



Contents lists available at ScienceDirect

Soil Biology & Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Review

Size does matter: Application-driven approaches for soil metagenomics

Kavita S. Kakirde, Larissa C. Parsley, Mark R. Liles*

Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

ARTICLE INFO

Article history:

Received 30 April 2010

Received in revised form

24 July 2010

Accepted 26 July 2010

Available online xxx

Keywords:

Soil

Metagenomics

Metagenomic library construction

PCR

Pyrosequencing

16S rRNA

High molecular weight DNA

Cosmid

Fosmid

BAC

ABSTRACT

Metagenomic analyses can provide extensive information on the structure, composition, and predicted gene functions of diverse environmental microbial assemblages. Each environment presents its own unique challenges to metagenomic investigation and requires a specifically designed approach to accommodate physicochemical and biotic factors unique to each environment that can pose technical hurdles and/or bias the metagenomic analyses. In particular, soils harbor an exceptional diversity of prokaryotes that are largely undescribed beyond the level of ribotype and are a potentially vast resource for natural product discovery. The successful application of a soil metagenomic approach depends on selecting the appropriate DNA extraction, purification, and if necessary, cloning methods for the intended downstream analyses. The most important technical considerations in a metagenomic study include obtaining a sufficient yield of high-purity DNA representing the targeted microorganisms within an environmental sample or enrichment and (if required) constructing a metagenomic library in a suitable vector and host. Size does matter in the context of the average insert size within a clone library or the sequence read length for a high-throughput sequencing approach. It is also imperative to select the appropriate metagenomic screening strategy to address the specific question(s) of interest, which should drive the selection of methods used in the earlier stages of a metagenomic project (e.g., DNA size, to clone or not to clone). Here, we present both the promising and problematic nature of soil metagenomics and discuss the factors that should be considered when selecting soil sampling, DNA extraction, purification, and cloning methods to implement based on the ultimate study objectives.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Previous cultivation-based studies have proven soils to be an excellent resource for the discovery of novel microbial natural products (Schatz and Waksman, 1944). The discrepancy between the numbers of microorganisms visible via microscopy and the colonies obtained from laboratory cultivation is several orders of magnitude for most soils, and overcoming the “great plate count anomaly” (Staley and Konopka, 1985) in order to access a greater diversity of bacteria has become one of the most significant challenges and opportunities in the field of molecular microbial ecology. Many studies have demonstrated that the phylogenetic and functional diversity of microorganisms in various habitats, including soil, vastly exceeds the diversity of prokaryotic phyla known from cultivation (Ward et al., 1990; Hugenholtz et al., 1998; Rondon et al., 2000; Breitbart et al., 2003; Dinsdale et al., 2008). Fortunately, the recent development of metagenomic and other culture-independent approaches has enabled investigation of the

functional genetic diversity of soil microorganisms without the inherent biases of cultivation.

Metagenomics can be defined as the genomic analysis of the collective microbial assemblage found in an environmental sample (Handelsman et al., 1998). There are many variants on metagenomic approaches, which initially were dependent upon cloning of DNA from an environmental sample (Healy et al., 1995; Stein et al., 1996), but more recently many metagenomic approaches have relied upon high-throughput sequencing (Edwards et al., 2006). One of the main advantages of functional metagenomics is its ability to identify gene products from as-yet-uncultured microbes, many with no significant homolog within the GenBank database. Studies have applied a metagenomic approach to a number of different environments, such as soils (Rondon et al., 2000; Voget et al., 2003; Tringe et al., 2005), the complex microbiome of the rumen (Brulc et al., 2009), planktonic marine microbial assemblages (Beja et al., 2000a; Breitbart et al., 2002), deep sea microbiota (Sogin et al., 2006), an acid mine site (Tyson et al., 2004), arctic sediments (Jeon et al., 2009) and the Sargasso Sea (Venter et al., 2004).

This review focuses on metagenomic approaches for exploring the phylogenetic and functional diversity of soil microorganisms.

* Corresponding author. Tel.: +1 334 844 1656; fax: +1 334 844 1645.
E-mail address: lilesma@auburn.edu (M.R. Liles).

Despite the promise of metagenomics as a strategy for the identification of novel natural bioactive products, xenobiotic pathways, and other metabolic processes, soils present a unique set of technical challenges for the successful isolation and analysis of metagenomic DNA. Many of the methods are labor- and cost-intensive, and the full extent of the project should be considered before embarking on a metagenomic study of a soil sample(s). A key strategic decision will be whether to adopt a sequence-only strategy or one that involves cloning of metagenomic DNA. This will be dependent upon the nature of the gene(s) or gene product(s) that are targeted, the degree of knowledge concerning these genetic loci within extant microorganisms and sequence databases, and the interest in identifying biological functions that may not be recognized from a purely sequence-driven approach.

2. Exploring the soil environment

Soil is the major component of most terrestrial environments and is considered to be the most diverse ecosystem on Earth, with respect to its native microbial populations. One gram of soil is estimated to contain millions of bacteria, archaea, viruses, and eukaryotic microorganisms (Torsvik and Ovreas, 2002; Fierer et al., 2007; Wommack et al., 2008), of which only a small percentage has been cultivated in the laboratory (Hugenholtz et al., 1998; Curtis and Sloan, 2005). From phylogenetic surveys of soil ecosystems it is known that the number of prokaryotic species in a single soil sample exceeds known cultured prokaryotes. The soil environment is an abundant yet under-characterized source of genetic diversity that has great potential to enrich our understanding of soil microbial ecology and provide enzymes and bioactive compounds useful to human society.

2.1. Soil composition affects microbial diversity

Soils are dynamic and heterogeneous environments in which bacteria, fungi, protozoa, and other eukaryotes compete for nutrients and space. Often, this competition leads to the production of secondary metabolites with antimicrobial activity, which may explain why the majority of previously characterized antibiotics originated from soil microbes (Burgess et al., 1999; Garbeva and de Boer, 2009). Microbes are subjected to both biotic stress (e.g., competition, parasitism) and abiotic stress (e.g., fluctuations in temperature, moisture levels, etc.), leading to a dynamic ecosystem that fosters a variety of microbial interactions and functions.

Microbial activity and growth in soils is affected by its physical, chemical, and biological properties, and as a result of microbial processes, the soil environment is dramatically transformed in terms of its structure and chemistry (e.g., nitrogen fixation, organic matter decomposition). The physical composition (e.g., loamy, sandy, clay) of the soil will greatly influence its microbial population, as will its chemical characteristics, such as organic matter content and pH (Hassink et al., 1993). For example, the extent of bacterial diversity and number of bacteria present has been observed to be inversely related to the soil's particle size (Sessitsch et al., 2001). Although many soil microbes are mesophilic, more extreme environmental conditions or the presence of unusual contaminants may select for a distinct group of organisms, thus altering the overall community structure of that particular soil sample (George et al., 2009). In addition, the geographic location of the soil will affect phylogenetic composition and microbial growth, as temperature and moisture content will vary widely among different regions. Selecting a sampling site and method(s) is an important factor to consider when beginning a metagenomic analysis of soil microorganisms.

2.2. Soil sampling considerations

The depth of the soil sample will affect the number and types of microbes that are collected, as cell density is generally greater in surface soils when compared to subsurface soils. In addition, surface soils will contain phototrophic microorganisms (e.g., from the division *Cyanobacteria*) that will not be present at lower soil horizons (Veluci et al., 2006). In consideration of these variations, it is advisable to take multiple samples and pool the samples prior to analysis. Pooling is beneficial when a representative sample that encompasses diverse microorganisms is desired, but can be a disadvantage if the objective is to target a specific microbial population. In the latter case it is important that the sampling site and method be selected accordingly, and to assess the presence of the targeted population by cultivation or via specific molecular probes.

The depth of sampling and cross-contamination are also factors that should be considered. Soil augers are well suited for sampling because of their precision over using a shovel. Since sampling equipment may become contaminated with microbes from other layers before reaching the targeted depth, the top and/or outer layer of the sampled soil may be discarded. To prevent contamination between sampling runs, utilize a separate auger for each sample type, or the equipment can be treated with ethanol, bleach, and sterile water. After sampling it is critical to freeze or place samples on ice and process them as quickly as possible or store them at -80°C . Soil samples that have been stored desiccated are not recommended for use, due to lower yields of cells and/or DNA.

2.2.1. Extraction and purification of soil microbial metagenomic DNA

When extracting metagenomic DNA from a soil sample, the first consideration is DNA size. If the goal of the study is high-throughput sequencing, PCR amplification, or small-insert clone libraries, then a harsh extraction method that results in substantially sheared, yet highly purified metagenomic DNA will be sufficient (Section 4.2.2). Large-insert clone libraries will require adoption of an alternative DNA extraction protocol to provide sufficiently intact metagenomic DNA (Section 4.3.2). For any application, it is critical to isolate DNA from diverse microorganisms that are representative of the microbial assemblage; otherwise, downstream analyses may be biased against or in favor of a particular group of microorganisms (Liles et al., 2003; Feinstein et al., 2009). However, biased metagenomic libraries may be preferred, if one is targeting a consortium of microorganisms enriched for a specific functional activity (Healy et al., 1995), in which case the relative abundance of targeted microbial taxa during the enrichment and metagenomic library construction process may be monitored.

