the optical electrode, and this distance is approximately equal to the size of the immune complex (74). Immunosensors that use other detection systems have been described previously (84).

Biosensors need not be overly sophisticated: bacterial cells can be immobilized on the tip of an oxygen electrode (78) or within an online column (57). Occasionally, whole yeast cells (6) can be used. The electrode then senses activity on the basis of the change in oxygen concentration. With this type of biosensor, it is important to design a reference standard that distinguishes background activity from the activity of interest; otherwise, specificity for the analyte is lost. Both electrochemical and optical electrodes are very useful for the detection of signals from attached enzymes. The enzymatic reaction can cause a potential change that is detected by the electrochemical electrode or can cause a change in one of the components of the enzyme system that can be detected by the optical electrode. Scheper et al. (73) used the latter concept for their biosensor. They attached the glucose-fructose oxidase from *Zymomonas mobilis* to a fiber-optic cable attached to a fluorimeter. The enzyme complex contained bound $NADP^+$. When the enzyme oxidizes glucose, it reduces $NADP^+$ to NADPH, which is a fluorescent molecule. The change in fluorescence is therefore proportional to the concentration of glucose. There is also a critical weakness associated with this assay, in that it is dependent on the availability of $NADP^+$. When the supply is exhausted, the biosensor no longer functions. Resupplying the biosensor with an essential cofactor is often technically challenging but is required for long-term monitoring of the environment. A clever modification of the optical sensor was reported by Zhou et al. (91). They used conventional fluorescent antibodies to find either labeled microspheres or bacterial cells, but they then used a standing acoustic wave to concentrate the signal in one known area. This results in an increase in signal because very little is outside the range of the fluorescent sensor.

Nucleic acid biosensors depend on the ability of a single-stranded nucleic acid to hybridize with another fragment of DNA by complementary base pairing. The growing field of DNA microarray technology is testament to the strength of this type of biosensor. This same technology on a smaller scale can be used for detection of very specific nucleic acids. The biosensor described by Eggers et al. (25) integrates microelectronics, molecular biology, and computational science in an optical electrode format. Their device can detect hybridization and report on the spatial configuration of the hybridization signal on a glass surface (25) or a silicon wafer (41), to which the DNA probes are attached. Several different DNA oligomers can be attached to the optical electrode at different locations. The DNA on the biosensor is then hybridized to DNA that is free in solution. The free DNA must be labeled, usually with a fluorescent, luminescent, or radioisotope decay (^{32}P) signal. The signal is detected by a CCD camera, which is extremely sensitive. The computer identifies the location of the affected pixels and forms the signal into a recognizable array. Not only is this technology suitable for rapid DNA sequencing, but it is also applicable to the rapid detection of many different gene sequences from DNA extracted from a consortium. Because of the labeling requirement for standard microarrays, evanescent-wave technology may be more attractive. A nucleic acid biosensor that uses evanescent-wave technology has been described by Graham et al. (29). They used short fragments of nucleic acids that are small enough to reside within the field of the evanescent wave. They were able to detect fluorescein-labeled DNA hybridizing to their complementary immobilized probes in a flow cell. Fluorescence was monitored and reported as a change in the output voltage. As with many other technologies, biosensors are being miniaturized. This presents special challenges arising from the use of novel

materials and assembly techniques which are beyond the scope of this chapter. However, a good review is available (37).

EMERGING TECHNOLOGIES

Other technologies are now available whose use in environmental microbiology is speculative, although the opportunities are clearly apparent. Xu et al. (88) used silver nanoparticles in a study of membrane transport. These particles, 10 to 80 nm in diameter, were used to measure the porosity of membranes under various environmental conditions, as a surrogate for antibiotic uptake. Silver appears very bright under dark-field microscopy, allowing tracking of the particles through the cells. Won et al. (87) modified the surface of a commercial magnetic nanoparticle with a small peptide and a fluorescein tagant and were able to introduce these nanoparticles into HeLa cells. Although this is a eukaryotic cell, the application is intriguing. The nanoparticles could be pulled through the cell by means of a magnet, with the progress visualized using the fluorescent tag and confocal microscopy. The application was created to determine which host molecules will interact with the attached peptide. While these applications are somewhat limited, they demonstrate that many experiments might be possible if more versatile tagants were available.

