Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Metagenomic small molecule discovery methods

Zachary Charlop-Powers, Aleksandr Milshteyn and Sean F Brady

Metagenomic approaches to natural product discovery provide the means to harvest bioactive small molecules synthesized by environmental bacteria without the requirement of first culturing these organisms. Advances in sequencing technologies and general metagenomic methods are beginning to provide the tools necessary to unlock the unexplored biosynthetic potential encoded by the genomes of uncultured environmental bacteria. Here, we highlight recent advances in sequence-based and functional-based metagenomic approaches that promise to facilitate antibiotic discovery from diverse environmental microbiomes.

Addresses
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 (sbrady@rockefeller.edu, sbrady@mail.rockefeller.edu)

Current Opinion in Microbiology 2014, **19**:70–75

This review comes from a themed issue on **Novel technologies in microbiology**

Edited by **Emmanuelle Charpentier** and **Luciano Marraffini**

For a complete overview see the Issue and the Editorial

Available online 5th July 2014

http://dx.doi.org/10.1016/j.mib.2014.05.021

1369-5274/© 2014 Published by Elsevier Ltd. All right reserved.

Introduction

Many important antibiotic compounds have been isolated from cultured bacteria; however, the vast majority of bacteria remain recalcitrant to culturing [1]. It is estimated that soil contains as many as 10^5 unique species per gram and that uncultured microorganisms outnumber cultured ones by two to three orders of magnitude [2–4]. Metagenomics is a culture-independent approach that seeks to access the biosynthetic capacity of the 'uncultured majority' of bacterial species. By directly capturing DNA from the environment (environmental DNA, eDNA) and subsequently identifying, isolating, and expressing biosynthetic gene clusters in heterologous hosts, metagenomics has the potential to provide a complete toolkit for bringing biosynthetic diversity from the environment into drug discovery pipelines. Two general approaches are employed for interrogating and exploiting metagenomic eDNA for the production of small molecules. Sequence-based approaches profile the biosynthetic content of metagenomic samples, identify high-value targets, and aid in the targeted recovery of complete biosynthetic pathways from eDNA cosmid libraries. These recovered clusters often require genetic manipulation to activate small molecule production in a heterologous host. In contrast, function-based approaches aim to identify clones that are already biosynthetically active in a heterologous host by detecting a clone-induced phenotype in a host organism. This review covers recent technological and experimental advances that are accelerating metagenomic small molecule discovery efforts with a focus on (a) sequence homology-based techniques that facilitate metagenome profiling and gene cluster recovery and (b) advances in function-based methods that expedite the identification of bioactive clones.

Sequence-based metagenomic studies

The precipitous reduction of DNA sequencing cost is transforming the process of natural product drug discovery. Whereas classic, culture-based studies required isolation of compounds in the search for novel bioactivity, the availability of sequence data has driven the development of bioinformatic tools that can streamline the identification of target gene clusters without requiring chemical isolation. The methods used to identify gene clusters of interest in metagenomes generally fall into one of two categories: shotgun sequencing or PCR-based sequence tag approaches.

Shotgun studies

Genome-based approaches to natural product discovery stand to benefit from the proliferation of sequencing technologies and the accompanying bioinformatic analyses they enable. The torrent of genome sequences of cultured bacteria (>500/month at NCBI [5]) is sparking renewed interest in natural product discovery. This is due in large part to identification of previously unknown gene cluster in many organisms, including those that have been thoroughly studied [6]. Computational tools that scan assembled genomes and identify biosynthetic gene clusters, such as AntiSMASH and np.searcher, are now able to predict the expected natural products encoded by these clusters [7•,8]. Application of such tools to all newly sequenced genomes is becoming a routine part of new genome analysis, providing a way to identify and rank new clusters for genome mining. These tools can also be applied to assembled contigs generated from metagenomic sources and used to identify clusters from uncultured organisms, a