Two general approaches exist for environmental metagenomic DNA extraction, 1) DNA is directly extracted from the environmental sample; or 2) microbial cells are recovered from the environmental sample prior to lysis and DNA purification (i.e., "indirect extraction"). Direct extraction of metagenomic DNA has many advantages, including its decreased processing time and that it provides a greater DNA yield compared to other methods (Ogram et al., 1987). Unfortunately, this method often results in the isolation of a higher percentage of non-bacterial DNA (Ogram et al., 1987; Tsai and Olson, 1991; Tebbe and Vahjen, 1993). Indirect DNA extraction overcomes some limitations of the direct extraction method because it results in less non-bacterial DNA (Osborn and Smith, 2005) and, like direct extraction methods, can yield DNA from phylogenetically diverse origins (Gabor et al., 2003). However, indirect extraction methods are more time-consuming, in general provide lower DNA yields, and may bias against microorganisms

that are not easily dissociated from the environmental matrix or lysed via chemical and enzymatic treatment. Selecting which extraction method to adopt depends greatly on the desired downstream application. The decrease in genomic DNA fragment size resulting from harsh direct extraction and purification methods is typically not a problem in PCR-based or pyrosequencing studies since the targeted genetic loci are of relatively small size (e.g., less than a few kilobase pairs). Conversely, the indirect extraction method is generally used when the size of extracted DNA fragments must be maintained for use in constructing large-insert metagenomic libraries, and/or a high proportion of bacterial DNA template is desired prior to the molecular application.

Because of soil's physical and chemical heterogeneity, DNA isolated from soils is often co-isolated with organic compounds that can inhibit downstream applications such as PCR and metagenomic library construction. Depending on the composition of the soil, these contaminants may include humic acids, polyphenols, polysaccharides, and nucleases, which can also degrade DNA (Tebbe and Vahjen, 1993; Zhou et al., 1996; Frostegard et al., 1999; Sylvia, 2005). The removal of these co-isolated contaminants is critical to successful DNA manipulation, and extraction and purification methods should be selected to yield DNA suitable for the ultimate metagenomic application.

3. Metagenomic applications

Microorganisms in natural environments may contain genes that encode and express biosynthetic or biodegradative pathways of interest that have never been identified using culture-dependent methods. One strength of the metagenomic approach is in enabling researchers to investigate the phylogenetic and functional diversity of microorganisms at the community level, independent from cultivation-associated biases (Schloss and Handelsman, 2003; Cowan et al., 2005).

3.1. Natural product discovery: enzymes

Enzymes expressed from cultured soil microorganisms have been harvested and used commercially for many decades. High-throughput screening of environmental metagenomic DNA libraries has led to the discovery of many novel enzymes that are of great use in industrial applications. Indeed, the very first metagenomic study involved the identification of cellulases from a bioreactor "zoolibrary" (Healy et al., 1995). There are many examples of enzymes discovered via a metagenomic approach, such as a multifunctional glycosyl hydrolase identified from a rumen metagenomic library (Palackal et al., 2007), low pH, thermostable α -amylases discovered from deep sea and acidic soil environments (Richardson et al., 2002), pectinolytic lyases from soil samples containing decaying plant material (Solbak et al., 2005), agarases from soil (Voget et al., 2003) and lipolytic enzymes such as esterases and lipases (Rondon et al., 2000; Voget et al., 2003; Lee et al., 2004; Ferrer et al., 2005). In another study, 137 unique nitrilases were discovered from screening environmental (terrestrial and aquatic) DNA libraries using high-throughput and culture-independent methods (Robertson et al., 2004). A novel β -glucosidase gene isolated by screening a metagenomic library derived from alkaline polluted soil was found to be a first member of a novel family of β -glucosidase genes (Jiang et al., 2009). The discovery of a diverse set of genes that encode enzymes for cellulose and xylan hydrolysis from the resident bacterial flora of the hindgut paunch of a wood-feeding 'higher' termite (*Nasutitermes* sp.) and from moths was a result of metagenomic analysis (Brennan et al., 2004; Warnecke et al., 2007). In each of these studies, it should be noted that the rate of discovery is generally less than one clone with

activity per 1000 clones screened; therefore, the anticipated "hit rate" for any enzymatic activity should be considered prior to initiating metagenomic library screening. These are just a sampling of the many enzymatic activities discovered from metagenomes, providing ample evidence of the potential of this approach for the discovery of novel biocatalysts from the environment.

Mining for biocatalysts from metagenomic libraries usually involves three different strategies: 1) homology-driven metagenome mining based on high-throughput sequencing, 2) substrate-induced gene expression (see Section 3.4), or 3) function-based screening. Unlike chemical synthesis, biocatalysis does not include the use of toxic chemical reagents. The discovery of novel enzymes through these approaches is an economical and potentially environmentally responsible way to decrease the use of toxic chemicals traditionally used in many industries. This approach for enzyme discovery can help improve the efficiency of existing techniques and also enable novel processes for the production of various chemicals that serve as precursors in the synthesis of pharmaceuticals, insecticides, fertilizers, herbicides, etc.

3.2. Natural product discovery: antibiotics

As-yet-uncultured microorganisms are an untapped reservoir for the discovery of secondary metabolites such as antibiotics (Gillespie et al., 2002). The biosynthetic pathways encoding the secondary metabolites can be captured by cloning large fragments of contiguous metagenomic DNA into heterologous hosts that are easier to manipulate *in vitro*, such as *Escherichia coli* (Rondon et al., 2000; Gillespie et al., 2002; Liles et al., 2004). Many low molecular weight molecules are produced during specific growth phases such as during developmental stages or starvation (Clardy and Walsh, 2004) and exhibit bioactive properties. A diverse class of secondary metabolites is the polyketides (Moffitt and Neilan, 2003; Ginolhac et al., 2004; Schirmer et al., 2005; Wawrik et al., 2005), produced by modular enzymatic pathways with phenomenal structural heterogeneity and yet with some conserved DNA sequences that allow their identification via nucleic acid probes (see Section 5.2).

The adoption of heterologous hosts besides *E. coli* permits expression of cloned DNA from diverse sources. *Streptomyces* species and other *Actinobacteria* have been used as screening hosts for soil DNA libraries because of their ability to express diverse polyketide and other bioactive secondary metabolites and their relative ease of genetic manipulation (Martinez et al., 2005). For example, the antibiotic terragline with anti-*Mycobacterium* activity was discovered via heterologous expression of metagenomic clones within a *Streptomyces lividans* host (Wang et al., 2000). Another study introduced Type II PKS pathways, recovered from a metagenomic library, into *S. lividans* and *Staphylococcus albus* hosts, resulting in the production of clone-specific metabolites (King et al., 2009). Beyond the well-characterized metabolites of *Actinobacteria*, many other bacterial divisions may also prove to be prodigious producers of antibiotics, and serve as alternative hosts. In a study that expressed metagenomic libraries in multiple *Proteobacteria* hosts, the antimicrobial products detected in each host were distinct, supporting the contention that each heterologous host may yield a novel range of expressed metabolites from a given metagenomic library (Craig et al., 2010). Other studies have discovered metagenomic clones producing long-chain N-acyltyrosine antibiotics (Brady et al., 2004), antifungal agents (Chung et al., 2008), and triaryl cation antibiotics turbomycin A and B (Gillespie et al., 2002). The latter study, while using *E. coli* expression, is an example of the unique chemistry that may be derived from the combination of host metabolites (i.e., *E. coli* produced indole) and metagenomic clone chemistry (i.e., melanin pigment production).

3.3. Bioremediation

Xenobiotics include compounds such as antibiotics, pesticides, hormones, and other foreign biological or chemical contaminants that can affect a microbial community. Other examples of xenobiotics include aromatic compounds and their derivatives, and polychlorinated biphenyls (PCBs), anthropogenic chemical pollutants that persist in the environment and are recalcitrant to complete removal. Xenobiotic degradation can be achieved by biotic and/or abiotic reactions, and may be accelerated by harnessing microbial degradative activities to biostimulate or bioaugment the natural attenuation of environmental contaminants (Vogel, 1996; Cosgrove et al., 2010). The application of metagenomics may aid in the isolation of novel catabolic pathways for degradation of xenobiotic compounds, indicating the functional genetic capacity for contaminant degradation and providing molecular tools useful for identification of the microbial taxa encoding the biodegradative gene(s).

A combined approach using metagenomics and other molecular techniques is commonly used to study microorganisms useful for bioremediation of environmental contaminants. Labeled substrates have been used to target and recover genes from populations involved in the degradation process (Sul et al., 2009). This group used [¹³C]-labeled biphenyl to identify biphenyl dioxygenase genes from bacteria capable of growing in PCB-contaminated river sediments. Other metagenomic studies have identified catabolic pathways that encode nitrilases, which play an important role in both biosynthetic and catabolic reactions (Robertson et al., 2004), as well as enzymes with catalytic properties that degrade organic contaminants (Kim et al., 2007).

3.4. Strategies to improve the isolation of biosynthetic and catabolic pathways

Extraction of total metagenomic DNA and cloning to construct libraries requires extensive labor, time, and resources. The number of positive clones obtained from screening these libraries for the presence or expression of a specific gene or function is often very low because the target pathways comprise a small percentage of the total cloned DNA, and only a subset of the cloned genes may be expressed in a given heterologous host. There are various strategies that can be employed prior to library construction and/or screening that can improve the frequency of biosynthetic or catabolic pathway isolation. Although these methods may result in a loss of considerable diversity from the environmental sample, they also have the power to select for a particular population or function(s) of interest. The loss of diversity can be mitigated by altering the degree of the selective pressure criteria used.

A commonly applied strategy is to enrich the environmental sample for microbial populations capable of growth on certain substrates or for survival under different physicochemical conditions. Use of a selective medium will result in favorable growth and enrichment of the targeted population due to specific substrate utilization, as well as potentially other metabolically co-dependent microbial populations. Direct cloning from enrichment cultures enables studying metabolic activities of microbial assemblages and selection for specific microbes that produce an enzyme or compound of interest (Healy et al., 1995). This approach has also been used in the identification of biotin synthesis genes by isolation of clones carrying the biotin biosynthesis operon (Entcheva et al., 2001).