Quantum dots (also called Qdots) are nanometer-sized crystalline clusters. Typically they are made from semiconductor materials (e.g., CdTe, InP, and PbSe) at such small scales (1 to 10 nm in diameter) that they could possibly be used with bacterial cells. The construction of a Qdot is very complex, with a metal core surrounded by an inorganic "shell" material (e.g., zinc sulfide). Organic polymers can be attached to the shell, allowing complexing with various biomolecules such as antibodies or enzymes. To date the bulk of Qdot work in the biosciences has been with eukaryotic cells. Their use in environmental microbiology remains speculative, although they possess certain advantages that may make them valuable for specific tasks.

Qdots are used as fluorescent tagants, in much the same way that fluorescein is used. However, the spectral properties of Qdots are dependent on the actual size of the particle, as well as its shape and composition. With an increase in size, the color of the fluorescence shifts in a predictable manner. At first this was a disadvantage, but with refined manufacturing processes that create standard-sized particles, it becomes a great advantage. This unusual property means that Qdots might be available in an enormous range of fluorescent colors, giving workers multiple tags for environmental uses.

Qdots have several additional advantages. They have long fluorescent lifetimes (10 to 50 ns), allowing a fluorimeter to avoid the excitation wavelength and thus resulting in a more defined signal. Excitation is performed with a standard light source, eliminating the need for specific wavelengths of light. They emit exceptionally bright fluorescence, and, as already noted, they are available in a range of colors. Generally speaking, the smaller the Qdot, the more blue it is. Their emission spectra are typically very narrow and symmetric, avoiding the tailing of fluorescence seen in other tagants. They are also very stable molecules, and therefore long-term experiments can be carried out using the same tagant. In addition, they do not "photobleach" as other fluorescent compounds are prone to do.

However, Qdots also have certain disadvantages that must be acknowledged. They are nonbiological and cannot replace fluorescent bioreporter genes such as the GFP gene. Therefore they cannot be used for gene expression in vivo. They are also fabricated from heavy metals that may have toxic properties and are inherently hydrophobic. Both of these disadvantages are normally overcome by conjugating organic molecules to the surface as a means to attach reactive groups such as antibodies or biotin molecules. These modifications result in a water-soluble particle that can be more easily handled and is nontoxic. There are several protocols for attachment, many of which resemble the protocols for attachment to glass surfaces or silica particles (above). Even though they are in the nanometer range, they are still very large compared to biomolecules.

Most of the work with Qdots has been with eukaryotic cells, since they are larger and more complex internally. Using a biomolecule conjugated to a Qdot allows the tagant to either bind to a surface receptor and tag the cell for a long period or enter the cell and permit tracking of the biomolecule's progress through the cell. Excellent reviews of these applications are available (2, 35). The worker must have access to a suitable fluorescence imaging system, however, which requires skilled use (62).

Qdots are now available commercially. The first reported use of Qdots in microorganisms was by Kloepfer et al. (39). They used a CdSe Qdot that was conjugated to either a lectin or human transferrin, and they demonstrated the applicability of this technique with several different genera. As sizes and spectral properties become more standardized, the number of uses for Qdots will increase. It is suggested that they would be ideal for biofilm studies in which three-dimensional imaging of specific cells becomes important.

SUMMARY

As predicted in the previous edition of this chapter, the technologies described here have advanced and are becoming more commonplace in environmental microbiology. This is especially true for the biosensors, and the growth in commercially available products has certainly helped in this area. With the continued emphasis on multidisciplinary approaches to research topics, the use of these technologies will be invaluable. In the area of fluorescent proteins, a spectrum of excitation and emission wavelengths might soon be available that would allow the use of several bioreporter genes in one species or the use of bioreporters to distinguish individual species in a community. Multigene analysis will have a substantial impact on the understanding of genetic control. In the area of biosensors, the trend toward miniaturization and commercialization will continue. It is expected that fieldable biosensors will have a great impact on biowarfare monitoring and long-term ecological studies.

