Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Chemical Biology

Biocatalysts and small molecule products from metagenomic studies

Hala A Iqbal, Zhiyang Feng and Sean F Brady

The vast majority of bacteria present in environmental samples have never been cultured and therefore have not been exploited for the ability to produce useful biocatalysts or collections of biocatalysts generating interesting small molecules. Metagenomic libraries constructed using DNA extracted directly from natural bacterial communities offer access to the genetic information present in the genomes of these as yet uncultured bacteria. This review highlights recent efforts to recover both discrete enzymes and small molecules from metagenomic libraries.

Address
Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10065, United States

Corresponding author: Brady, Sean F (Sean.Brady@rockefeller.edu, sbrady@rockefeller.edu)

Current Opinion in Chemical Biology 2012, **16**:109–116

This review comes from a themed issue on
Biocatalysis and Biotransformation
Edited by Jon S Thorson and Ben Shen

Available online 26th March 2012

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2012.02.015

Introduction

It is estimated that up to 99% of bacteria in the environment are not readily cultured in the lab and as a result, most bacteria in the earth's biosphere have never been explored for the production of potentially useful products [1,2]. While it is not possible to characterize the enzymes and small molecules produced by uncultured bacteria using traditional microbiological methods, it is possible to extract microbial DNA directly from an environmental sample (environmental DNA, eDNA) and clone this DNA into easily cultured bacteria. This general strategy has been termed 'metagenomics' [3]. Metagenomic libraries constructed using DNA extracted directly from naturally occurring bacterial populations are now used extensively to screen for clones that have the genetic capacity to produce new biocatalysts as well as small molecule products. Whether looking for novel enzymes or new small molecules, most approaches for examining metagenomic libraries can be divided into two general categories: first, functional screening, which relies on the heterologous expression of eDNA in a model cultured host to yield a phenotype of interest; second, homology screening, which relies on DNA sequence similarity to identify clones containing a gene of interest (Figure 1). Here we present recent functional and homology-based metagenomic studies that have identified either novel bacterial enzymes (Table 1) or collections of enzymes (gene clusters) that encode the biosynthesis of interesting small molecules (Figure 2).

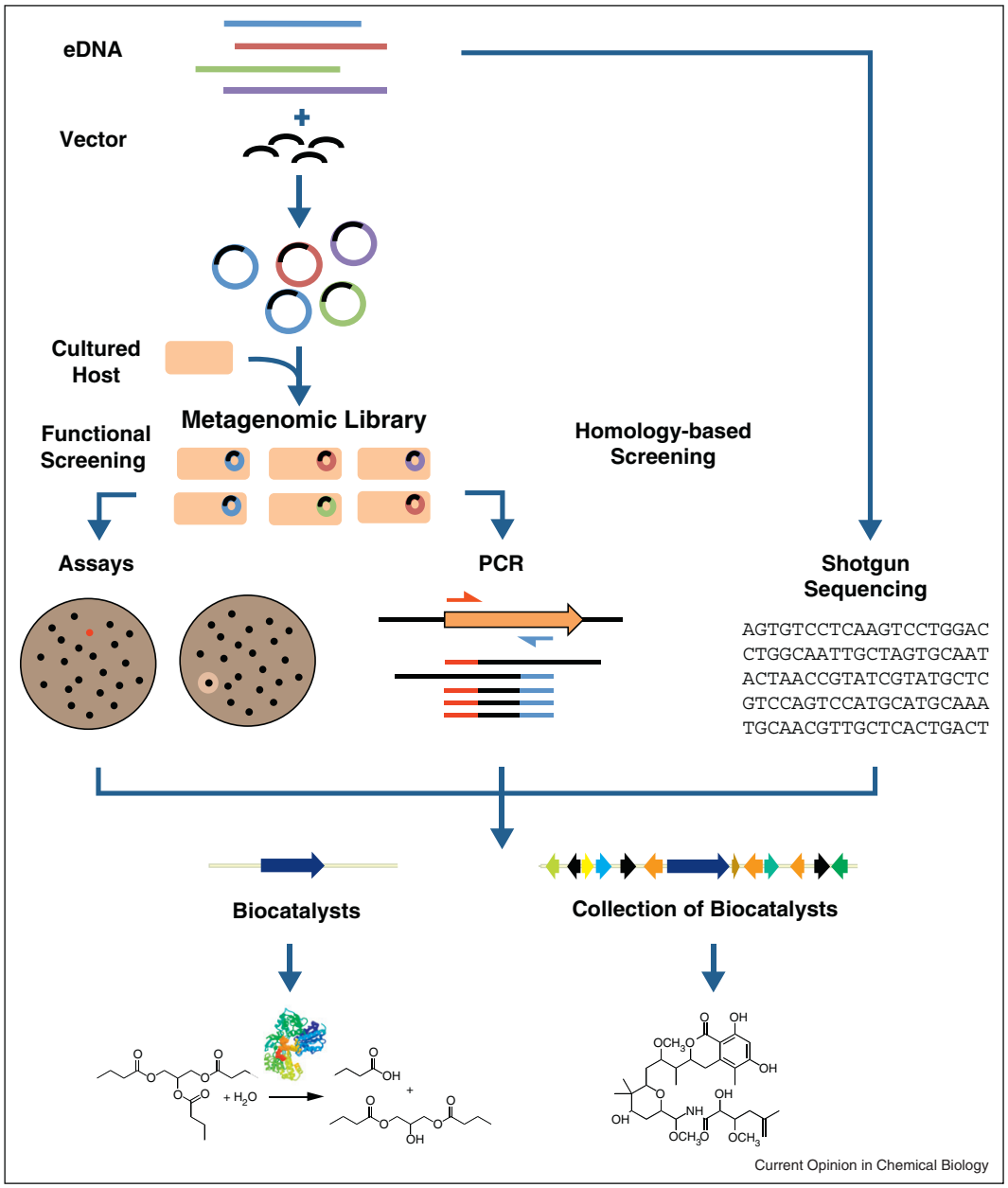
Functional screening for discrete biocatalysts