Stable isotope probing (SIP) is an approach that enriches the DNA (or RNA) of microorganisms that can utilize a stable isotope (e.g., ¹³C-glucose) and incorporate the isotope into newly synthesized nucleic acids (Radajewski et al., 2000; Dumont et al., 2006). The isolated “heavy” DNA from the treated environmental sample is subjected to density gradient centrifugation to separate the

¹³C-labeled DNA for analysis, which may then serve as DNA template within a PCR to identify the microorganisms that have incorporated the labeled substrate (Radajewski et al., 2000). Metagenomic analysis in conjunction with SIP can access a multitude of functional genes since the labeled DNA is enriched for the genomes of microbial populations with specific metabolic capabilities (Wellington et al., 2003). DNA-SIP has also been used to retrieve genomic fragments of an active population by cloning the ¹³C-labeled DNA without initial PCR amplification (Dumont et al., 2006). However, SIP has its limits and biases, such as dilution of the labeled substrate with unlabeled substrates and cross-feeding of ¹³C-labeled metabolic intermediates by other organisms (Radajewski et al., 2000). When using DNA-SIP for metagenomic analyses, the small amount of heavy DNA available can also be a hurdle to successful library construction. To overcome this challenge, methods such as multiple-displacement amplification (Dumont et al., 2006; Chen et al., 2008) and community growth enrichment by sediment slurries (Kalyuzhnaya et al., 2008) have been used to increase the amount of heavy DNA available for analysis. Despite these challenges, SIP coupled with metagenomics is an excellent culture-independent strategy to identify functional genes involved in the utilization of a variety of compounds or in degradation of environmental pollutants.

Another promising approach for identification of catabolic pathways has been described as substrate-induced gene expression screening (SIGEX). SIGEX identifies clones from an operon-trap metagenomic library that are induced in the presence of a specific substrate, resulting in green fluorescence protein expression that can be detected using fluorescence-activated cell sorting (Uchiyama et al., 2005). There are limitations of the SIGEX approach due to its dependence on *in cis* regulatory factors that are active within *E. coli* (de Lorenzo, 2005), so this rapid screen should not be considered an exhaustive survey, as is the case with any metagenomic analysis.

3.5. Metagenomics in quorum-sensing regulation studies

Quorum sensing (QS)-mediated bacterial responses to cell density are specific to each bacterial species, and are important in understanding bacterial pathogenesis and other bacterial phenotypes in natural environments (e.g., bioluminescence of *Vibrio fischeri* within the light organ of the *Euprymna* squid). The use of a metagenomic approach to study QS regulation in the soil environment was pioneered by Williamson et al. (2005) wherein they identified clones producing unknown molecules that activated QS-regulated genes. Clones of interest were identified using a high-throughput intracellular screen, i.e., the metagenomic DNA is within a host cell that contains a biosensor responsive to compounds inducing QS. Another study identified a clone that degraded N-acylhomoserine lactone (NAHL) from screening a pasture soil metagenomic library (Riaz et al., 2008). The identified gene was shown to encode a lactonase with NAHL degrading ability and the gene product efficiently quenched quorum-sensing-regulated pathogenic functions when expressed in *Pectobacterium carotovorum*. Metagenome-derived clones isolated in another study were found to encode novel lactonase family proteins interfering with QS (Schipper et al., 2009), that when expressed in *Pseudomonas aeruginosa* successfully inhibited motility and biofilm formation. Lastly, metagenomic libraries constructed with DNA isolated from activated sludge and soil have been screened using an *Agrobacterium* biosensor strain, resulting in the isolation of three unique clones with novel QS synthase genes (Hao et al., 2010).

4. Analyzing the soil metagenome

A variety of approaches may be employed for analyzing the soil metagenome, depending on the specific aims of the study.

The ultimate downstream application should dictate the methods used for soil sampling, DNA extraction and purification, and library construction and screening (if necessary). The biologically, chemically, and physically heterogeneous nature of soils presents many challenges to the successful characterization of its microbial metagenome. Representative coverage of the soil microbial community requires isolation and cloning of a large amount of DNA from a small sample, and depends on insert size and the number of clones. It has been estimated that the number of plasmid clones (5 kb average insert size) and BAC clones (100 kb average insert size) required for representative coverage of the diverse soil microbial community in 1 g of soil is 10^7 and 10^6 respectively (Handelsman et al., 1998). This is of course based on the assumption that all species in a soil environment are equally abundant. Since members of a community are rarely equally represented, the metagenomic library with minimum coverage is more likely to represent only the abundant species. In order to achieve substantial representation of the genomes from rare members of the soil community, a 100- to 1000-fold coverage of the metagenome is needed in library construction (Riesenfeld et al., 2004b). Since this translates to about 10,000 Gb of soil DNA, or 10^{11} BAC clones, it is not reasonable to suppose that bacterial taxa present in lower abundance will be represented within a metagenomic library unless an enrichment method is used. Also, when working with soil samples that have not been well-characterized, it is advisable to utilize a variety of different methods for DNA extraction and purification to empirically determine the ideal combination that will yield high-quality and high-diversity metagenomic DNA. Here, we discuss many of the metagenomic-based approaches used to study soil microbiology, as well as the approach-specific factors to consider when performing such analyses.

4.1. Sequencing

The use of PCR has become routine for molecular phylogenetic analysis based on ribotype diversity (Woese, 1987), often used in combination with community analysis methods such as denaturing gradient gel electrophoresis (e.g., Muyzer et al., 1993), 16S rRNA gene clone libraries (e.g., Chandler et al., 1997), or more recently microarrays (DeSantis et al., 2007; Liles et al., 2010). Although in many cases such studies are described as “metagenomic”, since indeed the template DNA used is derived from diverse genomes, such phylogenetic surveys of a single evolutionarily conserved gene are not truly metagenomic in nature and will not be further considered in this review.

Pyrosequencing and other next-generation approaches offer the capacity for massively parallel sequencing of metagenomic samples (Ronaghi, 2001). The accuracy of pyrosequencing is comparable to that achieved via Sanger sequencing (Huse et al., 2007), but it is more cost- and time-effective per sequenced nucleotide (Hugenholtz and Tyson, 2008), and sequencing read length has been gradually increasing with each iteration of sequencing technologies (Margulies et al., 2005). The increased availability of high-throughput sequencing technologies has made it possible for scientists to gain access to the genetic diversity within environmental communities (Sogin et al., 2006). Pyrosequencing has been used in the investigations of microbial diversity in soil (Roesch et al., 2007), deep sea ecosystems (Sogin et al., 2006) and phage populations from various environments (Dinsdale et al., 2008).

Because pyrosequencing relies on an amplification process, the same environmental contamination challenges that apply to PCR-based applications also apply to pyrosequencing. However, since pyrosequencing currently generates reads only 300–500 bp in length, obtaining intact, larger DNA is not critical (Metzker, 2005). Most commercially available soil DNA extraction methods yield

DNA suitable for pyrosequencing, and if needed, further DNA purification methods can be employed (see Section 4.2.1). Alternatively, the DNA template can be diluted (along with contaminants) to permit PCR amplification (Altshuler, 2006), or bovine serum albumin can be added to the reaction mixture to prevent humic acid-mediated inhibition (Kreider, 1996).

Read length is a critical factor in the probability that a metagenomic sequence will have a significant hit within GenBank or other database (Wommack et al., 2008). Even for a pure bacterial culture, it is not uncommon for a completely sequenced bacterial genome to have 35–45% of predicted open reading frames (ORFs) with no significant homolog in GenBank (Schwartz, 2000). This problem is only exacerbated with metagenomic sequences, with an even larger proportion of metagenomic sequences from soil and other environments having no significant BLAST homolog (Venter et al., 2004; Tringe et al., 2005; Pignatelli et al., 2008). Even with the difficulty in interpretation of much of the sequences within metagenomic datasets, substantial information related to the genomic composition, and predicted functions and metabolic pathways, of microbial communities has been unearthed from deep-sequencing approaches (Breitbart et al., 2002; Tyson et al., 2004).

4.2. Small-insert libraries

The construction and analysis of small-insert metagenomic libraries (less than ~10 kb average insert size) is a useful approach to identify gene product(s) encoded by a relatively small genetic locus, such as most enzymes, or genetic determinants of antibiotic resistance (Riesenfeld et al., 2004a; Parsley et al., 2010). Biases in cell lysis and cloning techniques may select against some prokaryotic taxa or gene products that are toxic to the host cell; therefore, it is important to select DNA extraction and cloning methods designed to yield a high proportion of DNA from the microorganisms of interest. Refer to Fig. 1 for a schematic representation of the overall steps involved in metagenomic library construction.

4.2.1. Selection of vectors and host organisms

Vectors used for the construction of small-insert libraries often possess a promoter for transcription of the cloned gene inserts and should be compatible with the host selected for screening. A vector with two promoter sites flanking the multi-cloning site facilitates gene expression that is independent of gene orientation and the promoters associated with inserts (Lammle et al., 2007). With the possibility of the expressed gene product having toxic effects on the host organism, it is important to regulate the expression levels of the cloned genes, which can be achieved by using vectors with inducible control over gene expression of the insert or plasmid copy number (Sukchawalit et al., 1999; Saida et al., 2006).

An additional issue to consider when selecting a vector is its ability to replicate in multiple hosts to enable heterologous expression of specific gene(s) of interest. Although the utility of using *E. coli* as a heterologous host for metagenomic library construction has been well-established (Rondon et al., 2000; Pfeifer and Khosla, 2001; Gillespie et al., 2002; Liles et al., 2004), other bacterial hosts may be more suitable for some applications, particularly if the percent G + C content of the cloned gene(s) are significantly different from that of *E. coli*, or if the regulatory factors required for expression or the biosynthetic capacity may be enhanced within another prokaryote.

