Methods in Microbiology

Volume 33
Functional Microbial Genomics

Edited by

Brendan Wren

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London

and

Nick Dorrell

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London

ACADEMIC PRESS

An imprint of Elsevier Science

Amsterdam Boston London New York Oxford Paris
San Diego San Francisco Singapore Sydney Tokyo

13 Cloning the Metagenome: Culture-independent Access to the Diversity and Functions of the Uncultivated Microbial World

Jo Handelsman, Mark Liles, David Mann, Christian Riesenfeld and Robert M Goodman
Department of Plant Pathology, University of Wisconsin, Madison, WI 53706, USA

◆◆

CONTENTS

◆◆◆◆◆◆ INTRODUCTION

For more than a century, the gold standard in microbiology has been to achieve pure cultures of microorganisms for laboratory study. The central role of cultivation in microbiology derives from the origins of the field, which were rooted in the study of disease etiology and the uses of microorganisms for processes related to making foods and beverages. The ecological study of microorganisms has advanced notably where a single microbial taxon performs an ecological role of major effect.

With the revelation over the past two decades of a vast diversity of previously unsuspected, and largely uncultivated, microflora inhabiting diverse natural environments throughout the biosphere, microbiology has undergone a revolution. The application of molecular biology methods

METHODS IN MICROBIOLOGY, VOLUME 33
ISBN 0–12–521533–9

first revealed convincing evidence for this new view of microbial diversity in the biosphere. Now molecular biology methods are providing powerful new approaches to address what now looms as one of the major challenges in microbiology, which is to understand the functions of microorganisms that have not been cultivated.

The use of molecular, culture-independent methods for assessing microbial diversity has shown that most microorganisms in the environment are not readily cultivated (Stahl *et al.*, 1985; Giovannoni *et al.*, 1990; Ward *et al.*, 1990; Pace, 1996; Suzuki *et al.*, 1997; Hugenholtz *et al.*, 1998), and that the diversity of the uncultivated majority (Whitman *et al.*, 1998) is vast (Stahl, 1993; Pace, 1996; Head *et al.*, 1998; Beja *et al.*, 2002b). For example, the readily cultivatable prokaryotes from soil represent 1% or less of the total (Griffiths *et al.*, 1996). Estimates indicate that soil contains more than 4000 species/g (Torsvik *et al.*, 1990, 1996), many of which fall into novel divisions of Bacteria and Archaea (Bintrim *et al.*, 1997; Borneman *et al.*, 1996). Similarly, sea water harbors diverse prokaryotes that defy cultivation attempts (Ward *et al.*, 1990; DeLong *et al.*, 1999). Analysis of both soil and sea water has revealed new groups of Archaea and an unanticipated abundance of Archaea, which may numerically dominate the open oceans (DeLong, 1992; DeLong *et al.*, 1994, 1999; McInerney *et al.*, 1995; Bintrim *et al.*, 1997; Simon *et al.*, 2000). The emerging image of the microbial world in natural environments is quite different from the one constructed by cultivation alone.

The challenge that long stumped microbiologists was to learn about the biology associated with the 16S rRNA sequences that indicate the high degree of diversity of uncultivated microorganisms. The solution lies, in part, in the application of genomics to assemblages of microorganisms using methods that circumvent the need for cultivation.

◆◆◆◆◆◆ OVERVIEW OF METAGENOMICS

Metagenomics is the genomic analysis of the collective genomes of an assemblage of organisms, or the 'metagenome'. The approach has also been termed 'environmental genomics' (Stein *et al.*, 1996; DeLong, unpublished) and 'community genomics' (Banfield, unpublished). In this chapter, we use the term 'metagenomics' to apply to all studies that fall within the definition, whether or not the original study used this term.

Metagenomic libraries containing DNA extracted directly from an environmental sample provide genomic sequences, and phylogenetic and functional information. Libraries can be screened for functions, and genomic sequence surrounding genes required for those functions can provide insight into the organism from which the function was derived. Phylogenetic markers such as 16S rRNA, *recA*, *radA* and genes encoding DNA polymerases or other conserved genes provide indications of the origin of the DNA (Schleper *et al.*, 1997; Eisen, 1998; Vergin *et al.*, 1998; Sandler *et al.*, 1999; Rondon *et al.*, 2000; Beja *et al.*, 2000b, 2002a). Linking phylogeny, sequence and functional analysis provides a multifaceted

approach to dissecting and reassembling the uncultivated microbial community.

Metagenomic libraries have been constructed from assemblages of microorganisms in seawater (Stein *et al.*, 1996; Cottrell *et al.*, 1999; Beja *et al.*, 2000a, b;), soil (Henne *et al.*, 1999, 2000; Rondon *et al.*, 2000; Entcheva *et al.*, 2001), and associated with a marine sponge (Schleper *et al.*, 1998). Metagenomic analysis has been applied to diverse problems in microbiology and has yielded insight into the physiology of uncultivated organisms as well as access to the potentially useful enzymes and secondary metabolites they produce (Table 13.1).

