Biotechnological prospects from metagenomics

Patrick D Schloss and Jo Handelsman*

The recognition that most microorganisms in the environment cannot be cultured by standard methods stimulated the development of metagenomics, which is the genomic analysis of uncultured microorganisms. Two types of analysis have been used to obtain information from metagenomic libraries: a function-driven approach, in which metagenomic libraries are initially screened for an expressed trait, and a sequence-driven approach, in which libraries are initially screened for particular DNA sequences. New antibiotics and enzymes are among the early discoveries from metagenomics. Future refinement of methods that enrich for genes of particular function will accelerate the rate of discovery of useful molecules.

Addresses
Department of Plant Pathology, 1630 Linden Drive, Russell Laboratories, University of Wisconsin, Madison, WI 53706, USA
*e-mail: joh@plantpath.wisc.edu

Current Opinion in Biotechnology 2003, **14**:303–310

This review comes from a themed section on
Environmental biotechnology
Edited by Ian M Head and Mark J Bailey

0958-1669/03/$ – see front matter

DOI 10.1016/S0958-1669(03)00067-3

Abbreviations

BrdU	bromodeoxyuridine
PCB	polychorinated biphenyl
PCR	polymerase chain reaction
PKS	polyketide synthase
SIP	stable-isotope probing

Introduction

Metagenomics is the culture-independent genomic analysis of microbial communities. The term is derived from the statistical concept of *meta*-analysis (the process of statistically combining separate analyses) and genomics (the comprehensive analysis of an organism's genetic material) [1•]. Metagenomics can be used to address the challenge of studying prokaryotes in the environment that are, as yet, unculturable and which represent more than 99% of the organisms in some environments [2]. This approach builds on recent advances in microbial genomics and in the polymerase chain reaction (PCR) amplification and cloning of genes that share sequence similarity (e.g. 16S rRNA, *nif*, *rec*A) directly from environmental samples [3]. Whereas PCR amplification requires prior knowledge of the sequence of the gene to design primers for amplification, direct extraction and cloning of DNA for genomics can theoretically access genes of any sequence or function. Direct genomic cloning offers the opportunity to capture operons or genes encoding pathways that may direct the synthesis of complex molecules, such as antibiotics. Sequence information about the genes flanking a gene of particular interest can also be obtained, potentially providing insight into the genomic environment of the gene or the phylogenetic affiliation of the organism from which it was derived. Moreover, a long-term goal of metagenomic analysis is to reconstruct the genomes of uncultured organisms by identifying overlapping fragments in metagenomic libraries and 'walking', clone to clone, to assemble each chromosome.

In this review, we discuss the two approaches to metagenomic analysis, designated the 'function-driven' and 'sequence-driven' analyses and review key discoveries derived from metagenomics. We close with an exploration of future technical and strategic improvements that will enhance the discovery rate and focus on types of genes or organisms of particular interest for biotechnological advances.

From ecology to biotechnology

Biotechnological applications from metagenomics will be fostered by the pursuit of fundamental ecological studies and focused screens for bioprospecting, just as both basic and applied approaches have contributed to the discovery of antibiotics and industrial enzymes from cultured microorganisms. The history of soil microbiology illustrates well the interdependence of basic and applied research and the role of serendipity in the discovery process. For example, the discovery of streptomycin and other bacterial antibiotics sprang from very basic studies of the taxonomy and ecology of actinomycetes in soil conducted by soil microbial ecologist, Selman Waksman, who did not intend to study or cure human infectious disease when he discovered the antibiotics [4,5]. Similarly, turbomycin, one of the first antibiotics discovered by metagenomics, was identified by accident in a clone that had hemolytic activity as part of a basic study directed toward rationalizing the prevalence of hemolysins among cultured soil bacteria [1•,6]. Therefore, to obtain the maximum number of biotechnological applications, it is essential that both basic biology and utility screens be pursued as part of the new field of metagenomics.

Analysis of metagenomic libraries

The function-driven analysis

Two approaches, the function-driven analysis and the sequence-driven analysis, have emerged to extract

biological information from metagenomic libraries (Figure 1). The function-driven analysis is initiated by identification of clones that express a desired trait, followed by characterization of the active clones by sequence and biochemical analysis. This approach quickly identifies clones that have potential applications in medicine, agriculture or industry by focusing on natural products or proteins that have useful activities. The limitations of the approach are that it requires expression of the function of interest in the host cell and clustering of all of the genes required for the function. It also depends on the availability of an assay for the function of interest that can be performed efficiently on vast libraries, because the frequency of active clones is quite low. Many approaches are being developed to mitigate these limitations. Improved systems for heterologous gene expression are being developed with shuttle vectors that facilitate screening of the metagenomic DNA in diverse host species and with modifications of *Escherichia coli* to expand the range of gene expression.