4.2.2. Preparation of DNA for cloning

The preparation of DNA for small-insert libraries is similar to that used for PCR- or pyrosequencing-based applications. A sufficient

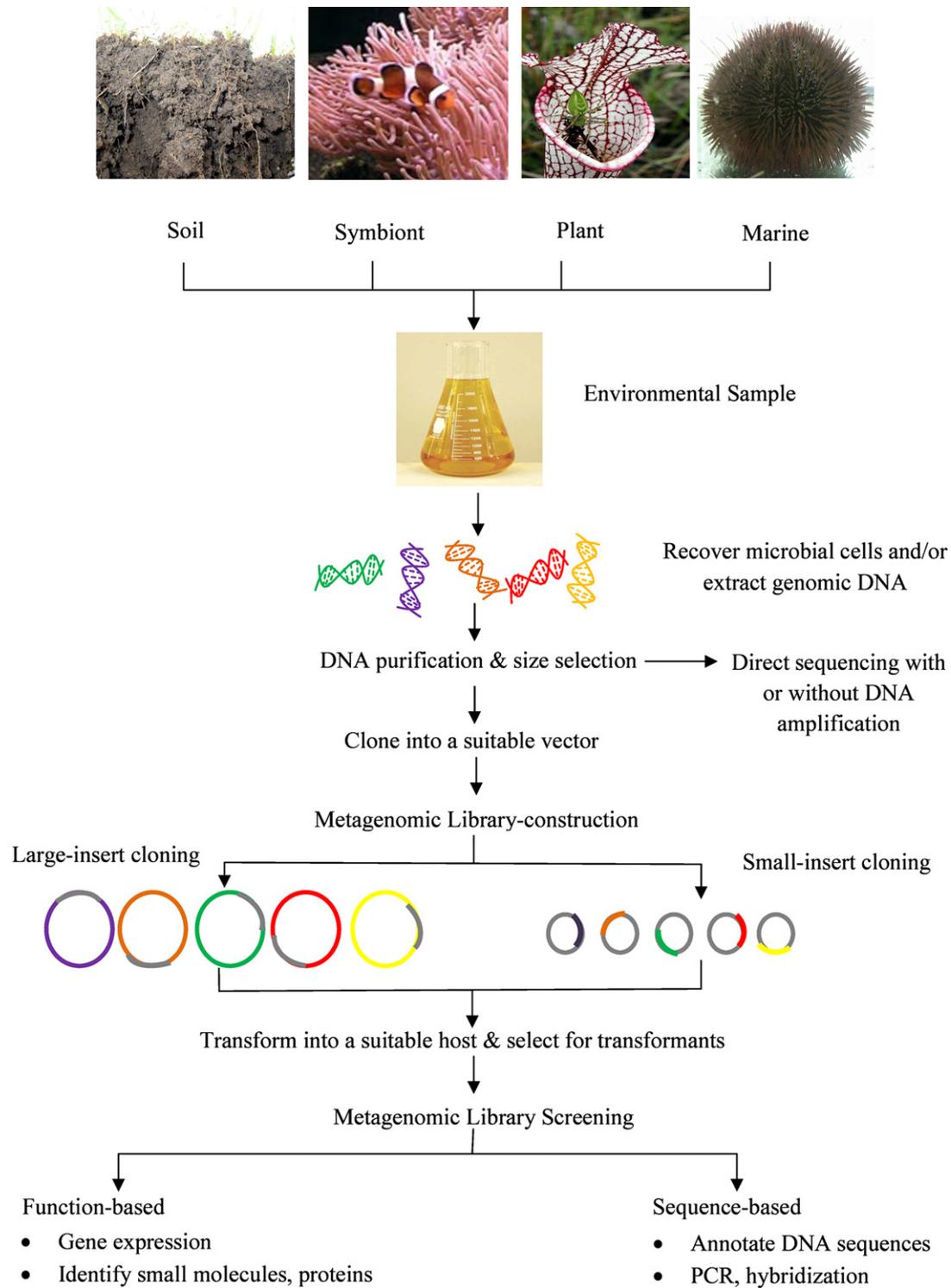


Fig. 1. Schematic representation of the steps involved in construction of metagenomic libraries. (Plant and marine photos courtesy of Dr. Debbie Folkerts and Dr. Kevin Fielman, respectively).

yield of DNA is necessary for successful library construction, and soil contaminants co-isolated with the DNA such as humic acids can interfere with efficient cloning. DNA extraction and purification conditions should be harsh enough to lyse a variety of microbes and remove the majority of contaminants, while the degree of DNA fragmentation that is permissible will depend on the desired average insert size of the library. If the desired average insert size is less than

20 kb, a commercial kit (e.g., MoBio Laboratories, Qiagen) may provide a useful method for obtaining DNA of sufficient size, purity, and yield for small-insert cloning. One study using Antarctic top soil used two separate commercial kits to further purify the DNA after cell lysis for construction of small-insert libraries (Cieslinski et al., 2009). In cases when commercial kits are not suitable, such as soils with high clay content, it may be advisable to adopt cell-based

("indirect extraction") methods such as sucrose/Percoll density gradient centrifugation or Nycodenz treatment, which have been shown to generate DNA appropriate for small-insert cloning (Bakken and Lindahl, 1995).

Regardless of which DNA extraction method is used, it is possible that further purification will be required for efficient cloning. Many DNA purification methods may be effective in yielding DNA suitable for cloning, such as phenol and chloroform extraction, and/or treatment with hexadecyltrimethylammonium bromide (CTAB) or polyvinylpyrrolidone (PVPP), which may be combined with CsCl density centrifugation or hydroxyapatite column chromatographic purification (Holben et al., 1988; Selenska and Klingmuller, 1991; Knaebel and Crawford, 1995; Roose-Amsaleg et al., 2001; Lee et al., 2004). However, it has been shown that many of these methods (i.e., PVPP addition, CsCl density centrifugation, and hydroxyapatite column chromatographic purification) resulted in a decreased DNA yield (Steffan et al., 1988). In the case of indirect extraction methods, some studies have found that a washing step prior to cell lysis is useful for the removal of soluble inhibitors and extracellular DNA (Xia et al., 1995; Harry et al., 1999). Unfortunately, many soil samples require a combination of these purification steps, which significantly increases processing time and can lead to an even greater loss of DNA. For example, one study compared DNA extracted from five different soils with various organic matter contents and found that the samples with the highest organic matter content required five purification steps to yield sufficiently pure DNA (Van Elsas et al., 1997). Following extraction and purification of the DNA, it may be physically sheared or partially restriction digested and then size-selected by extracting the DNA in the desired size range from an agarose gel (Riesenfeld et al., 2004b; Lammle et al., 2007). Because the size-selected DNA will likely be less than 20 kb, it can be column-purified, the gel slices may be treated with GELase enzyme (Epicentre), or the DNA may be electroeluted from the gel prior to cloning (Osoegawa et al., 1998).

4.3. Large-insert libraries

Large-insert metagenomic libraries contain large, contiguous DNA fragments that have the potential to contain intact biosynthetic pathways involved in the synthesis of antimicrobial compounds, multiple enzymes with catabolic activity, or operons encoding other complex metabolic functions. However, along with potential advantages for some applications, large-insert cloning from soil microorganisms also presents many technical challenges in order to obtain and screen high-quality metagenomic libraries containing DNA from representative microorganisms.

4.3.1. Selection of vectors and host organisms

Because the applications appropriate for large-insert metagenomic libraries depend on their ability to capture large, intact genetic pathways, the selection of an appropriate cloning vector is critical to the maintenance and expression of the cloned pathways. Several vector options exist for cloning HMW DNA from environmental samples, such as cosmids, fosmids, and BACs. The cosmid, a hybrid plasmid that contains *cos* sequences from the λ phage genome; was one of the first vectors used for cloning (Collins and Hohn, 1978). The packaging capacity of cosmids varies depending on the size of the vector itself but usually lies around 40–45 kb. While typical plasmids can maintain inserts of 1–20 kb, cosmids are capable of containing DNA inserts of about 30 kb up to 40 kb. The size limits ensure that vector self-ligation resulting in empty clones is not a problem. Both broad host range cosmids and shuttle cosmids are available (Craig et al., 2009). Cosmids can replicate like plasmids when they contain a suitable origin of replication and

they commonly possess selective genes such as antibiotic resistance to facilitate screening of transfected cells. Fosmid vectors, which are similar to cosmids but are based on the *E. coli* F-factor replicon, were developed for constructing stable libraries from complex genomes (Kim et al., 1992). The low copy number of fosmid vectors offers higher stability than comparable high-copy number cosmids. A low copy number is optimal for long-term survival of the plasmid in a host. Also, plasmid copy number determines gene dosage. Recombinant clones from large-insert libraries may express gene products that are toxic to the host and hence it is important to maintain libraries in single copy until screening for a function. Fosmid copy number is tightly regulated in *E. coli* to 1–2 copies per cell, and fosmids can typically accommodate cloned inserts between 40 and 50 kb. BAC vectors are based on the same F-factor replicon but have the capacity to maintain large inserts in excess of 100 kb (Shizuya et al., 1992). Along with the long-term stability conferred by the F-factor for maintenance, a modified BAC vector also containing an RK2 origin of replication is capable of inducible copy number, alternating between single-copy and high-copy BAC maintenance (Wild et al., 2002). The inducible-copy phenotype can have significant advantages for the yield of DNA from metagenomic clones, and potentially for expression of cloned genes.

Although fosmid vectors are limited in insert size compared to BAC vectors, their significantly higher cloning efficiency enables construction of metagenomic libraries with many thousands of transductants. Conversely, BAC vectors even though capable of accommodating higher insert sizes have lower cloning efficiency than that of fosmid vectors. As mentioned previously, HMW DNA for fosmid-based cloning may be treated with harsher extraction and purification methods, which could yield a higher concentration of DNA from more diverse microorganisms than that of DNA isolated for BAC-based cloning. However, because BAC vectors can stably maintain cloned inserts hundreds of kilobases in size, they offer a greater chance of isolating intact pathways or of linking phylogenetic and functional genetic information (Stein et al., 1996). Therefore, the predicted size of the pathway of interest, its native level of activity, and its relative abundance within the community must be considered when choosing a suitable cloning vector for large-insert metagenomic library construction.

