Metagenomic approaches to natural products from free-living and symbiotic organisms†

Sean F. Brady,*[a] Luke Simmons,[a] Jeffrey H. Kim[a] and Eric W. Schmidt*[b]

Covering: up to the end of January 2009

Bacterial cultivation has been a mainstay of natural products discovery for the past 80 years. However, the majority of bacteria are recalcitrant to culture, providing an untapped source for new natural products. Metagenomic analysis provides an alternative method to directly access the uncultivated genome for natural products research and for the discovery of novel, bioactive substances. Applications of metagenomics to diverse habitats, such as soils and the interior of animals, are described.

1   Introduction

Extensive surveys of microbial 16S rRNA gene sequences derived directly from environmental samples suggest that only a small fraction of the bacteria present in nature can be readily cultured in the laboratory.[1–8] In fact, less than 1% of the bacteria present in most environmental samples are thought to be readily cultured using current fermentation technologies. More than 80 major bacterial divisions have now been identified in culture-independent analyses of environmental samples, and fewer than half of these contain cultured isolates.[9–11] This diverse collection of as-yet uncultured environmental bacteria promises to be a rich, although challenging, source of novel natural products.

Like their free-living relatives, symbiotic bacteria living in close association with other organisms are largely uncultivated and have the genetic potential to synthesize diverse, novel natural products. In addition, symbiotic bacteria are known or suspected to synthesize numerous natural products previously isolated from whole animals and other organisms. A series of recent reviews in this journal and elsewhere documents the importance of "symbiotic" natural products, some of which are clinically used drugs or drug leads.[12–14]

In an attempt to circumvent the challenges associated with culturing environmental bacteria, whether free-living or symbiotic, culture-independent small-molecule discovery strategies have been explored by a number of groups. In these studies, DNA is extracted directly from environmental samples and cloned into model cultured bacterial hosts (Fig. 1). Environmental DNA (eDNA) or metagenomic DNA are collective terms used to describe DNA extracted directly from a given environment (*e.g.* soil, sediment, water, aquatic, plant, *etc.*) independent of host cultivation or fermentation. Clone libraries derived from eDNA are of potential interest to natural products chemists because all of the genes required for the biosynthesis of a natural product, including genes that code for biosynthetic, regulatory and resistance enzymes are typically found clustered on bacterial chromosomes. Large insert eDNA clones therefore have the potential to contain functionally intact biosynthetic gene clusters that might confer the production of metabolites to an easily cultured bacterial host.

This review covers culture-independent studies that have focused on the discovery of natural products from either free-living bacterial communities or from bacteria closely associated with other organisms (symbionts).

2   Metagenomics of free-living bacterial communities

Two general strategies, phenotypic and homology (DNA) based screening, have been employed to discover new natural products

[a]Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, NY, 10021, USA. E-mail: sbrady@rockefeller.edu
[b]Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT, 84112, USA. E-mail: ews1@utah.edu
† This article is part of a themed issue on genomics.
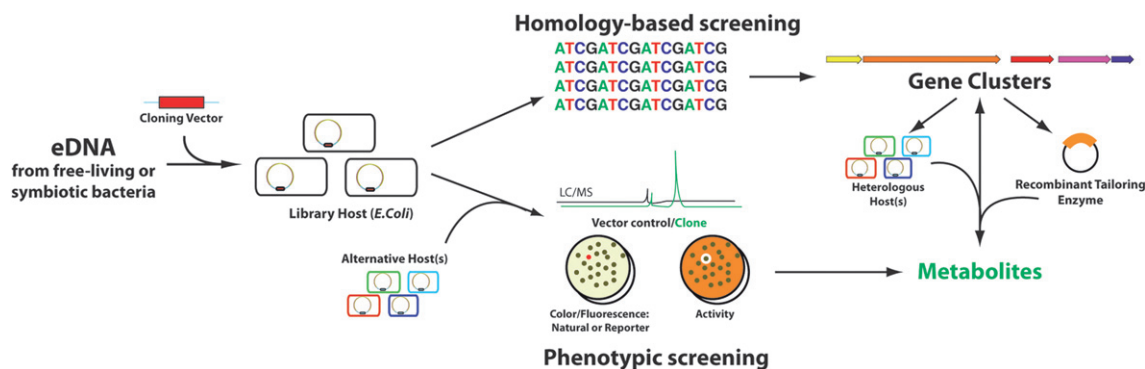
Fig. 1 Overview of culture independent strategies used for the discovery of natural products and natural product gene clusters from eDNA libraries.

from large eDNA libraries. With phenotypic screening, eDNA libraries are initially examined in simple screens designed to identify clones that exhibit properties traditionally associated with the production of small molecules (*i.e.* color production, antibacterial activity or signaling phenomena). These clones are then examined in detail for the production of clone-specific small

Sean F. Brady

Sean F. Brady graduated with a degree in molecular biology in 1993 from Pomona College in Claremont, California. He received his Ph.D. in organic chemistry from Cornell University in 2001, where he studied under Jon Clardy. In 2002, he moved to Harvard Medical School as a fellow in the Institute of Chemistry and Cell Biology. He was named an instructor in the department of biological chemistry and molecular pharmacology at Harvard Medical School in 2004. In 2006 he moved to The Rockefeller University as an assistant professor.

Jeffrey H. Kim

Jeffrey H. Kim graduated with a degree in molecular biology in 2003 from Princeton University. In 2003, he moved to The Rockefeller University where he studied under Tarun Kapoor. In 2007, he began his doctorate studies at The Rockefeller University with Sean Brady, where he is focusing on the discovery of novel genetically encoded small molecules from uncultured bacteria.

Luke Simmons

Luke was born and raised in Frederick, Maryland. Following undergraduate studies in biology and chemistry at the University of Oregon, Luke joined the Natural Products Division of the National Cancer Institute. In 2007 he received a Ph.D. from the Scripps Institution of Oceanography, UCSD, where he studied the bioorganic and natural products chemistry of marine cyanobacteria with Professor William H. Gerwick. In 2008 Luke joined the laboratory of Sean F. Brady at the Rockefeller University to investigate enediyne biosynthetic gene clusters from environmental DNA. Luke is currently a Humboldt Postdoctoral Fellow in the laboratory of Professor Rolf Müller at the Helmholtz Institute for Pharmaceutical Research in Saarland, Germany.

Eric W. Schmidt

Eric W. Schmidt received a B.S. in Chemistry from the University of California, San Diego. He earned his Ph.D. studying marine natural products chemistry with the late D. John Faulkner at Scripps Institution of Oceanography. As a postdoctoral fellow, he studied the biosynthesis of aflatoxin with Craig Townsend at The Johns Hopkins University. In 2001, he moved to the Department of Medicinal Chemistry, University of Utah, where he is currently Associate Professor. Research interests include symbiosis, biosynthesis, and the chemistry and biology of natural products.

molecules. In homology-based screening, degenerate primers based on conserved natural product biosynthetic genes of interest are used to PCR amplify probe sequences from an eDNA library. The eDNA-specific probes are then used to recover individual clones from the library and the biosynthetic enzymes found on these clones are used to generate novel small molecules. Results from both phenotype and homology-based screening of eDNA libraries are discussed here. This section of the review covers all reports of small molecules discovered from the heterologous expression of environmental DNA up to the end of January 2009.

3 Phenotypic screening of metagenomic libraries

3.1 The terragines, molecules from the functional expression of eDNA clones in *Streptomyces lividans*

In 2000 a collaboration between the Anderson and Davies groups isolated terragines A–E (**1**–**5**) from soil eDNA clones hosted in *Streptomyces lividans*.[15] Prior reports of metagenomic library construction and phenotypic screening had appeared in the literature but none of these studies identified small-molecule-producing clones.[16,17] Using DNA isolation techniques developed earlier by Davies and co-workers, environmental DNA was recovered directly from soil samples collected in British Columbia, Canada.[18] This DNA was then used to construct a small collection of cosmid clones in *Escherichia coli* using a shuttle vector that would enable the facile intergenic transfer of environmental clones from the library host (*E. coli*) into *S. lividans*. Culture broth extracts from each of the resulting *S. lividans* recombinants were screened by HPLC-ESIMS, and two clones were found to produce clone specific metabolites. The characterization of the metabolites produced by these clones resulted in the identification of terragines A–E (**1**–**5**) (Fig. 2) as well as the re-isolation of the known microbial siderophore norcardamine (**6**).[19]

Clones that produce terragine-like metabolites were subsequently found at a very high frequency in this library. Eighteen of the approximately 1000 *S. lividans* recombinants screened were found to produce members of the terragine/norcardamine families. None of these clones were sequenced and therefore the biosynthetic origin of the terragines is not known. Although terragines did not display any notable biological activity, this work represents the first example of small molecules being characterized from eDNA clones.

3.2 Molecules from the functional expression of eDNA clones in *Escherichia coli*

Metagenomic library construction methods have been extensively optimized for *E. coli*, and the creation of genetically diverse cosmid- and fosmid-based libraries is now easily attainable.[16,20,21] Despite the potential limitations of using *E. coli* as an expression host, several studies have shown that novel metabolites can be identified using this host.

3.2.1 *N*-Acyl amino acids.
The first novel bioactive small molecules derived from the heterologous expression of eDNA were a series of long-chain *N*-acyl amino acid antibiotics reported in 2000.[22] In this study, an agar overlay assay against *Bacillus subtilis* was used to screen a cosmid library for antibacterially active clones. The antibacterial activity exhibited by one of the clones found in this screen was traced to the production of a family of *N*-acyltyrosine derivatives with saturated and monounsaturated fatty acid side chains 8–18 carbons in length (**7**). Sequencing and random transposon mutagenesis showed that a single open-reading frame (ORF) was necessary for the production of long-chain *N*-acyltyrosines (CSL12.1; Genbank No. AF324335), and subsequent subcloning of this ORF confirmed it was also sufficient for the production of these metabolites in *E. coli*. The closest relative identified in a BLAST search was a hypothetical *N*-acyl transferase (MJ1207) from the cultured bacteria *Methanococcus janaschii*.

