PART A

NEW APPROACHES TO FINDING AND MODIFYING ENZYMES

1

FUNCTIONAL METAGENOMICS AS A TECHNIQUE FOR THE DISCOVERY OF NOVEL ENZYMES AND NATURAL PRODUCTS

LUKE A. MOE

Department of Plant and Soil Sciences, University of Kentucky, Lexington, Kentucky

MATTHEW D. MCMAHON, AND MICHAEL G. THOMAS

Department of Bacteriology, University of Wisconsin–Madison, Madison, Wisconsin

I. INTRODUCTION

The industrial use of biocatalysts and the recent reemphasis on the isolation of natural products with desired biological activities are driving the search for new mechanisms for accessing the metabolic potential of microorganisms. This emphasis on microorganisms comes from an appreciation of the enormous biodiversity found within them [1–4] and the understanding that traditional culturing techniques to isolate these organisms have only enabled us to access approximately 1% of the microbial population present in a soil environment [5–7]. Further, it is reasonable to assume that this low level of culturability will be found in many other ecological niches. Due to this limited access to the full metabolic potential in a targeted environment, there is a clear interest in developing techniques allowing us access to the metabolic potential of the remaining 99% of microorganisms.

One approach that aims to circumvent the limitations of culturability is the use of metagenomics. The term *metagenome* was introduced to define the combined

Enzyme Technologies: Metagenomics, Evolution, Biocatalysis, and Biosynthesis,
Edited by Wu-Kuang Yeh, Hsiu-Chiung Yang, and James R. McCarthy

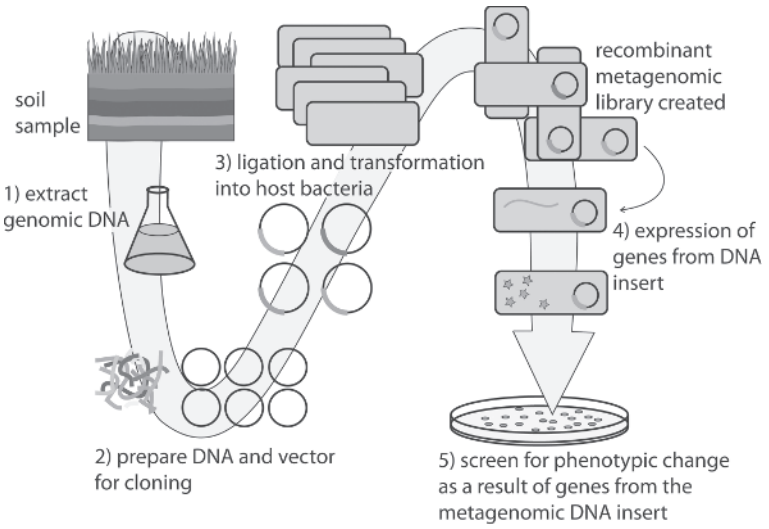

FIGURE 1 Construction and functional screening of a metagenomic DNA library from soil. (*See insert for color representation of the figure.*)

genomes of all the organisms in a particular environment, with *metagenomics* referring to a collection of techniques allowing access to this genomic information [8]. The shared features of metagenomic techniques are that the total DNA of the microbial population is extracted from the environment and cloned into appropriate bacterial cloning vectors in a culture-independent manner (Fig. 1); thus, this technique has the potential to access the 99% of the metabolic potential that has previously been inaccessible.

Once metagenomic DNA has been extracted from an environment of interest and the associated DNA library has been constructed, there are two mechanisms for analyzing the metabolic potential present. One approach is to use sequence-based methods whereby the library is sequenced and analyzed for genes of interest. Although this approach has been successful in identifying new homologs of enzymes of interest (e.g., bacteriorhodopsin homologs [9]) and in gaining an understanding of total microbial community present [10], it does have limitations regarding the identification of new and novel biocatalysts. Because sequence-based identification of enzymes relies on relatively high levels of sequence identity, only those enzymes displaying significant homology to enzymes known to catalyze a specific reaction will be recognized. Thus, those enzymes displaying low sequence identity to known functional classes, and novel enzymes that exhibit the desired activity but show no homology to known functional classes, will not be identified using sequence-based techniques alone.

Alternatively, functional metagenomics is based on detecting a desired biological activity that has been introduced into the host organism by virtue of the recombinant metagenomic clone. Using this function-based approach, metagenomic clones encoding a variety of desired activities have been discovered; these

activities include esterases [11,12] and nitrile hydratases [13], enzymes responsible for the production of natural products (e.g., antibiotics [11] and pigments [14]), and antibiotic resistance genes [15,16]. The power of this approach is derived from the ability to identify clones encoding the desired biological activity in a large metagenomic library (often more than 10^6 recombinant clones) without prior sequence knowledge and without requiring a large-scale sequencing project.

Although successful, this approach to metagenomic DNA analysis also has its limitations. The most obvious limitation is the ability to devise a phenotypic screen to identify clones of interest. The second limitation deals with the heterologous host itself. Although *Escherichia coli* has proven to be an extraordinarily compliant host, it is not always able to express genes and produce functional proteins from distantly related microorganisms. This limitation can be due to *E. coli* lacking the necessary regulatory elements to turn on expression of the gene clusters or the failure of *E. coli* to recognize the gene promoters due to a lack of the appropriate sigma factors. Additional issues arise when screening for the production of natural products, since *E. coli* may lack the necessary cofactors or precursors required for production of the metabolite. The promise of functional metagenomics is emphasized in this chapter and we provide the reader with information on the steps that various researchers have taken to circumvent the limitations discussed above.

The field of metagenomics has grown enormously recently and it is nearly impossible to adequately summarize all the advances in sequence-based and functional genomics. Instead, what we have chosen to do here is to provide the reader with information on, and references for construction of, a metagenomic DNA library from a specific environment (e.g., soils/sediments) and to provide options for how such a library can be screened using functional genomics approaches.

In discussing the construction and functional screening of a soil metagenome, we have divided this chapter into several sections. First, we discuss briefly the choice of source material, which can be important in enrichment for a desired biological activity. Second, we provide details on how to isolate metagenomic DNA from soils/sediments and prepare it for cloning. Third, we discuss the construction of a metagenomic DNA library. Fourth, we provide examples of functional metagenomic screens used for identifying clones with various biological activities. Finally, we provide a brief summary and some thoughts on what challenges lay ahead for metagenomics.

II. CONSTRUCTION OF METAGENOMIC DNA LIBRARIES

A. Choice of Source Material

Our research has focused on functional metagenomics from soil-derived DNA libraries. Our focus on this environment is driven by two factors. First, it is now quite evident that the soil environment has enormous biological and metabolic diversity. One gram of soil has between 4×10^7 and 2×10^9 prokaryotic cells [17,18]. Furthermore, Torsvik and colleagues have estimated that this vast number of prokaryotic cells consists of 3000 to 11,000 different genomes [5,6]. This

estimate is likely to be an underestimate based on the likelihood that there will be a number of rare and underrepresented members of this population. Thus, there is enormous metabolic potential from a single gram of soil.

Although we have focused on mesophilic soils/sediments, there are a number of examples of other soil environments and other ecological niches that have been investigated. For soils/sediments these include extreme environments such as high temperatures [19], high and low pH [20], and high salt [21], in addition to cultivated and uncultivated fields [11,15,22], and different types of soils (e.g., clay [23], sandy [24], loam [15]). There is also a trans-European project that focuses on metagenomic analysis of disease-suppressive soils to identify components produced by soil inhabitants that can be exploited for agricultural purposes [25].

Although there has been a significant focus on soils/sediments, this has not been the only environment of focus. Other environments include such diverse areas as animal gastrointestinal tracts for the identity of β-glucanases [26], marine sponges to identify natural products [27], and the resident microbiota of the gypsy moth for resistance determinants [16]. In the end, the choice of starting material can influence the success of a functional metagenomic screen.

B. Isolating DNA from Soils and Sediments

There are two general methods by which DNA is obtained from soil samples: the direct lysis and cell separation (or cell extraction) methods. As their names imply, they differ according to whether bacterial cells are separated from the environmental matrix prior to lysis. In the following section we provide overviews of both methods, including advantages and disadvantages of each. However, for the purposes of this chapter, we believe the reader may be more interested in the direct lysis method, and we give a more extensive outline of a protocol in use in our labs. Although we do not offer a comprehensive review of methods for the isolation of DNA from environmental samples, we refer the reader to a number of references that address these techniques in detail [28–33].

Direct Lysis Method The direct lysis method for obtaining DNA reportedly results in larger quantities of DNA isolated from samples [30] and thus a more representative picture of the genetic material present in a sample. However, because lysis is done in the presence of the soil matrix, isolated DNA extracts typically include contaminants that may hinder downstream DNA manipulation steps. Further, the harsh chemical and mechanical steps involved in processing samples typically result in smaller DNA fragments available for cloning [29]. Nevertheless, direct lysis is the more popular method of the two for researchers interested in functional screening, due primarily to its relative simplicity (commercial kits are available for small-insert DNA library cloning), ease in sample processing, limited time input, and demonstrated success in producing libraries from which enzymes of interest can be identified (Table 1). A number of papers describe individual steps meant to enhance the viability of the DNA during the processing steps; these papers are noted during the description of the protocol.