Table 13.1. Applications of metagenomic analysis

Characterization of phylogenetic diversity
Characterization of new genome organizations
Elucidation of new biochemical pathways for primary metabolism or energy transduction
Identification of reservoirs of resistance to environmental pollutants
Discovery of enzymes
Discovery of secondary metabolites and other biologically active small molecules
Discovery of polymers
Elucidation of trophic structure of microbial assemblages

◆◆◆◆◆◆ ISOLATION OF ENVIRONMENTAL DNA

The physical characteristics of a metagenomic library, such as insert size, number of clones and storage method, are dictated by the kind of natural products to be identified from the library. The method used to isolate metagenomic DNA from a natural environment must reflect the desired physical characteristics as well as the factors that will affect recovery of DNA from the targeted microorganism(s). For example, to identify lipases (or other enzymes) present within a microbial community, a library of DNA inserts less than 10 kb in size would be sufficient, but the number of clones must be sufficiently large to compensate for the relatively low frequency of clones that are active on indicator media (Henne *et al.*, 2000). For such small insert metagenomic libraries, harsh extraction methods (e.g. bead-beating) may be used to achieve genomic DNA from diverse microorganisms in soil or other natural environments (Krsek and Wellington, 1999).

The recovery of high molecular weight metagenomic DNA from natural environments can be a significant challenge, especially from soil or sediments. Metagenomic libraries containing large inserts from marine picoplankton have been prepared by lysing bacteria embedded in agarose (Stein *et al.*, 1996; Beja *et al.* 2000b), which is the same method that is commonly used to construct BAC libraries from eukaryotes or prokaryotes in pure culture (Wang and Warren, 1996; Rondon *et al.*, 1999b). The use of

agarose plugs to protect the DNA from shearing during cell lysis and restriction digestion facilitates construction of libraries containing large inserts, although we have found that cell lysis within agarose plugs containing extracted soil microbes is inefficient, at least from certain soils (unpublished data). The choice of method for DNA extraction will balance maintaining the DNA in large pieces with the diversity of the community to be represented in the library and the resistance of the targeted population to lysis (Krsek and Wellington, 1999; Stach *et al.*, 2001). We have found that multiple rounds of freezing and thawing combined with hot phenol extraction is effective for recovery of high molecular weight genomic DNA from diverse soil microorganisms (Zhou *et al.*, 1996; Rondon *et al.*, 2000). Efficiency of restriction digestion is enhanced by further purification of DNA by separation of high molecular weight DNA on, and recovery from, a preparative agarose gel. Typically, the largest genomic DNA recovered in this manner is less than 400 kb, and after restriction digestion the majority of restricted genomic DNA is sized less than 50 kb, but size selection can produce inserts as large as 80–155 kb (Beja *et al.*, 2000b; Rondon *et al.*, 2000). We are investigating cloning methods that do not require restriction digestion and methods for DNA recovery from cells physically separated from soil (Lindahl, 1996) to produce metagenomic libraries with inserts greater than 100 kb.

◆◆◆◆◆◆ VECTORS

Large-insert vectors

The design and choice of vectors for construction of metagenomic libraries is driven largely by the desired insert size and the relative importance of heterologous gene expression. For some analyses, isolation of the DNA is sufficient and gene expression is not required. Libraries constructed with DNA isolated from sea water (Stein *et al.*, 1996; Vergin *et al.*, 1998; Beja *et al.*, 2000b, 2002a) and soil (Rondon *et al.*, 2000) have provided a source of genomic DNA linked to 16S rRNA genes. The rRNA genes were identified by polymerase chain reaction (PCR) analysis, and the flanking DNA was analyzed by sequencing, thereby obviating the need for gene expression until after the gene of interest was identified. For these applications, large inserts are an advantage because larger fragments enhance the probability of the same clone containing a phylogenetic marker and a gene encoding an activity of interest (Stein *et al.*, 1996) and provide insights into genome organization. Large inserts are also required to capture large genes or operons encoding biosynthetic pathways for secondary metabolite or for genomic analysis that requires large segments of genomes (Handelsman *et al.*, 1998; Rondon *et al.*, 1999a; August *et al.*, 2000). These applications are facilitated by bacterial artificial chromosomes (BAC vectors), which were developed for use in eukaryotic genomics and can stably maintain inserts as large as 350 kb (Shizuya *et al.*, 1992). If medium-sized fragments are needed, then fosmids (Stein *et al.*, 1996) and cosmids (Brady and Clardy, 2000; Brady *et al.*, 2001) are

appropriate. An advantage of using these vectors is that the efficiency of cloning in fosmids and cosmids is typically 100–1000-fold greater than BACs, but they accept fragments up to 40 kb. BACs and cosmids are maintained in single copy in *E. coli*, which is an advantage when cloning genes that may be detrimental to the host cell when maintained in high copy. Genes for antibiotic synthesis have been cloned successfully from soil in the BAC vector, pBeloBAC11, and the cosmid, superCos1 (Brady and Clardy, 2000; Brady *et al.*, 2001; MacNeil *et al.*, 2001; Gillespie *et al.*, 2002).