Subsequent publications have reported eDNA clones that produce a variety of long-chain *N*-acylated metabolites with antibacterial activity (Fig. 3). These include clones that produce long *N*-acyltyrosines (**7**), *N*-acylphenylalanines (**8**), *N*-acyltryptophans (**9**), *N*-acylarginines (**10**) and *N*-acylputrescines (**11**).[23–25] Transposon mutagenesis and sequence analysis indicated that in each case a single open-reading frame was responsible for the biosynthesis of each family of metabolites.[23–25] In many cases the *N*-acyl amino acid synthases (NAS) characterized from these clones show only limited sequence identity to each other.

A detailed analysis of culture broth extracts from one *N*-acyltyrosine-producing clone indicated that it produced two additional families of clone specific metabolites, long-chain enamides (**12**) and long-chain enol esters (**13**).[26] Sequence analysis of the regions surrounding the *N*-acyl synthase (NAS) from this clone revealed a biosynthetic gene cluster containing 13 ORFs, designated *fee*A–M (fatty acid enol ester; GenBank No. AY128669). The characterization of intermediates that accumulate in key transposon mutants along with a detailed
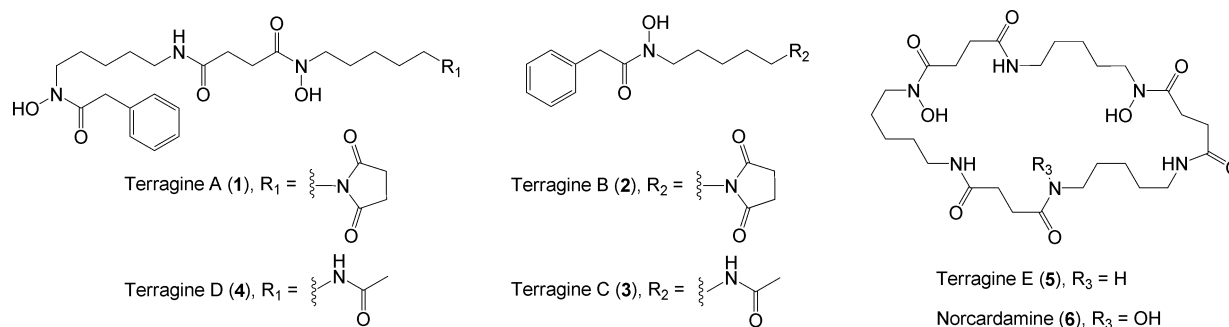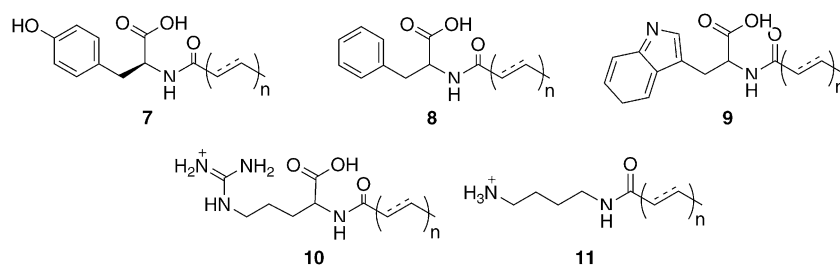

Fig. 2 The terragines (**1**–**5**) and norcardamine (**6**) were characterized from *S. lividans* transformed with soil eDNA cosmids.

This journal is © The Royal Society of Chemistry 2009

Fig. 3 Families of *N*-acyl amino acid antibiotics characterized from eDNA clones. In each case a collection of metabolites with both saturated and monounsaturated side chains of various lengths are produced by the antibacterially active eDNA clone.
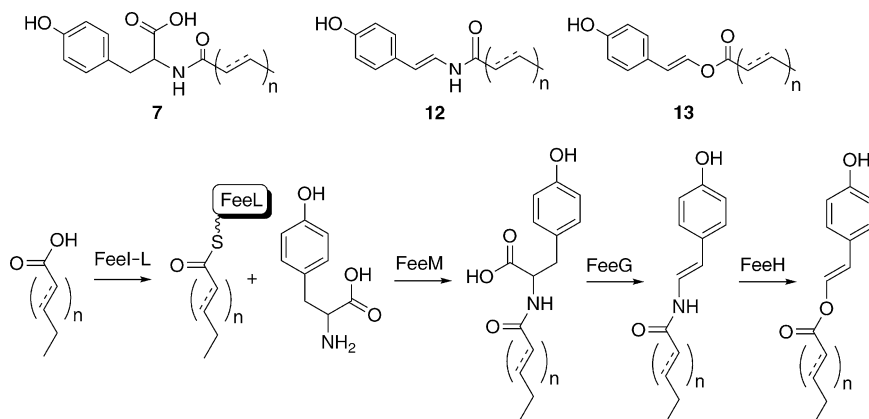
Fig. 4 Long chain *N*-acyltyrosine antibiotics (**7**) have been characterized from a number of different antibacterially active eDNA clones. One of these clones was found to produce additional families of metabolites (**8** and **9**). In this clone the *fee* gene cluster is responsible for the biosynthesis of all three families of compounds.

BLAST analysis revealed the presence of a fatty acid desaturase (*fee*F), an oxidative decarboxylase (*fee*G), an *N,O*-transferase (*fee*H), an acyl carrier protein (*fee*L), a predicted fatty acid loading system (*fee*I–K) and the NAS (*fee*M).[26] From these studies the authors proposed the biosynthetic scheme shown in Fig. 4.

3.2.2 Violacein and deoxyviolacein. The appearance of color in a bacterial colony is often an indication of small-molecule biosynthesis. Simple visual inspection can thus be used as an initial screen to identify clones that might be producing clone-specific natural products. In 2001 the visual screening of a cosmid library constructed from soil eDNA inserts yielded several blue-colored *E. coli* clones. Extensive HRMS and NMR analysis of the colored metabolites produced by one of these clones found that the color resulted from the production of violacein (**14**) and deoxyviolacein (**15**) (Fig. 5). Violacein and deoxyviolacein are tryptophan dimers originally isolated from
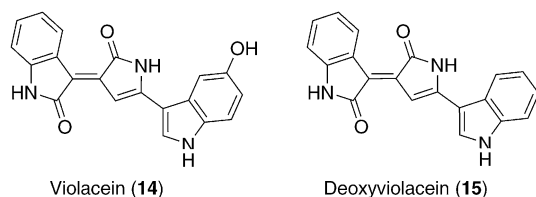

Fig. 5 The previously reported metabolites violacein (**14**) and deoxy-violacein (**15**) are produced by a purple eDNA cosmid clone.

the cultured bacterium *Chromobacterium violaceum*. Sequencing of transposon insertions that knock out color production, increase color production, or result in change in color, identified a cluster of four genes (*vio*A–D, GenBank No. AF367409) that resembled the violacein gene cluster sequenced from *C. violaceum*.[27] While the gene organization of the violacein cluster from *C. violaceum* and eDNA is identical, the individual violacein biosynthetic enzymes (*i.e.* VioA–D) only show 48, 62, 71 and 69% amino acid identity, respectively.[27,28]

3.2.3 Turbomycins A and B. In a 2002 study, the visual screening of eDNA libraries prepared in *E. coli* yielded three pigmented clones. All three clones produced a brown pigment suggestive of melanin.[29] One of the clones was examined in detail for the potential production of clone-specific small molecules. While no clone-specific metabolites were identified in this study, acid precipitate from the melanin-producing clones was shown to contain an increased concentration of two colored metabolites. High-resolution mass spectrometric (HRMS) and nuclear magnetic resonance (NMR) spectroscopic analysis of the colored material isolated from this acid precipitate indicated that the colored metabolites were a previously reported triaryl cation given the name turbomycin A (**16**) and a new analog given the name turbomycin B (**17**) (Fig. 6). Turbomycin A (**16**) was originally described as a fungal metabolite from *Saccharomyces cerevisiae* in 1972.[30]

Random transposon mutagenesis and subcloning showed that a single ORF (58% G + C) predicted to encode a 40 kDa protein
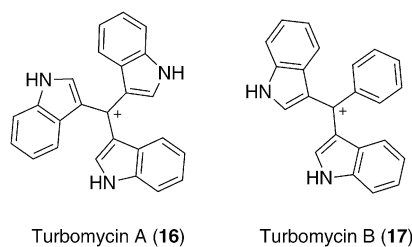
Fig. 6 Turbomycin A and B are tri-aryl cations that accumulate in the acid precipitate of melanin-producing eDNA clones at a higher level than is observed in acid precipitates from vector control cultures.

was necessary and sufficient for the increased production of the turbomycins. BLAST analysis indicated the putative protein was homologous to several known enzymes; 53% identity to legiolysin (*L. pneumophila*); 54% identity to hemolysin from *Vibrio vulnificus*; and 49% identity to 4-hydroxyphenylpyruvate dioxygenase (4HPPD) from *Pseudomonas* sp. (P80064).[29] 4HPPD-related enzymes are known to catalyze the production of homogentisic acid (HGA), which under aerobic conditions undergoes spontaneous polymerization into HGA-melanin.[31] The majority of the HGA undergoes spontaneous polymerization to melanin, which by an unknown mechanism appears to enhance the formation of the turbomycins under acidic conditions.[29] The biosynthetic origin of **16** and **17** in this system is not clear.