strategy that has been particularly useful in the elucidation of the small molecule producing clusters of uncultured endosymbionts of marine [9•,10,11] and terrestrial [12•] metazoans. The assembly of symbiont genomes from metagenomic samples has been used to identify the gene clusters encoding a potent cytotoxin, patellazole, a novel polyketide, nosperin, as well as to guide the discovery of a genus of bacterial symbionts, *Entotheonella*, with promising biosynthetic potential [9•,12•,13]. By coupling deep-sequencing with other tools like whole genome amplification [14] and single-cell isolation, extensive biosynthetic information can be gleaned from otherwise difficult-to-access organisms, as is the case with the recent elucidation of the apratoxin cluster from a marine cyanobacterium [15]. The application of whole-genome sequencing is a useful tool for the characterization of endosymbionts and other relatively small metagenomes; however, other techniques are necessary for the complex metagenomes found in many natural environments like soil.

Sequence tag tools

A typical soil metagenome may contain 10^4–10^5 unique species [2,3]. Shotgun assembly of such metagenomes is still very challenging. Fortunately, substantial information about biosynthetic genes can be obtained through the use of simpler, PCR-based sequence tag approaches [16]. Sequence tags are PCR amplicons generated using primers targeting conserved biosynthetic genes that can be used for phylogenetic analysis. They take advantage of the modularity of biosynthetic systems, which have evolved for horizontal transfer of useful phenotypes (*e.g.* small molecule production), while facilitating the creation of chemical novelty through well-established genetic mechanisms [17,18]. Sequence tags can be mapped to related sequences within known biosynthetic clusters, which is the basis of the eSNAPD and NaPDoS programs [19•,20]. At high degrees of sequence similarity (eSNAPD: *E*-value $<10e^{-40}$, NaPDoS: 90% sequence identity), short sequence tags of only several hundred base pairs can effectively match a read to a reference gene cluster. Remarkably, the general structure of an entire gene cluster can generally be inferred from the tag, as validated by the recovery of eDNA clones predicted to encode novel glycopeptide, lipopeptide, and bis-intercalator natural products [19•]. The true utility of these approaches is not in the identification of known gene clusters but instead in rapidly identifying gene clusters encoding congeners of valuable compounds, or in finding potentially novel gene clusters that have remained undetected in the environment [19•].

Earlier applications of sequence tags to natural product characterization were for genotyping of strains of the prolific marine *Actinomcyete* genus *Salinispora* [21,22]. It allowed a quick, inexpensive way to profile biosynthetic potential of cultured organisms without requiring chemical isolation. The use of 454-pyrosequencing enabled scaling of this approach to profile entire metagenomes. Pyrosequencing of ketosynthase and condensation domains from marine sponge metagenomes uncovered hundreds of previously unknown sequences, including several clades that had not been previously observed by extensive Sanger sequencing of these sponge metagenomes [23•]. Furthermore, a head-to-head comparison of shotgun sequencing of the same samples demonstrated that PCR-based approaches were often 10–100 times more sensitive in identifying unique sequences from a metagenome of interest [23•]. Even greater biosynthetic biodiversity was observed in desert soil microbiomes, where 1000s of unique adenylation domain sequences were detected with only a fraction of them shared among distinct microbiomes. A similar pattern was observed for amplicons derived from Type I (ketosynthase) and Type II (ketosynthase alpha) polyketide biosynthesis, suggesting that soil metagenomes are likely to be a rich source of novel bioactive compounds [24,25]. While purified metagenomic DNA is suitable for profiling the biosynthetic diversity, arrayed, large-insert libraries are needed to facilitate isolation of the identified pathways [26]. Prior to choosing a sample for use in library construction, sequence tag-based methods can be employed to screen eDNA to identify the most biosynthetically rich environments. These same methods can then be used to identify and guide the isolation of specific clones from eDNA libraries. Clones recovered from metagenomic libraries in this manner have been heterologously expressed to yield new bioactive Type II polyketide antibiotics [25,27]; new tryptophan-based cytotoxins [28–30]; the marine-derived siderophores bisucaberin and vibrioferrin [31,32]; modified versions of the antitumor compound pederin [33]; cyanobacterial-derived cyclic peptides [34]; and new members of the microviridin family of ribosomally synthesized peptides [35,36].