Functional screening of metagenomic libraries for industrially relevant enzymes has often been conducted using the desired enzymatic activity as a direct readout. In recent years, functional screening efforts have expanded beyond these simple assays to include the use of reporter genes and complementation as tools for identifying metagenomic clones encoding enzymes of interest.

In simple direct readout assays, libraries are plated on media containing a substrate for an enzyme of interest, and the appearance of either a halo or color is then used to identify clones encoding the product of the desired enzyme. For example, plates containing 5-bromo-4-chloro-3-indolyl caprylate (X-caprylate) or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), both of which yield a blue precipitate upon hydrolysis, have been used to recover novel esterases and β-galactosidases from metagenomic libraries [4,5]. In another recent study, Hu *et al.* screened an Antarctic desert soil metagenomic library for esterases using agar plates containing the substrate triglyceride tributyrin [6]. From a clone surrounded by a clear halo, which indicated tributyrin hydrolysis, they characterized a cold-active esterase only distantly related to reported lipases. Other examples of identifying novel enzymes from metagenomic libraries using this approach include screening for proteases on plates supplemented with skim milk, amylases on starch plates and cellulases on carboxymethylcellulose plates [7–9].

Unfortunately, most enzymes are not amenable to identification by simply screening for changes in colony appearance. Uchiyama and Miyazaki have developed 'gene-expression' reporter assays in an attempt to overcome these limitations. In their product-induced gene-expression (PIGEX) assay, a reporter gene is coupled to a product-sensitive transcription factor such that the presence of the product of a desired enzymatic reaction leads to the transcription of the reporter gene [10••]. Their

Figure 1



Overview of metagenomic screening methods.

proof of principle study used green fluorescent protein (GFP) placed under the control of a benzoate-sensitive transcription factor to screen for new amidases. In this assay, amidases convert benzamide to benzoate resulting in the expression of GFP by the benzoate-sensitive transcription factor. Using a wastewater sludge metagenomic library, the authors found 11 unique amidases, three of which were distantly related to known amidases. Conceptually similar assays, including a substrate-induced gene-expression (SIGEX) assay designed to identify enzymes involved in the catabolism of a compound of interest and a lacZ-based reporter assay designed to identify attenuators of quorum sensing, have also been used to screen metagenomic libraries [11–13].

In addition to reporter gene assays, complementation has been used as a strategy for isolating novel biocatalysts from metagenomic libraries. In these studies, the ability

Table 1

Summary of representative biocatalyst directed metagenomic screens.