3.2.4 Isocyanide metabolites. Antibacterial screening of an eDNA library constructed from soil collected in Massachusetts identified a clone that produces compound **18**, a new isocyanide-functionalized indole-based antibiotic.[32] Transposon mutagenesis of the cosmid isolated from this clone suggested that two eDNA-derived open-reading frames, *isnA* and *isnB* (GenBank No. DQ084328), were required for the production of this new metabolite in *E. coli* (Fig. 7).[32] While the isocyanide functional group had been seen previously in bacterial secondary metabolites, its biosynthesis and the source of its component atoms was not known.[33,34] The cloning and heterologous expression of a predicted eDNA-derived isonitrile biosynthetic enzyme in *E. coli* made it possible for well-controlled feeding experiments to reveal the origin of these atoms in **18**.[32,35] Extensive "inverse labeled" feeding studies, using *E. coli* strains with mutations in key primary metabolic steps, were designed to systemically interrogate the *E. coli* metabolome for the source of the isocyanide carbon. These studies revealed that the isocyanide carbon was derived from the C2 carbon of five-carbon sugars from the pentose phosphate pathway. Additional feeding studies suggested that tryptophan is the source of the nitrogen atom

in **18**. *In vitro* reconstitution experiments using purified IsnA, IsnB, tryptophan and ribulose-5-phosphate confirmed the hypothetical origins of these isocyanide atoms (Fig. 7).[32,35] In a subsequent DNA-based screening study, additional gene clusters harboring *isnA* from both cultured and uncultured bacteria were identified. This study led to the identification of 12 predicted *isnA* homologs from eDNA and sequenced bacterial genomes. Expression of these isonitrile synthase genes in various hosts ultimately yielded nine clone-specific small molecules.[36]

3.2.5 Indigo- and indirubin-producing clones. The most common classes of clones found when screening eDNA libraries for natural products have been antibacterially active clones that produce long chain *N*-acyl metabolites (discussed in Section 3.2.1), red-pigmented antibacterially active clones that express aminolevulinic acid synthetases (*hemA*), brown melanin-producing clones (discussed in Section 3.2.3), and blue-pigmented clones that produce mixtures of indigo (**19**) and indirubin (**20**) (Fig. 8).[23–25,29,37–39] Three reports of metagenomic clones that produce indigo and indirubin have appeared in the literature.[37–39] A BAC clone derived from New England soil eDNA (GenBank No. DQ000460) was found to contain two predicted indole dioxygenases, each sufficient to produce the observed pigments.[37] A fosmid clone constructed using DNA extracted from Korean soil (GenBank No. EF569599) was found to contain a single predicted monooxygenase that is responsible for the observed pigments.[38] A plasmid clone harboring eDNA from the gypsy moth mid-gut (GenBank No. AR053980) contained a two-component flavin-dependent monooxygenase system (MoxZ/Y).[39] While MoxY alone is sufficient for color production, the color production is enhanced by MoxZ, a predicted NADH:flavin oxidoreductase.

The pigment clones from the New England and Korean soil libraries were identified in simple color screens while the clone from the gypsy moth library was found in an intracellular assay designed to identify small-molecule quorum-sensing mimics produced by clones in eDNA libraries (metabolite regulated
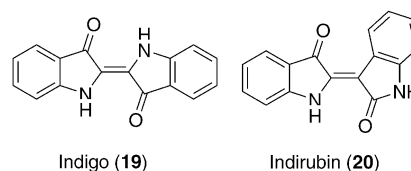
Fig. 8 Three different metagenomic studies have identified eDNA clones that produce indigo and indirubin.
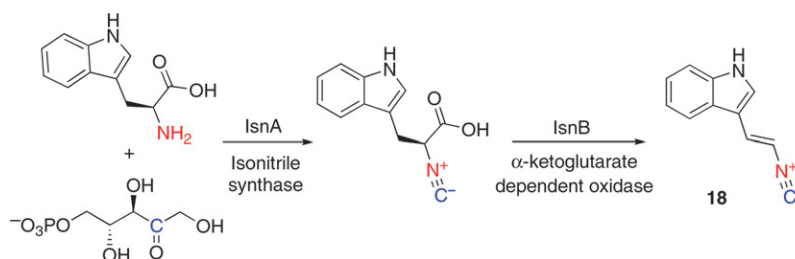

Fig. 7 The isonitrile-functionalized indole derivative **18** is produced in *E. coli* from tryptophan and a five-carbon sugar using the eDNA-derived enzymes IsnA and IsnB.

expression, METREX).[38] In the METREX assay, activated LuxR (LuxR bound to an *N*-acylhomoserine lactone or an *N*-acylhomoserine lactone mimic) induces the expression of green fluorescent protein (GFP). Two METREX active eDNA clones that produce small molecules have been reported in the literature. One produces an *N*-acylhomoserine lactone that is reported to closely resemble *N*-(3-oxohexanoyl)-L-homoserine lactone and the other is the blue gypsy moth mid-gut clone described above. This blue clone produces indigo, indirubin and a METREX- active metabolite that was not structurally characterized.[39,40] In the METREX assay, heterologously produced metabolites do not have to exit the cell to be detected. Intracellular assays like METREX are therefore potentially more sensitive and could help increase the frequency at which small-molecule-producing clones are found in large eDNA libraries.

3.3 Molecules from the functional expression of eDNA clones in *Ralstonia metallidurans*

Historically, the heterologous expression of eDNA has been carried out in two model systems, *E. coli* and *Streptomyces*. The development of a more phylogenetically diverse collection of potential library hosts will likely be needed to fully access the genetic diversity available in any given metagenomic sample. Recent work with the β-proteobacterium *Ralstonia metallidurans* indicates that, as predicted, expanding the host repertoire is likely to increase the number of small-molecule-producing clones identified from eDNA library screening.[41] In this study,
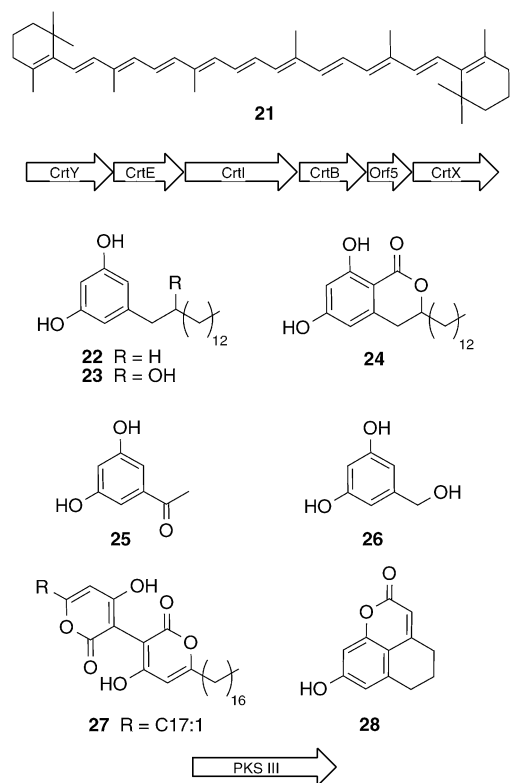
Fig. 9 Both known (**21–26**) and novel (**27** and **28**) small molecules were characterized from clones found in phenotypic screens of *R. metallidurans*-based libraries.

phenotypic screening of an eDNA cosmid library hosted in *R. metallidurans* yielded two clones, one yellow-pigmented clone (RM3), and one antibacterially active clone (RM57) that produced clone-specific small molecules (Fig. 9). While the library used in this study was screened in *R. metallidurans*, it was constructed in a new broad-host-range RK2-based cosmid (pJWC1) that should permit eDNA libraries to be screened in a phylogenetically diverse collection of bacterial hosts.[42]

Cosmids from RM3 and RM57 were isolated and sequenced. The sequencing of RM3 (GenBank No. FJ151553) revealed a six-gene operon with homology to carotenoid gene clusters, and the yellow metabolite purified from RM3 cultures was found to be spectroscopically identical to β-carotene (**21**). Analysis of the RM57 sequence (GenBank No. FJ151552) did not suggest an obvious biosynthetic source of the antibacterial activity produced by this clone. Transposon mutagenesis of the cosmid from RM57 indicated that the antibiotic activity was dependent on a putative type III polyketide synthase. The closest relatives of this type III polyketide synthase are from Acidobacteria, a largely uncultured group of bacteria. Seven small molecules were isolated from cultures of clone RM57: **22–28** were elucidated as long- and short-chain substituted resorcinol-like derivatives, while **27** was deduced as a pyrone heterodimer with both saturated and monounsaturated side chains. Compound **28** was determined to be a novel tricyclic isocoumarin-based carbon skeleton. The novel compounds **27** and **28** were reported to exhibit antibacterial activity against *Staphylococcus aureus* and *Bacillus subtilis*.[41]

4 DNA-based screening of metagenomic libraries

4.1 Type II ketosynthase genes from eDNA: PCR-based study

The first report of attempting to access small molecules from the genetic information present in the genomes of uncultured bacteria did not rely on the recovery of intact gene clusters to access small molecules but instead used a PCR-based strategy to recover novel type II polyketide (PKS) ketoacyl synthase sequences.[43] Bacterial type II PKS biosynthetic gene clusters encode a minimal-PKS that consists of three enzymes: two β-ketoacyl synthases, KS_α, which catalyzes the Claisen-type condensation between the growing polyketide and incoming acyl-CoA subunits, and KS_β, which controls polyketide chain length (*i.e.* chain length factor), and an acyl carrier protein (ACP), which anchors the growing polyketide chain during chain elongation. In type II PKS systems, the minimal PKS gene cassette is sufficient to generate the parent polyketide, while additional enzymes (*e.g.* ketoreductases (KR), cyclases, aromatases) are required to fold and cyclize the poly-β-ketoacyl intermediate into the final structure. Structural diversity in type II PKS natural products therefore originates from variation within the minimal-PKS domains and their combination with assorted modification enzymes. Following the principle that structural diversity can be generated by constructing new combinations of type II PKS biosynthetic elements, Hutchinson and co workers turned to eDNA as a source of novel genes.[43] Knowing that all three constituents of the minimal-PKS (KS_α-KS_β-ACP) are required for function, a system was designed to determine how heterologous $KS_{\alpha\beta}$ pairs affect overall polyketide chain length and structure.