REFERENCES

1. **Afendra, A. S., C. Vargas, J. J. Nieto, and C. Drainas.** 2004. Gene transfer and expression of recombinant proteins in moderately halophilic bacteria. *Methods Mol. Biol.* **267:**209–223.
2. **Alivisatos, A. P., W. Gu, and C. Larabell.** 2005. Quantum dots as cellular probes. *Annu. Rev. Biomed. Eng.* **7:**55–76.
3. **Andersen, J. B., C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin.** 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64:** 2240–2246.

4. **Anderson, M. T., I. M. Tjioe, M. C. Lorincz, D. R. Parks, L. A. Herzenberg, G. P. Nolan, and A. Herzenberg.** 1996. Simultaneous fluorescence activated cell sorter analysis of two distinct transcriptional elements within a single cell using engineered green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **93:**8508–8511.
5. **Arvanitis, N., C. Vargas, G. Tegos, A. Perysinakis, J. J. Nieto, A. Ventosa, and C. Drainas.** 1995. Development of a gene reporter system in moderately halophilic bacteria by employing the ice nucleation gene of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* **61:**3821–3825.
6. **Baronian, K. H. R.** 2004. The use of yeast and moulds as sensing elements in biosensors. *Biosens. Bioelectron.* **19:** 953–962.
7. **Bevis, B. J., and B. S. Glick.** 2002. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nat. Biotechnol.* **20:**83–87.
8. **Blodgett, K. B., and I. Langmuir.** 1937. Build-up films of barium stearate and their optical properties. *Phys. Rev.* **51:** 964–982.
9. **Boivin, R., F. P. Chalifour, and P. Dion.** 1988. Construction of a Tn5 derivative encoding bioluminescence and its introduction in *Pseudomonas*, *Agrobacterium*, and *Rhizobium*. *Mol. Gen. Genet.* **213:**50–55.
10. **Buckle, P. E., R. J. Davies, T. Kinning, D. Yeung, P. R. Edwards, and D. Pollard-Knight.** 1993. The resonant mirror: a novel optical sensor for direct sensing of bimolecular interactions. II. Applications. *Biosens. Bioelectron.* **8:** 355–363.
11. **Burlage, R. S., and C. Kuo.** 1994. Living biosensors for the management and manipulation of microbial consortia. *Annu. Rev. Microbiol.* **48:**291–309.
12. **Burlage, R. S., Z. K. Yang, and T. Mehlhorn.** 1996. A transposon for green fluorescent protein transcriptional fusions: application for bacterial transport experiments. *Gene* **173:** 53–58.
13. **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. *Science* **263:**802–805.
14. **Cherian, S., R. K. Gupta, B. C. Mullin, and T. Thundat.** 2003. Detection of heavy metal ions using protein-functionalized microcantilever sensors. *Biosens. Bioelectron.* **19:**411–416.
15. **Christensen, P. D., L. P. Nielsen, N. P. Revsbech, and J. Sorensen.** 1989. Microzonation of denitrification activity in stream sediments as studied with a combined oxygen and nitrous oxide microsensor. *Appl. Environ. Microbiol.* **55:**1234–1241.
16. **Cody, C. W., D. C. Prasher, W. M. Westler, F. G. Prendergast, and W. W. Ward.** 1993. Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* **32:**1212–1218.
17. **Crameri, A., E. A. Whitehorn, E. Tate, and W. P. C. Stemmer.** 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14:**315–319.
18. **Cush, R., J. M. Cronin, W. J. Stewart, C. H. Maule, J. Molloy, and N. J. Goddard.** 1993. The resonant mirror: a novel optical sensor for direct sensing of biomolecular interactions. I: Principle of operation and associated instrumentation. *Biosens. Bioelectron.* **8:**347–353.
19. **DeAngelis, K. M., P. Ji, M. K. Firestone, and S. E. Lindow.** 2005. Two novel bacterial biosensors for detection of nitrate availability in the rhizosphere. *Appl. Environ. Microbiol.* **71:**8537–8547.