Future application of sequencing technology to metagenomes

Several promising technologies may extend the power of whole genome sequencing to metagenomes. Nano-pore based sequencing boasts long read lengths that alleviate the problem of assembling repetitive regions within the genome and is quickly becoming the method of choice for bacterial genome sequencing [37,38]. Long read sequencing technologies are of particular interest when sequencing natural product gene clusters due to the highly repetitive nature of some biosynthetic gene clusters. Additionally, single-cell and microdroplet-based methods can now obtain sequence data from single cells, without requiring the generation of an eDNA library, and can facilitate the sequencing of the rare biosphere [39•,40]. These sequencing methods will expedite the *in silico* characterization of naturally occurring biosynthetic gene

clusters and will push the molecule-discovery bottleneck downstream to the activation of gene cluster expression.

Function-based metagenomics

Sequence-based metagenomics takes full advantage of the information gained through advances in DNA sequencing. Unfortunately, pathways recovered by sequence-based methods often require genetic refactoring to activate clusters in a heterologous host. Functional metagenomics provides a complementary approach that bypasses the refactoring steps by screening for, and isolating, clones that are already active in the heterologous host strain. A variety of functional screens have been developed to date that rely on phenotypic detection using: pigmentation, enzymatic or antibiotic activity [41–43]; selection of complementation-dependent reporters [44,45]; and substrate induction (SIGEX and METREX) [46,47]. Direct screening of fermentation broths has also been used, although this approach becomes impractical with increasing library size [48]. While functional metagenomics presents a powerful set of tools for identifying bioactive compounds encoded by environmental microbiomes, the size and the heterogeneous nature of eDNA libraries pose a number of challenges that are currently being addressed through a combination of technical advances and new screening methods.

Library creation and maintenance

Recent advances in eDNA cloning and in broad-host range vector design can facilitate the creation of eDNA libraries and allow a single library to be screened in a variety of hosts. In contrast to genome-based cosmid libraries, creation of eDNA libraries can be challenging due to difficulties associated with obtaining sufficient quantities of high molecular weight (HMW) DNA free of environmental inhibitors that can interfere with cloning. A newly developed technology, synchronous coefficient of drag alteration (SCODA), has enabled recovery of HMW eDNA from virtually any source by removing interfering contaminants while concentrating dilute samples [49•,50]. The cloning of eDNA into a broad-host range shuttle vector facilitates transfer of eDNA libraries into different hosts, where orthogonal collections of genes are likely to be expressed. Vectors based on ΦC31 and ΦBT1 phage integrase systems for transfer of libraries into diverse *Streptomyces* spp. have existed for some time [51,52]. More recently, RK2-derived vectors have been constructed to allow movement of libraries into diverse alpha, beta, and gamma proteobacterial species [53–55]. Creating libraries in shuttle vectors allows for conjugative transfer of eDNA into a wide range of hosts, including bacteria belonging to biosynthetically rich phyla like *Streptomyces* and beta-proteobacteria [56].

Host improvement and new detection methods

It is thought that potential transcriptional, translational and biochemical blocks can hinder the expression of an exogenous gene cluster in an individual host. Using multiple heterologous hosts for functional screening maximizes the chance of identifying bioactive molecules by matching eDNA-derived clusters with native host biochemistries. It is also possible to imagine adapting hosts to more permissively express eDNA-derived biosynthetic gene clusters. *Streptomyces* appear to be one of the most prolific secondary metabolite producers and, as such, a number of strain improvement efforts have focused on this genus [56]. Ribosome engineering and use of mutant RNA polymerases have been employed in *Streptomyces* spp., as well as myxobacteria and fungi, in an effort to facilitate activation of cryptic pathways and improve generic gene expression [57]. Other efforts to activate silent gene clusters have focused on deleting global negative regulators of biosynthesis, such as DasR, or have used overexpression of positive regulators such as LAL [58,59]. Similarly, overexpression of the alternative sigma factor σ^{54} has been used to facilitate expression of a Type II polyketide synthase (PKS) gene cluster in *Escherichia coli* [60]. In general, these host-manipulations result in more promiscuous transcription and may therefore allow for a higher percentage of cloned eDNA gene clusters to be expressed in high throughput functional screens.