SEK4 (**29**)   SEK4b (**30**)   UWM1 (**31**)   Tetracenomycin (Tcm) F2 (**32**)
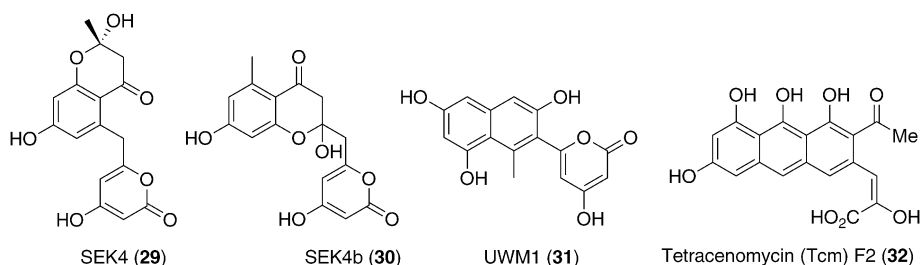
Fig. 10 KS_β sequences were PCR-amplified from eDNA and these eDNA-derived KS_β sequences were used to construct hybrid minimal PKSs. The resulting hybrid minimal PKS cassettes were then shuttled into *Streptomyces* and found to produce compounds **29–32**.

Degenerate PCR primers based on conserved regions within KS_α and ACP genes in type II PKSs known to produce 16- and 24-carbon chains were used to amplify KS_β sequences directly from soil DNA. The four KS_β genes found in this study were distinct from any previously described KS_β sequences. By swapping each new KS_β gene for the corresponding gene in the tetracenomycin (TCM) and actinorhodin (ACT) biosynthetic pathways from *Streptomyces glaucescens* and *Streptomyces coelicolor*, respectively, hybrid PKS cassettes were constructed. *S. lividans* and *S. glaucescens* were then employed as heterologous expression hosts. Ultimately, the known octaketide shunt products UWM1 (**29**), SEK4 (**30**), SEK4b (**31**) and the decaketide TCM F2 (**32**) (Fig. 10) were characterized from these hybrid systems.[43] Although this study did not result in the discovery of novel small molecules from eDNA, it was the first study to demonstrate the potential utility of the vast genetic diversity encoded within the genomes of uncultured bacterial communities for the discovery of small molecules.

4.2 Type I PKS

The dienic alcohol isomers **33** and **34** were derived from a study that used both DNA- and phenotype-based strategies to screen a 5000-member eDNA cosmid clone library (Fig. 11). In this study the cosmid library was initially screened by PCR for the presence of type I polyketide ketosynthase (KS) domains, and then extracts from a subset of clones were screened by HPLC.[44] Eleven novel PKS I-like sequences were identified in a screen of the library using two different sets of ketosynthase (KS)-specific degenerate primers. The eDNA-derived KS sequences showed similarity to known type I KS sequences from a variety of bacteria including myxobacteria (*Stigmatella aurantiaca*), cyanobacteria (*Microcystis* and *Nostoc*) and various mycobacteria. One KS-containing clone was fully sequenced and found to contain six predicted NRPS and PKS megasynthases. These included a truncated NRPS gene, a mixed NRPS–PKS gene, three full-length PKS genes and a truncated PKS gene. Individual clones from this library (both KS-containing clones as well additional randomly selected clones) were individually mobilized into *Streptomyces lividans* TK24 for heterologous expression studies. Each *S. lividans* clone was initially screened for antibacterial activity, but no antibacterially active clones were detected. Extracts from cultures of 40 *S. lividans* recombinants grown under a variety of culture conditions were subsequently screened by HPLC for the presence of clone-specific metabolites. Two clones that yielded identical chromatographic
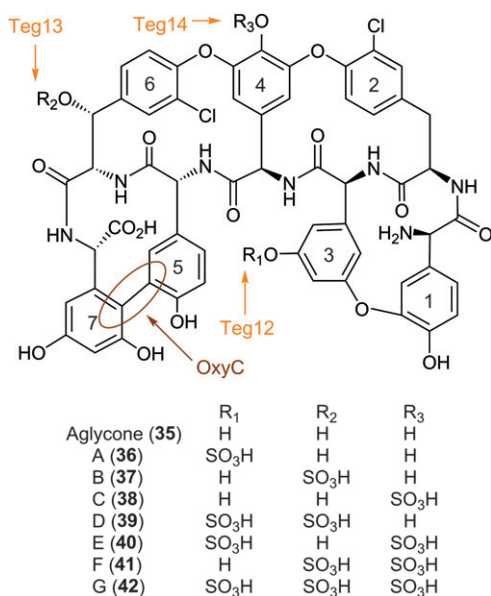

33   **34**
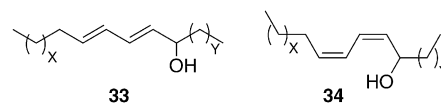
Fig. 11 Diene alcohols (**33** and **34**) found during the analysis of culture broth extracts from cultures of *S. lividans* transformed with a soil eDNA cosmid clone (X + Y = 12).

profiles were found to produce six related clone-specific metabolites. Liquid chromatography–mass spectrometry analysis of the clone metabolites found in organic extracts from these *S. lividans* exoconjugants led to the identification of the novel aliphatic diene alcohol isomers **33** and **34**.[44]

4.3 Targeting OxyC genes

In addition to screening eDNA libraries for completely novel small molecules, it also possible to imagine screening large eDNA libraries for new derivatives of many pharmacologically important natural products. With this in mind, a recent study set out to identify eDNA-derived gene clusters that might encode the biosynthesis of new glycopeptide congeners related to the clinically relevant antibiotics vancomycin and teicoplanin.[45] Vancomycin and teicoplanin are nonribosomal peptide synthetase (NRPS) derived, oxidatively cross-linked, heptapeptide antibiotics. OxyC is a conserved cytochrome P450 that catalyzes an oxidative C–C coupling between the hydroxyphenylglycine at position 5 and the dihydroxyphenylgylcine at position 7 in both vancomycin and teicoplanin.[46] PCR screening of a 10 000 000-membered eDNA mega-library using degenerate primers based on known OxyC sequences yielded two OxyC-like amplicons. These sequences were then used as probes to recover OxyC containing cosmid clones from the mega-library. Successive rounds of library screening were carried out to recover overlapping cosmid clones representing two eDNA-derived glycopeptide gene clusters.

Bioinformatic analysis suggested that the VEG (vancomycin-like eDNA gene cluster, Genbank No. EU874252) and TEG (teicoplanin-like eDNA gene cluster, Genbank No. EU874253) pathways both encode the biosynthesis of new glycopeptide congeners (Fig. 12). The VEG pathway contains the complete complement of genes required for the biosynthesis of a cross-linked, glycosylated heptapeptide (hydroxyphenylglycine (HPG), betahydroxytyrosine (BHT), HPG, HPG, HPG, BHT-dihydroxyphenylglycine (DPG)).[45] The TEG cluster encodes a predicted heptapeptide that only differs from the VEG product by the substitution of DPG for HPG at position 3. While

	R_1	R_2	R_3
Aglycone (35)	H	H	H
A (36)	SO$_3$H	H	H
B (37)	H	SO$_3$H	H
C (38)	H	H	SO$_3$H
D (39)	SO$_3$H	SO$_3$H	H
E (40)	SO$_3$H	H	SO$_3$H
F (41)	H	SO$_3$H	SO$_3$H
G (42)	SO$_3$H	SO$_3$H	SO$_3$H

Fig. 12 The VEG and TEG glycopeptide gene clusters were cloned from a soil eDNA mega library. Sulfo-teicoplanins A–G (36–42) were produced from the teicoplanin aglycone (35) using three sulfotransferases (Teg13, 14, and 15) found in the eDNA-derived TEG gene cluster.

the VEG pathway is rich in methyl- and glycosyltransferase-finishing enzymes, the TEG pathway contains three predicted sulfotransferase genes (*Teg*12, 13 and 14). The presence of these sulfotransferases suggests TEG might encode the biosynthesis of an unprecedented polysulfated glycopeptide congener (Fig. 12). The TEG pathway is predicted to produce a heptapeptide that only differs from the teicoplanin aglycone (35) by the substitution of BHT for the tyrosine at residue 2. Based on this predicted structural homology, the authors elected to use the three TEG sulfotransferases in an attempt to modify the teicoplanin substrate with novel sulfation patterns *in vitro*. In total, seven sulfated glycopeptide congeners, the sulfo-teicoplanins A–G (36–42), were generated in this study (Fig. 12).[45] Sulfo-teicoplanins A–G showed activity against MRSA strain USA 300 and VRSA strain JH2. A general trend of increased MIC was observed with the addition of more sulfates to the teicoplanin aglycone (Fig. 12).[45]

5 Metagenomics of symbiotic bacteria

In the previous section, we focused on pathway cloning from environments such as soils, waters, and so on. Here, the focus is on metagenomic approaches that examine organisms as a habitat – in short, metagenomic approaches to symbiosis. There are also many other ways to examine natural products from symbioses, the most important being cultivation-based approaches to obtain pure strains of bacteria. These will not be discussed, as they are covered by an excellent recent series of reviews in this journal.[12,13]

Symbiosis is often broadly defined as the intimate association of two or more different organisms. For example, mutualistic symbioses in which each partner benefits are obvious examples of the classical definition of symbiosis, but casual and pathogenic associations are also sometimes considered to be symbioses. The

definition of symbiosis is subject to some debate, but for practical reasons, any situation in which the source of genes is a whole (mixed) organism will be considered a symbiosis in this review. Bacteria are arguably the most important 'symbionts' from the natural products perspective, although symbiotic fungi, dinoflagellates, and other organisms can also produce important natural products. Bacteria live in virtually every animal and plant, as well as in many other types of organisms such as other bacteria, fungi, and other eukaryotic microbes.