While low copy number is an advantage for preventing cell death due to genes that are toxic in high copy and for stability of large inserts, it is a disadvantage for vector purification, insert sequencing and detection of products from poorly expressed genes. To combine the advantages of high and low copy plasmids, we developed a new vector, superBAC1 (Figure 13.1). This plasmid is maintained at low copy number in strain JW366, a genetically modified version of *E. coli* strain DH10B (Wild *et al.*, 1998, 1999) and can be induced with arabinose to 50–100 copies per cell. superBAC carries the pBeloBAC11 backbone with its *oriS* origin of replication as well as *oriV* from the IncP plasmid RK2. *oriV* requires the transcription factor TrfA, and the *trfA* gene is under control of an arabinose-inducible promoter in strain JW366. Therefore, replication of superBAC1 in the absence of arabinose is mediated by the *oriS* origin, which maintains the plasmid at one copy per cell. When arabinose is added to the medium, *trfA* is expressed and replication is mediated by *oriV*. As expression from the arabinose-inducible promoter can be modulated by titrating glucose and arabinose concentrations, the copy number can be modulated between 1 and 100 copies per cell. The low copy plasmid facilitates construction and maintenance of large insert libraries, while the ability to attain high copy has several advantages for sequence and expression analysis of the resulting libraries. The difficulties associated with the purification of large, low copy BACs for sequencing, subcloning and other manipulations is alleviated by the induction of high copy. In addition, activities undetected on a low copy BAC may be identified more readily when maintained at high copy. Increased levels of antimicrobial agent synthesis might be obtained if the genes encoding the pathway were maintained in high cell copy number, thereby facilitating detection of activity in a functional assay that might be below the level of detection if the genes were maintained at single copy. However, the clones can be grown in the absence of arabinose prior to screening, thereby reducing the likelihood of cell death due to high level production of a toxic product. If a copy of the clone bank is always maintained without arabinose, then self-poisoning by the copy of a clone that produces high levels of an antimicrobial agent under arabinose induction does not prevent further study of the clone.

Expression vectors

Although there is ample evidence that diverse genes from diverse organisms are expressed in *E. coli* (Ding and Yelton, 1993; Ferreya *et al.*,

1993; Black *et al.*, 1995; Chávez *et al.*, 1995), expression of some traits in metagenomic libraries requires transcription or translation machinery or chemical substrates that are lacking in *E. coli*. While the genetic code is nearly universal, transcription factors are specialized for organism and gene type, and translation is affected by organism-specific codon usage, which is co-ordinated with the availability of the cognate tRNAs in the cell. Moreover, biosynthesis of some secondary metabolites is dependent on products of primary metabolism that are not universal.

Expression vectors can be used to overcome the barrier of recognition of foreign promoters by *E. coli*'s transcription machinery (Newman and Fuqua, 1999). To avoid the detrimental effects of constitutive expression of toxic products, many expression vectors contain inducible promoters whose expression can be regulated by amendments to the medium. Expression vectors typically accept small fragments of DNA and the promoter in the vector will drive only genes adjacent to it and in the correct orientation. Cloning small fragments of DNA from environmental sources is appropriate when the goal is isolation of single genes or very small operons. Many expression vectors are available for cloning small fragments (<10 kb) that are maintained at low, medium or high copy number. For example, libraries constructed with soil DNA cloned into pBluescript SK(+) contain clones expressing lipase, esterase, and 4-hydroxybutyrate dehydrogenase activities (Henne *et al.*, 1999, 2000).

Shuttle vectors

Although expression vectors effectively address limitations on transcription, many other factors are required for expression of traits in a foreign host. Translation and post-translational processing must be faithfully conducted (Joseph *et al.*, 2001; Karlin *et al.*, 2001). To perform new biosynthetic functions, the substrates must be available in the cell. If the gene product must be secreted to be detected, then secretion signals must be recognizable to *E. coli*. Lack of co-ordination of any of these processes will result in no observable activity.

For certain pathways, expression of the authentic product is unlikely. For example, synthesis of polyketides in *E. coli* with genes derived from actinomycetes is unlikely because of the differences in promoters and codon usage and the lack of the appropriate building blocks (Cane *et al.*, 1998; Tang *et al.*, 2000; Pfeifer *et al.*, 2001). To provide a different cellular environment and, therefore, facilitate a different range of gene expression, an alternative approach is to deliver the cloned genes to a different host species. This requires a vector that can be mobilized from *E. coli* to another species.