TABLE 1 Enzymatic Activities Identified Using Functional Megagenomics

Enzyme Class and *Enzyme*	Screening Method	Vector (Average Insert Size)	Amount of DNA Screened[a]	Positive Clones Recovered	Ref.
Lipase/esterase					
Lipase/esterase	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with tributyrin; positive clones identified by presence of clear halo around colony	pEpiFOS-5 (35 kb)	NR (estimate 1.2 Gb)	8	[60]
	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with α-naphthyl acetate, then reacted with Fast Blue RR to detect esterase activity; reaction with α-naphthol results in formation of a diazo dye and characteristic color around colony	λZAP (5.5 kb)	NR (estimate 77 Mb)	11	[20]
	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with 3% Bacto Lipid; positive clones identified by presence of clear halo around colony	pBeloBAC11 (27 kb)	~100 Mb	2	[11]
Polysaccharide hydrolysis					
Cellulase	Recombinant *E. coli* metagenomic library plated on LB agar, plates then overlaid with top agar containing 0.1% Ostazin Brilliant Red hydroxyethyl cellulose; positive clones identified by a yellow halo around the colony	pBeloBAC11 (27 kb)	~100 Mb	0	[11]

(*Continued overleaf*)

TABLE 1 (*Continued*)

Enzyme Class and *Enzyme*	Screening Method	Vector (Average Insert Size)	Amount of DNA Screened[a]	Positive Clones Recovered	Ref.
	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with 0.1% carboxymethylcellulose; after growth for 7 days the plates were flooded with Congo Red; positive clones identified by a clearing zone around the colony	pSuperCOS1 (NR)	NR (estimate 2.8 Gb)	1	[69]
Amylase	Recombinant *E. coli* metagenomic library plated on Bacto Starch agar, plates were flooded with Bacto-stabilized Gram iodine after 3 days; positive clones identified by an orange halo around the colony	pBeloBAC11 (27 kb)	~100 Mb	8	[11]
Cellobiohydrolase	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with 4-methylumbelliferyl-β-D-cellobioside; positive clones identified by the fluorescent reaction product	pWEB-TNC (35.1 kb)	1.14 Gb	0	[56]
β-*Glucosidase*	Recombinant *E. coli* metagenomic library plated with esculin hydrate (a β-glucoside) and ferric ammonium citrate; positive clones identified by the black precipitate resulting from the reaction of the hydrolytic product of esculin with ferric ion	pWEB-TNC (35.1 kb)	1.14 Gb	7	[56]

Chitinase	Recombinant *E. coli* metagenomic library grown with fluorogenic substrate 4-methylumbelliferyl-β-D-*N*, *N*′-diacetylchitobioside; activity scored by presence of fluorescent hydrolysis product	Reported as "Epicentre CopyControl fosmid library kit," presumably pCC1FOS (40 kb)	NR (estimate 240 Mb)	4	[55]
Xylanase	Recombinant *E. coli* metagenomic library plated and overlaid with top agar containing 0.1–0.3% azo dye-linked xylan; positive clones identified by disappearance of the dye color	NR (3–6 kb)	~1 Mb	4	[70]
	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with 1% Remazol Brilliant Blue xylan; positive clones identified by a clearing zone around the colony	pMBM803 (7–10 kb)	NR (estimate 204 Mb)	1	[57]
Oxygenase					
Polyphenol oxidase	Recombinant *E. coli* metagenomic library plated on NZY soft agar supplemented with 50 μM syringaldazine, positive clones identified by the presence of a purple halo surrounding the colony	λZAP (5.5 kb)	NR	1	[54]
Monooxygenase	Recombinant *E. coli* metagenomic library plated on LB agar; appearance of blue color indicates formation of indigo from indole by hydroxylation followed by spontaneous dimerization	pZErO-2 (5.5 kb)	NR (estimate 358 Mb)	1	[71]

(*Continued overleaf*)

TABLE 1 (*Continued*)

Enzyme Class and *Enzyme*	Screening Method	Vector (Average Insert Size)	Amount of DNA Screened[a]	Positive Clones Recovered	Ref.
Extradiol dioxygenase	Recombinant *E. coli* metagenomic clones grown in LB broth, then lysed; the lysate was clarified and catechol was added to 0.5 mM; positive clones identified based on formation of a yellow ring-cleavage product	pCC1FOS (33 kb)	3.2 Gb	91	[59]
4-Hydroxyphenylpyruvate dioxygenase	Recombinant *E. coli* metagenomic clones grown in LB broth supplemented with 600 μg/mL L-tyrosine; positive clones identified as those that produce a brown pigment	pSuperCos1 (40 kb)	NR (estimate 1.2 Gb)	5	[72]
Acylhomoserine lactone (AHL)–mediated quorum sensing activation and inhibition					
AHL synthase	Metagenomic DNA clones transformed into an *E. coli* biosensor strain engineered to sense AHLs; production of AHLs by a metagenomic clone results in a GFP-positive phenotype for the colony	pCC1FOS (39 kb), pCC1BAC (13–47 kb), pSuperBAC (5–33 kb)	653 Mb	12	[44]

	Recombinant *E. coli* metagenomic clones assayed on an agar plate for the ability to restore AHL-dependent pigment production to a *Chromobacterium* strain that has had its AHL synthase knocked out	NR	NR	0	[73]
AHL lactonase	Metagenomic DNA clones transformed into an *E. coli* biosensor strain engineered to sense AHLs; the recombinant clones then grown in the presence of an AHL known to activate the reporter; positive clones identified as those clones that did not fluoresce	pCC1FOS (39 kb), pCC1BAC (13–47 kb), pSuperBAC (5–33 kb)	653 Mb	2	[44]
	Recombinant *E. coli* metagenomic clones assayed on an agar plate for the ability to interfere with an AHL-sensing *Chromobacterium* biosensor strain in the presence of AHLs	NR	NR	1	[73]
Other					
Protease	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with 1% (w/v) skim milk; positive clones identified by the presence of a clear halo surrounding the colony	pCC1FOS (NR)	NR	1	[74]

(*Continued overleaf*)

TABLE 1 (*Continued*)

Enzyme Class and *Enzyme*	Screening Method	Vector (Average Insert Size)	Amount of DNA Screened[a]	Positive Clones Recovered	Ref.
Nitrilase	Recombinant *E. coli* metagenomic library grown in M9 medium lacking a nitrogen source, supplemented with a nitrile substrate; growth of recombinant clones on this medium indicated that the metagenomic DNA insert conferred the ability to liberate ammonia from the nitrile substrate through a nitrilase reaction, allowing the clone to use ammonia as the sole nitrogen source	λZAP (1–10 kb)	NR (between 0.65 Gb and 6.5 Tb)	> 200	[43]
Phosphatase	Recombinant *E. coli* metagenomic library plated on nutrient agar supplemented with 40 μg/mL 5-bromo-4-chloro-3-indolyl phosphate (X-phosphate); positive clones identified based on blue colony color	pJOE930 (3.2 kb)	NR (estimate 102 Mb)	13	[53]
Menadione resistance	Recombinant *E. coli* metagenomic clones cultured in LB broth supplemented with a concentration of menadione toxic to the host strain (0.6 mM); positive clones identified as those that grew	pCC1FOS (33 kb)	NR (estimate 3.2 Gb)	4	[61]

β-*Lactamase*	Recombinant *E. coli* metagenomic library plated on LB agar supplemented with a concentration of β-lactam antibiotic toxic to the host cell (concentration varies depending on the β-lactam); positive clones identified as those that grew	pCF430 (5–10 kb), pCC1FOS (30 kb), pCC1BAC (8 kb)	12.4 Gb	13	[16]
5-Enoylpyruvylshikimate-3-phosphate synthase (EPSPS)	Metagenomic DNA clones assayed in an *E. coli* strain deficient in EPSPS activity for the ability to recover growth; because EPSPS is the target for the herbicide glyphosate; screening took place at varying concentrations of glyphosate to select for glyphosate-resistant enzymes, allowing growth of the host strain	pACYC (2–6 kb)	NR (estimate 1 Mb)	3	[62]
Benzoylformate decarboxylase (BFD)	Metagenomic DNA clones transformed into a BFD-deficient strain of *Pseudomonas putida*; positive clones identified as those that conferred the ability to grow on benzoylformate as the sole carbon source	pBBR1MCS (2–10 kb)	NR (estimate ~100 Mb)	1	[51]

(*Continued overleaf*)

TABLE 1 (*Continued*)

Enzyme Class and *Enzyme*	Screening Method	Vector (Average Insert Size)	Amount of DNA Screened[a]	Positive Clones Recovered	Ref.
Glycerol dehydratase	Metagenomic DNA clones transformed into an *E. coli* strain engineered to contain all the genes required for anaerobic growth on glycerol as a sole carbon source with the exception of glycerol dehydratase; positive clones identified as those that conferred the ability to grow under these conditions	pBluescript SK(+) (3.3 kb, 5 kb, 3.4 kb)	NR (estimate 1.3 Gb)	2	[63]
Alcohol dehydrogenase	Recombinant *E. coli* or *Rhizobium leguminosarum* plated on minimal media with 0.1% (v/v) ethanol as the sole carbon source; positive clones identified as those that conferred the ability to grow under these conditions	pLAFR3 (25 kb)	2.75 Gb	1	[37]
Amidase	Metagenomic clones selected in an *E. coli* leucine auxotroph for the ability to recover growth on medium containing various amide-linked leucine derivatives; positive clones identified as those that grew under these conditions	pZErO-2 (5.2 kb)	NR	6	[42]

D-3-hydroxybutyrate dehydrogenase	Recombinant *Sinorhizobium meliloti* and *E. coli* strains were selected for the ability to use D-3-hydroxybutyrate as a sole carbon source; positive clones identified as those that grew under these conditions	pRK7813 (33 kb)	1.5 Gb	34	[38]
DNAse	Recombinant *E. coli* metagenomic clones were plated on Bacto DNAse methyl green agar; positive clones identified by an orange halo surrounding the colony	pBeloBAC11 (27 kb)	~100 Mb	1	[11]

[a]If NR (not reported), estimates were made by multiplying the number of clones screened by the average metagenomic DNA insert size when these figures were reported.

The protocol described below has been used in our laboratories to obtain DNA sufficient for small- and large-insert libraries; it was developed by Lynn Williamson (University of Wisconsin–Madison, Department of Bacteriology), drawing from collaborative research and experimentation with a number of published protocols.