There are two primary ways of looking at organisms as a habitat in the pursuit of natural products. The first, a gene-directed method, is essentially identical philosophically and methodologically to homology-based and phenotypic approaches described in previous sections of this review. For example, natural product biosynthetic genes can be cloned and expressed from uncultivated bacteria inhabiting a whole animal, just as they can be cloned and expressed from soils. A good example of this close relationship can be found in the gypsy moth story earlier in this review (see Section 3.2.5).[39] The second approach involves looking at natural products first, then engaging in directed gene cloning. For this reason, it is chemistry-directed (also called "structure-directed") rather than gene-directed. For example, there are many potently bioactive marine natural products isolated from animals, and often these compounds are thought to have a microbial origin. Biogenetic hypotheses based upon these chemical structures can be used to clone or identify biosynthetic genes. In terms of technology, chemistry-directed approaches are being used to solve the so-called "supply problem" of marine natural products. Instead of harvesting animals from reef environments to obtain compounds, genetic techniques are used to synthesize the compounds in transgenic bacteria. Chemistry-directed methods also have certain biotechnological advances, described in Section 8.

## 6 Chemistry-directed approaches to symbiosis

Often, a known chemical structure from a whole organism, such as an animal or plant, is suspected to be of microbial origin. In these cases, metagenomics can be used in the discovery and expression of the biosynthetic pathway to the known compound. This bears an exact analogy to pathway cloning from cultured bacteria, and the underlying methods are similar. A major difference is that metagenomes in animals are enormously complicated, with hundreds or thousands of bacterial strains sometimes inhabiting a single animal. This increased complexity has required the development of new techniques and ideas, the basics of which will be reviewed in this section.

Primarily, the complexity of metagenomes in whole organisms has necessitated approaches that simplify the problem. As the cost of sequencing decreases and bioinformatics becomes more powerful, it is increasingly possible to avoid these simplifying schemes, using shotgun sequencing and informatics to identify candidate clones, but this remains challenging. Two basic simplifying approaches have been described: 1) Identification of candidate producing microbial strains prior to cloning; 2) Identification of genes leading to extremely highly homologous compounds or biochemical modifications in simpler systems. Application of these methods has led to the major successes in chemistry-directed metagenomics thus far.
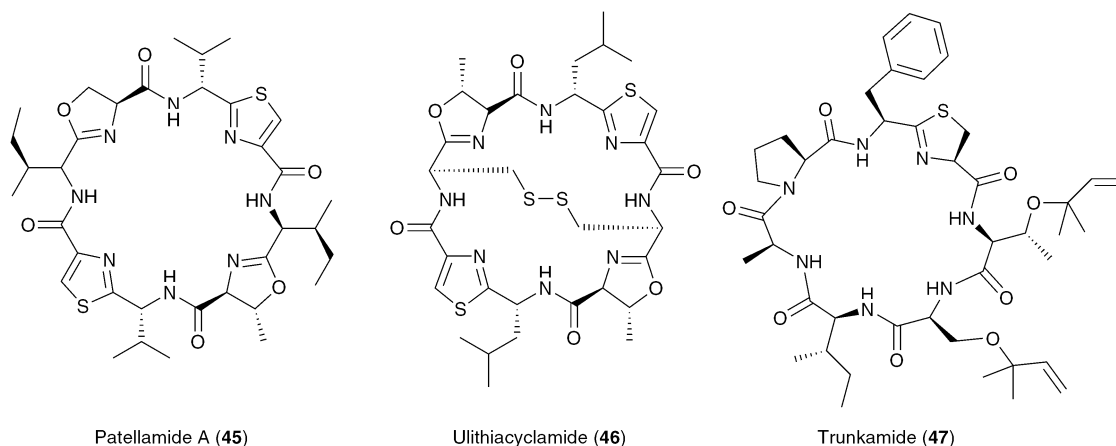
6.1 Candidate strain approaches

6.1.1 Bryostatins. The first genes to be identified using simplifying approaches were those to the highly potent anti-cancer compounds, bryostatins (*e.g.*, **43**) (Fig. 13). In this example, candidate producing bacteria were identified prior to gene cloning. Bryostatins were isolated in low quantities from the marine bryozoan *Bugula neritina*, where they serve to chemically defend the animal and its larvae from predation.[47] It was known that *B. neritina* maintained a single bacterial strain using vertical transmission.[48] That is, the larvae of *B. neritina* contained this strain, ensuring passage from one generation to the next. The bacteria have yet to be cultivated, but they were identified by methods including 16S gene sequence analysis and *in situ* hybridization to be '*Candidatus* Endobugula sertula'.[49] Because it was clear that these were the major (and essentially only) bacteria
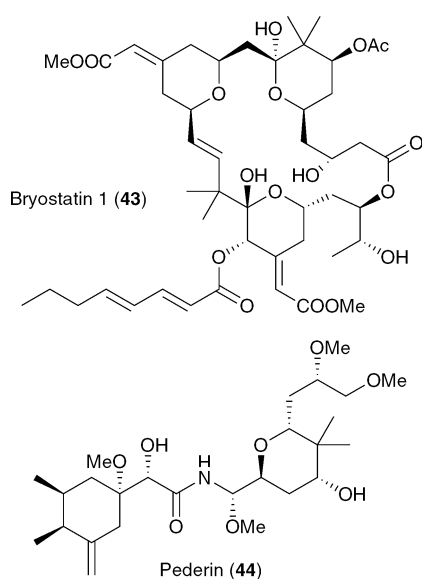
transmitted in the larvae, larvae were obtained and used to greatly enrich the uncultivated bacterium. Subsequently, bacterial type I PKS degenerate primers were used to clone candidate ketosynthase (KS) genes.[50] Later, these genes enabled identification and sequencing of the whole gene cluster.[51] Fragments of this gene cluster were functionally expressed in *E. coli*, adding to the probability that the correct cluster had been identified.[52] The identification of a single candidate strain, and subsequent enrichment of that strain prior to gene cloning, greatly enabled whole-pathway identification.

6.1.2 Pederin and *trans*-AT genes. Candidate uncultivated bacteria were also identified in *Paederus fuscipes* beetles, greatly enabling pathway cloning. These beetles secrete a noxious substance containing a highly potent cytotoxin, pederin (**44**) (Fig. 13). Microbiological data implicated an uncultivated bacterium, *Pseudomonas* sp., in the synthesis of these compounds, and the beetles themselves were relatively enriched in this uncultivated strain.[53–55] This identification enabled cloning of candidate biosynthetic genes using a PCR-based approach. KS-specific degenerate primers were used to identify a candidate cluster from whole-beetle metagenomic DNA, leading to the first publication of a "symbiotic" biosynthetic gene cluster in a groundbreaking study.[56] Because the genes matched the expected domain architecture for pederin, as type I modular PKS genes, it seemed highly likely that the correct genes were identified. More recently, direct biochemical evidence for this pathway function has been obtained.[57] A difficulty in pathway analysis was that the pederin PKS genes lacked acyltransferase (AT) domains, indicating that they belong to a growing family of "*trans*-AT" PKS genes. The first group of bacterial type I modular PKSs to be cloned were all "*cis*-AT", having the AT domain *in cis* with the remainder of the PKS polypeptide. By contrast, the *trans*-AT group has a distinct set of AT polypeptides that dock with the rest of the PKS polypeptide at specific docking sites. At the time of the initial pederin pathway discovery, no such other genes were in the literature, complicating analysis. However, the discovery proved to be a boon to chemistry-directed cloning approaches, as described below.

Fig. 13 Genes for the biosynthesis of the above two polyketides were more readily obtained because of guiding microbiology that simplified the problem.

Fig. 14 Representative cyanobactins are shown above. Their biosynthetic genes were cloned from enriched samples of *Prochloron* symbiotic bacteria living in ascidian animals.

6.1.3 Cyanobactins. The importance of obtaining a simplified system, in which candidate producing bacteria can be enriched, is clearly demonstrated by the above studies. A more recent series of studies relying on enriched uncultivated symbionts involved the relationship between didemnid family ascidians and *Prochloron* spp. cyanobacteria.[58] These ascidians inhabit shallow tropical waters, where they often contain high concentrations of cytotoxic cyclic peptides, such as patellamide A (**45**), ulithiacyclamide (**46**), and trunkamide (**47**) (this family of compounds has recently been named "cyanobactins") (Fig. 14).[59] Didemnids were also well known for harboring *Prochloron* spp., which have been extensively studied for their biological properties. For example, they contribute nutritionally to the survival of their host animals.[58] Because of this relationship, the first structure paper on cyanobactins identified *Prochloron* as possible cyanobactin producers.[60] Later cell-separation evidence further implicated *Prochloron* in cyanobactin biosynthesis, although conflicting evidence was obtained.[61–63] *Prochloron* is easily enriched by simply squeezing the host. In our hands, enriched samples containing 50–90% *Prochloron* DNA are obtained by this method within a 10-minute period.

Using this advantage, two different approaches were employed to identify cyanobactin-producing genes. A large-insert library of *Prochloron*-enriched DNA was transferred to *E. coli* and screened for presence of cyanobactins, leading to identification of clones that were capable of cyanobactin synthesis.[64] A chemistry-directed cloning approach was also used. In principle, the cyanobactins could be synthesized either ribosomally or nonribosomally. PCR and fosmid analysis of *Prochloron* metagenomes from 17 samples made the NRPS-based pathway extremely unlikely.[65] Therefore, the *Prochloron* metagenome was directly shotgun-sequenced, leading to identification of ribosomal precursor peptides encoding cyanobactins. Clones containing these sequences were expressed in *E. coli*, leading to production of the predicted cyanobactins.[66] These studies marked the first pathway expression of biosynthetic genes from uncultivated symbionts. With the sequence context available from metagenome sequencing, the producing organism was confirmed to be *Prochloron*. This is a simplified system, since the ribosomal peptide sequences are directly encoded and thus readily verified, and the DNA is greatly enriched for *Prochloron*, but it shows the power of direct sequencing in metagenomics applications. The availability of a gene cluster has also enabled

biosynthetic pathway engineering, as discussed later in this review.