As with small-insert libraries, *E. coli* is the preferred host for the construction of large-insert metagenomic libraries due to its high cloning efficiency. This host has been successfully used to express many bioactive enzymes and compounds in metagenomic studies (Handelsman et al., 1998; Heath et al., 2009). In addition, *S. lividans* has been used as a heterologous host for library screening, and it has more stringent promoter recognition and regulation properties when compared to *E. coli* (Martinez et al., 2005). Because large-insert libraries may contain clones that express gene products that are toxic to the library host, it is important to maintain libraries in single copy until screening for a function and to consider the use of multiple hosts to increase the probability of identifying and characterizing the function(s) of interest. It has been shown that clones positive for a specific activity detected using one host may not be detected in a different host and vice-versa (Li and Qin, 2005; Wang et al., 2006; Craig et al., 2009, 2010). A range of Gram-positive and -negative bacteria can be used as hosts for heterologous expression, and the corresponding vectors selected should be compatible with those hosts (Sosio et al., 2000; Martinez et al., 2004; Hain et al., 2008). Vector systems such as pRS44 enable shuttling into other Gram-negative hosts and have higher potential for function-based screening across species barriers and heterologous gene expression (Aakvik et al., 2009). Several other factors are necessary for successful expression of the cloned pathways (e.g., co-factors,

post-translation modification enzymes, inducers, chaperones etc.), which may be provided by the vector or the host organism.

4.3.2. Preparation of DNA for cloning

Large-insert metagenomic libraries are the most challenging to construct, but also can provide significant advantages for some applications since they enable identification and characterization of intact functional pathways encoded on large, contiguous DNA fragments (Stein et al., 1996; Beja et al., 2000b; Rondon et al., 2000; Courtois et al., 2003). All of the considerations discussed previously regarding the selection of DNA extraction and purification methods apply to large-insert cloning, along with an additional critical issue: the construction of large-insert metagenomic libraries depends on obtaining sufficiently pure DNA of high molecular weight (in excess of ~100 kb). However, most extraction and purification methods result in DNA significantly smaller than this size (Tien et al., 1999; Wellington et al., 2003; Miller and Day, 2004). Although a few methods can yield DNA from soil greater than 1 Mbp in size (Berry et al., 2003; Liles et al., 2004), it has been demonstrated that these indirect extraction methods can result in inefficient cloning due to contaminants that may be co-isolated with the metagenomic DNA and require further purification.

The successful recovery of high molecular weight (HMW) metagenomic DNA from soil microorganisms presents many extraction and purification challenges. A primary goal is to obtain DNA from an assemblage of diverse bacterial cells that are representative of the soil microbial community DNA. However, the harsh extraction methods (i.e., bead-beat lysis) typically employed for PCR or small-insert cloning applications will result in substantially fragmented DNA that is much too small for large-insert cloning. The use of indirect DNA extraction methods can somewhat alleviate this dilemma by first separating the cells from the soil sample, embedding them in an agarose plug, and then carefully lysing the cells and purifying the resulting DNA rather than performing the extraction *in situ*. Repeated homogenization and differential centrifugation are often sufficient to separate the cells from the soil sample (Faegri et al., 1977; Hopkins et al., 1991), although other dispersion methods include the use of cation-exchange resin (Macdonald, 1986; Jacobsen and Rasmussen, 1992) and incubating the soil with sodium deoxycholate or polyethylene glycol (Liles et al., 2008). Another novel method that is capable of selectively concentrating DNA within a gel while rejecting high concentrations of contaminants is SCODA (Pel et al., 2009), but the quantities of DNA capable of being extracted may not be sufficient for cloning without further amplification.

The choice of extraction and purification method also depends on which cloning vector will be employed, such as a fosmid or bacterial artificial chromosome (BAC). Metagenomic libraries constructed in a fosmid vector are introduced into their heterologous host using a λ phage-based packaging system, which limits the clone insert size to 40–50 kb. Although DNA isolated for fosmid libraries must be treated carefully to prevent excessive shearing of DNA, using a fosmid vector does allow the use of harsher extraction and purification methods than those that may be used for BAC cloning. Also, during fosmid library construction, the DNA is typically size-selected by physically shearing the DNA into fragments of a desired length rather than by restriction digestion. This “direct size-selection” method eliminates the need for gel extraction (which can lead to DNA loss) and the possibility of DNA degradation due to over-digestion. An alternative to the physical shearing method was proposed by Quaiser and colleagues, who constructed fosmid libraries containing soil metagenomic DNA contaminated with humic and fulvic acids by embedding the DNA in agarose, electrophoresing the DNA through agarose containing PVP, and then combining the subsequent removal of the PVP with the size-selection step which resulted in purified,

“clonable” DNA in the 30–100 kb size range (Quaiser et al., 2002). In combination with other purification steps, the inclusion of a formamide plus NaCl treatment was shown to significantly increase the efficiency of cloning of large DNA fragments into fosmid or BAC vectors (Liles et al., 2008). Factors that have been demonstrated to affect the size of recovered DNA include not only the DNA extraction method used but also the microbial growth status and chemical composition of the soil (Bertrand et al., 2005). In general, DNA extracted from bacterial cells is significantly larger than DNA directly extracted from soil but is also found in lower yields (Liles et al., 2008); however, this loss can be reduced by using wide-bore pipette tips to prevent shearing of DNA, performing multiple rounds of indirect extraction on each soil sample, minimizing the amount of agarose that is retained during size selection, or using electroelution as an alternative to extraction of DNA from the agarose gel (Osoegawa et al., 1998).

5. Metagenomic library screening

The analysis of metagenomic libraries involves two main strategies, function-based or sequence-based screening. The choice of screening method depends on many factors, including the type of library constructed, the genetic loci or functional activity of interest, and the time and resources available to characterize the library. Both approaches offer advantages and disadvantages, which will be discussed here.

5.1. Function-based screening

Function-based methods involve screening a metagenomic library to detect the expression of a particular phenotype conferred on the host by cloned DNA (Henne et al., 1999). Because the frequency of discovering active pathways from metagenomic libraries is often low, high-throughput screening of library clones is the most efficient approach for function-based detection of activity. By screening on indicator media, *E. coli* recombinant clones that express a novel phenotype (not already encoded on the *E. coli* genome) may be recognized. As opposed to high-throughput screening methods, a direct selection for a positive clone that has acquired resistance to an antibiotic or heavy metal can be performed by excluding microorganisms that are unable to grow in the presence of these selective compounds (Riesenfeld et al., 2004b; Mirete et al., 2007; Parsley et al., 2010).

Another approach for functional screening of metagenomic libraries is to use host strains or mutants of host strains that require heterologous complementation for growth under selective conditions (Simon and Daniel, 2009). Growth is exclusively observed in the case of recombinant clones that possess the gene of interest and produce an active product. This strategy has been applied for the detection of enzymes involved in poly-3-hydroxybutyrate metabolism (Wang et al., 2006), DNA polymerase I (Simon et al., 2009), operons for biotin biosynthesis (Entcheva et al., 2001), lysine racemases (Chen et al., 2009), glycerol dehydratases (Knietzsch et al., 2003) and naphthalene dioxygenase (Ono et al., 2007).

Screening can also be performed by detecting a specific phenotypic characteristic, in which individual clones are assayed for a particular trait. Incorporation of specific substrates in the growth medium will allow the identification of the corresponding enzymatic activity encoded by a metagenomic clone(s). Examples include the identification of esterases (Elend et al., 2006; Chu et al., 2008) by formation of a clear halo around a colony on the indicator medium and the identification of extradiol dioxygenases by the production of a yellow compound (Suenaga et al., 2007). Metagenomic clones expressing an antimicrobial activity may be detected by growth inhibition assays of a suitable tester organism using soft agar

overlays over the clone colonies or a microtiter plate assay using the supernatant extracts from the clone cultures (Rondon et al., 2000; Courtois et al., 2003; Brady et al., 2004; Craig et al., 2009). As discussed previously, SIGEX is an additional functional screening approach in order to identify genes for substrate catabolism.

Although function-based screening is a powerful tool to identify novel natural products or metabolic activities from as-yet-uncultured organisms, it is often limited by a number of obstacles that may be difficult to overcome. Detecting a recombinant clone that expresses a gene product will depend upon successful gene transcription, translation, protein folding, and secretion from the host organism. By adopting high-throughput screening protocols and multiple heterologous expression hosts, the probability of discovering the function(s) of interest may be improved.