Enzymes found	Screening method	eDNA source	Hits	Clones screened	Vector	Insert size (kb)	Reference
Esterase	Functional	Cotton field soil	1	92,000	Plasmid	3.5	[4]
β-Galactosidase	Functional	Oil field soil	3	12,000	Plasmid	4.8	[5]
Esterase	Functional	Antarctic soil	1	10,000	Fosmid	30	[6]
Protease	Functional	Desert soil	1	30,000	Plasmid	6	[7]
			16	17,000	fosmid	32	
Amylase	Functional	Marine sediment	1	20,000	Fosmid	30–40	[8]
Cellulase	Functional	Compost soil	4	100,000	Cosmid	33	[9]
Amidase	Gene reporter assay	Wastewater sludge	11	96,000	Fosmid	30–40	[10]
Phenol degradation	Gene reporter assay	Groundwater	62	152,000	Plasmid	7	[11]
Oxidoreductase	Gene reporter assay	Soil	1	8000	Phagemid	2.5–6	[12]
DNA polymerase	Complementation	Glacial ice	9	230,000	Plasmid	4	[13]
Histidine biosynthesis	Complementation	Forest soil	1	13,000	Plasmid		[14]
Protease	PCR homology	Grassland soil	2	11,520	Fosmid	40	[15]
		Grassland soil	3	30,494	Fosmid	40	
		Wastewater	5	26,800	Cosmid	30	
Herbicide degradation	PCR homology	Agricultural and forest soil	437				[16]
Copper P-type ATPase	PCR homology	Copper waste-exposed sediment	14				[17]
Cellulase	Shotgun sequencing	Cow rumen	27,755				[19••]
Methyl halide transferase	Synthetic metagenomics	NCBI database	89				[20]
Cellulase	Shotgun sequencing	Compost	800				[21]

of metagenomic clones to restore, or complement, a mutation in a reporter strain, is used to detect the expression of an enzyme of interest. This approach was used by Simon *et al.* to identify new DNA polymerases from uncultured environmental bacteria. In this study, a glacial ice metagenomic library was transformed into a temperature-sensitive strain of *Escherichia coli* harboring a cold-inactive polymerase mutation that is lethal below 20 °C [14]. When the library was shifted to 18 °C only *E. coli* containing clones capable of complementing the temperature-sensitive polymerase mutation could grow. In total, nine complementing clones were recovered and they were found to encode either DNA polymerases or domains typical of polymerase enzymes.