20. **DeBeer, D., and J. C. Van den Heuvel.** 1988. Response of ammonium-selective microelectrodes based on the neutral carrier nonactin. *Talanta* **35:**728–730.
21. **Decher, G., B. Lehr, K. Lowack, Y. Lvov, and J. Schmitt.** 1994. New nanocomposite films for biosensors: layer-by-layer adsorbed films of polyelectrolytes, proteins or DNA. *Biosens. Bioelectron.* **9:**677–684.
22. **Delagrave, S., R. E. Hawtin, C. M. Silva, M. M. Yang, and D. C. Youvan.** 1995. Red-shifted excitation mutants of the green fluorescent protein. *Bio/Technology* **13:**151–154.
23. **DeLorenzo, V., M. Herrero, U. Jakubzik, and K. N. Timmis.** 1990. Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative bacteria. *J. Bacteriol.* **172:**6568–6572.
24. **Dickson, R. M., A. B. Cubitt, R. Y. Tsien, and W. E. Moerner.** 1997. On/off blinking and switching behavior of single molecules of green fluorescent protein. *Nature* **388:** 355–358.
25. **Eggers, M., M. Hogan, R. K. Reich, J. Lamture, D. Ehrlich, M. Hollis, B. Kosicki, T. Powdrill, K. Beattie, S. Smith, R. Varma, R. Gangadharan, A. Mallik, B. Burke, and D. Wallace.** 1994. A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups. *BioTechniques* **17:**516–524.
26. **Escher, A., D. J. O'Kane, J. Lee, and A. A. Szalay.** 1989. Bacterial luciferase αβ fusion protein is fully active as a monomer and highly sensitive *in vivo* to elevated temperature. *Proc. Natl. Acad. Sci. USA* **86:**6528–6532.
27. **Gauglitz, G.** 2005. Direct optical sensors: principles and selected applications. *Anal. Bioanal. Chem.* **381:**141–155.
28. **Goldman, E. R., and D. C. Youvan.** 1992. An algorithmically optimized combinatorial library screened by digital imaging spectroscopy. *Bio/Technology* **10:**1557–1561.
29. **Graham, C. R., D. Leslie, and D. J. Squirrell.** 1992. Gene probe assays on a fibre-optic evanescent wave biosensor. *Biosens. Bioelectron.* **7:**487–493.
30. **Hansen, M. C., R. J. Palmer, C. Udsen, D. C. White, and S. Molin.** 2001. Assessment of GFP fluorescence in cells of *Streptococcus gordonii* under conditions of low pH and low oxygen concentration. *Microbiology* **147:**1383–1391.
31. **Hassler, C. S., and M. R. Twiss.** 2006. Bioavailability of iron sensed by a phytoplanktonic Fe-bioreporter. *Environ. Sci. Technol.* **40:**2544–2551.
32. **Heim, R., D. C. Prasher, and R. Y. Tsien.** 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **91:**12501–12504.
33. **Hirmo, S., E. Artursson, G. Puu, T. Wadstrom, and B. Nilsson.** 1998. Characteristics of *Helicobacter pylori* interactions with sialylglycoconjugates using a resonant mirror biosensor. *Anal. Biochem.* **257:**63–66.
34. **Hirmo, S., E. Artursson, G. Puu, T. Wadstrom, and B. Nilsson.** 1999. *Helicobacter pylori* interactions with human gastric mucin studied with a resonant mirror biosensor. *J. Microbiol. Methods* **37:**177–182.
35. **Hotz, C. Z.** 2005. Applications of quantum dots in biology. *Methods Mol. Biol.* **303:**1–17.
36. **Jensen, K., N. P. Revsbech, and L. P. Nielsen.** 1993. Microscale distribution of nitrification activity in sediment determined with a shielded microsensor for nitrate. *Appl. Environ. Microbiol.* **59:**3287–3296.
37. **Jianrong, C., M. Yuqing, H. Nongyue, W. Xiaohua, and L. Sijiao.** 2004. Nanotechnology and biosensors. *Biotechnol. Adv.* **22:**505–518.
38. **Kalabat, D. Y., J. M. Froelich, T. K. Phuong, R. A. Forsyth, V. G. Newman, and J. W. Zyskind.** 1998. Chitobiase, a new reporter enzyme. *BioTechniques* **25:**1030–1035.
39. **Kloepfer, J. A., R. E. Mielke, M. S. Wong, K. H. Nealson, G. Stucky, and J. L. Nadeau.** 2003. Quantum dots as strain- and metabolism-specific microbiological labels. *Appl. Environ. Microbiol.* **69:**4205–4213.