The coupling of more sensitive natural product detection tools with libraries hosted in improved strains should lead to a greater number of eDNA-encoded compounds being identified in functional screens. In recent years, mass-spectrometry has emerged as a powerful and sensitive tool for natural product screening. Improved selective detection techniques, such as those for phosphonic acid and phosphonate containing compounds, are making it possible to detect previously undetectable small molecules [61]. Similarly, a suite of tools termed peptido-genomics and glyco-genomics, are able identify NRPS-derived peptides and O-/N-glycosyl containing sugar monomers directly from bacterial colonies on agar plates. Mass-spec data is used in combination with bioinformatic predictions of biosynthetic systems, providing a powerful cross-referencing system that can be used to predict the structure of potential metabolites [62,63]. So far, these techniques have been applied only to cultured bacteria; however, it is easy to imagine how such approaches could be applied to direct functional screening of metagenomic libraries.

Library enrichment

Analysis of sequenced genomes reveals that less than 2% of the genome is devoted to secondary metabolism [64]; consequently, metagenomic DNA libraries are sparsely populated with biosynthetic genes of interest. Gene cluster enrichment strategies can be used to simultaneously reduce the size and increase the biosynthetic density of eDNA libraries, thereby increasing the efficiency of functional metagenomic applications

[46,65–68]. Complementation of 4′-phosphopantetheinyl transferase (PPTase) activity has been recognized as a valuable tool for gene cluster mining applications. PPTases are required to generate holo-non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymes and have been used in a phage display strategy designed to recover NRPS/PKS sequences [69]. More recently, PPTase gene complementation has been successfully used for enriching *E. coli* and *Pseudomonas aeruginosa* hosted libraries for NRPS/PKS containing clones by linking it to production of either a colored indicator molecule or a siderophore for selective growth on low iron media [70,71•]. Using low-iron selection strategy in *E. coli*, 50-fold enrichment in NRPS/PKS gene content was achieved in just two rounds of selection [71•].

Future directions in functional metagenomics

While sequence-based metagenomic screening approaches are now quite robust and capable of targeting the discovery of diverse novel molecules from many different environments, functional metagenomic screening methods have not yet advanced to the extent that enables this approach to rapidly harvest molecules from the environment. We believe that three key advances are necessary to bring functional metagenomics to maturity. First, new model heterologous hosts must be identified and engineered that are able to promiscuously activate a more diverse set of biosynthetic gene clusters. Second, we need improved DNA cloning methods that enable capture of complete gene clusters on individual eDNA clones. Finally, new methods are needed for selectively enriching eDNA libraries for clones containing a variety of secondary metabolite genes. Together, these advances will facilitate a substantial increase in the frequency and diversity of small molecules with novel bioactivities that can be harvested from the environment using functional metagenomic approaches.

Conclusions

By taking a gene-based approach, metagenomics can exploit the sequencing revolution and bypass many of the traditional hurdles to drug discovery. While cultured organisms have yielded many of our most important antimicrobial agents, these organisms represent only a small fraction of total microbial diversity. Metagenomic methods provide a means to evaluate the biosynthetic potential of the bacterial majority, thereby providing an opportunity to find truly novel antimicrobials. While there remain bottlenecks in the metagenomic drug-discovery platform, such as the heterologous expression of metagenomic pathways, these problems are not unique to metagenomics and are also being tackled by the broader microbiology and synthetic biology communities. As the development of metagenomics-specific tools progresses and the most promising, high-throughput, genome based approaches are adopted by the field, metagenomics should play an increasingly important role in the future of antibiotic drug discovery.