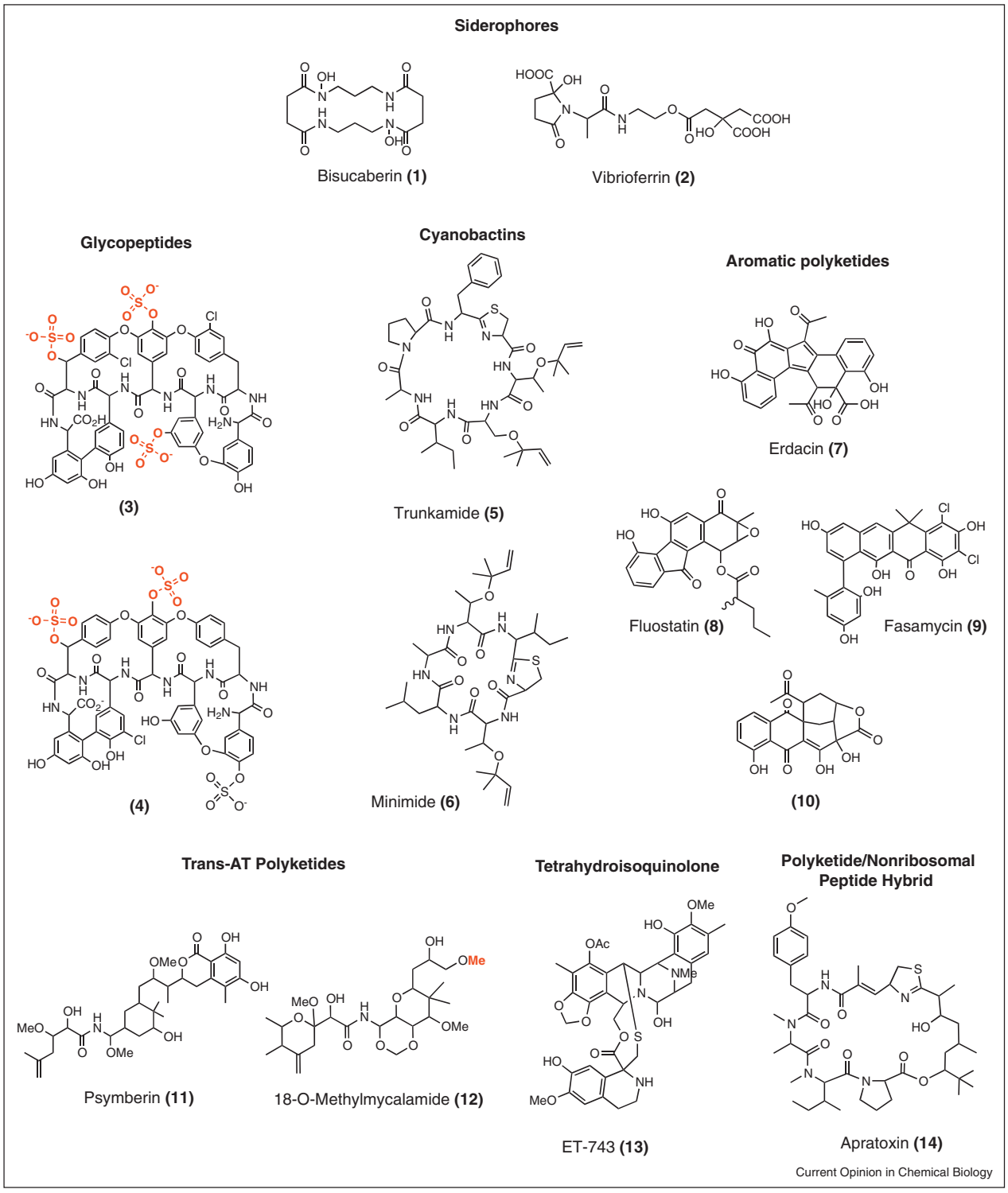
Homology-based screening

Homology-based screening has typically involved the use of PCR and/or colony hybridization to identify new members of a known gene family. PCR screening has been used, for example, to identify new proteases, herbicide-degrading genes and copper resistance enzymes [15–17]. In an extension of this strategy, Wang *et al.* generated a library of chimeric lipases with distinct substrate specificities by shuffling pools of eDNA derived PCR amplified lipase gene fragments [18].

With advances in next generation sequencing, it has now become possible to bypass experimental hybridization methods and instead use bioinformatics to detect conserved enzymatic sequence motifs in shotgun sequenced eDNA. This approach is intrinsically higher throughput and allows for more flexible homology searches than PCR. Hess *et al.* shotgun sequenced a 286 Gb cow rumen metagenomic library and searched this dataset for cellulolytic enzymes with potential applications in the biofuel industry [19••]. Their homology search revealed 28,000 potential carbohydrate active genes, of which 90 candidate genes were amplified from cow rumen eDNA, expressed and tested for activity against 10 different carbohydrate substrates. More than half of these enzymes were verified as active on at least one substrate.

The utility of sequence-based screening has also been extended to existing metagenomic datasets. Although the DNA encoding sequences found in public databases are not directly available for functional analysis, enzymes of interest can be characterized using what has been called 'synthetic metagenomics' [20]. In this approach, genes of interest are codon optimized, chemically synthesized, cloned and then expressed in a heterologous host. Bayer *et al.* bioinformatically mined methyl halide transferase (MHT) enzymes, which are used to produce agriculturally relevant fumigants, from the NCBI sequence database comprising both cultured and uncultured organisms. Their homology search revealed 89 potential MHT genes, which were chemically synthesized and then expressed in *E. coli*. All but five of the enzymes were verified to produce methyl halides in the presence of halide salts. A similar approach coupling sequence-based screening with gene synthesis, was used by Allgaier *et al.* to study glycoside hydrolases from metagenomic sequencing data derived from compost bacteria [21].

Figure 2



Representative small molecules studied using metagenomic methods. Enzymatic modifications to molecules are highlighted in red.

From discrete enzymes to biosynthetic pathways

Beyond industrially relevant enzymes, bacteria are the source of numerous small molecules with pharmaceutically important activities such as antibiosis, cytotoxicity and immunosuppression [22,23]. As with the search for discrete biocatalysts, the search for collections of biocatalysts that can biosynthesize small molecules of interest has been carried out using both functional and homology-based approaches.

Functional screening

Assays for detecting clones of interest have focused primarily on simple phenotypically identifiable traits such as antibiosis and color that are often associated with secondary metabolite production. The small number of high throughput assays available for detecting clones that produce small molecules has limited functional screening, thus new assays are needed. In an example of a novel screen, Fujita *et al.* reported the use of the indicator Chrome Azurol S (CAS), which changes from orange to blue in the presence of iron, to isolate clones encoding siderophores (iron chelators) from marine metagenomic libraries hosted in *E. coli*. In these studies they recovered gene clusters from the known siderophores Bisucaberin (**1**) and Vibrioferrin (**2**) [24,25]. Other functional studies have focused on expanding the phylogenetic diversity of bacterial hosts available for metagenomic based small molecule discovery efforts [26,27].