Although the genes encoding the enzymes required for synthesis of secondary metabolites are usually clustered on a contiguous piece of DNA, obtaining fragments of DNA large enough to contain the information required for synthesis of complex antibiotics, which can require over 100 genes, presents a challenge. This is a particular problem when the DNA is isolated from soil or other environments that contain high concentrations of contaminants that inhibit cloning. Methods to improve the preparation of large fragments of DNA that are clean enough to clone are being pursued vigorously in many laboratories. To address the challenge of detecting rare, active clones in large libraries, efforts are being directed toward the design of highly sensitive assays and robotic screens that efficiently detect low levels of activity in many samples. The most convenient traits to study are those that present a selectable phenotype, such as resistance to an antibiotic or growth on an unusual substrate, because selections are orders of magnitude more efficient than screens.

Functional screens of metagenomic libraries have identified both novel and previously described antibiotics [1•,6,7,8•,9–14], an antibiotic resistance gene [15], lipases [1•,16•], chitinases [17], membrane proteins [18], genes encoding enzymes for the metabolism of 4-hydroxybutyrate [19], and genes encoding the biotin synthetic pathways [20]. Sequencing of the flanking DNA of some clones reveals a gene or a group of genes that can be used to infer the phylogenetic affiliation of the organism from which the DNA in the clone was isolated. In the absence of a highly conserved gene — such as one encoding 16S rRNA, DNA polymerase or RecA — codon usage and sequence alignment of gene clusters with genes in the databases can be used to make phylogenetic inferences, although these are less robust than those based on conserved genes.