6.1.4 Fungal–bacterial symbiosis. Symbioses involving fungi have been examined using the chemistry-targeting approach. Rhizoxin (**48**) and rhizonin (**49**) are polyketide–peptide and peptide toxins, respectively, found in the fungus *Rhizopus microsporus* (Fig. 15). When attempting to clone biosynthetic genes for rhizoxin using a priming strategy, only bacteria-like PKS genes were identified.[67–70] By contrast, fungal-type PKS genes were not found. In fact, these PKSs were *trans*-AT genes that were localized to an endosymbiotic bacterium, '*Burkholderia rhizoxina*', living within the fungal mycelia. Later, rhizonins were also linked to endosymbiotic *Burkholderia*.[68] This symbiosis has since been found to be of global occurrence.[70]

6.1.5 Leaf-cutting ants. A final example of the chemistry-directed approach involves PCR of whole environments after discovery of pathways in cultivated bacteria. An example of this approach concerns symbiotic *Streptomyces* sp. living with leaf-cutting ants. These ants cultivate fungal gardens, in which leaves are digested first by fungi, then consumed by ants. Because these beneficial fungal monocultures are sometimes infected by other fungi, the ants have developed a strategy to control fungal infection. The ants have very specific symbiotic relationships with actinomycete bacteria that synthesize antifungal compounds, enabling the ants to maintain their fungal gardens.[71] From *Streptomyces* sp. isolated from various leaf-cutting ant species, the known antifungal agent candicidin A (**50**) was identified (Fig. 16).[72] This amphotericin-like polyene significantly inhibited pathogenic fungi, but not beneficial fungi. A metagenomics approach was then applied in which candicidin biosynthetic genes were amplified both from cultivated bacteria and from whole-ant DNA, clearly indicating the presence of these genes in a single worker ant.

6.1.6 Discodermolide – the need for simplicity. Clearly, identification of putative producing strains greatly accelerates metagenomic approaches in animals. An example of the difficulty of pathway discovery without prior candidate strain identification can be found in attempts to clone the discodermolide (**51**) gene cluster from the sponge *Discodermia dissoluta* (Fig. 16).[73] Discodermolide is an important lead anticancer compound of

Rhizoxin (**48**)   Rhizonin A (**49**)
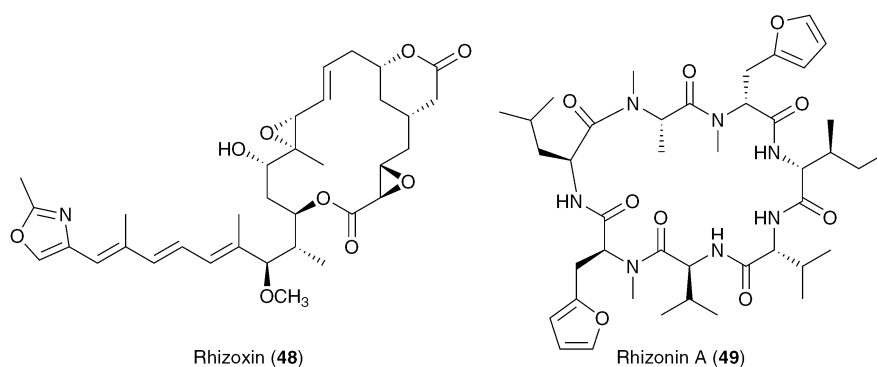
Fig. 15 Two metabolites originally isolated from fungi were later shown to originate in endosymbiotic bacteria.

complex, presumed polyketide, biosynthetic origin. Work at the Kosan company employed a unique hybridization array strategy to discover *hundreds* of different PKS and nonribosomal peptide synthetase (NRPS) gene clusters from the sponge, yet none could be definitively tied to discodermolide production. This study was very important because it revealed for the first time the true complexity of natural product genes in these metagenomes.

6.1.7 Enabling technologies for identification of producing strains. Clearly, as the above examples illustrate, relatively precise gene-targeting methods have been required to obtain the desired genes from symbiotic metagenomes. Below, some of the most important non-genetic advances in biosynthetic pathway targeting will be described. These targeting technologies have enormous potential to speed the rate of chemistry-directed cloning in symbiosis and metagenomic research. However, it should be kept in mind that as core genetic technologies continue to decrease in cost, simpler direct sequencing, DNA synthesis, and expression methods will likely compete favorably with these targeting strategies.

In the past 30 years, cell separation has been used to determine which cells contain the natural product of interest within complex symbiotic mixtures.[74] These studies impacted later metagenomic approaches, such as those used with *Prochloron*–ascidian and *Oscillatoria*–sponge symbioses. In this type of study, cells are physically separated by various methods, and then extracted with solvents and analyzed. These cell separation studies could be supplemented with genetic approaches that helped to focus in on the target organisms. An early example of this type of research connecting cell separation to genetics involved symbiosis between *Theonella swinhoei* (and related sponges) and filamentous heterotrophic bacteria. Cell separation studies implicated these filamentous bacteria in production of complex cyclic – and presumably nonribosomal – peptides such as theopalauamide (**52**) (Fig. 17).[75] Using 16S methodology, the bacteria were identified as '*Candidatus* Entotheonella palauensis', a deeply branching group distantly related to known δ-proteobacteria such as myxobacteria and sulfate reducers.[76] '*Candidatus* Entotheonella palauensis' has since been found as a symbiont of various sponge species, where it harbors a considerably diverse set of natural product biosynthetic genes.[73,77]

A substantial improvement to the cell separation procedure was recently reported, in which individual cells are directly examined by MALDI MS.[78] Since single cells can be readily obtained and sequenced or used for PCR, this technology promises to greatly accelerate metagenomics research. It could also greatly reduce errors associated with cell separation techniques, most of which derive from the ability of bacteria to secrete natural products and the difficulty of localizing extremely hydrophobic, membrane-associated compounds. These problems make misidentification of producing cells a very real possibility if cell separation is employed. By contrast, a lack of specificity is very readily ruled out by the MALDI method.

Another series of advances derive from immunohistochemical approaches to compound localization. The first such experiment localized latrunculin-family compounds (*e.g.*, **53**) to sponge cells (Fig. 17).[79] Later, the excitatory amino acid derivative, dysiherbaine (**54**) was localized to *Synechocystis* sp. cyanobacteria within the sponge *Ledenfeldia chondrodes* (formerly *D. herbacea*).[80] As has been found in other cyanobacterial symbioses in sponges and ascidians, natural product synthesis was only associated with a subset of *Synechocystis* cells, despite essentially identical 16S genes. A clear caveat to all types of localization studies – based on both histochemistry and mass spectrometry – was found in immunohistochemical examination of bryostatin location in *B. neritina* bryozoans.[81] Although the compounds are almost certainly produced by certain bacteria, the compounds did not co-localize with the bacteria and instead were distributed in host tissues. This study clearly demonstrates the complexity inherent in symbiotic natural products research.

6.2 Chemical and biochemical homology approaches

6.2.1 Onnamide and *trans*-AT. An alternative chemistry-directed approach involves identification of very precise biosynthetic gene relatives, usually to highly homologous compounds. For example, PKS probes sometimes identify hundreds or thousands of candidate genes in metagenomes without enrichment; this is simply too much to handle using current technology. Identification of much more precise chemical motifs allows the number of candidate genes to be cut down
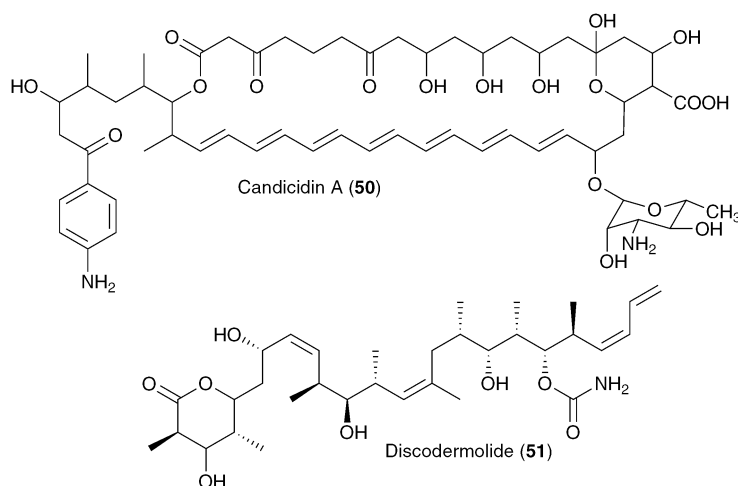