Homology-based screening

In homology-based small molecule discovery efforts, a metagenomic library, or even crude eDNA sequencing data, is probed to identify gene clusters containing conserved sequences that are predicted to be associated with the biosynthesis of a molecule of interest. Complete eDNA derived gene clusters and individual eDNA derived enzymes have been used to generate both new and known bioactive secondary metabolites.

Glycopeptides

Vancomycin and teicoplanin are clinically used glycopeptide antibiotics that exhibit activity against methicillin resistant Gram-positive bacteria. Banik and Brady used degenerate primers based on OxyC, a conserved oxidative coupling enzyme found in vancomycin and teicoplanin-like glycopeptide gene clusters, to identify and recover multiple predicted glycopeptide-encoding gene clusters from soil metagenomic libraries [28]. The recovery of complete biosynthetic gene clusters in these studies required the construction of megalibraries containing in excess of 10,000,000 unique cosmid clones. Recombinant sulfotransferases from one pathway were used *in vitro* to modify the teicoplanin aglycone, producing novel mono, di- and tri-sulfated glycopeptide derivatives (**3**). In a subsequent study, tailoring enzymes found in eDNA-derived glycopeptide biosynthetic clusters were expressed in *Streptomyces toyocaensis*, which naturally produces the mono-sulfated glycopeptide A47934 [29]. This resulted in new glycopeptide derivatives featuring methyl, sulfur and sugar substituents, which were further derivatized *in vitro* using sulfotransferases. In total 15 new anionic (sulfated) glycopeptide antibiotics were generated in these studies (**4**).

Cyanobactins

Cyanobactins are ribosomally produced cyclic peptides that are prevalent in extracts derived from marine samples, and they frequently display interesting cytotoxic activities [30]. In 2005, two separate groups reported the cloning and heterologous expression of biosynthetic gene clusters for the cyanobactins patellamide from metagenomic libraries of uncultured cyanobacterial symbionts associated with marine *Didemnidae* sponges [31,32]. Using end sequencing data and PCR primers based on conserved cyanobactin biosynthetic genes, gene clusters that encode both known and novel cyanobactins have subsequently been recovered from other marine symbiont metagenomic libraries. Donia *et al.* recovered the complete gene cluster for the known cyanobactin trunkamide (**5**) on a single fosmid found in an ascidian metagenomic library, and in a similar study, the same group cloned the gene cluster for minimide (**6**), which they predicted would be a novel cyanobactin [33,34••]. Upon re-engineering and optimization of this gene cluster, minimide was successfully heterologously expressed in *E. coli*. The Schmidt group took advantage of the fact that the structural diversity seen in cyanobactins largely arises from small changes in the gene encoding the ribosomally translated precursor peptide and employed a combination of orthogonal tRNAs loaded with unnatural amino acids, precursor peptide mutagenesis, and gene shuffling to generate a library of hybrid cyanobactins using the eDNA derived trunkamide biosynthetic machinery [35].

Type II polyketides

A structurally diverse collection of aromatic small molecules, including many antimicrobial and anticancer agents (*e.g.* tetracycline and doxorubicin), arise from iterative or Type II polyketide synthases [36]. While the gene clusters that code for the biosynthesis of these molecules are very different in their details, they all contain a minimal polyketide synthase that is composed of three highly conserved genes: two ketosynthases (KSs) and an acyl carrier protein. Both PCR studies and high throughput sequencing efforts have shown that eDNA samples are rich in novel minimal PKS genes [37,38]. In an effort to identify novel bioactive metabolites, Feng *et al.* used degenerate primers based on conserved sequences found in minimal polyketide synthase genes to recover polyketide biosynthetic gene clusters captured in soil eDNA libraries. Minimal PKS containing eDNA clones were introduced into model cultured *Streptomyces* hosts for heterologous expression studies. Characterization of the metabolites produced in

these studies identified a number of new metabolites with either previously unknown or rare carbon skeletons (**7**–**10**), one family of which exhibits activity against antibiotic resistant bacteria [39–40,41••].

Trans-acyltransferse (trans-AT) polyketides

A number of pharmacologically interesting polyketides isolated from uncultured marine symbionts are predicted to be biosynthesized using freestanding acyltransferases, or trans-ATs [42]. A productive strategy for identifying gene clusters encoding these metabolites has been to probe marine metagenomic libraries for trans-AT specific sequences. Trans-AT KS domains phylogenetically cluster in accordance with the specific substrate used by the KS domain. Using primers designed to recognize KS domains that utilize acetyl-derived starter units, Fisch *et al.* isolated a single amplicon present in a psymberin-producing marine sponge library that was absent in libraries from samples that did not produce the compound [43]. They then used this sequence to recover the psymberin (**11**) gene cluster from a *Psammocinia bulbosa* fosmid metagenomic library. Although no report of the heterologous expression of the complete psymberin gene cluster has yet appeared in the literature, the Piel group has reported the use of eDNA derived tailoring enzymes to modify trans-AT polyketides *in vitro*. They used an O-methyltransferase from the pederin gene clusters, which they cloned a number of years ago from a beetle symbiont metagenomic library, to site-specifically methylate the mycalamide A resulting in the production of a hybrid compound 18-O-methylmycalamide (**12**) with enhanced antitumor activity [44,45].