Acknowledgements

This work was supported by National Institutes of Health Grant GM077516. S.F.B. 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 special interest
- •• of outstanding interest

1. Cragg GM, Newman DJ: **Natural products: a continuing source of novel drug leads**. *Biochim Biophys Acta* 2013, **1830**:3670-3695.

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

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

4. Torsvik V, Daae FL, Sandaa RA, Ovreas L: **Novel techniques for analysing microbial diversity in natural and perturbed environments**. *J Biotechnol* 1998, **64**:53-62.

5. NCBI: *New NCBI Handbook Chapters: Eukaryotic and Prokaryotic Genome Annotation Pipelines*. NCBI News; 2013.

6. Bentley SD *et al.*: **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)**. *Nature* 2002, **417**:141-147.

7. • Medema MH *et al.*: **AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences**. *Nucleic Acids Res* 2011, **39**:W339-W346.
AntiSMASH is becoming the *de facto* standard for identifying and characterizing biosynthetic gene clusters. In addition to identifying gene clusters, the recent version has a set of new features that also assist in classifying novelty based on their relatedness to know clusters.

8. Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH: **Automated genome mining for natural products**. *BMC Bioinform* 2009, **10**:185.

9. • Kwan JC *et al.*: **Genome streamlining and chemical defense in a coral reef symbiosis**. *Proc Natl Acad Sci U S A* 2012, **109**:20655-20660.
This study uses a genomics approach to identify an endosymbiotic bacterial producer of a potent polyketide, providing evidence for marine chemical-sybioses, a pattern than may come to be recognized as common.

10. Donia MS, Ruffner DE, Cao S, Schmidt EW: **Accessing the hidden majority of marine natural products through metagenomics**. *ChemBioChem: Eur J Chem Biol* 2011, **12**:1230-1236.

11. Schmidt EW, Donia MS, McIntosh JA, Fricke WF, Ravel J: **Origin and variation of tunicate secondary metabolites**. *J Nat Prod* 2012, **75**:295-304.

12. • Kampa A *et al.*: **Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses**. *Proc Natl Acad Sci U S A* 2013, **110**:E3129-E3137.
A parallel to the work of marine endosymbionts, this paper provides an excellent example of how genome sequencing can augment traditional natural product characterization, in this case to describe a novel polyketide produced by members of a lichen microbiome.

13. Wilson MC *et al.*: **An environmental bacterial taxon with a large and distinct metabolic repertoire**. *Nature* 2014, **506(7486)**:58-62.

14. Hosono S *et al.*: **Unbiased whole-genome amplification directly from clinical samples**. *Genome Res* 2003, **13**:954-964.

15. Grindberg RV *et al.*: **Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage**. *PLoS ONE* 2011, **6**:e18565.

16. Udwary DW *et al.*: **Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica***. *Proc Natl Acad Sci U S A* 2007, **104**:10376-10381.

17. Fischbach MA, Walsh CT, Clardy J: **The evolution of gene collectives: how natural selection drives chemical innovation**. *Proc Natl Acad Sci U S A* 2008, **105**:4601-4608.

18. Wilson MC, Gulder TA, Mahmud T, Moore BS: **Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the sare1259-encoded cytochrome P450**. *J Am Chem Soc* 2010, **132**:12757-12765.

19. • Owen JG *et al.*: **Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products**. *Proc Natl Acad Sci U S A* 2013.

This paper demonstrates the utility of the sequence tag approach to identify and recover biosynthetic clones from the environment. By arraying eDNA as a library and using sequence tag interrogation, small amplicons were able to guide the recovery of full pathways from a number of biosynthetic systems.

20. Ziemert N *et al.*: **The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity**. *PLoS ONE* 2012, **7**:e34064.

21. Prieto-Davo A *et al.*: **Targeted search for actinomycetes from nearshore and deep-sea marine sediments**. *FEMS Microbiol Ecol* 2013, **84**:510-518.