Fig. 16 Polyketides from ant symbionts (**50**) and, putatively, a sponge symbiont (**51**).
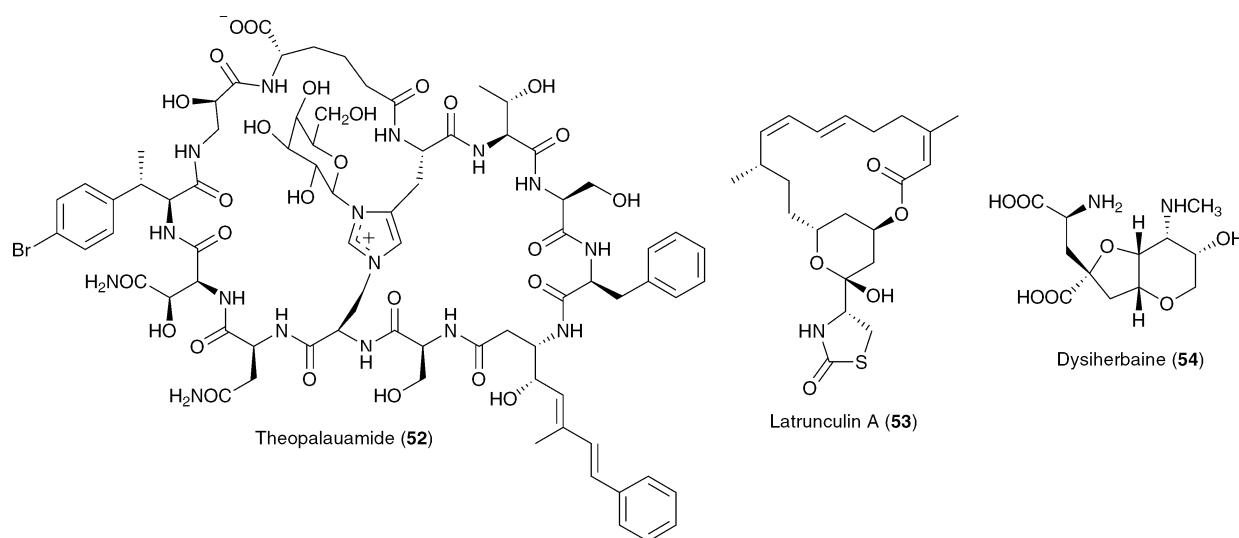
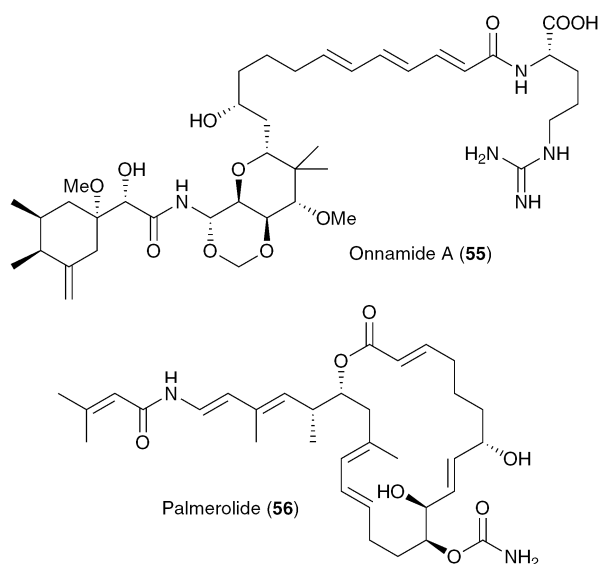**Fig. 17** Compounds that helped with enabling technologies.

Fig. 18 Some polyketides from sponges and ascidians.

greatly, potentially even to single hits. This approach has been successfully applied to several symbiotic metagenome problems. The first example of such an approach was the cloning of the onnamide gene cluster. Onnamide A (**55**), found in marine sponges of the genus *Theonella*, is very similar in structure to pederin (**44**) from *Paederus* beetles (Fig. 18). In fact, much of the polyketide portion of this molecule is essentially identical to that of pederin. A major difference is the addition of a terminal arginine, potentially added by NRPS domains, in the sponge compounds. In the initial identification of pederin biosynthetic genes, NRPS domains potentially activating arginine were identified.[56] However, the gene cluster was 'interrupted' by an oxidase domain, making it very likely that the NRPS domains were in fact not functional in the pederin pathway. The very clear homology between pederin and onnamide genes were used to identify the candidate onnamide gene cluster.[82] Briefly, the *trans*-AT feature of this gene cluster was used in the cloning

strategy. Specific primers for KS domains were applied. Subsequent phylogenetic analysis allowed identification of *trans*-AT PKSs, which were just a small subset of the total number of identified PKS gens in the metagenome. Fosmid clones were then screened to provide a partial candidate onnamide cluster. The fosmids also contained sequences for other bacterial gene types, indicating that these 'sponge' compounds were more probably of bacterial origin.

The onnamide study also led to the development of yet another enabling technology. The very complex, whole-sponge metagenome of *Theonella* was analyzed using a new pooling method, in which pools of clones were partitioned into semi-liquid media.[83] The pools were then screened by PCR, leading to identification of pools containing the target gene of interest. Pools were diluted, then re-screened by PCR, allowing identification of pure fosmids containing the target gene cluster in just a few steps. This widely applicable method allows the rapid, PCR-based identification of fosmids from highly complex mixtures. By contrast, other methods examining large libraries previously relied on hybridization, a much more expensive and potentially even problematic approach in some cases.

6.2.2 Palmerolide. A recent study of the palmerolide (**56**)-containing Antarctic ascidian *Synoicum adareanum* followed a similar strategy in the first report of ascidian metagenomic PKS genes (Fig. 18).[84,85] Palmerolide is a potent antitumor natural product of presumed hybrid PKS-NRPS origin. KS primers were used to identify PKS gene fragments clustering most strongly with the *trans*-AT group. *cis*-AT genes were not cloned in this study, and the identified KS genes shared similar features such as GC percentage, indicating that they could derive from similar or identical bacteria or gene clusters. These results are very promising for the ultimate identification of palmerolide biosynthetic genes, although much remains to be done to determine whether the correct genes have been found.

6.2.3 Halogenated peptides from sponge symbionts. An early example of genetically targeting a precise chemical motif can be
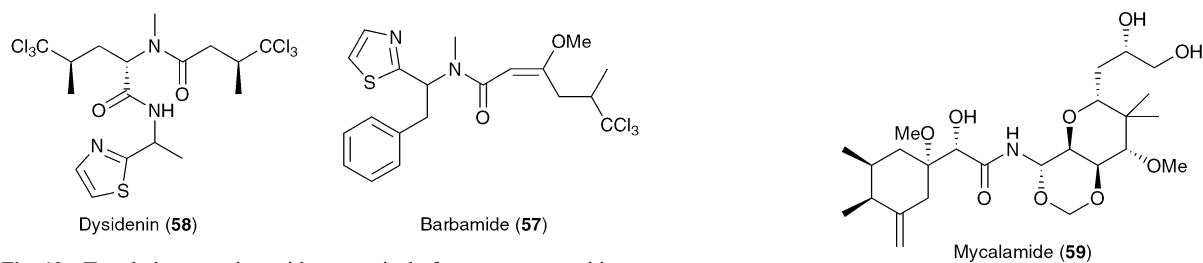
Dysidenin (**58**)　　　　Barbamide (**57**)

Fig. 19　Two halogenated peptides, putatively from sponge symbionts.

Mycalamide (**59**)

Psymberin (**60**)

Fig. 20　Further probable *trans*-AT products from sponges.

found in the barbamide (**57**)/dysidenin (**58**) group of non-ribosomal peptide natural products (Fig. 19). The dysidenin family was isolated from marine sponges of the genera *Dysidea* and *Lamellodysidea*, which harbor photosynthetic symbiotic cyanobacteria of the genus *Oscillatoria*. Cell separation studies implicated *O. spongeliae* in the synthesis of these nonribosomal peptides,[86] but cell separation methods can at times be misleading. Barbamide (**57**) was isolated from *Lyngbya majuscula*, a tropical marine cyanobacterium that forms mats containing diverse bacteria. The identification of the barbamide biosynthetic gene cluster from *L. majuscula* facilitated identification of candidate dysidenin biosynthetic genes within marine sponges.[87] This family of compounds contains a very rare chemical motif, a multiply chlorinated leucine residue. By sequence alignment and analysis, it proved possible to design specific primers aimed at this chemical motif, targeting the halogenase gene. In two separate studies, this feature was used to isolate candidate halogenases from marine sponges containing the dysidenin family of compounds. In one study, phylogenetic analysis and comparison was used to identify halogenases that were present only in tandem with the dysidinen-group compounds.[88,89] In another study, the identified halogenase fragments were hybridized with mRNA expressed in *Oscillatoria* sp. within whole sponge tissue, using a method known as catalyzed reporter deposition-fluorescence *in situ* hybridization.[90] This study closed the loop between pathway expression and cell separation studies for the first time. Presumably, these genetic features could possibly be used to obtain intact dysidenin family biosynthetic gene clusters, although this has not been reported. Through these studies, it was clear that the sponges contain multiple strains of *Oscillatoria*, some of which contained halogenase genes and some of which did not despite some otherwise identical genetic features.

6.2.4　*trans*-AT genes in sponges. Very recently, a twist on the *trans*-AT approach was described. A major difference between *trans*-AT and *cis*-AT pathways is that in the *trans*-AT group, phylogenetic analysis of KS domains can be used to predict chemistry.[91] Conversely, certain chemical features imply conserved sequence features within KS domains. When analyzing a group of sponges, Fisch *et al.* found that general KS primers amplified too many diverse PKS gene sequences, making it very difficult to identify the correct gene cluster.[92] A nested PCR strategy was then applied. First, the degenerate KS primers were used. Subsequently, a *trans*-AT specific PCR primer pair was applied. The primers were designed such that one primer was specific to 84% of all *trans*-AT KSs, while the second primer was specific for certain chemical motifs in the resulting polyketide
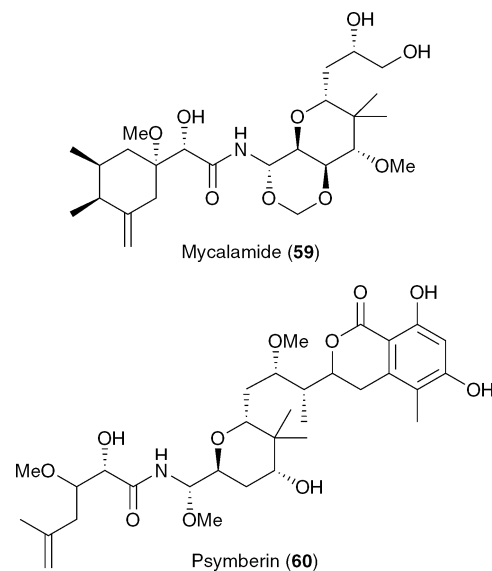
product. Using this strategy, probable gene clusters were identified for pederin relatives, mycalamide (**59**) and psymberin (**60**), both from marine sponge metagenomes (Fig. 20). In the case of psymberin, over 400 000 clones were screened using the pooling strategy described above, leading to rapid identification of 5 fosmids containing the putative psymberin cluster.

7　Gene-directed approaches to symbiosis

In the above examples, chemical structures were used to guide metagenomic approaches to gene cloning. Relatively less information is available concerning gene-directed approaches in whole organisms. Recently, it has become extremely common to perform metagenomic analysis on whole animals, plants, and other organisms. However, application of this approach to natural products has relatively rarely been discussed.