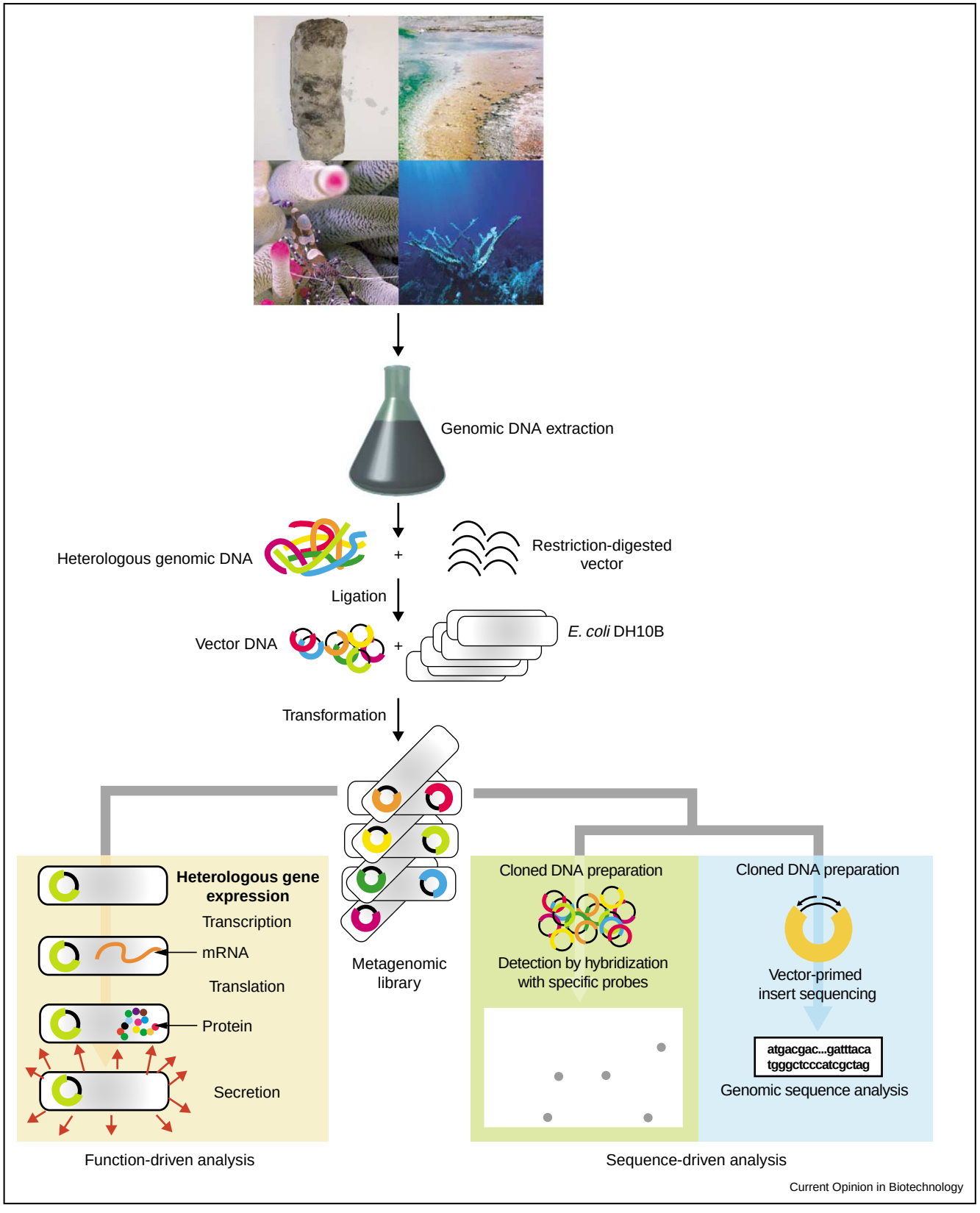
The sequence-driven analysis

Sequence-driven analysis relies on the use of conserved DNA sequences to design hybridization probes or PCR primers to screen metagenomic libraries for clones that contain sequences of interest. Significant discoveries have also resulted from random sequencing of metagenomic clones. Sequencing of clones carrying phylogenetic anchors, such as the 16S rRNA gene and the Archaeal DNA repair gene *rad*A [21,22•,23,24,25•,26–30], has led to functional information about the organisms from which these clones were derived. In what is arguably the most dramatic discovery from metagenomics to date, sequencing of a clone isolated from seawater that was initially identified because it carried a bacterial 16S rRNA gene revealed a gene with high similarity to bacteriorhodopsin genes. This result provided the first indication that rhodopsins are not limited to the Archaea, as previously thought [22•]. Subsequent heterologous expression of the bacteriorhodopsin gene in *E. coli* produced a functional biochemical characterization of the protein, completing the full spectrum of studies that link phylogeny to function (Figure 1).

There has been disagreement about the utility of random sequencing of metagenomic clones. Although some members of the field view this approach as too undirected to yield biological understanding, others stress that there is so little known about some divisions of Bacteria that any genomic sequence is helpful in guiding the design of experiments to reveal their biology. Sequencing many clones derived from one division may provide sufficient data to begin to discern patterns to develop hypotheses. The Acidobacterium division offers a prime candidate for metagenomic sequencing because it contains few cultured members, despite its apparent ubiquity and abundance across the globe. A collection of 73 metagenomic clones carrying 16S rRNA has genes that align with the Acidobacterium division have been assembled and full-length sequencing is underway [31]. In aggregate, these clones carry approximately 3 Mb of Acidobacterium DNA, the sequence of which will provide a window on the biology of this unknown group of Bacteria. A similar analysis has been initiated in the Crenarchaeota of soil,

(Figure 1 Legend) Construction and analysis of metagenomic libraries from soil, hot spring outflow, deep ocean tube worms and sponge tissue using function-driven and sequence-driven approaches. This strategy was originally proposed by Pace *et al*. [3] and executed by Stein *et al*. [28]. (The photograph of non-phototrophic organisms flanked by phototrophs in hot spring outflow from the Porcelain Basin Area of the Norris Geyser Basin at Yellowstone National Park is reproduced with permission from CLE Fisher. Licensed for use, ASM Microbe Library linked to http://www.microbelibrary.org.)

Figure 1


Genomic DNA extraction
Heterologous genomic DNA
+
Restriction-digested vector
Ligation
Vector DNA
+
E. coli DH10B
Transformation
Metagenomic library
Heterologous gene expression
Transcription
mRNA
Translation
Protein
Secretion
Function-driven analysis
Cloned DNA preparation
Detection by hybridization with specific probes
Cloned DNA preparation
Vector-primed insert sequencing
atgacgac...gatttaca
tgggctcccatcgctag
Genomic sequence analysis
Sequence-driven analysis
Current Opinion in Biotechnology

using 16S rRNA genes to identify clones derived from the Archaea and employing sequencing to compile genomic information about them [25•].