5.2. Sequence-based screening

Sequence-based screening involves direct sequencing of metagenomic DNA, either with or without cloning prior to sequencing and then subjecting the sequences to bioinformatic analyses (Kunin et al., 2008; Sleator et al., 2008). Practically speaking, a sequence-only approach to metagenomics involves significantly less laboratory bench work, relative to cloning-based approaches. Recent developments in next-generation sequencing (NGS) technologies have made available a number of methods that can be used for sequencing, although with varying costs and capabilities. GS20, the first instrument based on the 454 pyrosequencing technology was shown to sequence up to 25 million bases of a bacterial genome in a 4 h run, with average read lengths of 110 bp and 96% raw read accuracy (Margulies et al., 2005). A current model 454 GS-FLX sequencer using Titanium chemistry can achieve read lengths of up to 500 bp, with future improvements in read length expected. By comparison, the Illumina Solexa platform based on fluorescently labeled sequencing by synthesis generates 35–76 bp on average. The latest version of the short read sequencer from Applied Biosystems, called the SOLiD4, generates 100 Gb per run with read length of 50 bp. Though NGS technologies provide good overall coverage for single genomes, the short read lengths can be a serious limitation for efficient assembly of metagenomic sequences. The cost per megabase is highest for 454 sequencing at approximately \$10 followed by Solexa and SOLiD at about \$5 and \$2, respectively (Rothberg and Leamon, 2008). With the rapid growth and developments in this field it is very likely that the cost and read estimates will keep changing as NGS technology advances. The selection of a metagenomic strategy should be informed by the degree to which the gene(s) of interest are expected to be identified from a sequence-only approach; the interest (or lack thereof) in obtaining functional cloned genes, and the availability of time and resources for the project. As the cost per base pair of sequence has dropped dramatically through adoption of NGS technology, this has enabled large scale sequencing efforts accessible to individual academic researchers. Still, sequence data analysis can consume more time and resources than are initially anticipated. Fortunately, bioinformatics approaches to analyze metagenomic datasets have been developed that allow rapid comparative analyses.

With the enormous amount of sequence data generated by these different approaches, it is very important to have bioinformatics tools for such high-throughput sequence pipelines. Metagenomic studies must first curate the sequence reads to obtain data of sufficient quality, eliminating ambiguous base pairs and any vector or adaptor sequences. The edited sequences can then be used for gene prediction, and if desired, contig assembly. Given the non-exhaustive nature of most metagenomic sequence datasets, especially for analysis of soil communities, it is expected that contig assembly will be of limited benefit. For very diverse microbial assemblages, contig

sizes will be relatively short, and chimeric contigs will likely be present at a high frequency. Once high-quality metagenomic sequences are available, they can be deposited within sequence databases (e.g., GenBank env) and compared against other environmental metagenomic datasets. A useful tool for accessing metagenomic information is CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis), developed to serve the needs of the microbial ecology research community by creating a data repository and a bioinformatics resource to facilitate metagenomic sequence data storage, access, analysis, and synthesis (Smarr, 2006). A freely available open source system that can process metagenome sequence data is the metagenomics RAST server (MG-RAST) (Meyer et al., 2008). The MG-RAST server compares protein as well as nucleotide databases for functional assignments of sequences in the metagenome accompanied by a phylogenetic summary. Just like next-generation sequencing technology enabled generation of vast amount of sequence data, tools like MG-RAST have enabled high-performance computing for annotation and analysis of metagenomes.

There are available bioinformatics tools for gene prediction, such as MEGAN (MEtaGenome ANalyzer), a program that compares a set of DNA reads (or contigs) against databases of known sequences using comparative tools such as BLAST algorithms. MEGAN can then be used to compute and interactively explore the taxonomical content of the dataset by using NCBI taxonomy to summarize and order the results (Huson et al., 2007). Once a dataset of metagenomic sequences with significant GenBank hits has been assembled, these sequences can then be categorized by a subsystems approach using SEED to organize predicted gene functions according to related biological processes (Overbeek et al., 2005). SEED enables rapid annotation of metagenomic sequences according to similarity to previously known gene products. The predicted genes may also be assigned a phylogenetic classification using *Treephyler* for rapid taxonomic profiling of metagenomic sequences (Schreiber et al., 2010).

With each of these bioinformatics tools and approaches, it should be acknowledged that the predictive power of the sequence analysis is limited by the previously described gene functions available in public databases and that many putative functions may be inaccurately annotated. While this potential source of bias does affect the utility of a sequence-based approach to metagenomics, such intensive sequence-driven surveys of natural environments have profoundly affected our collective view of prokaryotic diversity and the extent of functional genetic diversity that has yet to be understood in terms of biological functionality (Venter et al., 2004).

6. Conclusions

The development of metagenomic approaches has provided an unprecedented level of access to microbial genomes from many different environments, making it possible to characterize the phylogenetic and functional diversity of as-yet-uncultured microorganisms from various biomes of interest. Because of its complex and dynamic nature, soil presents unique challenges for metagenomic applications. Selecting the most suitable combination of soil sampling, DNA extraction and purification, cloning and/or sequencing method that is most appropriate for the metagenomic study should begin with consideration of the ultimate desired outcome, for an application-driven approach to soil metagenomics.

The use of cutting-edge metagenomic-based technologies to access soil microbial communities has led to a remarkable increase in the discovery of pathways that encode diverse gene products, such as enzymes and antimicrobial compounds. Soils are expected to be a continuing rich resource of novel genetic and functional pathways of use and interest to academia and industry.

References

- Aakvik, T., Degnes, K.F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Volker, U., Ellingsen, T.E., Valla, S., 2009. A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiology Letters* 296, 149–158.
- Altshuler, M.L., 2006. PCR Troubleshooting, the Essential Guide, first ed. Caister Academic Press, Wymondham, Norfolk (UK).
- Bakken, L.R., Lindahl, V., 1995. Recovery of bacterial cells from soil. In: Trevors, J.T., Van Elsland, J.D. (Eds.), *Nucleic Acids in the Environment*. Springer-Verlag, Berlin, pp. 13–27.
- 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., DeLong, E.F., 2000a. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology* 2, 516–529.
- Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., DeLong, E.F., 2000b. Bacterial rhodopsin, evidence for a new type of phototrophy in the sea. *Science* 289, 1902–1906.
- Berry, A.E., Chiochini, C., Selby, T., Sosio, M., Wellington, E.M., 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiology Letters* 223, 15–20.
- Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M., Simonet, P., 2005. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *Journal of Microbiological Methods* 62, 1–11.
- Brady, S.F., Chao, C.J., Clardy, J., 2004. Long-chain N-acetyltyrosine synthases from environmental DNA. *Applied and Environmental Microbiology* 70, 6865–6870.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., Rohwer, F., 2002. Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America* 99, 14250–14255.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology* 185, 6220–6223.
- Brennan, Y., Callen, W.N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernandez, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G.P., Mathur, E.J., Short, J.M., Robertson, D.E., Steer, B.A., 2004. Unusual microbial xylanases from insect guts. *Applied and Environmental Microbiology* 70, 3609–3617.
- Brulc, J.M., Antonopoulos, D.A., Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., White, B.A., 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* 106, 1948–1953.
- Burgess, J.G., Jordan, E.M., Bregu, M., Meams-Spragg, A., Boyd, K.G., 1999. Microbial antagonism, a neglected avenue of natural products research. *Journal of Biotechnology* 70, 27–32.
- Chandler, D.P., Brockman, F.J., Fredrickson, J.K., 1997. Use of 16S rDNA clone libraries to study changes in a microbial community resulting from *ex situ* perturbation of a subsurface sediment. *FEMS Microbiology Reviews* 20, 217–230.
- Chen, Y., Dumont, M.G., Neufeld, J.D., Bodrossy, L., Stralis-Pavese, N., McNamara, N.P., Ostle, N., Briones, M.J., Murrell, J.C., 2008. Revealing the uncultivated majority, combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands. *Environmental Microbiology* 10, 2609–2622.
- Chen, I.C., Lin, W.D., Hsu, S.K., Thiruvengadam, V., Hsu, W.H., 2009. Isolation and characterization of a novel lysine racemase from a soil metagenomic library. *Applied and Environmental Microbiology* 75, 5161–5166.
- Chu, X., He, H., Guo, C., Sun, B., 2008. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Applied Microbiology and Biotechnology* 80, 615–625.
- Chung, E.J., Lim, H.K., Kim, J.C., Choi, G.J., Park, E.J., Lee, M.H., Chung, Y.R., Lee, S.W., 2008. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Applied and Environmental Microbiology* 74, 723–730.
- Cieslinski, H., Dlugolecka, A., Kur, J., Turkiewicz, M., 2009. An MTA phosphorylase gene discovered in the metagenomic library derived from Antarctic top soil during screening for lipolytic active clones confers strong pink fluorescence in the presence of rhodamine B. *FEMS Microbiology Letters* 299, 232–240.
- Clardy, J., Walsh, C., 2004. Lessons from natural molecules. *Nature* 432, 829–837.
- Collins, J., Hohn, B., 1978. Cosmids, a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage lambda heads. *Proceedings of the National Academy of Sciences of the United States of America* 75, 4242–4246.
- Cosgrove, L., McGeechan, P.L., Handley, P.S., Robson, G.D., 2010. Effect of bio-stimulation and bioaugmentation on degradation of polyurethane buried in soil. *Applied and Environmental Microbiology* 76, 810–819.
- Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerinneau, M., Jeannin, P., Simonet, P., Pernodet, J.L., 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology* 69, 49–55.
- Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., Wittwer, P., 2005. Metagenomic gene discovery, past, present and future. *Trends in Biotechnology* 23, 321–329.
- Craig, J.W., Chang, F.Y., Brady, S.F., 2009. Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chemical Biology* 4, 23–28.
- Craig, J.W., Chang, F.Y., Kim, J.H., Obajulu, S.C., Brady, S.F., 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and Environmental Microbiology* 76, 1633–1641.
- Curtis, T.P., Sloan, W.T., 2005. Microbiology. Exploring microbial diversity – a vast below. *Science* 309, 1331–1333.
- de Lorenzo, V., 2005. Problems with metagenomic screening. *Nature Biotechnology* 23, 1045–1046.
- DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubietta, I.X., Piceno, Y.M., Andersen, G.L., 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microbial Ecology* 53, 371–383.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L.L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., Olson, R., Paul, J., Brito, B.R., Ruan, Y.J., Swan, B.K., Stevens, R., Valentine, D.L., Thurber, R.V., Wegley, L., White, B.A., Rohwer, F., 2008. Functional metagenomic profiling of nine biomes. *Nature* 452, 629–633.
- Dumont, M.G., Radajewski, S.M., Miguez, C.B., McDonald, I.R., Murrell, C., 2006. Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. *Environmental Microbiology* 8, 1240–1250.
- Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., Saar, M.O., Alexander, S., Alexander Jr., E.C., Rohwer, F., 2006. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7, 57.
- Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J.D., Steele, H.L., Reymond, J.L., Jaeger, K.E., Streit, W.R., 2006. Isolation and biochemical characterization of two novel metagenome-derived esterases. *Applied and Environmental Microbiology* 72, 3637–3645.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T., Streit, W.R., 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology* 67, 89–99.
- Faegri, A., Torsvik, V., Goksoyr, J., 1977. Bacterial and fungal activities in soil, separation of bacteria and fungi by a rapid fractionated centrifugation technique. *Soil Biology and Biochemistry* 9, 105–112.
- Feinstein, L.M., Sul, W.J., Blackwood, C.B., 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Applied and Environmental Microbiology* 75, 5428–5433.
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chemistry and Biology* 12, 895–904.
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F., Jackson, R.B., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology* 73, 7059–7066.
- Frostegard, A., Courtois, S., Ramière, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X., Simonet, P., 1999. Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology* 65, 5409–5420.
- Gabor, E.M., Vries, E.J., Janssen, D.B., 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiology Ecology* 44, 153–163.
- Garbeva, P., de Boer, W., 2009. Inter-specific interactions between carbon-limited soil bacteria affect behavior and gene expression. *Microbial Ecology* 58, 36–46.
- George, I.F., Liles, M.R., Hartmann, M., Ludwig, W., Goodman, R.M., Agathos, S.N., 2009. Changes in soil *Acidobacteria* communities after 2,4,6-trinitrotoluene contamination. *FEMS Microbiology Letters* 296, 159–166.
- Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M., Handelsman, J., 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology* 68, 4301–4306.
- Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., Bertrand, H., Vogel, T.M., Perriere, G., Simonet, P., Nalin, R., 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Applied and Environmental Microbiology* 70, 5522–5527.
- Hain, T., Otten, S., von Both, U., Chatterjee, S.S., Technow, U., Billion, A., Ghai, R., Mohamed, W., Domann, E., Chakraborty, T., 2008. Novel bacterial artificial chromosome vector pUVBAC for use in studies of the functional genomics of *Listeria* spp. *Applied and Environmental Microbiology* 74, 1892–1901.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes, a new frontier for natural products. *Chemistry and Biology* 5, R245–R249.
- Hao, Y., Winans, S.C., Glick, B.R., Charles, T.C., 2010. Identification and characterization of new LuxR/LuxI-type quorum sensing systems from metagenomic libraries. *Environmental Microbiology* 12, 105–117.
- Harry, M., Gambier, B., Bourezgui, Y., Garnier-Sillam, E., 1999. Evaluation of purification procedures for DNA extracted from organic rich samples, interference with humic substances. *Analysis* 127, 439–442.
- Hassink, J., Bouman, L.A., Zwart, K.B., Bloem, J., Brussaard, L., 1993. Relationships between soil texture, physical protection of organic matter, soil biota, and C and N mineralization in grassland soils. *Geoderma* 57, 105–128.

- Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., Shanmugam, K.T., 1995. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied Microbiology and Biotechnology* 43, 667–674.
- Heath, C., Hu, X.P., Cary, S.C., Cowan, D., 2009. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. *Applied and Environmental Microbiology* 75, 4657–4659.
- Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G., 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Applied and Environmental Microbiology* 65, 3901–3907.
- Holben, W.E., Jansson, J.K., Chelm, B.K., Tiedje, J.M., 1988. DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Applied and Environmental Microbiology* 54, 703–711.
- Hopkins, D.W., Macnaughton, S.J., O'Donnell, A.G., 1991. A dispersion and differential centrifugation technique for representatively sampling microorganisms from soil. *Soil Biology and Biochemistry* 23, 217–225.
- Hugenholtz, P., Goebel, B.M., Pace, N.R., 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* 180, 4765–4774.
- Hugenholtz, P., Tyson, G.W., 2008. Microbiology, metagenomics. *Nature* 455, 481–483.
- Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* 8, R143.
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Research* 17, 377–386.
- Jacobsen, C.S., Rasmussen, O.F., 1992. Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Applied and Environmental Microbiology* 58, 2458–2462.
- Jeon, J.H., Kim, J.T., Kang, S.G., Lee, J.H., Kim, S.J., 2009. Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Marine Biotechnology* (New York, N.Y.) 11, 307–316.
- Jiang, C., Ma, G., Li, S., Hu, T., Che, Z., Shen, P., Yan, B., Wu, B., 2009. Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *Journal of Microbiology* 47, 542–548.
- Kalyuzhnaya, M.G., Lapidus, A., Ivanova, N., Copeland, A.C., McHardy, A.C., Szeto, E., Salamov, A., Grigoriev, I.V., Suci, D., Levine, S.R., Markowitz, V.M., Rigoutsos, I., Tringe, S.G., Bruce, D.C., Richardson, P.M., Lidstrom, M.E., Chistoserdova, L., 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotechnology* 26, 1029–1034.
- Kim, U.J., Shizuya, H., de Jong, P.J., Birren, B., Simon, M.I., 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research* 20, 1083–1085.
- Kim, B.S., Kim, S.Y., Park, J., Park, W., Hwang, K.Y., Yoon, Y.J., Oh, W.K., Kim, B.Y., Ahn, J.S., 2007. Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. *Journal of Applied Microbiology* 102, 1392–1400.
- King, R.W., Bauer, J.D., Brady, S.F., 2009. An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. *Angewandte Chemie* 48, 6257–6261 (International ed. in English).
- Knaebel, D.B., Crawford, R.L., 1995. Extraction and purification of microbial DNA from petroleum-contaminated soils and detection of low numbers of toluene, octane and pesticide degraders by multiplex polymerase chain reaction and Southern analysis. *Molecular Ecology* 4, 579–591.
- Knietsch, A., Bowien, S., Whited, G., Gottschalk, G., Daniel, R., 2003. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Applied and Environmental Microbiology* 69, 3048–3060.
- Kreader, C.A., 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology* 62, 1102–1106.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., Hugenholtz, P., 2008. A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews* 72, 557–578 (Table of Contents).
- Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., Rupp, S., 2007. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology* 127, 575–592.
- Lee, S.W., Won, K., Lim, H.K., Kim, J.C., Choi, G.J., Cho, K.Y., 2004. Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Applied Microbiology and Biotechnology* 65, 720–726.
- Li, X., Qin, L., 2005. Metagenomics-based drug discovery and marine microbial diversity. *Trends in Biotechnology* 23, 539–543.
- Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J., Goodman, R.M., 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology* 69, 2684–2691.
- Liles, M.R., Williamson, L.L., Goodman, R.M., Handelsman, J., 2004. Isolation of high molecular weight genomic DNA from soil bacteria for genomic library construction. In: Kowalchuk, G.A., et al. (Eds.), *Molecular Microbial Ecology Manual*, second ed. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 839–852.
- Liles, M.R., Williamson, L.L., Rodbumer, J., Torsvik, V., Goodman, R.M., Handelsman, J., 2008. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Applied and Environmental Microbiology* 74, 3302–3305.
- Liles, M.R., Turkmen, O., Manske, B.F., Zhang, M., Rouillard, J.M., George, I.F., Balsler, T., Billor, N., Goodman, R.M., 2010. A phylogenetic microarray targeting 16S rRNA genes from the bacterial division *Acidobacteria* reveals a lineage-specific distribution in a soil clay fraction. *Soil Biology and Biochemistry* 42, 739–747.
- Macdonald, R.M., 1986. Sampling soil microflora-dispersion of soil by ion-exchange and extraction of specific microorganisms from suspension by elutriation. *Soil Biology and Biochemistry* 18, 399–406.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P.,irage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- Martinez, A., Kolvek, S.J., Yip, C.L., Hopke, J., Brown, K.A., MacNeil, I.A., Osburne, M.S., 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology* 70, 2452–2463.
- Martinez, A., Kolvek, S.J., Hopke, J., Yip, C.L., Osburne, M.S., 2005. Environmental DNA fragment conferring early and increased sporulation and antibiotic production in *Streptomyces* species. *Applied and Environmental Microbiology* 71, 1638–1641.
- Metzker, M.L., 2005. Emerging technologies in DNA sequencing. *Genome Research* 15, 1767–1776.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386.
- Miller, R.V., Day, M.J., 2004. Microbial Evolution, Gene Establishment, Survival, and Exchange. ASM Press, Washington, D.C.
- Mirete, S., de Figueras, C.G., Gonzalez-Pastor, J.E., 2007. Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. *Applied and Environmental Microbiology* 73, 6001–6011.
- Moffitt, M.C., Neilan, B.A., 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *Journal of Molecular Evolution* 56, 446–457.
- Muyzer, G., de Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology* 59, 695–700.
- Ogram, A., Sayler, G.S., Barkay, T.J., 1987. DNA extraction and purification from sediments. *Journal of Microbiological Methods* 7, 57–66.
- Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Applied Microbiology and Biotechnology* 74, 501–510.
- Osborn, A.M., Smith, C.J., 2005. *Molecular Microbial Ecology* Taylor & Francis. Abingdon [England], New York.
- Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., de Jong, P.J., 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1–8.
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goemann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* 33, 5691–5702.
- Palackal, N., Lyon, C.S., Zaidi, S., Luginbuhl, P., Dupree, P., Goubet, F., Macomber, J.L., Short, J.M., Hazlewood, G.P., Robertson, D.E., Steer, B.A., 2007. A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Applied Microbiology and Biotechnology* 74, 113–124.
- Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper Jr., W.F., Liles, M.R., 2010. Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Applied and Environmental Microbiology* 76, 3753–3757.
- Pel, J., Broemeling, D., Mai, L., Poon, H.L., Tropini, G., Warren, R.L., Holt, R.A., Marziali, A., 2009. Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules. *Proceedings of the National Academy of Sciences of the United States of America* 106 (35), 14796–14801.
- Pfeifer, B.A., Khosla, C., 2001. Biosynthesis of polyketides in heterologous hosts. *Microbiology and Molecular Biology Reviews* 65, 106–118.
- Pignatelli, M., Aparicio, G., Blanquer, I., Hernandez, V., Moya, A., Tamames, J., 2008. Metagenomics reveals our incomplete knowledge of global diversity. *Bioinformatics* 24, 2124–2125.
- Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., Eck, J., Sensen, C.W., Schleper, C., 2002. First insight into the genome of an uncultivated crenarchaeote from soil. *Environmental Microbiology* 4, 603–611.