40. **Koblizek, M., J. Masojidek, J. Komend, T. Kucera, R. Pilloton, A. K. Mattoo, and M. T. Giardi.** 1998. A sensitive photosystem II-based biosensor for detection of a class of herbicides. *Biotechnol. Bioeng.* **60:**664–669.
41. **Lamture, J. B., K. L. Beattie, B. E. Burke, M. D. Eggers, D. J. Ehrlich, R. Fowler, M. A. Hollis, B. B. Kosicki, R. K. Reich, S. R. Smith, R. S. Varma, and M. E. Hogan.** 1994. Direct detection of nucleic acid hybridization on the surface of a charge coupled device. *Nucleic Acids Res.* **22:**2121–2125.
42. **Langridge, W. H. R., K.. J. Fitzgerald, C. Koncz, J. Schell, and A. A. Szalay.** 1989. Dual promoter of *Agrobacterium tumefaciens* mannopine synthase genes is regulated by plant growth hormones. *Proc. Natl. Acad. Sci. USA* **86:**3219–3223.
43. **Liedberg, B., C. Nylander, and I. Lundstrom.** 1983. Surface plasmon resonance for gas detection and biosensing. *Sens. Activators* **4:**299–304.
44. **Lu, H., H. Chen, Y. Lin, and J. Lin.** 2000. A reusable and specific protein A-coated piezoelectric biosensor for flow injection immunoassay. *Biotechnol. Prog.* **16:**116–124.
45. **Maskos, U., and E. M. Southern.** 1992. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ. *Nucleic Acids Res.* **20:**1679–1684.
46. **Meighen, E. A.** 1991. Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* **55:**123–142.
47. **Meighen, E. A.** 1994. Genetics of bacterial bioluminescence. *Annu. Rev. Genet.* **28:**117–139.
48. **Miller, W. G., M. T. Brandl, B. Quinones, and S. E. Lindow.** 2001. Biological sensor for sucrose availability: relative sensitivities of various reporter genes. *Appl. Environ. Microbiol.* **67:**1308–1317.
49. **Miller, W. G., J. H. Leveau, and S. E. Lindow.** 2000. Improved GFP and *inaZ* broad-host-range promoter-probe vectors. *Mol. Plant-Microbe Interact.* **13:**1243–1250.
50. **Miura, N., M. Sasaki, K. V. Gobi, C. Kataoka, and Y. Shoyama.** 2003. Highly sensitive and selective surface plasmon resonance sensor for detection of sub-ppb levels of benzo[a]pyrene by indirect competitive immunoreaction method. *Biosens. Bioelectron.* **18:**953–959.
51. **Moaz, R., L. Netzer, J. Gun, and J. Sagiv.** 1988. Self-assembling monolayers in the construction of planned supramolecular structures and as modifiers of surface properties. *J. Chim. Phys.* **85:**1059–1065.
52. **Mullett, W., E. P. C. Lai, and J. M. Leung.** 1998. Immunoassay of fumonisins by a surface plasmon resonance biosensor. *Anal. Biochem.* **258:**161–167.
53. **Nelson, B. P., M. R. Liles, K. B. Frederick, R. M. Corn, and R. M. Goodman.** 2002. Label-free detection of 16S ribosomal RNA hybridization on reusable DNA arrays using surface plasmon resonance imaging. *Environ. Microbiol.* **4:**735–743.
54. **Nenninger, G. G., J. B. Clendenning, C. E. Furlong, and S. S. Yee.** 1998. Reference-compensated biosensing using a dual channel surface plasmon resonance sensor system based on planar lightpipe configuration. *Sens. Actuators B* **51:**38–45.
55. **Nielsen, N., L. H. Larsen, M. S. M. Jetten, and N. P. Revsbech.** 2004. Bacterium-based NO_2^- biosensor for environmental applications. *Appl. Environ. Microbiol.* **70:**6551–6558.
56. **Nivens, D. E., T. E. McKnight, S. A. Moser, S. J. Osborn, M. L. Simpson, and G. S. Sayler.** 2004. Bioluminescent bioreporter integrated circuits: potentially small, rugged and inexpensive whole-cell biosensors for remote environmental monitoring. *J. Appl. Microbiol.* **96:**33–46.