22. Gontang EA, Gaudencio SP, Fenical W, Jensen PR: **Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria**. *Appl Environ Microbiol* 2010, **76**:2487-2499.

23. • Woodhouse JN, Fan L, Brown MV, Thomas T, Neilan BA: **Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges**. *ISME J* 2013, **7**:1842-1851.

A fine example of the application of the sequence tag approach to the characterization of biosynthetic diversity in marine environments.

24. Feng Z, Chakraborty D, Dewell SB, Reddy BV, Brady SF: **Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis**. *J Am Chem Soc* 2012, **134**:2981-2987.

25. Reddy BV *et al.*: **Natural product biosynthetic gene diversity in geographically distinct soil microbiomes**. *Appl Environ Microbiol* 2012, **78**:3744-3752.

26. Brady SF: **Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules**. *Nat Protoc* 2007, **2**:1297-1305.

27. Kang HS, Brady SF: **Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes**. *Angew Chem* 2013, **52**:11063-11067.

28. Chang FY, Ternei MA, Calle PY, Brady SF: **Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment**. *J Am Chem Soc* 2013, **135**:17906-17912.

29. Chang FY, Brady SF: **Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening**. *Proc Natl Acad Sci U S A* 2013, **110**:2478-2483.

30. Chang FY, Brady SF: **Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017**. *J Am Chem Soc* 2011, **133**:9996-9999.

31. Fujita MJ *et al.*: **Cloning and heterologous expression of the vibrioferrin biosynthetic gene cluster from a marine metagenomic library**. *Biosci Biotechnol Biochem* 2011, **75**:2283-2287.

32. 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.

33. Zimmermann K, Engeser M, Blunt JW, Munro MH, 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.

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

35. Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E: **Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides**. *Appl Environ Microbiol* 2010, **76**:3568-3574.

36. Gatte-Picchi D, Weiz A, Ishida K, Hertweck C, Dittmann E: **Functional analysis of environmental DNA-derived microviridins provides new insights into the diversity of the tricyclic peptide family**. *Appl Environ Microbiol* 2014, **80**:1380-1387.

37. Mavromatis K *et al.*: **The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation**. *PLoS ONE* 2012, **7**:e48837.

38. Koren S *et al.*: **Reducing assembly complexity of microbial genomes with single-molecule sequencing**. *Genome Biol* 2013, **14**:R101.

39. • Fitzsimons MS *et al.*: **Nearly finished genomes produced using gel microdroplet culturing reveal substantial intraspecies genomic diversity within the human microbiome**. *Genome Res* 2013, **23**:878-888.

By combining micro-droplet cultivation with whole-genome amplification this paper demonstrates the feasibility of sequencing the genomes of low-abundance or rare bacterial species.

40. Lasken RS: **Genomic sequencing of uncultured microorganisms from single cells**. *Nat Rev Microbiol* 2012, **10**:631-640.

41. Lim HK *et al.*: **Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli***. *Appl Environ Microb* 2005, **71**:7768-7777.

42. MacNeil Ia *et al.*: **Expression and isolation of antimicrobial small molecules from soil DNA libraries**. *J Mol Microbiol Biotechnol* 2001, **3**:301-308.

43. Lee MH, Lee S-W: **Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities**. *Genomics Inform* 2013, **11**:114-120.

44. Schipper C *et al.*: **Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa***. *Appl Environ Microbiol* 2009, **75**:224-233.

45. Kellner H, Luis P, Portetelle D, Vandenbol M: **Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants**. *Microbiol Res* 2011, **166**:360-368.

46. 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.

47. Williamson LL, Borlee BR: **Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor**. *Appl Environ Microbiol* 2005, **71**:6335-6344.

48. Wang GY *et al.*: **Novel natural products from soil DNA libraries in a streptomycete host**. *Org Lett* 2000, **2**:2401-2404.

49. • Engel K, Pinnell L, Cheng J, Charles TC, Neufeld JD: **Nonlinear electrophoresis for purification of soil DNA for metagenomics**. *J Microbiol Methods* 2012, **88**:35-40.