The sequence conservation of regions of phylogenetic anchors facilitates their isolation without prior knowledge of the full gene sequence. By contrast, the sequences of most genes of practical importance are far too divergent to make it possible to identify new homologues by PCR or hybridization. However, a few classes of genes contain sufficiently conserved regions to facilitate their identification by sequence instead of function. Two hotly pursued examples are the genes encoding polyketide synthases (PKSs) and peptide synthetases, which contribute to synthesis of complex antibiotics. The PKSs are modular enzymes with repeating domains containing divergent regions that provide the variation in chemical structures of the products. These regions are flanked by highly conserved regions, which have provided the basis for designing probes to identify PKS genes among metagenomic clones [13]. Mixing and matching PKS domains from different sources has yielded new antibiotics, stimulating interest in the discovery of new PKS genes [32]. Similar approaches have identified genes encoding light-harvesting complexes from uncultured bacteria [33•] and *radA* and DNA polymerase genes from uncultured Archaea [26,34]. Thus, application of this approach is likely to be broadened as patterns emerge that define more gene families containing highly conserved features.

Large-scale sequencing projects, such as the one led by The Institute for Genomic Research (TIGR) and Monterey Bay Coastal Ocean Microbial Observatory, will generate vast databases of end sequences of metagenomic clone inserts that carry phylogenetic anchors. These projects will thus expand the available genomic information about some of the more obscure microbial divisions (http://www.tigr.org/tdb/MBMO/). These undirected sequencing efforts complement approaches that require prior knowledge of gene sequence, which are necessarily limited in the scope of genes they will discover. The genes most distantly related to the consensus for a class of genes are of great interest for biotechnological applications, but are the least likely to be detected by probes or primers designed based on the database entries. The undirected approach is likely to unmask homologues of useful genes that would not have been detected by directed studies.

Enrichment strategies to improve the odds

One of the sustained frustrations with analysis of metagenomic libraries is the low frequency of clones of a desired nature. To increase the proportion of active clones in a library, several strategies have been designed to enrich for the sequences of interest before cloning. The potential power of this strategy is evident in the elegant genomics performed on uncultured Bacteria and Archaea that are highly enriched in associations with their hosts. The first complete genome sequence obtained for an uncultured bacterium is for *Buchnera aphidicola*, an obligate symbiont of aphids. The prokaryotic cells can be effectively separated from insect tissue to produce relatively pure microbial DNA, making it feasible to sequence and reassemble the genome despite the inability to grow the bacterium in culture [35,36•,37]. Genomic analysis has also been successfully conducted on the uncultured Archaea *Cenarchaeum symbiosum*, which is a symbiont of a marine sponge. *C. symbiosum* is highly enriched in the sponge – it is the only archaeal phylotype found in the sponge and represents 65% of the prokaryotic cells [38]. DNA was prepared from archaeal cells that were further enriched by differential centrifugation. Fosmid libraries prepared from this DNA contain an estimated 15 copies of the genome, providing sufficient material to draw inferences about genome structure and variability [27]. The successful assembly of the *B. aphidicola* genome [36•] as well as the *Vibrio cholerae* genome, which contains two chromosomes [39], provides a strong foundation for developing biological and computational methods for assembling more complex genome assemblages. However, the first and essential step is to enrich for the genomes of interest.