1. Sieve soil through a fine (ca. 2 mm, sterile) mesh to eliminate roots and particulates.
2. Split 100 g of the sieved soil into two 50-g samples, and add each sample to a sterile 250-mL centrifuge bottle. To each bottle add 75 mL of Z buffer [at pH 8.0: 100 mM Tris-HCl; 100 mM sodium phosphate; 100 mM ethylenediaminetetraacetic acid (EDTA); 1.5M NaCl; 1% w/v cetrimonium bromide (CTAB)].
3. Lyse the soil samples using two freeze–thaw cycles; freezing must be complete and can be accomplished using liquid nitrogen or a dry ice/ethanol bath. The samples can be thawed by incubation in a 65°C water bath. Freezing in a dry ice bath will take about 40 min; complete thawing will take about the same amount of time.
4. To the final thaw at 65°C, add 9 mL of 20% sodium dodecyl sulfate and 4.5 mL of 5 M guanidinium isothiocyanate. Mix by gentle inversion.
5. Incubate at 65°C for 2 h with occasional mixing.
6. Centrifuge at 10°C, 20 min at $15{,}000 \times g$.
7. Pipette off the supernatant, containing DNA, into two clean, sterile 250-mL centrifuge bottles. Be sure to use wide-bore pipette tips when handling the DNA in this step, and in the remaining steps, to avoid shearing.
8. To the DNA-containing solution, add 25 mL of chloloform/isoamyl alcohol (24 : 1) and mix gently for 10 min at room temperature.
9. Centrifuge at 10°C, 20 min at $15{,}000 \times g$.
10. Pipette off the supernatant, containing DNA, into two clean, sterile 250-mL centrifuge bottles.
11. Precipitate DNA by adding isopropanol to 70% (about 40 mL per bottle) and mixing gently for 5 min. Let this sample incubate without mixing at room temperature for an additional 20 min.
12. Centrifuge at 10°C, 40 min at $15{,}000 \times g$.
13. A visible brownish pellet should form—this is the DNA. Carefully pour off the supernatant, and remove all of the remaining liquid from the bottle.
14. Resuspend the DNA pellet gently in a minimal amount of $T_{10}E_{10}$ (at pH 8.0: 10 mM Tris-HCl; 10 mM EDTA) using wide-bore pipette tips. 1 to 2 mL of $T_{10}E_{10}$ per tube should suffice. Using wide-bore pipette tips, aliquot the resuspended DNA solution equally into four 1.5-mL Eppendorf tubes.
15. Extract the DNA with an equal volume of Tris-buffered phenol/chloroform (commercially available, pH 8.0), invert to mix, and centrifuge at $16{,}000 \times g$ for 10 min at room temperature.

16. Remove the aqueous layer (top) containing DNA, and extract with chloroform/isoamyl alcohol as above.
17. Verify the presence of DNA by running a small fraction on an agarose gel. The DNA can be stored at 4°C for a few days, but should be stored at −80°C for long-term storage.

Cell Separation Method Separating bacterial cells from the soil matrix prior to lysis is done primarily to obtain larger DNA fragments, eliminate extracellular DNA, and minimize difficulties associated with using the DNA in downstream applications. This method, detailed by Liles et al. [29], relies on homogenization of the soil sample, followed by differential centrifugation and filtration to separate bacterial cells from eukaryotic cells, soil particulates, and other contaminants. Additional components may be added during the process, such as poly(vinylpolypyrrolidone) and deoxycholate, to minimize humic contamination and enhance separation of live bacterial cells from the soil matrix, respectively. Following the final centrifugation step to pelletize the bacterial cells, they can be further purified using a Nycodenz cell density gradient, which enhances lysis in the following steps. The bacterial cells are embedded in an agarose plug that is incubated in a lysis buffer, then transferred to an EDTA- and protease-containing buffer meant to eliminate nuclease activity in the lysate plug. The lysate plug is then treated with protease inhibitors, washed, and stored in TE buffer. To remove DNA from the agarose plug, the plug is embedded in an agarose gel, and the DNA is electrophoresed into this gel. The electrophoresis step is also a method by which the DNA can be verified and size-selected. Appropriately sized DNA can be extracted from the gel and stored for further use. A number of references describe cell separation and extraction protocols in greater detail [28–30,33].

C. Preparation of DNA for Cloning

Proceeding from the direct lysis protocol, the DNA can be verified, size-selected, and purified by agarose gel electrophoresis or pulsed-field gel electrophoresis. Appropriately sized DNA is carefully excised in a thin band, resulting in an agarose "noodle." The DNA can be extracted from the agarose by electroelution or by enzymatic digestion of the agarose using a commercially available enzyme (e.g., GELase, Epicentre, Madison, WI) [29]. Prior to removal from the agarose, however, an optional treatment of the excised agarose noodle with formamide and NaCl may improve suitability of the DNA for further cloning, presumably by preventing nuclease activity that remains associated with the DNA [29].

Formamide Treatment Procedure

1. To a 15-mL Falcon tube, add the agarose noodle and cover it with the formamide solution (80% v/v formamide; 0.8 M NaCl; 20 mM Tris-HCl, pH 8.0). Incubate this mixture at 14°C for 24 h.

2. Following this treatment, decant the formamide solution and dialyze the noodle in 1 L of $T_{10}E_1$ buffer (at pH 8.0: 10 mM Tris-HCl; 1 mM EDTA) at 4°C with gentle stirring for 24 to 48 h with one buffer exchange.

Removal of the DNA from the Agarose Noodle by Electroelution

1. Cut small pieces of dialysis tubing to a size slightly larger than the agarose noodle; prepare by rinsing in cold, sterile Tris–acetate–EDTA (TAE) buffer. Store in TAE buffer on ice.
2. Clip one end of the dialysis tubing and fill with sterile TAE buffer at a concentration slightly lower than 1× (to prevent the accumulation of water inside the tubing). Load gel pieces of about 3 cm into their respective dialysis tubes, aligning the long side of the gel parallel with the long side of the tubing. Squeeze out the bubbles and remove as much liquid as possible, then clip the remaining end of the tubing. Keep the tubing with the gel piece wet at all times.
3. Place the tubing with gel in an electrophoresis chamber, with the long side of the gel facing toward the positive electrode. Run for 4 h at 5 V/cm at 4°C in 1 × TAE. At the end of the run, reverse the current for about 1 min to free the DNA from the wall of the dialysis tubing.
4. Rinse the membranes in chilled, sterile milliQ (MQ) H_2O before opening, then remove the clip from one end. Using a wide-bore pipette tip, remove all of the buffer from the inside of the membrane; this contains the eluted DNA. Store the DNA at 4°C. The presence of DNA can be verified by running an agarose gel on a portion of the eluted DNA. This also serves to confirm that the DNA is of appropriate size. If necessary, the DNA can be concentrated by ethanol precipitation; however, this step will result in some DNA loss.

Conversely, DNA can be removed from the agarose noodle by treatment with commercial enzymes that catalyze removal of the agarose matrix (in this case, low-melting-point agar is recommended); examples of these enzymes include GELase (Epicentre) and agarase (multiple sources). Note, however, that borate ion (a component of TBE buffer) is inhibitory to GELase and that the GELase buffer supplied can be inhibitory to lambda phage packaging reactions, which may be used in further steps.

Because environmental and cellular contaminants can have a significant effect on the success of enzymatic manipulation of the DNA, hybridization of the DNA, and polymerase chain reaction (PCR) amplification of the DNA [31], it may be worthwhile to test the cleanliness of the DNA at this point. The purity can be tested spectrophotometrically, by observation of specific absorbance ratios (A_{260}/A_{230} compares DNA/humic materials and A_{260}/A_{280} compares DNA/protein [30]). Further, subjecting a small portion of the DNA to enzymatic digestion using a restriction endonuclease, or using the DNA in a PCR inhibition assay [32], are additional methods to determine if the DNA is clean enough for downstream applications.

D. Cloning the DNA into an Appropriate Vector

There are a number of considerations to be taken into account when deciding which cloning vector to be used in library construction, the most evident being the size of the DNA inserts being cloned. The protocol outlined above is designed to generate DNA fragments with sizes suitable for cloning into fosmid, cosmid, or bacterial artificial chromosome (BAC) vectors: called *large-insert vectors*. For the purposes of this chapter, we describe further manipulations to generate fosmid metagenomic libraries using DNA that has been size-selected at greater than 20 kb.

Large-insert libraries are less likely to result in truncated open reading frames (ORFs) and typically contain enough DNA to make generalizations about the source from which the DNA came, offering the possibility of linking phylogeny with function. Large inserts will also be necessary if you wish to identify multiple genes simultaneously: for example, those in a biosynthetic or catabolic operon. Popular vectors for large-insert libraries include the fosmids pCC1FOS and pCC2FOS (Epicentre, Madison, WI) and the BAC pCC1BAC (Epicentre). Epicentre markets a kit designed for generating fosmid-based genomic libraries in the vectors pCC1FOS or pCC2FOS; these kits work well for generating fosmid-based metagenomic libraries as well. Other common cosmid vectors are available from the companies Stratagene and Invitrogen. Small-insert libraries are popular due to their ease in cloning and manipulation. DNA isolated using commercial kits is of appropriate size and purity for small-insert cloning. Many functional screens looking for relatively small enzymes (e.g., lipases) have been successful in using such libraries. Commercially available kits for construction of metagenomic DNA libraries include Mo Bio Laboratories' PowerSoil DNA kit and Qbiogene's FastDNA SPIN kit. Both kits use physical methods for lysis, which will result in some shearing of genomic DNA, but our laboratories have had success using the PowerSoil DNA kit to generate small-insert metagenomic libraries up to about 10 kb in size.

There are many useful features that can be included in vectors to simplify issues with DNA maintenance and gene expression, such as sequence information required for integration into the chromosome, which may increase stability and reduce the need for antibiotic selection [34,35]. Other features can include the ability to induce copy number, induce gene expression through addition of promoters upstream and downstream of the multicloning site, inclusion of recognition sites for recombinases useful in conjunction with transposons for later deletions, and addition of sites that allow retrofitting with helper plasmids for movement and stability in heterologous hosts [36]. Because phage transfection is common in metagenomic DNA library construction, COS sites are common components of vectors, allowing phage packaging and transduction into *E. coli*. Phage packaging size selects DNA to produce average insert sizes varying from 30 to 40 kb, depending on the vector [34,36], and is an efficient way to generate a large-insert library in *E. coli* from many sources.

Although *E. coli* is by far the most popular host organism for functional metagenomic screening, metagenomic libraries can be screened in many other

hosts. As such, vectors designed specifically for the host are often required. As screening of metagenomic libraries in alternate hosts becomes more common, the diversity of vectors and available hosts continues to grow. Currently, vectors are available for cosmid libraries in a broad range of proteobacteria, including *Rhizobium leguminosarum* [37], *Sinorhizobium meliloti* [38], *Ralstonia metallidurans* [39], and *Pseudomonas* [40]. Vectors have also been designed for *Streptomyces* sp. and *Thermus thermophilus* [34,35]. BAC (bacterial artificial chromosome) vectors have also been developed for screening of metagenomic libraries in *Pseudomonas putida* and *Streptomyces lividans* [41].