The SCODA technology provides a powerful tool for purifying high molecular weight DNA for the creation of metagenomic cosmid libraries. This paper is a head-to-head comparison of SCODA with alternative isolation procedures.

50. Pel J *et al.*: **Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules**. *Proc Natl Acad Sci U S A* 2009, **106**:14796-14801.

51. Gregory MA, Till R, Smith MCM: **Integration site for streptomyces phage ΦBT1 and development of site-specific integrating vectors**. *J Bacteriol* 2003, **185**:5320-5323.

52. Kuhstoss S, Rao RN: **Analysis of the integration function of the streptomycete bacteriophage phi C31**. *J Mol Biol* 1991, **222**:897-908.

53. Kakirde KS, Wild J, Godiska R, Mead DA: **Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries**. *Gene* 2011, **475**:57-62.

54. Craig JW, Chang FY, Brady SF: **Natural products from environmental DNA hosted in *Ralstonia metallidurans***. *ACS Chem Biol* 2009, **4**:23-28.

55. Aakvik T *et al.*: **A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species**. *FEMS Microbiol Lett* 2009, **296**:149-158.

56. Donadio S, Monciardini P, Sosio M: **Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics**. *Nat Prod Rep* 2007, **24**:1073-1109.

57. Ochi K, Hosaka T: **New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters**. *Appl Microbiol Biotechnol* 2013, **97**:87-98.

58. Laureti L *et al.*: **Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens***. *Proc Natl Acad Sci U S A* 2011, **108**:6258-6263.

59. van Wezel GP, McDowall KJ: **The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances**. *Nat Prod Rep* 2011, **28**:1311-1333.

60. Stevens DC *et al.*: **Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli***. *PLoS ONE* 2013, **8**:e64858.

61. Evans B, Zhao C, Gao J, Evans C: **Discovery of the antibiotic phosacetamycin via a new mass spectrometry-based method for phosphonic acid detection**. *ACS Chem Biol* 2013, **8**:908-913.

62. Kersten RD *et al.*: **A mass spectrometry-guided genome mining approach for natural product peptidogenomics**. *Nat Chem Biol* 2011, **7**:794-802.

63. Kersten RD *et al.*: **Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules**. *Proc Natl Acad Sci U S A* 2013, **110**:E4407-E4416.

64. Garcia JAL, Fernández-Guerra A, Casamayor EO: **A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes**. *Mol Phylogenet Evol* 2011, **61**:650-658.

65. Kalyuzhnaya MG *et al.*: **Fluorescence in situ hybridization-flow cytometry-cell sorting-based method for separation and enrichment of type I and type II methanotroph populations**. *Appl Environ Microbiol* 2006, **72**:4293-4301.

66. Meyer QC, Burton SG, Cowan DA: **Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome**. *Biotechnol J* 2007, **2**:36-40.

67. Morimoto S, Fujii T: **A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking**. *Appl Microbiol Biotechnol* 2009, **83**:389-396.

68. Zhang K, He J, Yang M, Yen M, Yin J: **Identifying natural product biosynthetic genes from a soil metagenome by using T7 phage selection**. *ChemBioChem: Eur J Chem Biol* 2009, **10**:2599-2606.

69. Yin J *et al.*: **Genome-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display**. *Chem Biol* 2007, **14**:303-312.

70. Owen JG, Robins KJ, Parachin NS, Ackerley DF: **A functional screen for recovery of 4′-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries**. *Environ Microbiol* 2012, **14**:1198-1209.

71. Charlop-Powers Z, Banik JJ, Owen JG, Craig JW, Brady SF:
• **Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis**. *ACS Chem Biol* 2013, **8**:138-143.

Metagenomic libraries can be quite large while only a fraction of clones may contain biosynthetic systems of interest. This study demonstrates that functional complementation of biosynthetic proteins can be engineered to as a tool to enrich eDNA libraries for biosynthetic gene clusters.