To take advantage of potential host-specific abilities for heterologous expression of metagenomic DNA constructs, we have further modified superBAC1 to make it amenable to transfer and maintenance of BAC libraries in alternative host species. The inclusion of the origin of transfer, *oriT*, from RK2 on superBAC1 affords conjugal mating of the plasmid into alternative host species, including *Bacillus*, *Enterococcus*, *Staphylococcus*, *Streptococcus* and *Listeria* strains when the *tra* gene is supplied *in trans* (Guiney and Yakobsen, 1983; Yakobson and Guiney, 1984; Trieu-Cuot *et*

al., 1987). Conjugal transfer circumvents transformation, which occurs at low efficiency in many Gram-positive species. To maintain the transferred BACs in an alternative host, the inclusion of a replication origin that is functional in the chosen host species is required. Our initial shuttle vector is designed to transfer to *Bacillus* species, using *rep60* from *Bacillus* plasmid pTA1060 to mediate replication (Uozumi *et al.*, 1980) and a chloramphenicol-resistance gene from *Bacillus* plasmid pC194 for selection (Leonhardt and Alonso, 1988; Steinmetz and Richter, 1994).

Wang *et al.* constructed metagenomic libaries in an *E. coli–Streptomyces* shuttle vector and discovered an antibiotic biosynthetic pathway that was likely derived from a relative of *Streptomyces* (Wang *et al.*, 2000). Other replication origins that would expand the host range for BAC libraries include the broad host range origin host range pAMβ1, (3–5 copies per cell) (Trieu-Cuot *et al.*, 1987; Poyart and Trieu-Cuot, 1997) and the *repA* gene from pSH71 for streptococci (deVos, 1987; deVos *et al.*, 1997).

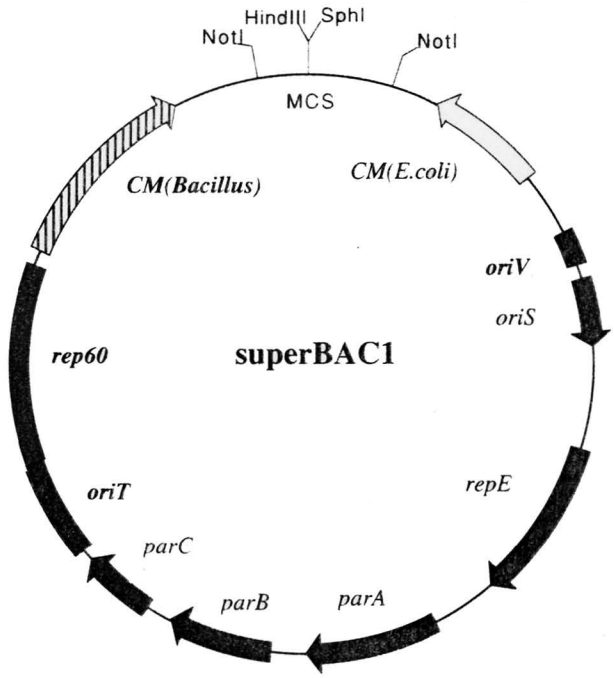

Figure 13.1. superBAC1: a shuttle vector with inducible copy number in *E. coli* strain JW366.

◆◆◆◆◆◆ HOST CELLS

The choice of the host cell in which to construct the library is guided in part by the requirements for transformation and stability of the vector. Choice of the host cell is also determined by the phenotypes of interest in the library. For functional screening or selections, it is essential that the host cell itself does not express the characteristic of interest.

◆◆◆◆◆◆ STORAGE AND CLONE MANAGEMENT

The intrinsic complexity of a metagenome, whether derived from a few species or from thousands, requires that metagenomic libraries contain many clones. The resulting storage and management challenges can be daunting if addressed with the standard techniques typically used for maintenance of genomic libraries of organisms in pure culture. Libraries are ideally maintained with each clone in pure culture in a separate well of a 96-well. The cells can be preserved with cryoprotectants, such as glycerol or dimethyl sulfoxide and stored at –80°C. This format is amenable to high throughput screening with manual or robotic replicators that are designed for the 96-well format. However, for study of complex microbial assemblages in typical laboratories that lack facilities to handle more than 10 000–20 000 clones, this format is impractical. For example, the soil is thought to contain at least 4000 different genomes of the size and complexity of that of *E. coli*. A metagenomic library containing clones with inserts of an average of 80 kb would require 250 000 clones to provide singlefold coverage of the metagenome. To obtain adequate representation of all sequences in the metagenome would require 1–10 million clones. Therefore, for many applications, maintaining libraries in pools is the only feasible approach for laboratories lacking extensive robotic capabilities.

◆◆◆◆◆◆ CHARACTERIZING METAGENOMIC LIBRARIES

Phylogenetic analysis

Once a library has been constructed, it is important to determine which microbial genomes are represented within the library. Phylogenetic information, in the form of 16S rRNA or other highly conserved gene sequences, reveal potential biases in the preparation or cloning of genomic DNA, and may suggest alternative hosts for functional screening. Although a small proportion of the clones in a library is likely to carry a rRNA operon, each one that does provides the opportunity to link phylogenetic and functional information, and through genomic sequence annotation and analysis permits a partial assessment of the metabolic capabilities for uncultured microorganisms (Stein *et al.*, 1996; Vergin *et al.*, 1998; Beja *et al.*, 2000a, b, 2002; Rondon *et al.*, 2000).