E. Plasmid Library Production

Methods for production of metagenomic libraries in plasmid vectors are much more variable than fosmid/cosmid or BACs, as numerous vectors are suitable for construction, and small pieces of DNA can be efficiently transformed into a host cell using a variety of techniques, including heat shock, electroporation, conjugation, and phage transfection. In general, DNA insert sizes should be smaller than 10 kb for stability and efficiency in transformation, yet large enough to harbor full-length genes. Most plasmid libraries fall in the range 3 to 5 kb; however, this fragment size is dictated largely by the size of DNA fragments resulting from the method of DNA isolation. Successful combinations of vector with transformation protocol include pZErO-2 with electroporation [42] and LambdaZAP with phage transfection [43]. The pZErO-2 cloning vector (Invitrogen) is small (3.3 kb) and utilizes positive selection to prevent self-ligation. The multiple cloning site (MCS) is acceptable for blunt-end cloning and falls within the *E. coli*–lethal *ccdB* gene, and kanamycin is used for positive selection. The LambdaZAP vector (Stratagene) contains a pBluescript phagemid within a reconstructed phage genome. Metagenomic DNA up to 10 kb can be ligated into the pBluescript MCS, then the resulting mixture containing DNA to about 50 kb can be packaged into phage for transfection into *E. coli*, where the phagemid will be excised, producing a small-insert metagenomic library in *E. coli*.

F. Fosmid/Cosmid Library Production

Fosmid and cosmid vectors are designed for construction of genomic DNA libraries. They are capable of carrying large DNA inserts (40 kb range) and can contain DNA features appropriate for packaging and transfection using commercially available bacteriophage. We describe here a process for construction of a fosmid library from the purified metagenomic DNA. We have had success using Epicentre's copy control fosmid library construction kit; literature can be found at epicentre.com. Because phage packaging is most efficient with DNA of about 50 kb, DNA should be size-selected to about 40 kb for efficient fosmid library construction in Epicentre's pCC1FOS or pCC2FOS vectors (about 8 kb). Size selection can be done by pulsed-field gel electrophoresis followed by gel

extraction or by random shearing (the size must then be verified by agarose gel electrophoresis).

Proceeding from the DNA isolation step with appropriately sized DNA, the cleaned DNA will need to be end-repaired prior to ligation. This step serves to blunt the DNA fragments and provide a 5′ phosphorylation site for ligation. Following end repair, the reaction mixture should be quenched and the DNA purified by isopropanol precipitation. At this point, proceed to the ligation reaction with pCC1FOS or pCC2FOS (Epicentre) as the vector. These vectors differ by the arrangement of their respective PCR priming sites used for sequencing the resulting DNA insert. These vectors can be purchased as linearized, dephosphorylated "ligation-ready" products meant to eliminate self-ligation in the ensuing reaction. Epicentre recommends that 0.25 μg of size-selected DNA be used in the ligation with 0.5 μg of the ligation-ready fosmid backbone. Following the ligation reaction, the DNA can be stored or you can proceed directly to the phage packaging reaction and transfection. This protocol involves incubation of the ligation mixture with phage Lambda packaging extract, then transfecting the resulting mixture into a culture of growing *E. coli* cells. The recombinant metagenomic library is then selected on media containing chloramphenicol, which is the selectable marker for the fosmid backbone. The Epicentre fosmid kit provides the necessary reagents and protocols for end repair, ligation, phage packaging, and transfection.

G. BAC Library Production

BAC (bacterial artificial chromosome) libraries offer the advantage of stabilizing the largest pieces of DNA (about 150 to 350 kb). These vectors are useful for large-scale DNA sequencing projects and genomic libraries, but have found little success in functional screening of environmental metagenomic DNA libraries. This situation appears to be due to the inability to obtain and clone DNA fragments of appropriate size from environmental samples (Lynn Williamson, personal communication). Studies using BACs as metagenomic vectors have reported DNA insert sizes that are similar to those found in fosmid and cosmid vectors, and the total clone output is lower than those typically found with fosmid and cosmid libraries. Nonetheless, improvements in metagenomic DNA preparation protocols may render BAC-based libraries more useful in future endeavors. Published studies using BAC metagenomic libraries in functional metagenomic screens include those of Rondon et al., Allen et al., and Williamson et al. (Table 1) [11,16,44].

H. Plating and Storage of the Recombinant *E. coli* Metagenomic Library

Things to be considered prior to screening or selecting the recombinant *E. coli* metagenomic library include efficiency in transformation and desired plating density. The Epicentre fosmid library production kit provides control DNA from

which it is possible to determine an overall efficiency in fosmid packaging and *E. coli* transfection. For plasmid libraries using electroporation or heat shock as a means of transformation, commercially available strains should be provided with a transformation efficiency value. Further, this can be determined experimentally using a control plasmid, preferably one with a size appropriate to that seen with the metagenomic DNA library. When screening, desired plating density is entirely dependent on the nature of the screen or selection. For those screens dependent on individual colonies being observed on a standard agar petri plate, it is desirable to plate at a density such that the phenotype being observed [e.g., green fluorescent protein (gfp)−positive colony, or a colored halo surrounding a colony] can be readily differentiated (i.e., 100 to 500 CFU per plate). Conversely, when growth of the recombinant clone is dependent on the presence of a fully functional gene being expressed from the metagenomic library (i.e., a selection-based strategy, see below), the plating density should be dependent on the estimated likelihood of the particular gene being present. For example, the study of Allen et al. [16] was directed at identifying genes from a metagenomic library that conferred resistance to antibiotics on the recombinant *E. coli* host. In this case, the library was plated at about 500,000 CFU per plate for a selection-based strategy. In many cases it is possible to conduct dilution plating of the transformation mix, while storing this mixture overnight, followed by full-scale plating at the desired dilution the next day. For library storage purposes, the entirety of the library can be plated on appropriate selective media at a density low enough to ensure efficient selection. From these plates, colonies can be scraped from the media using a sterile microspatula, pooled in an appropriate storage tube, and stored at -80°C in the presence of 10% glycerol or another cryoprotectant.

I. Alternative Hosts for Screening of Metagenomic Libraries

The greatest challenges with detection of bioactivity using functional metagenomics are gene expression, production of functional protein, and metabolite production in a host that may be evolutionarily distant from the original organism. Choice of host dictates the precursors, sigma factors, rare tRNAs, and cofactors available for natural product biosynthesis, gene expression, and enzyme or molecule synthesis [15]. One study comparing expression of antibiotic biosynthesis clusters contained on BAC constructs in the hosts *Streptomyces lividans, E. coli* DH10B, and *Pseudomonas putida* found that expression of the same antibiotic clusters varied among the three hosts [41]. The host plays a critical role in determining what activities are found when functionally screening metagenomic libraries.

As discussed previously, the common host for functional metagenomics, *E. coli*, has proven suitable for many selections and screens. But while *E. coli* is an excellent choice when looking for enzymes or compounds commonly found in Proteobacteria, the ability to produce functional proteins or compounds for the degradation of unusual carbon sources or assembly of natural products can rely on components that are not found in *E. coli*.

One approach to increase detection of functional clones during screening of metagenomic libraries is the use of alternative hosts [15]. A suitable heterologous host should have several attributes, including simplicity of use, favorable growth characteristics, genetic tools, and cellular machinery needed for protein production and activity [45].

Which host is the best choice depends largely on the source of DNA and intended results. For example, when screening for biosynthesis of natural products from a soil-derived metagenomic library, a natural product–rich actinomycete may be the optimal host, as they are prominent members of the soil microbial community. One notable example of using an actinomycete for functional metagenomics is the study by Wang et al. [46]. In this study, *S. lividans* was used as an alternative host, and 1020 clones were screened using a rapid HPLC-ESIMS screening method. A total of 18 recombinants that produced members of the terragine/norcardamine natural product families, including five novel compounds, were found (Fig. 2). With almost 2% of the metagenomic clones producing natural products of this family, the hit rate is significantly higher than in previously published screens using *E. coli* as a host [47].

Several other studies have used functional screening of metagenomic libraries using hosts other than *E. coli* to discover that host specificity is common. Li et al. used defined tryptophan auxotrophs of the heterologous hosts *Rhizobium leguminosarum* and *E. coli* to look for cosmids in a wastewater metagenomic library that complemented mutations in one of three mutations in the three separate *trp* operons [48]. Several cosmids were found that complemented auxotrophy of one or more *R. leguminosarum* but not the corresponding *E. coli* strains. Further, one cosmid complemented all *E. coli* auxotrophs but none of the corresponding *R. leguminosarum* auxotrophs.

Host specificity in screening metagenomic libraries for bioactivities was further demonstrated by Wang et al., who selected a library from activated sludge and soil bacterial communities for complementation of a *Sinorhizobium meliloti bdhA* mutant, responsible for critical D-3-hydroxybutyrate dehydrogenase activity, allowing growth on D-3-hydroxybutyrate [38]. Clones that conferred D-3-hydroxybutyrate utilization on *E. coli* were also isolated. The selection found 25 unique clones in *Sinorhizobium* and nine in *E. coli*. Of the 25, one clone was able to confer D-3-hydroxybutyrate utilization on *E. coli*, and none of the clones isolated in *E. coli* were able to complement *Sinorhizobium*. The results show a surprising level of host specificity among the genes discovered.