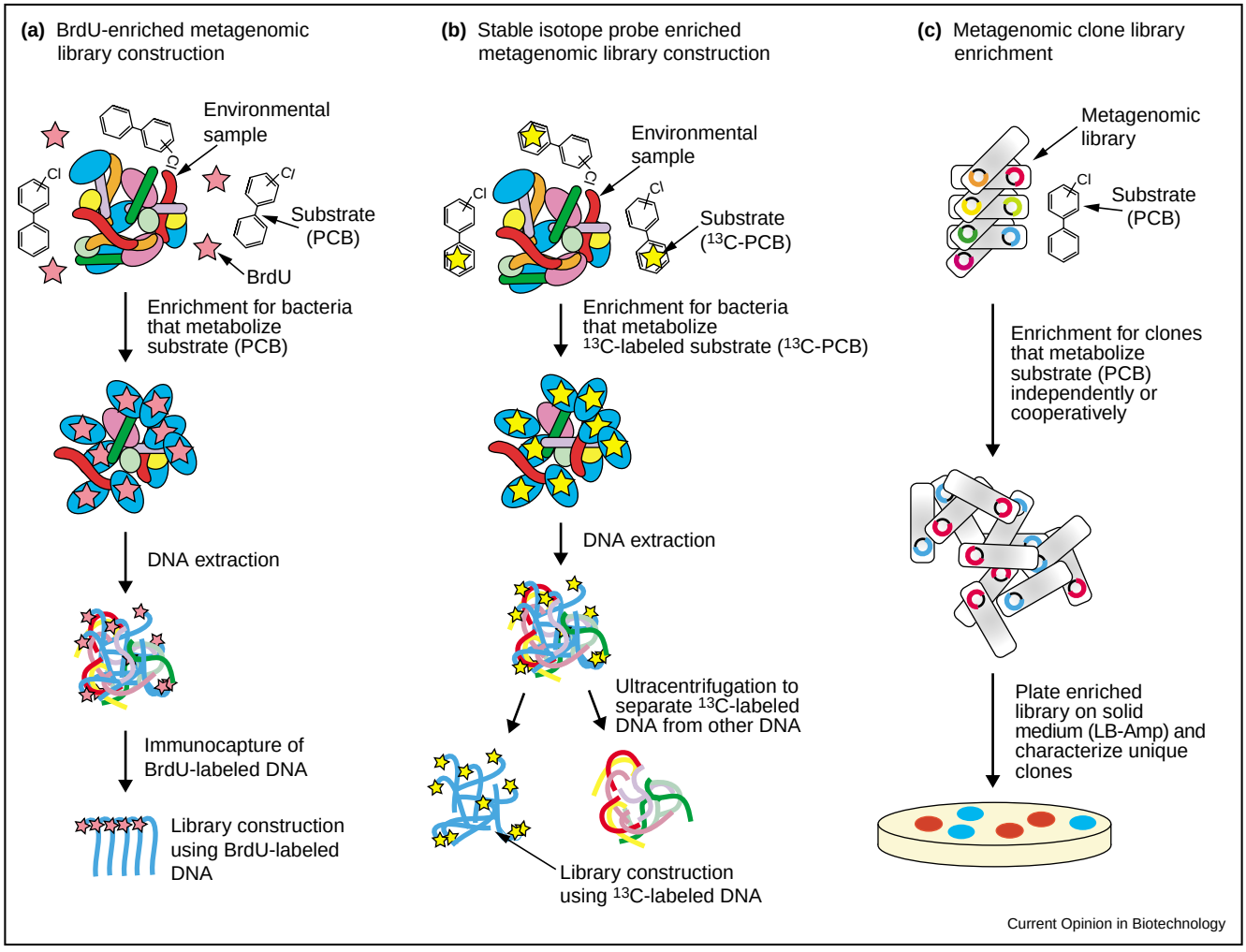
Enriching for genomes from metabolically active cells

The symbionts of aphids and the marine sponge represent the simplest type of community, containing a single species with some genetic variation in the population. At the other extreme of complexity is the soil, containing an estimated 10^9 prokaryotes and more than 2000 genome types per gram of soil, with an average genome type representing less than 0.05% of the soil community [40]. With today's technology it would be difficult to obtain complete coverage of all of the genomes in a soil community. Therefore, carving out slices of the community, selected for a common feature, reduces the complexity of the task and brings into sight the possibility of complete coverage of a subset of the soil community.

A simple enrichment is for GC content of the genomes. As many organisms that have a high GC content in their DNA are of particular interest (e.g. Actinomycetes and Acidobacteria), DNA can be extracted from the soil and then subjected to ultracentrifugation to enrich for high GC content DNA. Although this is a fairly crude approach and will not provide a complete separation, it will certainly increase the representation of certain genomes in a library.

A more elegant separation method is bromodeoxyuridine (BrdU) enrichment (Figure 2a). The principle underlying this type of enrichment is that metabolically active organisms will incorporate a labeled nucleotide into their DNA, which can then be isolated on the basis of the incorporated label. BrdU can be fed to bacteria in soil samples

Figure 2



Enrichment for specialized DNA from environmental samples using **(a)** BrdU-enrichment, **(b)** stable-isotope probing and **(c)** metagenomic library enrichment using PCBs as a model substrate.

and the labeled DNA recovered by immunocapture [41,42]. 16S rRNA analysis performed on the DNA identifies phylogenetic groups that were metabolically active in the original sample. The addition of selective substrates with the BrdU further discriminates among the members of the microbial community, enriching metagenomic libraries for those that grow on the added nutrient [43]. This strategy could be used to enrich for organisms that grow on xenobiotics or on substrates such as starch, cellulose and proteins to find amylases, cellulases and proteases, respectively, or other enzymes of interest in metagenomic libraries.