Identification of rRNA gene-containing BAC clones has been accomplished by several methods. Arrays of BAC clones have been probed with Archaea-specific probes (Stein *et al.*, 1996), and pools of BAC clones have been screened by PCR-based techniques using universal primer sets (Beja *et al*, 2000b; Rondon *et al.*, 2000). A principal challenge in the PCR amplification of rRNA genes from BAC library template is the contamination of host chromosomal DNA in each BAC preparation. The presence of non-*E. coli* rRNA genes within a pool of BAC clones may be demonstrated by

rRNA intergenic spacer analysis (Beja *et al.*, 2000b), denaturing gradient gel electrophoresis or restriction fragment polymorphism analysis (Rondon *et al.*, 2000). Furthermore, the amplification of host rRNA genes may be prevented by exonuclease digestion of chromosomal DNA (often only partially effective) and by incorporating host template-specific, terminally modified oligonucleotides ('terminator' oligos) within a PCR, allowing preferential amplification of non-host rRNA genes (Rondon *et al.*, 2000). When screening for rRNA genes from a pool of BAC clones representing an entire library, a combination of exonuclease treatment and incorporation of terminator oligos provides sufficient reduction of interference by *E. coli* rRNA genes. When screening large numbers of BAC clones for unique rRNA genes, division- or taxa-specific primer sets provide selectivity for particular phylogenetic groups, eliminating the problem of interference of the host cell's rRNA genes. The use of superBAC1, described above, which has inducible copy permits a greater ratio of BAC vector to host chromosomal DNA within BAC preparations.

Alternative phylogenetic markers can be used to assess the diversity of microorganisms that contributed DNA to a metagenomic library. Genes that have revealed phylogenetic relationships and unexpected diversity include genes involved in DNA repair (*recA* and *radA*), replication and photosynthesis (Schleper *et al.*, 1997; Sandler *et al.*, 1999; Beja *et al.*, 2002b). To identify evolutionarily conserved genes, a strategy must be developed so that the *E. coli* copies of the genes do not complicate the screen. In some cases, it is possible to use a functional complementation, taking advantage of *E. coli* mutations. Currently, we are screening metagenomic libraries for the presence of *recA* genes, which can be identified on the basis of sequence conservation (Eisen, 1998; Sandler *et al.*, 1999) as well as functional complementation of the *E. coli recA* gene, selecting BAC clones by their ability to confer resistance to ultraviolet radiation or chemical mutagenesis. Using a combination of markers provides a larger arsenal of tools that assess microbial diversity, potentially identifying BAC clones that might not be identified by functional or sequence-based analysis that are of particular interest owing to the identity of the organism from which there were derived.

Sequence analysis

The tremendous amount of genomic sequence information contained within metagenomic libraries can provide a window into the 'black box' of natural environments. The steep increase in sequencing capacity in recent years has reduced the sequencing of microbial genomes to a daily routine, and holds the key to unlocking the mysteries of the natural microbial world. At present, the sheer scale of metagenomic libraries necessary to represent a microbial assemblage in soil prevents the sequencing of entire BAC libraries, although for less complex natural environments this is a viable option. For the moment, generation of sequence data from soil metagenomic BAC clones relies upon a variety of screening methods to identify BAC clones that have a phylogenetic and/or functional linkage,

as well as random BAC end-sequencing to acquire a crude census of the BAC clones within a metagenomic library. Sequence-based analysis has led to surprising discoveries, including the identification of a bacterial rhodopsin gene in a bacterial species, which contrasted with all work on cultured organisms that indicated that these genes were confined to Archaea (Beja *et al.*, 2000a). Analysis of clones from uncultivated Archaea has indicated a functional diversity that is not necessarily reflected in the 16S rRNA genes (Beja *et al.*, 2002a).

Functional analysis

Most of the functional screening of metagenomic libraries reported to date has involved traditional low-technology screens. Enzymes, such as lipases, amylases and esculin hydrolases, have been identified with colorimetric or other visually discernible indicators in the medium on which the clones are plated (Henne *et al.*, 1999, 2000; Brady and Clardy, 2000; Rondon *et al.*, 2000; Brady *et al.*, 2001; MacNeil *et al.*, 2001). Pigmented antibiotics have been identified by visual inspection of arrayed libraries followed by antibiosis assays of selected clones (Brady *et al.*, 2001; MacNeil *et al.*, 2001; Gillespie *et al.*, 2002). Antibiotics that lack visible color have been identified by visualizing zones of inhibition around clones plated on sensitive indicator organisms (Brady and Clardy, 2000). Clones that either lyse or alter the appearance of red blood cells have been identified by replicating libraries on to lawns of sheep red blood cells (Gillespie *et al.*, 2002).