The well-studied genus *Pseudomonas* has also been utilized as an alternative host for functional metagenomics. Ono et al. utilized *Pseudomonas putida* in a functional selection to discover naphthalene catabolism genes in a metagenomic library derived from oil-contaminated soil. This study used complementation of auxotrophs for naphthalene dioxygenase genes (*nahAc*) [40]. *P. putida* was chosen because of unpublished data showing that expression of the complete set of naphthalene-catabolic (*nah*) genes in *E. coli* does not confer the ability to use naphthalene as a sole carbon source. Moreover, *P. putida* has approximately 20 types of sigma factor compared to the seven

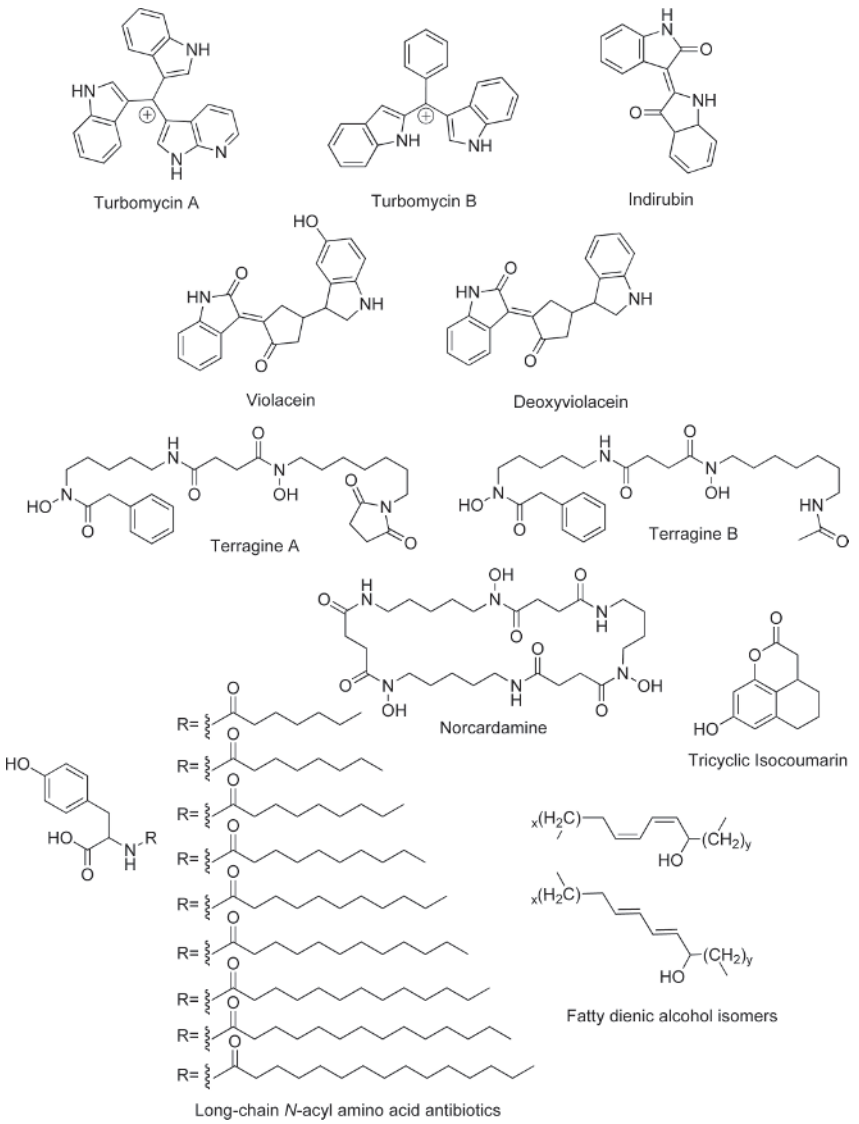

FIGURE 2 Chemical structures of natural products isolated from metagenomic DNA libraries.

found in *E. coli*, which could suggest that *P. putida* is a more suitable host for efficient expression of foreign genes [49,50]. Using a *P. putida nahAc* auxotroph, two *nahAc*-complementing cosmids were discovered in *P. putida* transformants grown on M9 agar plates supplemented with naphthalene as the sole carbon source.

P. putida was also used by Henning et al. to select a small insert soil-derived metagenomic library for clones encoding benzoylformate decarboxylases that grow on benzaldehyde as the sole carbon source [51]. One recombinant clone harbored an open reading frame whose deduced protein exhibits 72% identity to the amino acid sequence of a putative benzoylformate decarboxylase from *Polaromonas naphthalenivorans*.

The beta-Proteobacteria *Ralstonia metallidurans* has also been utilized for the discovery of pigment and antibiosis from soil-derived metagenomic libraries. A recent study found a yellow pigmented clone and another producing a novel putative type III polyketide that inhibits growth of both *Bacillus subtilis* and *Staphylococcus aureus* [39]. When transformed into *E. coli*, both cosmids did not confer the production of any detectable clone-specific small molecules.

Alternative hosts to *E. coli* for functional screening of metagenomic libraries have expanded to include diverse organisms, including the extreme thermophile *T. thermophilus* [35]. After construction of a cosmid vector allowing transformation of libraries into *T. thermophilus* and initial screening for cosmids conferring xylanase activity, libraries were screened in both *E. coli* and *T. thermophilus*, resulting in 12 and 20 active clones, respectively. When these clones were tested in the other host, 10 of the 12 clones (83%) found in *E. coli* conferred xylanase activity on *T. thermophilus*, whereas 10 of the 20 clones (50%) found in *T. thermophilus* conferred activity on *E. coli*.

Discoveries made using alternative hosts provide further support for their utilization, as these studies show that host specificity on enzyme or small-molecule detection is a critical component of functional screening. Future studies should carefully consider the source of DNA and intended results before deciding on an appropriate host or vector for functional screening.

J. Identification of Enzymes Through Functional Screening of Metagenomic Libraries

While we use the generic term *screen* when referring to the process of identifying genes of interest, functional metagenomics relies on both selection-based and screening strategies to identify genes of interest. In the following sections we point out the advantages and disadvantages of each and describe a number of screens and selections that have yielded enzymes of interest from metagenomic DNA (see Table 1).

III. FUNCTIONAL SCREENING

Screening strategies rely on a measurable difference in phenotype among those recombinant metagenomic clones carrying the gene of interest. These strategies require that all recombinant clones grow under the same conditions, after which the differences in phenotype become apparent either by visual scoring or other assaying techniques. Because all recombinant clones are grown and analyzed by the same technique, these strategies are typically time and resource

intensive. Further, phenotype scoring can become cumbersome as the number of plates/cultures increases. Depending on the nature of the screen, unforeseen problems can result in recovery of false positives as well. In these cases, metagenomic DNA inserts can harbor genes that have an unanticipated effect on the host strain, causing it to appear as a positive phenotype. One example will be given below regarding a screen for protease enzymes. Nonetheless, screening strategies are more popular than selection strategies due to a number of perceived advantages, including the relative ease involved in most screens (they typically do not require genetic manipulation of the host strain) and because many enzyme classes of interest already have well-characterized chromogenic and fluorogenic substrates available. Further, existing substrates can often be easily modified for the identification of a different enzymatic activity.

A. Screening Strategies

The in vivo use of chromogenic substrates is a common practice in molecular biology research. For example, the widespread use of blue–white colony screening is based on hydrolysis of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) by recombinant β-galactosidase inside a host cell. In this case, the X-gal is present in the medium and is taken up by growing cells. Hydrolysis of X-gal liberates 5-bromo-4-chloro-3-hydroxyindole, which spontaneously dimerizes to produce a blue dye, ultimately giving the colony a characteristic blue color (Fig. 3). Similar principles can be applied in metagenomic screens—in fact, both X-gal and the structurally related X-phosphate have been used to identify enzymes in a metagenomic library [52,53]. The most straightforward screens utilize a substrate that is cell permeable and uncolored; substrate is provided in the growth medium with the expectation that a colony (or culture) with the color indicative of the enzymatic activity harbors a gene encoding the enzyme of interest.

In addition to the screen in which β-galactosidase enzymes were identified based on cleavage of X-gal, a number of fairly straightforward chromogenic screening techniques have proved to be successful. One technique relies on use of the compound syringaldazine (Fig. 3) as a means to identify phenol oxidase enzymes [54]. Upon oxidation of syringaldazine, a reaction characteristic of a number of phenol oxidase–type enzymes, the compound turns a deep purple color. Beloqui et al. were able to identify a novel polyphenol oxidase from a Bovine rumen metagenomic library using a simple agar plate–based screen.

The strategies used for chromogenic screens can be applied with fluorogenic substrates as well. For example, Le Cleir et al. screened a metagenomic library for enzymes able to cleave 4-methylumbelliferyl-β-D-*N*,*N*′-diacetylchitobioside, a reaction characteristic of chitinase enzymes. Chitinase activity was scored by the appearance of a fluorescence phenotype, indicating cleavage of the substrate to produce 4-methylumbelliferone [55]. Similarly, the cellobioside analog of this substrate has been used to screen for cellobiohydrolases [56].

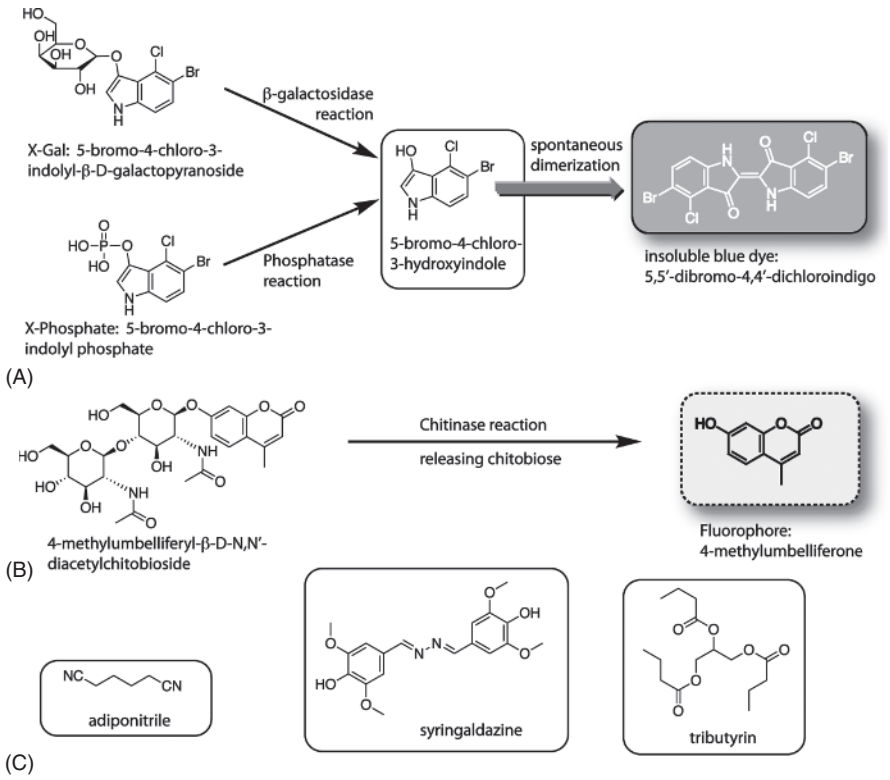

FIGURE 3 Assorted reactions and structures described in the text: (A) screening for β-galactosidase and phosphatase enzymes based on the principles of the X-gal reaction; (B) screening for chitinase enzymes based on release of the fluorophore 4-methylumbelliferone from 4-methylumbelliferyl-β-D-*N*-*N*′-diacetylchitobioside; (C) structures of compounds used in various selections and screens described in the text. (*See insert for color representation of the figure.*)