Another enrichment method is stable-isotope probing (SIP, Figure 2b; see the article by Radjewski, McDonald and Murrell in this issue for a more complete description), which involves providing a ^{13}C-labeled substrate to soil bacteria. The bacteria that can use the substrate incorporate the ^{13}C into their DNA, making it more dense than normal DNA containing ^{12}C. SIP has been successfully used for labeling and separating DNA [44–47,48•] and RNA [49,50]. Density gradient centrifugation cleanly separates the labeled from unlabled nucleic acids, which can then be used either for PCR-based analysis or direct cloning to construct metagenomic libraries. The method has enormous potential for subdividing microbial communities into functional units to simplify analysis and will offer broad opportunity to study community functions if it can be expanded to stable isotopes of other elements, such as phosphorus or nitrogen.

Although these approaches provide a significant step in refining metagenomic library construction by enriching for DNA from a subset of the community, both methods have limitations that need to be addressed by future research. The first potential problem is cross-feeding.

Any bacterium that is metabolically active will take up BrdU and will therefore be represented in a library that is supposed to be enriched for organisms that utilize an added substrate. An additional issue is timing. The more prolonged the substrate feeding, the higher the probability that the substrate will be recycled in the community and the basis of the enrichment will break down. Finally, the immunocapture method and density gradient centrifugation may shear the DNA, making it difficult to retrieve pathways encoded on large fragments of DNA.

RNA-based SIP reduces the cross-feeding problems associated with DNA-based SIP [49,50], but thus far, has been used only to construct 16S rRNA gene libraries. If the isotope-labeled RNA presents a viable substrate for making cDNA, then libraries could be constructed from cDNA representing genes that are actively transcribed in the presence of the labeled compound. Metagenomic clone libraries could be screened in hybridization arrays using the cDNA clones as probes. This approach has the attractive feature that one library constructed with non-selected DNA from the entire community could be screened repeatedly using different probes, each developed from a distinct subset of the community.

Enrichments for metagenomic clones forming metabolic consortia

Research on metagenomic libraries has focused on activities of individual clones. A key direction for development of this technology is to enrich for consortia of clones that together accomplish a desired function (Figure 2c). The approach is analogous to the isolation of bacteria that cooperate to degrade complex pools of polychlorinated biphenyl (PCB) compounds in soil, a process mediated by two or more species of bacteria in which each species contributes part of a 'mosaic pathway' [51]. Such mosaic pathways indicate that it will not be possible to capture the genes for certain pathways on a contiguous piece of DNA; thus, approaches to study multiple clones simultaneously are essential to the long-term utility of metagenomics. Whereas most work in the past has focused on screening metagenomic libraries on solid medium, isolation of clone consortia will be facilitated by liquid enrichments in which substrates and products can diffuse freely among members of the mixture.

Conclusions

Metagenomics is a young and exciting technique that has broad application in biology and biotechnology. Although many advances in heterologous gene expression, library construction, vector design, and screening will improve it, the current technology is sufficiently powerful to yield products for solving real world problems, including the discovery of new antibiotics and enzymes. Approaches that enrich for a portion of the microbial community or for a collection of metagenomic clones will enhance the power of metagenomic analysis to address targeted questions in microbial ecology and to discover new biotechnological applications.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C *et al.*: **Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms**. *Appl Environ Microbiol* 2000, **66**:2541-2547.

The authors report the construction and functional screening of a metagenomic library from soil containing 1 Gb of DNA in a bacterial artificial chromosome vector. The library contained diverse functions and was derived from diverse taxa, as indicated by analysis of 16S rRNA genes.

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

3. Pace NR, Stahl DA, Lane DJ, Olsen GJ: **Analyzing natural microbial populations by rRNA sequences**. *ASM News* 1985, **51**:4-12.

4. Waksman SA, Foster JW: **Associative and antagonistic effects of microorganisms: II. Antagonistic effects of microorganisms grown on artificial substrates**. *Soil Sci* 1937, **43**:69-76.

5. Waksman SA, Curtis RE: **The actinomyces of the soil**. *Soil Sci* 1916, **1**:99-134.

6. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J: **Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA**. *Appl Environ Microbiol* 2002, **68**:4301-4306.