The barriers of heterologous expression and perhaps other factors have resulted in a low frequency of metagenomic clones expressing any given characteristic for which libraries have been screened. For example, in one study of 700 000 clones, 65 were found to produce antimicrobial activity (Brady and Clardy, 2000); in a library of 3648 clones, one expressed DNase, one antibacterial, two lipase and eight amylase activity (Rondon *et al.*, 2000); and in another study, a library of 286 000 clones yielded three lipolytic clones (Henne *et al.*, 2000). Therefore, the ability to screen many clones is essential to discovery in metagenomic libraries. Use of selectable phenotypes enhances discovery because only clones expressing the desired characteristic will grow in culture. Selections can be designed to complement a mutation for a normal function in *E. coli*, such as was done to identify new antiporters (Majernik *et al.*, 2001), or to identify a new function, such as antibiotic resistance (Riesenfeld *et al.*, unpublished).

◆◆◆◆◆◆ CONCLUDING REMARKS

Microbiology has seen three major revolutions, each coupled to development of a new powerful technique for 'seeing' microorganisms. The invention of the microscope, followed by improvements in optics and manufacturing, first revealed the existence of a microbial world. The microscope remains a microbiologist's indispensable tool. *In vitro* cultivation opened up a vast suite of powerful experimental approaches for

microbiology. The diversity of microorganisms (based on features such as colony morphology, pigmentation and growth rates) that were responsive to cultivation from many environments was an immediate result of the cultivation technologies that today we call plating. Especially in etiology, isolation in pure culture of microorganisms that were candidate pathogens and their reintroduction as inoculum into healthy hosts resulting in disease (followed by reisolation – the four canonical steps of Kochs' Postulates), was a major influence on microbiology that continues to dominate microbiology 150 years after its introduction. The third revolution stemmed from the insight of Carl Woese and his colleagues that the sequences of ribosomal RNA genes revealed evolutionary relationships among organisms, and the application of Woese's insight by Norm Pace and his colleagues to microbial ecology, confirming previous hints that environmental samples contained previously unknown microbial diversity. In each of these revolutions, the technology innovation provided direct access to a previously inaccessible view of microorganisms. It is a statement that can rarely be made about technological innovations in science to note that none of these innovative technologies replaced a previous less powerful technology, but instead each added to the toolbox of the microbiologist.

Each of the three innovations also led to further developments that built upon the power of the technology. In the case of the application of molecular biology tools to evolutionary and ecological microbiology, pioneered by Woese and Pace, the further developments have included powerful new tools for phylogenetic inference based on ribosomal RNA gene sequences, and *in situ* methods (marrying molecular biology to microscopy) for detection and ecological analysis of microorganisms in the environments, and at the scales, in which they function.

We view metagenomics as a further example of a powerful new tool that derives from the application of molecular methods to microbial ecology. Metagenomics presents an opportunity to derive new insights about the microbial world by uniting the power of genomics with the study of the vast phylogenetic and functional diversity of microorganisms in natural environments.

Acknowledgements

We thank Laurie Luther for assistance in preparing the manuscript. This work was supported by the National Science Foundation, the National Institutes of Health, the McKnight Foundation, the Packard Foundation, and the University of Wisconsin–Madison College of Agricultural and Life Sciences. DAM was supported by the Damon Runyon Cancer Research Foundation Fellowship, DRG1594.

References

August, P. R., Grossman, T. H., Minor, C., Draper, M. P., MacNeil, I. A., Pemberton, J. M., Call, K. M., Holt, D. and Osburne, M. S. (2000). Sequence analysis and functional characterization of the violacein biosynthetic pathway from *Chromobacterium violaceum*. *J. Mol. Microbiol. Biotechnol.* **2**, 513–519.

Beja, O., Aravind, L., Koonin, E. V., Suzuki, M., Hadd, A., Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A., Spudich, J. L., Spudich, E. N. and DeLong, E. F. (2000a). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906.

Beja, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A., Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S. B., Feldman, R. A. and DeLong, E. F. (2000b). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529.

Beja, O., Suzuki, M. T., Heidelberg, J. F., Nelson, W. C., Preston, C. M., Hamada, T., Eisen, J. A., Fraser, C. M. and DeLong, E. F. (2002b). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**, 630–633.

Beja, O., Koonin, E. V., Aravind, L., Taylor, L. T., Seitz, H., Stein, J. L., Bensen, D. C., Feldman, R. A., Swanson, R. V. and DeLong, E. F. (2002a). Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68**, 335–345.

Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P. and Goodman, R. M. (1997). Molecular phylogeny of Archaea from soil. *Proc. Natl Acad. Sci. USA* **94**, 277–282.

Black, C., Fyfe, J. A. M. and Davies, J. K. (1995). A promoter associated with the neisserial repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. *J. Bacteriol.* **177**, 1952–1958.

Borneman, J., Skroch, P. W., O'Sullivan, K. M., Palus, J. A., Rumjanek, N. G., Jansen, J. L., Nienhuis, J. and Triplett, E. W. (1996). Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* **62**, 1935–1943.

Brady, S. F. and Clardy, J. (2000). Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* **122**, 12903–12904.