Many chromogenic/fluorogenic substrates are too large to cross a cell membrane; this includes polymeric substrates approximating polysaccharides, such as Ostazin brilliant red hydroxyethyl cellulose [11] and Remazol brilliant blue xylan [57]. In these cases, screens must either rely on export of the enzymes of interest from the host cell or permeabilization of the cell membrane to allow the enzyme access to the substrate. For example, a study by Yun et al. [58] used a functional screen for amylase enzymes that required successful secretion of the enzymes to detect the degradation of starch. Recombinant metagenomic clones were plated on starch-containing medium and grown to form colonies. The plates were overlaid with D-cycloserine to permeablize the cell membrane, and then flooded with Gram's iodine solution. Positive clones demonstrated a bright halo surrounding colonies upon illumination with fluorescent light. Liquid culture–based screens

have also used cell lysis to optimize enzyme access to substrate [59]. In a liquid culture–based screen to detect catechol dioxygenase activity, Suenaga et al. grew recombinant metagenomic clones in 96-well plates. Following growth to sufficient density, the plates were centrifuged to pellet cells, and the cells were resuspended in buffer and lysed using a mild detergent. Catechol dioxygenase activity was detected, after adding catechol, by monitoring yellow color formation as a result of catechol ring opening to form 2-hydroxymuconate semialdehyde [59].

A number of functional screens use formation of "zones of clearing" on agar plates supplemented with a substrate that imparts an opaque character to the media. Perhaps the most straightforward screen for lipase enzymes is a zone-of-clearing assay in which a lipid or ester compound is present in the solid medium: for example, the compound tributyrin (Fig. 3). Following colony growth, positive clones are identified based on formation of a clear zone surrounding the colony, indicating hydrolysis of the lipid or ester substrate [60]. One further example of a zone-of-clearing assay involves the use of skim milk agar to identify proteases. Skim milk is included in the solid medium at a minimal percentage and, following colony growth, positive clones are identified based on a clear zone surrounding the colony [60]. However, careful analysis of this screen by Jones et al. has established that the skim milk screen will give false-positive clones as well as clones with authentic protease activity [52]. In this study, the authors identified a number of other factors that result in a clearing zone using skim milk agar; these include production of recombinant glycoside hydrolases from the metagenomic DNA insert and expression of genes that induce acid production by the host strain.

One particularly interesting area with significant opportunity for development comprises the use of engineered biosensor host strains for screening. In this scenario, metagenomic DNA libraries are screened in a host strain that has been genetically modified to sense and report a specific stimulus. The stimuli could be any number of things, such as the product or substrate of an enzymatic reaction. One example of such a screen is the METREX biosensor screen of Williamson et al., which was used to identify genes involved in acylhomoserine lactone–mediated quorum sensing (Fig. 4) [44]. In this screen, the *E. coli* host strain carries a biosensor plasmid that has been specifically engineered to sense acylhomoserine lactones. This plasmid carries three genes of interest: an antibiotic-resistance selectable marker, the *luxR* transcriptional activator, and a green fluorescent protein (*gfp*) variant gene directly upstream of the lux promoter. The *luxR* gene is constitutively expressed from the plasmid, and the LuxR protein, in the absence of acylhomoserine lactone, exists as an inactive monomer in the cell. Upon binding acylhomoserine lactone, however, the luxR protein dimerizes, forming an active complex that activates expression of the *gfp* gene from the *lux* promoter. Thus, in the presence of acylhomoserine lactone (either provided directly to the cell *in trans*, or produced via an acylhomoserine lactone synthase gene on the metagenomic DNA insert) the recombinant clone will exhibit a gfp-positive phenotype. Williamson et al. used this biosensor strain to identify 12 acylhomoserine lactone

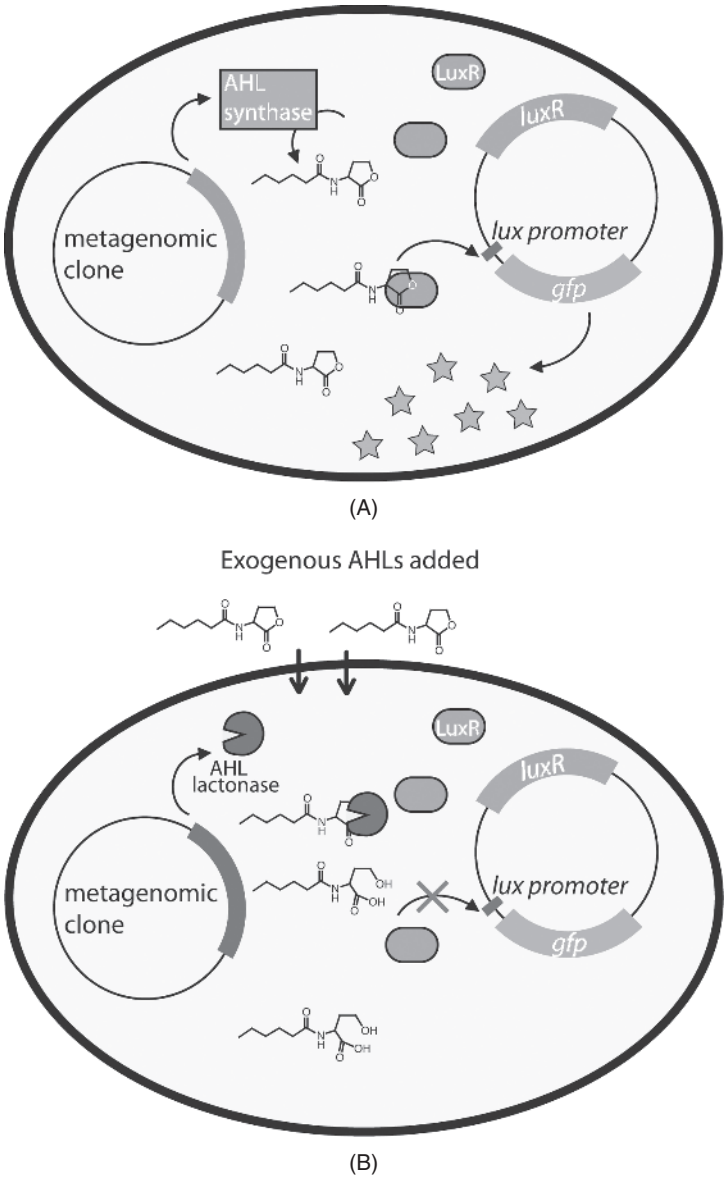

FIGURE 4 METREX biosensor screen based on acylhomoserine lactone–mediated quorum sensing. (A) In the presence of a metagenomic clone harboring an AHL synthetase, the recombinant LuxR binds the AHL and activates expression of the *gfp* gene from the *lux* promoter. (B) When exogenous AHLs are added, hydrolysis of the AHL by an AHL lactonase yields an inactive compound, giving rise to a gfp-negative phenotype. (From [44].) (*See insert for color representation of the figure.*)

synthase genes from a metagenomic DNA library by virtue of this gfp-positive phenotype [44]. Conversely, the screen can also be used to identify mechanisms for interfering with acylhomoserine lactone–mediated quorum sensing. In this screen, acylhomoserine lactone is provided to the cells *in trans*, and the screen is used to identify those clones that exhibit a gfp-negative phenotype. Williamson et al. were able to identify two acylhomoserine lactone hydrolases from this screen that catalyzed a lactonase reaction, rendering the acylhomoserine lactone inactive [44].

B. Selection-Based Strategies

Selection-based strategies rely on a gene or genes from the metagenomic DNA insert to bestow the ability of the recombinant clone to grow under the conditions specified. In contrast to the screening strategies described above, only the positive clones will grow, and as such, it is feasible to plate the recombinant clones at much higher density than would be practical for a screen. Thus, these strategies are typically less resource intensive than corresponding screens. Moreover, identification of positive clones does not require any additional instrumentation, as the positive clones are simply identified as those that grow under the specified conditions. Understandably, selection strategies are therefore limited to identifying activities that are necessary for growth; examples of these strategies are given below. However, some of these issues have been circumvented creatively by using specific auxotrophic host strains or by otherwise engineering the host strain.

As utilization of X-gal is a common screening practice in molecular biology, antibiotic resistance as a selectable marker is also a common practice among those using recombinant DNA technologies and provides an ideal example of selection-based identification of genes of interest from metagenomic libraries. An interest in identifying and characterizing antibiotic resistance genes from soil led Allen et al. to use a selection strategy to identify β-lactam resistance genes from a remote, pristine soil. By plating a recombinant *E. coli* metagenomic library on media containing lethal levels of β-lactam antibiotics, Allen et al. were able to identify 13 novel β-lactamase genes, including the first described bifunctional β-lactamase [16]. This strategy has been used to identify genes that confer resistance to a variety of antibiotics, including the β-lactams, aminoglycosides, and tetracycline [15,16].

Menadione is a polycyclic aromatic compound that is toxic to bacteria at high concentrations, due reportedly to its ability to generate reactive oxygen species in the cell. Using a strategy similar to that above, Mori et al. grew a recombinant *E. coli* metagenomic library under conditions of menadione toxicity and recovered four recombinant clones that allowed for growth under these conditions [61]. While previous studies had identified mechanisms of menadione resistance involving overexpression of genes that prevent buildup of excess reactive oxygen species in the cell, characterization of one of the positive clones from this study revealed that the resistance, somewhat surprisingly, was conferred by a recombinant UDP-glucose 4-epimerase. The authors went on to describe a potential

biological mechanism whereby the enhanced UDP-glucose 4-epimerase activity may decrease cell permeability, leading to a decrease in the menadione concentration in the cell. This example demonstrates one interesting aspect of functional metagenomics: that is, the recovery of a gene or genes that are shown to confer a particular function (e.g., menadione resistance), but whose mechanism of action is unclear at the time of discovery.