57. **Nomura, Y., K. Ikebukaro, K. Yokoyama, T. Takeuchi, Y. Arikawa, S. Ohno, and I. Karube.** 1998. Application of a linear alkylbenzene sulfonate biosensor to river water monitoring. *Biosens. Bioelectron.* **13:**1047–1053.
58. **O'Kane, D. J., M. Ahmad, I. B. C. Matheson, and J. Lee.** 1986. Purification of bacterial luciferase by high-performance liquid chromatography. *Methods Enzymol.* **133:**109–128.
59. **Ow, D. W., K. V. Wood, M. DeLuca, J. R. deWet, D. R. Helinski, and S. H. Howell.** 1986. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* **234:**856–859.
60. **Palomares, A. J., M. A. DeLuca, and D. R. Helinski.** 1989. Firefly luciferase as a reporter enzyme for measuring gene expression in vegetative and symbiotic *Rhizobium meliloti* and other gram-negative bacteria. *Gene* **81:**55–64.
61. **Perozzo, M. A., K. B. Ward, R. B. Thompson, and W. W. Ward.** 1988. X-ray diffraction and time-resolved fluorescence analyses of *Aequorea* green fluorescent protein crystals. *J. Biol. Chem.* **263:**7713–7716.
62. **Pinaud, F., X. Michalet, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Iyer, and S. Weiss.** 2006. Advances in fluorescence imaging with quantum dot bio-probes. *Biomaterials* **27:**1679–1687.
63. **Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier.** 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111:**229–233.
64. **Ramsing, N. B., M. Kuhl, and B. B. Jorgensen.** 1993. Distribution of sulfate-reducing bacteria, O_2, and H_2S in photosynthetic biofilms determined by oligonucleotide probes and microelectrodes. *Appl. Environ. Microbiol.* **59:**3840–3849.
65. **Revsbech, N. P.** 1989. An oxygen microsensor with a guard cathode. *Limnol. Oceanogr.* **34:**474–478.
66. **Revsbech, N. P., and B. B. Jorgensen.** 1986. Microelectrodes: their use in microbial ecology. *Adv. Microb. Ecol.* **9:**293–352.
67. **Ripp, S., P. Jegier, M. Birmele, C. M. Johnson, K. A. Daumer, J. L. Garland, and G. S. Sayler.** 2006. Linking bacteriophage infection to quorum sensing signaling and bioluminescent bioreporter monitoring for direct detection of bacterial agents. *J. Appl. Microbiol.* **100:**488–499.
68. **Rodriguez, J. F., D. Rodriguez, J. Rodriguez, E. B. McGowan, and M. Esteban.** 1988. Expression of the firefly luciferase gene in vaccinia virus: a highly sensitive gene marker to follow virus dissemination in tissues of infected animals. *Proc. Natl. Acad. Sci. USA* **85:**1667–1671.
69. **Rogers, K. R.** 1995. Biosensors for environmental applications. *Biosens. Bioelectron.* **10:**533–541.
70. **Rogowsky, P. M., T. J. Close, J. A. Chimera, J. J. Shaw, and C. I. Kado.** 1987. Regulation of the *vir* genes of *Agrobacterium tumefaciens* plasmid pTiC58. *J. Bacteriol.* **169:**5101–5112.
71. **Sanseverino, J., R. K. Gupta, A. C. Layton, S. S. Patterson, S. A. Ripp, L. Saidak, M. L. Simpson, T. W. Schultz, and G. S. Sayler.** 2005. Use of *Saccharomyces cerevisiae* BLYES expressing bacterial bioluminescence for rapid, sensitive detection of estrogenic compounds. *Appl. Environ. Microbiol.* **71:**4455–4460.
72. **Santegoeds, C. M., A. Schramm, and D. deBeer.** 1998. Microsensors as a tool to determine chemical microgradients and bacterial activity in wastewater biofilms and flocs. *Biodegradation* **9:**159–167.
73. **Scheper, T., C. Muller, K. D. Anders, F. Eberhardt, F. Plotz, C. Schelp, O. Thordsen, and K. Schugerl.** 1994. Optical sensors for biotechnological applications. *Biosens. Bioelectron.* **9:**73–83.