Brady, S. F., Chao, C. J., Handelsman, J. and Clardy, J. (2001). Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic Lett.* **3**, 1981–1984.

Cane, D. E., Walsh, C. T. and Khosla, C. (1998). Biochemistry – harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68.

Chávez, S., Reyes, J., Cahuvat, C. F., Florencio, F. J. and Candau, P. (1995). The NADP–glutamate dehydrogenase of the cyanobacterium *Synechocystis* 6803: cloning, transcriptional analysis and disruption of the *gdhA* gene. *Plant Mol. Biol.* **28**, 173–188.

Cottrell, M. T., Moore, J. A. and Kirchman, D. L. (1999). Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **65**, 2553–2557.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* **89**, 5685–5689.

DeLong, E. F., Wu, K. Y., Prezelin, B. B. and Jovine, R. V. M. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature (London)* **371**, 695–697.

DeLong, E. F., Taylor, L. T., Marsh, T. L. and Preston, C. M. (1999). Visualization and enumeration of marine planktonic archaea and bacteria using polyribonucleotide probes and fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **65**, 5554–5563.

deVos, W. M. (1987). Gene cloning and expression in lactic streptococci. *FEMS Microbiol. Rev.* **46**, 281–295.

deVos, W. M., Kleerebezem, M. and Kuipers, O. (1997). Expression systems for industrial Gram-positive bacteria with low guanine and cytosine content. *Curr. Opin. Biotechnol.* **8**, 547–553.

Ding, M. and Yelton, D. B. (1993). Cloning and analysis of the *leuB* gene of *Leptospira interrogans* serovar *pomona*. *J. Gen. Microbiol.* **139**, 1093–1103.
Eisen, J. A. (1998). Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* **8**, 163–167.
Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W. R. (2001). Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl. Environ. Microbiol.* **67**, 89–99.
Ferreyra, R. G., Soncini, F. C. and Viale, A. M. (1993). Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the prototrophic sulfur bacterium *Chromatium vinosum*. *J. Bacteriol.* **175**, 1514–1523.
Fiandt, M. (2000). Construction of an environmental genomic DNA library from soil using the EpiFOS fosmid library production kit. *Epicentre Forum* **7**, 6.
Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., Clardy, J., Goodman, R. M. and Handelsman, J. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* (in press).
Giovannoni, S. J., Britschgi, T. B., Meyer, C. L. and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63.
Griffiths, B. S., Ritz, K. and Glover, L. A. (1996). Broad-scale approaches to the determination of soil microbial community structure: application of the community DNA hybridization technique. *Microbial Ecol.* **31**, 269–280.
Guiney, D. G. and Yakobsen, E. (1983). Location and nucleotide sequence of the transfer origin of the broad range host plasmid RK2. *Proc. Natl Acad. Sci. USA* **80**, 3595–3598.
Handelsman, J., Rondon, M. R., Brady, S., Clardy, J. and Goodman, R. M. (1998). Molecular biology provides access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249.
Head, I. M., Saunders, J. R. and Pickup, R. W. (1998). Microbial evolution, diversity and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb. Ecol.* **35**, 1–21.
Henne, A., Daniel, R., Schmitz, R. A. and Gottschalk, G. (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* **65**, 3901–3907.
Henne, A., Schmitz, R. A., Bomeke, M., Gottschalk, G. and Daniel, R. (2000). Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 3113–3116.
Hugenholtz, P., Goebel, B. M. and Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774.
Joseph, P., Fantino, J. R., Herbaud, M. L. and Denizot, F. (2001). Rapid orientated cloning in a shuttle vector allowing modulated gene expression in *Bacillus subtilis*. *FEMS Microbiol. Lett.* **205**, 91–97.
Karlin, S., Mrazek, J., Campbell, A. and Kaiser, D. (2001). Characterizations of highly expressed genes of four fast-growing bacteria. *J. Bacteriol.* **183**, 5025–5040.
Krsek, M. and Wellington, E. M. H. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *J. Microbiol. Meth.* **39**, 1–16.
Leonhardt, H. and Alonso, J. C. (1988). S1 mapping of the pC194 encoded chloramphenicol gene in *Bacillus subtilis*. *Nucleic Acids Res.* **16**, 1618.
Lindahl, V. (1996). Improved soil dispersion procedures for total bacterial counts, extraction of indigenous bacteria and cell survival. *J. Microbiol. Meth.* **25**, 279–286.

MacNeil, I. A., Minor, T. C., August, P. R., Grossman, T. H., Loiacono, K. A., Lynch, B. A., Phillips, T., Narula, S., Sundaramoorthi, R., Tyler, A., Aldredge, T., Long, H., Gilman, M. and Osburne, M. S. (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* **3**, 301–308.

Majernik, A., Gottschalk, G. and Daniel, R. (2001). Screening of environmental DNA libraries for the presence of genes conferring $Na^+(Li^+)/H^+$ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J. Bacteriol.* **183**, 6645–6653.