In addition to selections based on growth in the presence of a toxin, selection-based strategies can include those that confer growth on unusual or complex sources of essential nutrients. The study of Robertson et al. identified a large collection of nitrilase enzymes by selecting for growth of recombinant *E. coli* metagenomic clones in minimal media supplemented with various nitriles as the sole source of nitrogen. The structure of one of these compounds, adiponitrile, is shown in Figure 3 [43]. Thus, nitrilase activity was required to liberate nitrogen, in the form of ammonia, for growth. Selection for growth recovery was also used in a study by Wang et al. in which the authors were able to identify enzymes from a metagenomic library that are used in metabolism of the natural microbial carbon storage compound poly-3-hydroxybutyrate [38]. *E. coli* does not naturally harbor the enzyme D-3-hydroxybutyrate dehydrogenase, which is required to convert the polymeric compound into the more easily metabolized compound acetoacetate. However, in the presence of the enzyme, *E. coli* can use the resulting product as a sole carbon source. From this *E. coli* selection, Wang et al. identified nine recombinant clones that were able to grow on poly-3-hydroxybutyrate [38].

A related selection strategy utilizes auxotrophy of the host strain as a means for obtaining enzymes of interest. For example, the study of Gabor et al. took advantage of leucine auxotrophy in their host *E. coli* TOP10 [Δ(ara-leu)7697] as a means for identifying enzymes that catalyze amidase reactions [42]. The recombinant metagenomic library was selected in the presence of various amide-linked leucine derivatives: To enable growth, amidase activity was required to liberate leucine from these compounds. Six amidase-positive recombinant clones were recovered in this way [42]. In another example of engineered *E. coli* auxotrophy used for selection, 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS) activity in *E. coli* was knocked out using a kanamycin resistance cassette that was inserted into the *aroA* gene [62]. This gene encodes the enzyme EPSPS, a key enzyme in the shikimate pathway required for aromatic amino acid biosynthesis. EPSPS is also the target for the herbicide glyphosate. Jin et al. constructed metagenomic DNA libraries from glyphosate-contaminated soils, and selected for growth of the recombinant metagenomic library in the presence of 50 mM glyphosate. Three recombinant clones were recovered that harbored EPSPS activity that was resistant to glyphosate. One of these clones conferred growth at glyphosate concentrations up to 150 mM [62].

Auxotrophy can also be included as part of an engineered catabolic pathway in the host strain, such is the case in the study of glycerol dehydratase activity in soil metagenomic DNA libraries by Knietsch et al. [63]. To identify glycerol dehydratase activity in the metagenomic libraries, the host *E. coli* strain was retrofitted with the anaerobic glycerol breakdown pathway of *Citrobacter*

freundii, minus the glycerol dehydratase genes. For the modified *E. coli* host strain to grow anaerobically on glycerol as the sole carbon source, a recombinant metagenomic clone would have to provide the glycerol dehydratase activity. This engineered catabolic selection strategy resulted in the recovery of two glycerol dehydratase-positive metagenomic clones [63].

Functional metagenomics is an expanding component of natural product discovery. However, functional screens for natural products rely on expression of a full natural product biosynthesis pathway, which presents several challenges for functional metagenomics. First, the vector must allow inserts large enough to encapsulate the entire pathway, which in many cases can be over 20 kb in length. For this reason, large insert libraries are a necessity in functional screens for natural products. Second, the heterologous host used in screening must be capable of expressing the foreign pathway and producing a fully functional metabolite. Third, the screen must be sensitive enough to detect the metabolite to a level high enough to separate it from the background. Finally, the screen must be easy and cheap enough to use at a large enough scale to locate functional natural product pathways among a large sample of metagenomic DNA.

The diversity of functional screens for detection of natural product gene clusters is limited only by the creativity of the scientists that are looking for them. Some of the most successful methods thus far have been the most simple and intuitive screens. Utilizing tester organisms to detect antimicrobial activity from a library has been performed using diverse screening hosts, target organisms, and library types [64]. These screens have generated many activities, some of which have been associated directly with natural products [11,22,39,65–67].

Another straightforward and effective screen for natural product gene clusters is detection of pigment production [14,22,39]. Pigments are often associated with the conjugated bond systems present in natural products and often represent an additional screen. Pigments also represent a beneficial property for purification of the natural product that can be used to determine purification protocols without additional assays to determine in which fraction it is present. Additionally, a screening method utilizing HPLC-ESIMS was used to locate individual clones from a metagenomic library expressed in *S. lividans*, producing four novel natural products [38].

In contrast to biosynthetic pathways, enzymes involved in catabolic pathways have been identified primarily through individual functional screens, such as those described above. Although this area has not been fully explored, one could, in theory, also identify catabolic pathways from a metagenomic library based on growth selection: for example, by selecting for the ability to grow on complex hydrocarbons as a sole carbon source. In a 2005 paper, Uchiyama et al. developed a screening technology designed to identify catabolic operons from metagenomic libraries, which they designated SIGEX (Substrate-induced gene expression) [68]. This screen takes advantage of a documented feature of catabolic operons: that is, that expression of such operons is often induced in the presence of the compound to be catabolized. Many catabolic pathways encode a transcriptional activator that is sensitive to the concentration of the relevant metabolite, such that at sufficient

concentrations, the regulator (which is often genetically proximal to the operon) will activate expression of the catabolic operon. Uchiyama et al. constructed an operon-trap gfp-based vector that was used for cloning metagenomic DNA. By screening the metagenomic library in the presence of benzoate or naphthalene, they were able to identify metagenomic clones that were genetically activated in the presence of these compounds based on a gfp-positive phenotype. From this screen, they identified 62 positive clones, 58 activated in the presence of benzoate and four activated in the presence of naphthalene [68].

IV. CONCLUSIONS

Metagenomics, and in particular functional metagenomics, is a relatively new field, but it has already demonstrated significant success identifying both enzymes of interest and bioactive natural products. Among the best known enzymes in use in biotechnology are those used in detergents and food processing (e.g., lipases/esterases, proteases), those used in industrial processes (e.g., cellulases, amylases), and those used in biosynthesis of fine and bulk chemicals (e.g., nitrilase, glycerol dehydratase). A number of screens and selections have been developed to identify such enzymes from metagenomic DNA libraries. Table 1 lists relevant functional metagenomics papers according to enzyme and describes the screening techniques for each, the approximate library size, and the outcome from the screen. Although a number of studies use similar screening techniques to identify the same enzymes, we have chosen one reference for each screen type to demonstrate a representative outcome from each. For example, there are numerous published studies in which lipase activity was detected on media supplemented with tributyrin; we chose to reference the Lee et al. paper because we could calculate an approximate metagenomic library size as the reported average insert size of the metagenomic clones multiplied by the number of clones screened [60]. From these data it is possible to calculate an approximate "hit rate" for each screen as the number of positive clones recovered per unit of metagenomic DNA screened.

Table 1 also indicates the vector and average insert size from those studies in which this information was published. We note that the size of the metagenomic libraries shown in Table 1 varies from the low Mb range to the high Gb—and even, potentially, Tb—range. In the case of proteases and lipases, one might expect all of the bacterial cells in an environment to harbor such genes. As such, these studies demonstrate success in screening small-insert metagenomic libraries in the low-Mb range. To identify less abundant enzymatic activity, we recommend constructing metagenomic libraries greater than 1 Gb prior to screening. For example, the study of Allen et al. identified 13β-lactamase enzymes from a 12-Gb soil metagenomic library, corresponding to about one gene per gigabase of metagenomic DNA for an activity that is not likely to be abundant in the native soil environment [16].

Although the hit rate is indeed dependent on the abundance of the desired genes in a metagenomic library, it is now recognized that a significant population of

these genes are not properly expressed in a standard laboratory *E. coli* screening strain and therefore would not be recovered in a functional screen. As we describe earlier, a number of studies have tried to address this issue by modification of the vector backbone or by screening metagenomic libraries in different host organisms; these studies have met with varying success. Nonetheless, the matter of enhanced functional expression will continue to be a significant issue as the field develops and certainly warrants further study.

REFERENCES

1. WB Whitman, DC Coleman, WJ Wiebe. Prokaryotes: the unseen majority. Proc Natl Acad Sci USA 95:6578–6583, 1998.
2. J Dunbar, SM Barns, LO Ticknor, CR Kuske. Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol 68:3035–3045, 2002.
3. TP Curtis, WT Sloan, JW Scannell. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA 99:10494–10499, 2002.
4. M Lunn, WT Sloan, TP Curtis. Estimating bacterial diversity from clone libraries with flat rank abundance distributions. Environ Microbiol 6:1081–1085, 2004.
5. V Torsvik, J Goksoyr, FL Daae. High diversity in DNA of soil bacteria. Appl Environ Microbiol 56:782–787, 1990.
6. V Torsvik, L Ovreås. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5:240–245, 2002.
7. RI Amann, W Ludwig, KH Schleifer. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169, 1995.
8. J Handelsman, MR Rondon, SF Brady, J Clardy, RM Goodman. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249, 1998.
9. O Beja, L Aravind, EV Koonin, MT Suzuki, A Hadd, LP Nguyen, SB Jovanovich, CM Gates, RA Feldman, JL Spudich, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289:1902–1906, 2000.
10. GW Tyson, J Chapman, P Hugenholtz, EE Allen, RJ Ram, PM Richardson, VV Solovyev, EM Rubin, DS Rokhsar, JF Banfield. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43, 2004.
11. MR Rondon, PR August, AD Bettermann, SF Brady, TH Grossman, MR Liles, KA Loiacono, BA Lynch, IA MacNeil, C Minor, et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–2547, 2000.
12. JN Kim, MJ Seo, EA Cho, SJ Lee, SB Kim, CI Cheigh, YR Pyun. Screening and characterization of an esterase from a metagenomic library. J Microbiol Biotechnol 15:1067–1072, 2005.
13. D Brady, A Beeton, J Zeevaart, C Kgaje, F van Rantwijk, RA Sheldon. Characterisation of nitrilase and nitrile hydratase biocatalytic systems. Appl Microbiol Biotechnol 64:76–85, 2004.