ET-743

Rath *et al.* recovered the biosynthetic cluster for the anticancer agent ET-743 (**13**) from a metagenomic library of uncultured tunicate bacterial symbionts [46••]. The parallels between ET-743 and other tetrahydroisoquinoline structures such as saframycin and safracin led the authors of this study to the hypothesis that ET-743 was of bacterial origin and encoded by a non-ribosomal peptide synthase similar to that seen in other tetrahydroisoquinoline gene clusters. In a cloning-independent strategy, DNA isolated directly from field collected bacterial symbionts found in a tunicate shown to produce ET-743 was 454 pyrosequenced. This data was assembled and candidate ET-743 related nonribosomal peptide synthetase (NRPS) genes were identified by their similarity to saframycin and safracin biosynthetic genes. One NRPS cluster found in these experiments was predicted to contain all of the biosynthetic genes necessary for the assembly of a tetrahydroisoquinoline core. The enzymatic activity of the predicted reductive termination domain seen in this cluster was subsequently confirmed *in vitro* using saframycin intermediates as substrates, linking the gene cluster to ET-743 biosynthesis in a cloning-independent manner.

Single cell genomics

Single cell genomics has been used to aid sequence-based screening efforts. In this strategy, single bacterial cells are isolated from complex microbial communities and then subjected to multiple displacement amplification (MDA) in order to obtain sufficient genomic DNA for sequencing [47]. In a study by Grindberg *et al.*, single cells of *Lyngbya bouillonii* were isolated from cyanobacterial filaments containing a consortium of symbiotic bacteria [48]. DNA from these cells was amplified by MDA, sequenced and confirmed to be *L. bouilonii* by 16S analysis. Biosynthetic genes of the known polyketide/nonribosomal peptide hybrid Apratoxin (**14**) were identified by screening *in silico* for genes predicted to be involved in the introduction of the β-alkylation seen in this metabolite. Clones containing the apratoxin biosynthetic gene cluster were subsequently recovered from a *L. bouillonii*-symbiont metagenomic genomic library. In other examples of this approach, fluorescence-activated cell sorting (FACS) has been used to obtain single cells prior to MDA and the sequence data obtained from these cells has shed light on the biosynthetic capacities of individual sponge symbionts [49,50].

Future prospects

The reservoir of potentially useful products encoded by the earth's microbiome is still largely underexplored, as only a small minority of bacterial species has been cultured in the laboratory. Metagenomic methods have begun to provide access to both biocatalysts and metabolites encoded within the genomes of these previously inaccessible bacteria. In the years to come, advances in sequencing technologies, bioinformatics prediction tools, heterologous expression methods and synthetic biology will undoubtedly increase the efficiency and utility of this general approach.

Acknowledgements

This work was supported by NIH GM077516. SFB is a Howard Hughes Medical Institute early career scientist.

References and recommended reading

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

- •• of outstanding interest

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

2. Rappé MS, Giovannoni SJ: **The uncultured microbial majority**. *Annu Rev Microbiol* 2003, **57**:369-394.

3. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM: **Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products**. *Chem Biol* 1998, **5**:R245-R249.

4. Yao J, Fan XJ, Lu Y, Liu YH: **Isolation and characterization of a novel tannase from a metagenomic library**. *J Agric Food Chem* 2011, **59**:3812-3818.

5. Wang K, Li G, Yu SQ, Zhang CT, Liu YH: **A novel metagenome-derived beta-galactosidase: gene cloning, overexpression,**

purification and characterization. *Appl Environ Microbiol* 2010, **88**:155-165.

6. Hu XP, Heath C, Taylor MP, Tuffin M, Cowan D: **A novel, extremely alkaliphilic and cold-active esterase from Antarctic desert soil**. *Extremophiles* 2011, **16**:79-86.

7. Neveu J, Regeard C, Dubow MS: **Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts**. *Appl Environ Microbiol* 2011, **91**:635-644.