McInerney, M. O., Wilkinson, M., Patching, J. W., Embley, T. M. and Powell, R. (1995). Recovery and phylogenetic analysis of novel archaeal rRNA sequences from a deep-sea deposit feeder. *Appl. Environ. Microbiol.* **61**, 1646–1648.

Newman, J. R. and Fuqua, C. (1999). Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. *Gene* **227**, 197–203.

Pace, N. R. (1996). New perspective on the natural microbial world: molecular microbial ecology. *ASM News* **62**, 463–470.

Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. and Khosla, C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792.

Poyart, C. and Trieu-Cuot, P. (1997). A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in Gram-positive bacteria. *FEMS Microbiol. Lett.* **156**, 193–198.

Rondon, M. R., Goodman, R. M. and Handelsman, J. (1999a). The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol.* **17**, 403–409.

Rondon, M. R., Raffel, S. J., Goodman, R. M. and Handelsman, J. (1999b). Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proc. Natl Acad. Sci. USA* **96**, 6451–6455.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Osburne, M. S., Clardy, J., Handelsman, J. and Goodman, R. M. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547.

Sandler, S. J., Hugenholtz, P., Schleper, C., DeLong, E. F., Pace, N. R. and Clark, A. J. (1999). Diversity of *radA* genes from cultured and uncultured *Archaea*: comparative analysis of putative RadA proteins and their use as a phylogenetic marker. *J. Bacteriol.* **181**, 907–915.

Schleper, C., Swanson, R. V., Mathur, E. J. and DeLong, E. F. (1997). Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **179**, 7803–7811.

Schleper, C., DeLong, E. F., Preston, C. M., Feldman, R. A., Wu, K. Y. and Swanson, R. V. (1998). Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **180**, 5003–5009.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797.

Simon, H. M., Dodsworth, J. and Goodman, R. M. (2000). Crenarchaeota colonize terrestrial plant roots. *Environ. Microbiol.* **2**, 495–505.

Stach, J. E. M., Bathe, S., Clapp, J. P. and Burns, R.G. (2001). PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol. Ecol.* **36**, 139–151.

Stahl, D. A. (1993). The natural history of microorganisms. *ASM News* **59**, 609–613.
Stahl, D. A., Lane, D. J., Olsen, G. J. and Pace, N. R. (1985). Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl. Environ. Microbiol.* **49**, 1379–1384.
Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. and DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599.
Steinmetz, M. and Richter, R. (1994). Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in *Bacillus subtilis*, through in vivo recombination. *Gene* **142**, 79–83.
Suzuki, M. T., Rappe, M. S., Haimberger, Z. W., Winfield, H., Adair, N., Strobel, J. and Giovannoni, S. J. (1997). Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. *Appl. Environ. Microbiol.* **63**, 983–989.
Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.
Torsvik, V., Goksøyr, J. and Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**, 782–787.
Torsvik, V., Sorheim, R. and Goksoyr, J. (1996). Total bacterial diversity in soil and sediment communities – a review. *J. Indust. Microbiol.* Biotech. **17**, 170–178.
Trieu-Cuot, P., Carlier, C., Martin, P. and Courvalin, P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. *FEMS Microbiol. Lett.* **48**, 289–294.
Uozumi, T., Ozaki, A., Beppu, T. and Arima, K. (1980). New cryptic plasmid of *Bacillus subtilis* and restriction analysis of other plasmids found by general screening. *J. Bacteriol.* **142**, 315–318.
Vergin, K. L., Urbach, E., Stein, J. L., DeLong, E. F., Lanoil, B. D. and Giovannoni, S. J. (1998). Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl. Environ. Microbiol.* **64**, 3075–3078.
Wang, G.L. and Warren, R. (1996). Construction of an Arabidopsis BAC library and isolation of clones hybridizing with disease-resistance, gene-like sequences. *Plant Mol. Biol. Reporter* **14**, 107.
Wang, G.-Y.-S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, J. and Davies, J. (2000). Novel natural products from soil DNA libraries in a Streptomycete host. *Organic Lett.* **2**, 2401–2404.
Ward, D. M., Weller, R. and Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**, 63–65.
Whitman, W. B., Coleman, D. C. and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583.
Wild, J., Sektas, M., Hradecna, Z. and Szybalski, W. (1998). Targeting and retrofitting pre-existing libraries of transposon insertions with FRT and oriV elements for in-vivo generation of large quantities of any genomic fragment. *Gene* **223**, 55–66.
Wild, J., Hradecna, Z. and Szybalski, W. (1999). On command amplification of large genomic fragments cloned in pBeloBAC/oriV vectors. In: *The 1999 Molecular Genetics of Bacteria and Phages Meeting*, Madison, WI, p. 44.
Yakobson, E. A. and Guiney, D. G. (1984). Conjugal transfer of bacterial chromosomes mediated by the RK2 plasmid transfer origin cloned into transposon Tn5. *J. Bacteriol.* **160**, 451–453.
Zhou, J., Bruns, M. A. and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322.