74. **Sethi, R. S.** 1994. Transducer aspects of biosensors. *Biosens. Bioelectron.* **9:**243–264.
75. **Shaw, J. J., and C. I. Kado.** 1987. Direct analysis of the invasiveness of *Xanthomonas campestris* mutants generated by Tn4431, a transposon containing a promoterless luciferase cassette for monitoring gene expression, p. 57–60. *In* D. P. S. Verma and N. Brisson (ed.), *Molecular Genetics of Plant-Microbe Interactions*. Martinus Nijhoff, Dordrecht, The Netherlands.
76. **Silcock, D. J., R. N. Waterhouse, L. A. Glover, J. I. Prosser, and K. Killham.** 1992. Detection of a single genetically modified bacterial cell in soil by using charge coupled device-enhanced microbiology. *Appl. Environ. Microbiol.* **58:**2444–2448.
77. **Skladal, P., N. O. Morozova, and A. N. Reshetilov.** 2002. Amperometric biosensors for detection of phenol using chemically modified electrodes containing immobilized bacteria. *Biosens. Bioelectron.* **17:**867–873.
78. **Svitel, J., O. Curilla, and J. Tkac.** 1998. Microbial cell-based biosensor for sensing glucose, sucrose or lactose. *Biotechnol. Appl. Biochem.* **27:**153–158.
79. **Sweerts, J. R. and D. DeBeer.** 1989. Microelectrode measurements of nitrate gradients in the littoral and profundal sediments of a meso-eutrophic lake (Lake Vechten, The Netherlands). *Appl. Environ. Microbiol.* **55:**754–757.
80. **Taguchi, K., Y. Tanaka, T. Imaeda, M. Hirai, S. Mohri, M. Yamada, and Y. Inoue.** 2004. Development of a genotoxicity detection system using a biosensor. *Environ. Sci.* **11:**293–302.
81. **Taitt, C. R., G. P. Anderson, and F. S. Ligler.** 2005. Evanescent wave fluorescence biosensors. *Biosens. Bioelectron.* **20:**2470–2487.
82. **Tsien, R.** 1998. The green fluorescent protein. *Annu. Rev. Biochem.* **67:**509–544.
83. **Van der Meer, J. R., D. Tropel, and M. Jaspers.** 2004. Illuminating the detection chain of bacterial bioreporters. *Environ. Microbiol.* **6:**1005–1020.
84. **Vo-Dinh, T., M. J. Sepaniak, G. D. Griffin, and J. P. Alarie.** 1993. Immunosensors: principles and applications. *Immunomethods* **3:**85–92.
85. **Well, M., M. Gosch, R. Rigler, H. Harms, T. Lasser, and J. R. van der Meer.** 2005. Ultrasensitive reporter protein detection in genetically engineered bacteria. *Anal. Chem.* **77:**2683–2689.
86. **Wise, D. L., and L. B. Wingard (ed.).** 1991. *Biosensors with Fiberoptics*. Humana Press, Clifton, N.J.
87. **Won, J., M. Kim, Y. Yi, Y. H. Kim, N. Jung, and T. K. Kim.** 2005. A magnetic nanoprobe technology for detecting molecular interactions in live cells. *Science* **309:**121–125.
88. **Xu, X., W. J. Brownlow, S. V. Kyriacou, Q. Wan, and J. J. Viola.** 2004. Real-time probing of membrane transport in living microbial cells using single nanoparticle optics and living cell imaging. *Biochemistry* **43:**10400–10413.
89. **Youvan, D.C.** 1994. Imaging sequence space. *Nature* **369:**79–80.
90. **Yu, J., S. Liu, and H. Ju.** 2003. Mediator-free phenol sensor based on titania sol-gel encapsulation matrix for immobilization of tyrosinase by a vapor deposition method. *Biosens. Bioelectron.* **19:**509–514.
91. **Zhou, C., P. Pivarnik, A. G. Rand, and S. V. Letcher.** 1998. Acoustic standing wave enhancement of a fiber-optic *Salmonella* biosensor. *Biosens. Bioelectron.* **13:**495–500.