8. Liu Y, Lei Y, Zhang X, Gao Y, Xiao Y, Peng H: **Identification and phylogenetic characterization of a new subfamily of α-amylase enzymes from marine microorganisms**. *Mar Biotechnol* 2011 doi: 10.1007/s10126-011-9414-3. [Epub ahead of print].

9. Pang H, Zhang P, Duan C-J, Mo X-C, Tang J-L, Feng J-X: **Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase**. *Curr Microbiol* 2009, **58**:404-408.

10. •• Uchiyama T, Miyazaki K: **Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes**. *Appl Environ Microbiol* 2010, **76**:7029-7035.
This paper describes a gene expression reporter assay for detecting enzymes via their products. A metabolite-sensitive transcription factor is coupled to a reporter gene to allow detection of the product. In this proof of principle study, PIGEX (product-induced gene expression) was used to find known and novel amidases.

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

12. Utpatel C, Bijtenhoorn P, Mayerhofer H, Mu J, Hornung C, Szesny M, Grond S, Thu A, Daniel R, Dierking K *et al.*: **A novel metagenomic short-chain dehydrogenase/reductase attenuates *Pseudomonas aeruginosa* biofilm formation and virulence on *Caenorhabditis elegans***. *PLoS One* 2011, **6**:e26278.

13. Schipper C, Hornung C, Bijtenhoorn P, Quitschau M, Streit WR, Grond S: **Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa***. *Appl Environ Microbiol* 2009, **75**:224-233.

14. Simon C, Herath J, Rockstroh S, Daniel R: **Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice**. *Appl Environ Microbiol* 2009, **75**:2964-2968.

15. Niehaus F, Gabor E, Wieland S, Siegert P, Maurer KH, Eck J: **Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases**. *Microb Biotechnol* 2011, **4**:767-776.

16. Zaprasis A, Liu Y-j, Liu S-j, Horn MA, Drake HL: **Abundance of novel and diverse tfdA-like genes, encoding putative phenoxyalkanoic acid herbicide-degrading dioxygenases, in soil**. *Appl Environ Microbiol* 2010, **76**:119-128.

17. De la Iglesia R, Valenzuela-Heredia D, Pavissich JP, Freyhoffer S, Andrade S, Correa Ja, González B: **Novel polymerase chain reaction primers for the specific detection of bacterial copper P-type ATPases gene sequences in environmental isolates and metagenomic DNA**. *Lett Appl Microbiol* 2010, **50**:552-562.

18. Wang Q, Wu H, Wang A, Du P, Pei X, Li H, Yin X, Huang L, Xiong X: **Prospecting metagenomic enzyme subfamily genes for DNA family shuffling by a novel PCR-based approach**. *J Biol Chem* 2010, **285**:41509-41516.

19. •• Hess M: **Metagenomic discovery of biomass-degrading genes and genomes from cow rumen**. *Science* 2011, **331**:463-467.
In this report, a sequence-based approach is used to uncover novel biocatalysts. Shotgun sequencing data from the microbial community of cow rumen was searched *in silico* for new cellulases.

20. Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA: **Synthesis of methyl halides from biomass using engineered microbes**. *J Am Chem Soc* 2009, **131**:6508-6515.

21. Allgaier M, Reddy A, Park JI, Ivanova NDP, Sapra R, Hazen TC, Simmons BA, Vandergheynst JS: **Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community**. *PLoS One* 2010, **5**:e8812.

22. Butler MS, Newman DJ: **Mother Nature's gifts to diseases of man: the impact of natural products on anti-infective, anticholestemics and anticancer drug discovery**. *Prog Drug Res* 2008, **65**:3-44.

23. Newman DJ, Cragg GM: **Natural products as sources of new drugs over the last 25 years**. *J Nat Prod* 2007, **70**:461-477.

24. Fujita MJ, Kimura N, Yokose H, Otsuka M: **Heterologous production of bisucaberin using a biosynthetic gene cluster cloned from a deep sea metagenome**. *Mol Biosyst* 2012, **8**:482-485.

25. Fujita MJ, Kimura N, Sakai A, Ichikawa Y, Hanyu T, Otsuka M: **Cloning and heterologous expression of the vibrioferrin biosynthetic gene cluster from a marine metagenomic library**. *Biosci Biotechnol Biochem* 2011, **75**:2283-2287.

26. Craig JW, Chang F-Y, Kim JH, Obiajulu SC, Brady SF: **Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria**. *Appl Environ Microbiol* 2010, **76**:1633-1641.

27. Kakirde KS, Wild J, Godiska R, Mead Da, Wiggins AG, Goodman RM, Szybalski W, Liles MR: **Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries**. *Gene* 2011, **475**:57-62.