7. August PR, Grossman TH, Minor C, Draper MP, MacNeil IA, Pemberton JM, Call KM, Holt D, Osburne MS: **Sequence analysis and functional characterization of the violacein biosynthetic pathway from *Chromobacterium violaceum***. *J Mol Microbiol Biotechnol* 2000, **2**:513-519.

8. • Brady SF, Chao CJ, Clardy J: **New natural product families from an environmental DNA (eDNA) gene cluster**. *J Am Chem Soc* 2002, **124**:9968-9969.

The authors describe the discovery of a new class of aminoacylated antibiotics from a metagenomic library constructed from soil.

9. Brady SF, Chao CJ, Handelsman J, Clardy J: **Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA**. *Org Lett* 2001, **3**:1981-1984.

10. Brady SF, Clardy J: **Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA**. *J Am Chem Soc* 2000, **122**:12903-12904.

11. MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R *et al.*: **Expression and isolation of antimicrobial small molecules from soil DNA libraries**. *J Mol Microbiol Biotechnol* 2001, **3**:301-308.

12. Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J: **Novel natural products from soil DNA libraries in a streptomycete host**. *Org Lett* 2000, **2**:2401-2404.

13. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS *et al.*: **Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products**. *Appl Environ Microbiol* 2003, **69**:49-55.

14. Piel J: **A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles**. *Proc Natl Acad Sci USA* 2002, **99**:14002-14007.

15. Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M, Mullany P: **Novel tetracycline resistance determinant from the oral metagenome**. *Antimicrob Agents Chemother* 2003, **47**:1430-1432.

16. Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R:
• **Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli***. *Appl Environ Microbiol* 2000, **66**:3113-3116.
The paper presents two screens for lipolytic activity, which were applied to metagenomic libraries constructed from soil. One screen led to the identification of one lipolytic clone out of 730 000 screened clones; the other led to the identification of three out of 286 000 clones, indicating the relatively low frequency of functionally active clones.

17. Cottrell MT, Moore JA, Kirchman DL: **Chitinases from uncultured marine microorganisms**. *Appl Environ Microbiol* 1999, **65**:2553-2557.

18. Majernik A, Gottschalk G, Daniel R: **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* 2001, **183**:6645-6653.

19. Henne A, Daniel R, Schmitz RA, Gottschalk G: **Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate**. *Appl Environ Microbiol* 1999, **65**:3901-3907.

20. Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR: **Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia**. *Appl Environ Microbiol* 2001, **67**:89-99.

21. Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB *et al.*: **Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage**. *Environ Microbiol* 2000, **2**:516-529.

22. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP,
• Jovanovich SB, Gates CM, Feldman RA, Spudich JL *et al.*: **Bacterial rhodopsin: evidence for a new type of phototrophy in the sea**. *Science* 2000, **289**:1902-1906.
The authors report the isolation of a new rhodopsin, the first derived from bacteria, that was discovered in a genomic library constructed from microorganisms in seawater.

23. Beja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF: **Proteorhodopsin phototrophy in the ocean**. *Nature* 2001, **411**:786-789.

24. Beja O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL, Bensen DC, Feldman RA, Swanson RV, DeLong EF: **Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces**. *Appl Environ Microbiol* 2002, **68**:335-345.

25. Quaiser A, Ochsenreiter T, Klenk HP, Kletzin A, Treusch AH, Meurer
• G, Eck J, Sensen CW, Schleper C: **First insight into the genome of an uncultivated crenarchaeote from soil**. *Environ Microbiol* 2002, **4**:603-611.
The authors identified clones carrying Crenarchaeota 16S rRNA genes in a soil metagenomic library and sequenced the flanking DNA to begin to compile genomic information about this poorly understood and, as yet, uncultured group of prokaryotes.

26. Schleper C, Swanson RV, Mathur EJ, DeLong EF: **Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum***. *J Bacteriol* 1997, **179**:7803-7811.

27. Schleper C, DeLong EF, Preston CM, Feldman RA, Wu KY, Swanson RV: **Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum***. *J Bacteriol* 1998, **180**:5003-5009.

28. Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF: **Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon**. *J Bacteriol* 1996, **178**:591-599.

29. Suzuki MT, Beja O, Taylor LT, Delong EF: **Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton**. *Environ Microbiol* 2001, **3**:323-331.

30. Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ: **Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales**. *Appl Environ Microbiol* 1998, **64**:3075-3078.

31. Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM: **A census of rRNA genes and linked genomic sequences within a soil metagenomic library.** *Appl Environ Microbiol* 2003: in press.

32. Cane DE, Walsh CT, Khosla C: **Harnessing the biosynthetic code: combinations, permutations, and mutations**. *Science* 1998, **282**:63-68.

33. Beja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM,
• Hamada T, Eisen JA, Fraser CM, DeLong EF: **Unsuspected diversity among marine aerobic anoxygenic phototrophs**. *Nature* 2002, **415**:630-633.
The paper presents an elegant study of the diversity of photosynthetic operons in a marine assemblage of prokaryotes using conserved sequences to screen a metagenomic library for *puf*L and *puf*M, the genes encoding the reaction center complex.

34. Sandler SJ, Hugenholtz P, Schleper C, DeLong EF, Pace NR, Clark AJ: **Diversity of *rad*A genes from cultured and uncultured archaea: comparative analysis of putative RadA proteins and their use as a phylogenetic marker**. *J Bacteriol* 1999, **181**:907-915.

35. Clark MA, Baumann L, Thao ML, Moran NA, Baumann P: **Degenerative minimalism in the genome of a psyllid endosymbiont**. *J Bacteriol* 2001, **183**:1853-1861.

36. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS,
• Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG: **50 million years of genomic stasis in endosymbiotic bacteria**. *Science* 2002, **296**:2376-2379.
The paper discusses the genomic sequence of an obligate endosymbiotic bacterium, *Buchnera aphidicola*. The sequence was obtained by constructing a genomic library from a single homogenized *Schizaphis graminum* female.

37. Wernegreen JJ, Ochman H, Jones IB, Moran NA: **Decoupling of genome size and sequence divergence in a symbiotic bacterium**. *J Bacteriol* 2000, **182**:3867-3869.

38. Preston CM, Wu KY, Molinski TF, DeLong EF: **A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov**. *Proc Natl Acad Sci USA* 1996, **93**:6241-6246.

39. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L *et al.*: **DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae***. *Nature* 2000, **406**:477-483.

40. Torsvik V, Goksoyr J, Daae FL: **High diversity in DNA of soil bacteria**. *Appl Environ Microbiol* 1990, **56**:782-787.

41. Urbach E, Vergin KL, Giovannoni SJ: **Immunochemical detection and isolation of DNA from metabolically active bacteria**. *Appl Environ Microbiol* 1999, **65**:1207-1213.

42. Borneman J: **Culture-independent identification of microorganisms that respond to specified stimuli**. *Appl Environ Microbiol* 1999, **65**:3398-3400.

43. Yin B, Crowley D, Sparovek G, De Melo WJ, Borneman J: **Bacterial functional redundancy along a soil reclamation gradient**. *Appl Environ Microbiol* 2000, **66**:4361-4365.

44. Radajewski S, Murrell JC: **Stable isotope probing for detection of methanotrophs after enrichment with $^{13}CH_4$**. *Methods Mol Biol* 2002, **179**:149-157.

45. Morris SA, Radajewski S, Willison TW, Murrell JC: **Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing**. *Appl Environ Microbiol* 2002, **68**:1446-1453.

46. Radajewski S, Webster G, Reay DS, Morris SA, Ineson P, Nedwell DB, Prosser JI, Murrell JC: **Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing**. *Microbiology* 2002, **148**:2331-2342.

47. Murrell JC, Radajewski S: **Cultivation-independent techniques for studying methanotroph ecology**. *Res Microbiol* 2000, **151**:807-814.

48. Radajewski S, Ineson P, Parekh NR, Murrell JC: **Stable-isotope**
• **probing as a tool in microbial ecology**. *Nature* 2000, **403**:646-649.

The paper describes the construction of 16S rRNA gene libraries using DNA extracted from ^{13}C-methanol-enriched soil. Density gradient centrifugation was used to separate the ^{13}C and ^{12}C labeled DNA fractions before cloning.

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

50. Manefield M, Whiteley AS, Ostle N, Ineson P, Bailey MJ: **Technical considerations for RNA-based stable isotope probing: an approach to associating microbial diversity with microbial community function**. *Rapid Commun Mass Spectrom* 2002, **16**:2179-2183.

51. Abraham WR, Nogales B, Golyshin PN, Pieper DH, Timmis KN: **Polychlorinated biphenyl-degrading microbial communities in soils and sediments**. *Curr Opin Microbiol* 2002, **5**:246-253.