14. SF Brady, CJ Chao, J Handelsman, J Clardy. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org Lett 3:1981–1984, 2001.
15. CS Riesenfeld, RM Goodman, J Handelsman. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ Microbiol 6:981–989, 2004.
16. HK Allen, LA Moe, J Rodbumrer, A Gaarder, J Handelsman. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. ISME J 3:243–251, 2009.
17. DD Richter, D Markewitz. How deep is soil ? Bioscience 45:600–609, 1995.
18. R Daniel. The metagenomics of soil. Nat Rev Microbiol 3:470–478, 2005.
19. JK Rhee, DG Ahn, YG Kim, JW Oh. New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. Appl Environ Microbiol 71:817–825, 2005.
20. M Ferrer, OV Golyshina, TN Chernikova, AN Khachane, D Reyes-Duarte, V Dos Santos, C Strompl, K Elborough, G Jarvis, A Neef, et al. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ Microbiol 7:1996–2010, 2005.
21. M Ferrer, OV Golyshina, TN Chernikova, AN Khachane, VA Martins Dos Santos, MM Yakimov, KN Timmis, PN Golyshin. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. Chem Biol 12:895–904, 2005.
22. DE Gillespie, SF Brady, AD Bettermann, NP Cianciotto, MR Liles, MR Rondon, J Clardy, RM Goodman, J Handelsman. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68:4301–4306, 2002.
23. A Ginolhac, C Jarrin, B Gillet, P Robe, P Pujic, K Tuphile, H Bertrand, TM Vogel, G Perriere, P Simonet, R Nalin. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol 70:5522–5527, 2004.
24. A Quaiser, T Ochsenreiter, HP Klenk, A Kletzin, AH Treusch, G Meurer, J Eck, CW Sensen, C Schleper. First insight into the genome of an uncultivated crenarchaeote from soil. Environ Microbiol 4:603–611, 2002.
25. JD van Elsas, R Costa, J Jansson, S Sjoling, M Bailey, R Nalin, TM Vogel, L van Overbeek. The metagenomics of disease-suppressive soils: experiences from the Metacontrol project. Trends Biotechnol 26:591–601, 2008.
26. B Singh, SK Gautam, V Verma, M Kumar. Metagenomics in animal gastrointestinal ecosystem: potential biotechnological prospects. Anaerobe 14:138–144, 2008.
27. J Kennedy, JR Marchesi, ADW Dobson. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. Appl Microbiol Biotechnol 75:11–20, 2007.
28. MR Liles, LL Williamson, RM Goodman, J Handelsman. Isolation of high molecular weight genomic DNA from soil bacteria for genomic library construction. In: GA Kowalchuk, FJ de Bruijn, IM Head, AD Akkermans, JD van Elsas, eds. *Molecular Microbial Ecology Manual*, 2nd ed. Dordrecht, The Netherlands: Kluwer Academic, 2004, pp. 839–852.
29. MR Liles, LL Williamson, J Rodbumrer, V Torsvik, RM Goodman, J Handelsman. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. Appl Environ Microbiol 74:3302–3305, 2008.

30. RJ Steffan, J Goksoyr, AK Bej, RM Atlas. Recovery of DNA from soils and sediments. Appl Environ Microbiol 54:2908–2915, 1988.
31. M Krsek, EMH Wellington. Comparison of different methods for the isolation and purification of total community DNA from soil. J Microbiol Methods 39:1–16, 1999.
32. DN Miller, JE Bryant, EL Madsen, WC Ghiorse. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl Environ Microbiol 65:4715–4724, 1999.
33. V Torsvik. Cell extraction method. In: AD Akkermans, JD van Elsas, FJ de Bruijn, eds. *Molecular Microbial Ecology Manual*. Dordrecht, The Netherlands: Kluwer Academic, 1995, pp. 1–15.
34. M Bierman, R Logan, K O'Brien, ET Seno, RN Rao, BE Schoner. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene 116:43–49, 1992.
35. A Angelov, M Mientus, S Liebl, W Liebl. A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. Syst Appl Microbiol 32:177–185, 2009.
36. AC Eliot, BM Griffin, PM Thomas, TW Johannes, NL Kelleher, H Zhao, WW Metcalf. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chem Biol 15:765–770, 2008.
37. M Wexler, PL Bond, DJ Richardson, AWB Johnston. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environ Microbiol 7:1917–1926, 2005.
38. CX Wang, DJ Meek, P Panchal, N Boruvka, FS Archibald, BT Driscoll, TC Charles. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. Appl Environ Microbiol 72:384–391, 2006.
39. JW Craig, FY Chang, SF Brady. Natural products from environmental DNA hosted in *Ralstonia metallidurans*. ACS Chem Biol 4:23–28, 2009.
40. A Ono, R Miyazaki, M Sota, Y Ohtsubo, Y Nagata, M Tsuda. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. Appl Microbiol Biotechnol 74:501–510, 2007.
41. A Martinez, SJ Kolvek, CL Yip, J Hopke, KA Brown, IA MacNeil, MS Osburne. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl Environ Microbiol 70:2452–2463, 2004.
42. EM Gabor, EJ de Vries, DB Janssen. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environ Microbiol 6:948–958, 2004.
43. DE Robertson, JA Chaplin, G DeSantis, M Podar, M Madden, E Chi, T Richardson, A Milan, M Miller, DP Weiner, et al. Exploring nitrilase sequence space for enantioselective catalysis. Appl Environ Microbiol 70:2429–2436, 2004.
44. LL Williamson, BR Borlee, PD Schloss, C Guan, HK Allen, J Handelsman. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl Environ Microbiol 71:6335–6344, 2005.

45. SC Wenzel, R Muller. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. Curr Opin Biotechnol 16:594–606, 2005.

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

47. J Handelsman. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685, 2004.

48. Y Li, M Wexler, DJ Richardson, PL Bond, AWB Johnston. Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. Environ Microbiol 7:1927–1936, 2005.

49. VA Dos Santos, S Heim, ER Moore, M Stratz, KN Timmis. Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. Environ Microbiol 6:1264–1286, 2004.

50. IT Paulsen, CM Press, J Ravel, DY Kobayashi, GS Myers, DV Mavrodi, RT DeBoy, R Seshadri, Q Ren, R Madupu, et al. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. Nat Biotechnol 23:873–878, 2005.

51. H Henning, C Leggewie, M Pohl, M Müller, T Eggert, KE Jaeger. Identification of novel benzoylformate decarboxylases by growth selection. Appl Environ Microbiol 72:7510–7517, 2006.

52. BV Jones, F Sun, JR Marchesi. Using skimmed milk agar to functionally screen a gut metagenomic library for proteases may lead to false positives. Lett Appl Microbiol 45:418–420, 2007.

53. K Laemmle, H Zipper, M Breuer, B Hauer, C Buta, H Brunner, S Rupp. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol 127:575–592, 2007.

54. A Beloqui, M Pita, J Polaina, A Martinez-Arias, OV Golyshina, M Zumarraga, MM Yakimov, H Garcia-Arellano, M Alcalde, VM Fernandez, et al. Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. J Biol Chem 281:22933–22942, 2006.

55. GR LeCleir, A Buchan, J Maurer, MA Moran, JT Hollibaugh. Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. Environ Microbiol 9:197–205, 2007.

56. Y Feng, CJ Duan, H Pang, XC Mo, CF Wu, Y Yu, YL Hu, J Wei, JL Tang, JX Feng. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. Appl Microbiol Biotechnol 75:319–328, 2007.

57. Y Hu, GM Zhang, AY Li, J Chen, LX Ma. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol 80:823–830, 2008.

58. J Yun, S Kang, S Park, H Yoon, MJ Kim, S Heu, S Ryu. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Appl Environ Microbiol 70:7229–7235, 2004.

59. H Suenaga, T Ohnuki, K Miyazaki. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ Microbiol 9:2289–2297, 2007.

60. SW Lee, K Won, HK Lim, JC Kim, GJ Choi, KY Cho. Screening for novel lipolytic enzymes from uncultured soil microorganisms. Appl Microbiol Biotechnol 65:720–726, 2004.

61. T Mori, H Suenaga, K Miyazaki. A metagenomic approach to the identification of UDP-glucose 4-epimerase as a menadione resistance protein. Biosci Biotechnol Biochem 72:1611–1614, 2008.

62. D Jin, W Lu, SZ Ping, W Zhang, J Chen, BQ Dun, RQ Ma, ZL Zhao, JY Sha, L Li, et al. Identification of a new gene encoding EPSPS with high glyphosate resistance from the metagenomic library. Curr Microbiol 55:350–355, 2007.

63. A Knietsch, S Bowien, G Whited, G Gottschalk, R Daniel. Identification and characterization of coenzyme B-12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl Environ Microbiol 69:3048–3060, 2003.

64. P Lorenz, K Liebeton, F Niehaus, J Eck. Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr Opin Biotechnol 13:572–577, 2002.

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

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

67. EJ Chung, HK Lim, JC Kim, GJ Choi, EJ Park, MH Lee, YR Chung, SW Lee. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. Appl Environ Microbiol 74:723–730, 2008.

68. T Uchiyama, T Abe, T Ikemura, K Watanabe. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol 23:88–93, 2005.

69. SJ Kim, CM Lee, BR Han, MY Kim, YS Yeo, SH Yoon, BS Koo, HK Jun. Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett 282:44–51, 2008.

70. Y Brennan, WN Callen, L Christoffersen, P Dupree, F Goubet, S Healey, M Hernandez, M Keller, K Li, N Palackal, et al. Unusual microbial xylanases from insect guts. Appl Environ Microbiol 70:3609–3617, 2004.

71. EW van Hellemond, DB Janssen, MW Fraaije. Discovery of a novel styrene monooxygenase originating from the metagenome. Appl Environ Microbiol 73:5832–5839, 2007.

72. CM Lee, YS Yeo, JH Lee, SJ Kim, JB Kim, NS Han, BS Koo, SH Yoon. Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome. Biochem Biophys Res Commun 370:322–326, 2008.

73. K Riaz, C Elmerich, D Moreira, A Raffoux, Y Dessaux, D Faure. A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases. Environ Microbiol 10:560–570, 2008.
74. DG Lee, JH Jeon, MK Jang, NY Kim, JH Lee, SJ Kim, GD Kim, SH Lee. Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. Biotechnol Lett 29:465–472, 2007.