28. Banik JJ, Brady SF: **Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary**. *Proc Natl Acad Sci U S A* 2008, **105**:17273-17277.

29. Banik JJ, Craig JW, Calle PY, Brady SF: **Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products**. *J Am Chem Soc* 2010, **132**:15661-15670.

30. Sivonen K, Leikoski N, Fewer DP, Jokela J: **Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria**. *Appl Environ Microbiol* 2010, **86**:1213-1225.

31. Long PF, Dunlap WC, Battershill CN, Jaspars M: **Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production**. *Chembiochem* 2005, **6**:1760-1765.

32. Schmidt EW, Nelson JT, Rasko Da, Sudek S, Eisen Ja, Haygood MG, Ravel J: **Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella***. *Proc Natl Acad Sci U S A* 2005, **102**:7315-7320.

33. Donia MS, Ravel J, Schmidt EW: **A global assembly line for cyanobactins**. *Nat Chem Biol* 2008, **4**:341-343.

34. •• Donia MS, Ruffner DE, Cao S, Schmidt EW: **Accessing the hidden majority of marine natural products through metagenomics**. *Chembiochem* 2011, **12**:1230-1236.
The authors report a library of novel cyanobactins that was produced using a combination of orthogonal tRNAs loaded with unnatural amino acids, precursor peptide mutagenesis, and gene shuffling.

35. Tianero MDB, Donia MS, Young TS, Schultz PG, Schmidt EW: **Ribosomal Route to small molecule diversity**. *J Am Chem Soc* 2012, **134**:418-425.

36. Hertweck C, Luzhetskyy A, Rebets Y: **Bechthold A: Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork**. *Nat Prod Rep* 2007, **24**:162-190.

37. Wawrik B, Kutliev D, Abdivasievna Ua, Kukor JJ, Zylstra GJ, Kerkhof L: **Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia**. *Appl Environ Microbiol* 2007, **73**:2982-2989.

38. Wawrik B, Kerkhof L, Zylstra GJ, Jerome J, Kukor JJ: **Identification of unique Type II polyketide synthase genes in soil**. *Appl Environ Microbiol* 2005, **71**:2232-2238.

39. King RW, Bauer JD, Brady SF: **An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin**. *Angew Chem Int Ed* 2009, **48**:6257-6261.

40. Feng Z, Kim JH, Brady SF: **Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived Type II PKS gene cluster**. *J Am Chem Soc* 2010, **132**:11902-11903.

41. •• Feng Z, Kallifidas D, Brady SF: **Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites**. *Proc Natl Acad Sci U S A* 2011, **108**:12629-12634.

In this paper, degenerate primers based on the minimal type II PKS were used to recover metagenomic clones containing polyketide gene clusters, and heterologous expression of these clusters led to the identification novel metabolites.

42. Piel J: **Biosynthesis of polyketides by trans-AT polyketide synthases**. *Nat Prod Rep* 2010, **27**:996-1047.

43. Fisch KM, Gurgui C, Heycke N, van der Sar Sa, Anderson Sa, Webb VL, Taudien S, Platzer M, Rubio BK, Robinson SJ *et al.*: **Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting**. *Nat Chem Biol* 2009, **5**:494-501.

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

45. Zimmermann K, Engeser M, Blunt JW, Munro MHG, Piel J: **Pederin-type pathways of uncultivated bacterial symbionts: analysis of o-methyltransferases and generation of a biosynthetic hybrid**. *J Am Chem Soc* 2009, **131**:2780-2781.

46. •• Rath CM, Janto B, Earl J, Ahmed A, Hu FZ, Hiller L, Dahlgren M, Kreft R, Yu F, Wolff JJ *et al.*: **Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743**. *ACS Chem Biol* 2011, **6**:1244-1256.

The biosynthetic cluster of ET-743 was identified in shotgun sequencing data obtained from field collected marine invertebrates.

47. Dean FB, Hosono S, Fang L, Wu X, Faruqi a F, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J *et al.*: **Comprehensive human genome amplification using multiple displacement amplification**. *Proc Natl Acad Sci U S A* 2002, **99**:5261-5266.

48. Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu W-t, Gerwick L, Dorrestein PC, Pevzner P, Lasken R *et al.*: **Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage**. *PLoS One* 2011, **6**:e18565.

49. Siegl A, Hentschel U: **PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification**. *Environ Microbiol Rep* 2010, **2**:507-513.

50. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T, Hentschel U: **Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges**. *ISME J* 2011, **5**:61-70.