- Radajewski, S., Ineson, P., Parekh, N.R., Murrell, J.C., 2000. Stable-isotope probing as a tool in microbial ecology. *Nature* 403, 646–649.
- Riaz, K., Elmerich, C., Moreira, D., Raffoux, A., Dessaux, Y., Faure, D., 2008. A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases. *Environmental Microbiology* 10, 560–570.
- Richardson, T.H., Tan, X., Frey, G., Callen, W., Cabell, M., Lam, D., Macomber, J., Short, J.M., Robertson, D.E., Miller, C., 2002. A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *The Journal of Biological Chemistry* 277, 26501–26507.
- Riesenfeld, C.S., Goodman, R.M., Handelsman, J., 2004a. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology* 6, 981–989.
- Riesenfeld, C.S., Schloss, P.D., Handelsman, J., 2004b. Metagenomics, genomic analysis of microbial communities. *Annual Review of Genetics* 38, 525–552.
- Robertson, D.E., Chaplin, J.A., DeSantis, G., Podar, M., Madden, M., Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., Short, J.M., 2004. Exploring nitrilase sequence space for enantioselective catalysis. *Applied and Environmental Microbiology* 70, 2429–2436.
- Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal* 1, 283–290.
- Ronaghi, M., 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11, 3–11.
- 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., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., Goodman, R.M., 2000. Cloning the soil metagenome, a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* 66, 2541–2547.
- Roose-Amsaleg, C.L., Garnier-Sillam, E., Harry, M., 2001. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology* 18, 47–60.
- Rothberg, J.M., Leamon, J.H., 2008. The development and impact of 454 sequencing. *Nature Biotechnology* 26, 1117–1124.
- Saida, F., Uzan, M., Odaert, B., Bontems, F., 2006. Expression of highly toxic genes in *E. coli*, special strategies and genetic tools. *Current Protein and Peptide Science* 7, 47–56.
- Schatz, A., Waksman, S.A., 1944. Effect of streptomycin and other antibiotic substances upon *Mycobacterium tuberculosis* and related organisms. *Proceedings of the Society for Experimental Biology and Medicine* 57, 244–248. Society for Experimental Biology and Medicine (New York, N.Y.).
- Schipper, C., Hornung, C., Bijtenhoorn, P., Quitschau, M., Grond, S., Streit, W.R., 2009. Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology* 75, 224–233.
- Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F., Hutchinson, C.R., 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Applied and Environmental Microbiology* 71, 4840–4849.
- Schloss, P.D., Handelsman, J., 2003. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* 14, 303–310.
- Schreiber, F., Gumrich, P., Daniel, R., Meinicke, P., 2010. TreePhyler, fast taxonomic profiling of metagenomes. *Bioinformatics* 26, 960–961.
- Schwartz, I., 2000. Microbial genomics, from sequence to function. *Emerging Infectious Diseases* 6, 493–495.
- Selenska, S., Klingmuller, W., 1991. Direct detection of nif-gene sequences of *Enterobacter agglomerans* in soil. *FEMS Microbiology Letters* 80, 243–246.
- Sessitsch, A., Weilharter, A., Gerzabek, M.H., Kirchmann, H., Kandler, E., 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Applied and Environmental Microbiology* 67, 4215–4224.
- Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M., 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America* 89, 8794–8797.
- Simon, C., Daniel, R., 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology* 85 (2), 265–276.
- Simon, C., Herath, J., Rockstroh, S., Daniel, R., 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Applied and Environmental Microbiology* 75, 2964–2968.
- Sleator, R.D., Shortall, C., Hill, C., 2008. Metagenomics. *Letters in Applied Microbiology* 47, 361–366.
- Smarr, L., 2006. The Ocean of Life, Creating a Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (a.k.a. CAMERA). Strategic News Service, Friday Harbor (Washington).
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences of the United States of America* 103, 12115–12120.
- Solbak, A.I., Richardson, T.H., McCann, R.T., Kline, K.A., Bartnek, F., Tomlinson, G., Tan, X., Parra-Gessert, L., Frey, G.J., Podar, M., Luginbuhl, P., Gray, K.A., Mathur, E.J., Robertson, D.E., Burk, M.J., Hazlewood, G.P., Short, J.M., Kerovuo, J., 2005. Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric. *The Journal of Biological Chemistry* 280, 9431–9438.
- Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A.M., Donadio, S., 2000. Artificial chromosomes for antibiotic-producing actinomycetes. *Nature Biotechnology* 18, 343–345.
- Staley, J.T., Konopka, A., 1985. Measurement of *in situ* activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology* 39, 321–346.
- Steffan, R.J., Goksoyr, J., Bej, A.K., Atlas, R.M., 1988. Recovery of DNA from soils and sediments. *Applied and Environmental Microbiology* 54, 2908–2915.
- Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., DeLong, E.F., 1996. Characterization of uncultivated prokaryotes, isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* 178, 591–599.
- Suenaga, H., Ohnuki, T., Miyazaki, K., 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology* 9, 2289–2297.
- Sukchawalit, R., Vattanaviboon, P., Sallabhan, R., Mongkolsuk, S., 1999. Construction and characterization of regulated α -arabinose-inducible broad host range expression vectors in *Xanthomonas*. *FEMS Microbiology Letters* 181, 217–223.
- Sul, W.J., Park, J., Quensen 3rd, J.F., Rodrigues, J.L., Seliger, L., Tsoi, T.V., Zylstra, G.J., Tiedje, J.M., 2009. DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Applied and Environmental Microbiology* 75, 5501–5506.
- Sylvia, D.M., 2005. Principles and Applications of Soil Microbiology, second ed. Pearson Prentice Hall, Upper Saddle River, N.J.
- Tebbe, C.C., Vahjen, W., 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and Environmental Microbiology* 59, 2657–2665.
- Tien, C.C., Chao, C.C., Chao, W.L., 1999. Methods for DNA extraction from various soils, a comparison. *Journal of Applied Microbiology* 86, 937–943.
- Torsvik, V., Ovreas, L., 2002. Microbial diversity and function in soil, from genes to ecosystems. *Current Opinion Microbiology* 5, 240–245.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P., Rubin, E.M., 2005. Comparative metagenomics of microbial communities. *Science* 308, 554–557.
- Tsai, Y.L., Olson, B.H., 1991. Rapid method for direct extraction of DNA from soil and sediments. *Applied and Environmental Microbiology* 57, 1070–1074.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43.
- Uchiyama, T., Abe, T., Ikemura, T., Watanabe, K., 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology* 23, 88–93.
- Van Elsas, J.D., Mantynen, V., Wolters, A.C., 1997. Soil DNA extraction and assessment of the fate of *Mycobacterium chlorophenicolum* strain PCP-1 in different soils by 16S ribosomal RNA gene sequence based most-probable-number PCR and immunofluorescence. *Biology and Fertility of Soils* 24, 188–195.
- Veluci, R.M., Neher, D.A., Weicht, T.R., 2006. Nitrogen fixation and leaching of biological soil crust communities in mesic temperate soils. *Microbial Ecology* 51, 189–196.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealon, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.
- Vogel, T.M., 1996. Bioaugmentation as a soil bioremediation approach. *Current Opinion in Biotechnology* 7, 311–316.
- Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E., Streit, W.R., 2003. Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology* 69, 6235–6242.
- Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J., Davies, J., 2000. Novel natural products from soil DNA libraries in a *Streptomyces* host. *Organic Letters* 2, 2401–2404.
- Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., Charles, T.C., 2006. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and Environmental Microbiology* 72, 384–391.
- Ward, D.M., Weller, R., Bateson, M.M., 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghasseman, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar, M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X., Hernandez, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P., Leadbetter, J.R., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565.
- Wawrik, B., Kerkhof, L., Zylstra, G.J., Kukor, J.J., 2005. Identification of unique type II polyketide synthase genes in soil. *Applied and Environmental Microbiology* 71, 2232–2238.
- Wellington, E.M., Berry, A., Krsek, M., 2003. Resolving functional diversity in relation to microbial community structure in soil, exploiting genomics and stable isotope probing. *Current Opinion Microbiology* 6, 295–301.

- Wild, J., Hradecna, Z., Szybalski, W., 2002. Conditionally amplifiable BACs, switching from single-copy to high-copy vectors and genomic clones. *Genome Research* 12, 1434–1444.
- Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H.K., Handelsman, J., 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Applied and Environmental Microbiology* 71, 6335–6344.
- Woese, C.R., 1987. Bacterial evolution. *Microbiological Reviews* 51, 221–271.
- Wommack, K.E., Bhavsar, J., Ravel, J., 2008. Metagenomics, read length matters. *Applied and Environmental Microbiology* 74, 1453–1463.
- Xia, X., Bollinger, J., Ogram, A., 1995. Molecular genetic analysis of the response of three soil microbial communities to the application of 2,4-D. *Molecular Ecology* 4, 17–28.
- Zhou, J., Bruns, M.A., Tiedje, J.M., 1996. DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology* 62, 316–322.