Perhaps the first example of a metagenomic approach applied to symbiosis involved lichens, which are symbiotic mixtures of fungi, photosynthetic algae or cyanobacteria, and other bacteria, where primarily fungal PKSs are targeted.[93,94] More recently, the metagenomic approach has been applied to marine sponges. Sponges are considered to be good targets of this approach because they contain abundant and highly diverse bacterial symbionts; in some cases up to 60% of their mass is bacteria.[74,94,95] The diversity of bacteria in sponges has been recently reviewed.[96] In several cases, sponge symbiont PKS and NRPS genes have been cloned by PCR without using chemical guidance. This approach gives a broad picture of the metabolic diversity and potential of marine sponges.

Chemistry-based screening of sponges using KS probes, as described in the preceding section, provided much insight into the diversity of PKS genes in sponges. Much has also been learned by gene-based approaches. An early example examined *Pseudoceratina clavata*, a sponge found on the Great Barrier Reef.[97] Using KS primers, 5 different KS-containing clones were identified. In addition, 5 fosmids containing additional KS

sequences were identified by multiplex PCR screening. Bacteria were cultivated from the same sponge, providing a comparison set. While the cultivated bacteria contained both *cis*- and *trans*-AT PKSs related to other bacterial genes, the 10 sequenced sponge metagenome clones branched in a new, sponge-specific bacterial clade. This new clade appeared to be associated with lipid biosynthesis.

A concurrent study reported that 11 out of 20 examined sponge species contain this sponge-specific clade, dubbed *sup*.[98,99] KS domains were amplified by PCR, and 150 clones were identified with <97% sequence identity and thus considered unique. Of these, 127 were part of the *sup* clade, while 6 were in another apparently sponge-specific group. The remaining sequences fell into the *cis*-AT, *trans*-AT, or PKS–NRPS groups that are associated with natural products. Based upon sequencing of the *sup* cluster, the sponge-specific group was predicted to synthesize methyl-branched fatty acids. The existence of such a large, unanticipated metabolic group that has not been found elsewhere underscores the value of examining animal metagenomes directly. The fact that only a small percentage of sponge PKS genes were predicted to lead to natural products was considered very useful to chemistry-directed approaches, since the *sup* genes could be discarded as candidates.

A recent study examined the metabolic potential of the sponge, *Haliclona simulans*.[100] Seven PKS genes and no NRPS genes were identified by PCR. All of the PKS genes appeared to be natural product-associated and did not fall in the sponge-specific clades.

8   Scientific and technological advances enabled by symbiosis metagenomics

It could be asked, given the number of natural product biosynthetic genes found in soil and other metagenomes, why even bother studying pathways from whole animals and other organisms? There are many scientific reasons to pursue such basic studies, but there is also a biomedical and technological rationale. For example, a widely pursued biomedical angle is solving the supply problem of compounds from marine and other animals.[74] By specifically accessing genes to important marine invertebrate compounds, it is possible to synthesize the genes in culture instead of collecting animals from coral reefs. There are many other technical advances enabled by examination of symbiotic metagenomes. As one example, examination of the cyanobactin biosynthetic pathways in marine ascidians and their symbionts has led to facile genome mining and rational engineering of new compounds for drug discovery.[101–103] Finally, certain types of natural product genes may be strongly associated with or essentially only found in certain animal habitats, as exemplified by the *sup* group.[99]

More than 60 cyanobactins have been isolated from ascidians harboring *Prochloron* bacteria.[59] These compounds differ by their amino acid sequences and lengths, as well as by the presence or absence of certain chemical features such as prenylation or oxidation. By studying pathways from nearly 100 different *Prochloron*–ascidian symbioses, single mutations were identified that led to
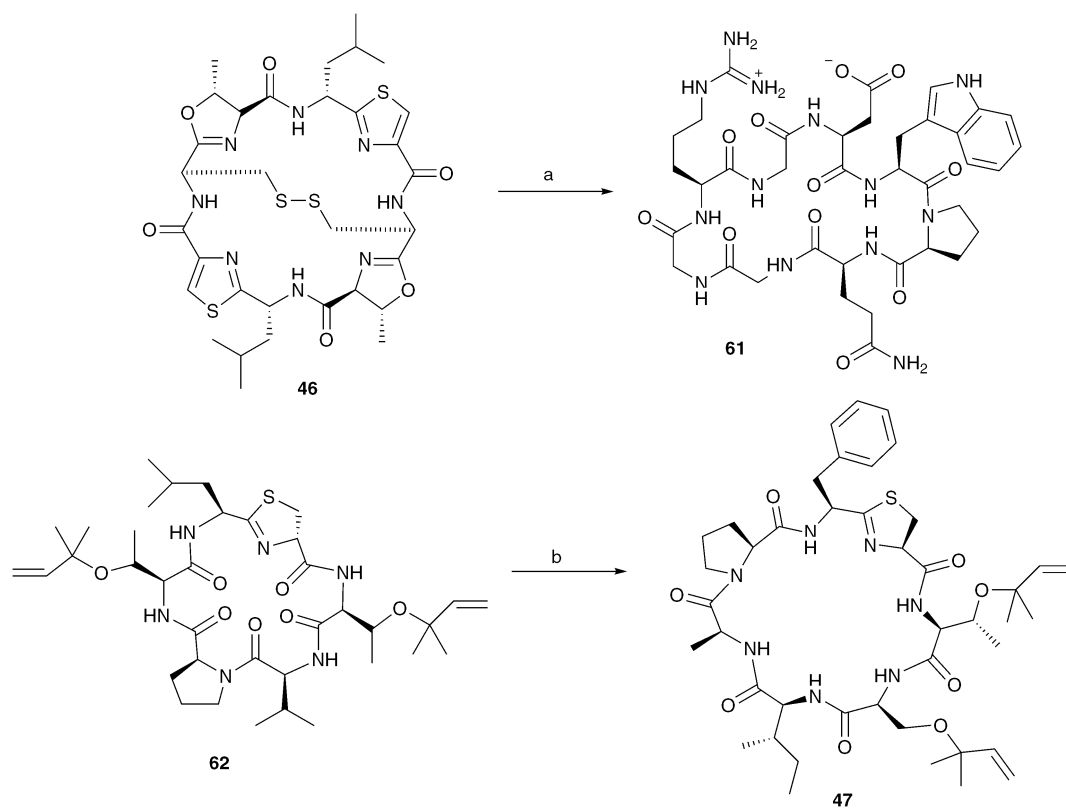
Fig. 21 Use of metagenomics in engineering. Top: The natural product ulithiacyclamide (**46**) was 'converted' to the wholly unnatural compound eptidemnamide (**61**) by (a) one-step PCR mutagenesis and synthesis in *E. coli*. Bottom: The natural product patellin 2 was converted to the rare natural product trunkamide (**47**) by (b) one-step recombination in yeast, followed by production in *E. coli*. These results and the natural selectivity of the enzymes indicate that large libraries of products can be synthesized by identical enzymes.

different chemical features.[101,103] By contrast, in most cultivated bacteria, there are numerous mutations even in pathways to nearly identical compounds. By examining the single mutations in cyanobactin pathways, a wholly unnatural natural product, epti-demnamide, was readily engineered for synthesis in *E. coli* (Fig. 21). Potentially, large numbers of derivatives could be synthesized genetically using this method. Indeed, the natural variation in tandem with the eptidemnamide structure suggests an extremely relaxed substrate selectivity of nearly identical modifying enzymes.[59] Another example of the application of this method is in drug supply. The rare cyanobactin derivative trunkamide has only been found in a small number of ascidians, potentially hindering its development as an anticancer drug lead. Using yeast recombination, genes for this compound were introduced into *E. coli*, where the mature product was synthesized.[103] These rapid genetic modifications take minimal work but introduce a great degree of structural variety into 'unnatural natural product' libraries.[59,101] As a final example, the enzymes underlying cyanobactin synthesis are of demonstrated broad substrate selectivity and have potential for the synthesis of fine chemicals *in vitro* or *in vivo*.[104]

Symbiosis in cultivated organisms has also yielded engineering insights. For example, the human pathogen *Mycobacterium ulcerans* (and other mycobacterial strains) produce mycolactone and derivatives, which cause debilitating skin ulcers in humans. By genomic comparison (albeit using cultivated bacteria), sequence features were identified that could enable rational, recombination-based synthesis of new polyketides.[105] Thus, symbiotic metagenomes are a promising area of study for biotechnological applications.

9   Conclusions and future directions

In recent years, proof of concept has been obtained for meta-genomic approaches to natural products from diverse habitats. Many of the limiting technical challenges have been overcome, leaving cost as the major potential barrier to progress. However, as DNA sequencing and synthesis costs rapidly decrease, eDNA methods are becoming mainstays of natural products discovery and research. These methods thus promise a bright future, both for the discovery of novel pharmaceuticals from nature and for the better understanding of the roles these compounds play in nature.

10   Acknowledgements

11   References

1   J. Gans, M. Wolinsky and J. Dunbar, *Science*, 2005, **309**, 1387–1390.
2   V. Torsvik, J. Goksoyr and F. L. Daae, *Appl. Environ. Microbiol.*, 1990, **56**, 782–787.
3   V. Torsvik, L. Ovreas and T. F. Thingstad, *Science*, 2002, **296**, 1064–1066.
4   H. W. Jannasch and G. E. Jones, *Limnol. Oceanogr.*, 1959, **4**, 128–139.
5   T. Kaeberlein, K. Lewis and S. S. Epstein, *Science*, 2002, **296**, 1127–1129.
6   S. G. Tringe, C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz and E. M. Rubin, *Science*, 2005, **308**, 554–557.
7   K. Zengler, G. Toledo, M. Rappe, J. Elkins, E. J. Mathur, J. M. Short and M. Keller, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15681–15686.
8   M. S. Rappe and S. J. Giovannoni, *Annu. Rev. Microbiol.*, 2003, **57**, 369–394.
9   J. R. Cole, B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity and J. M. Tiedje, *Nucleic Acids Res.*, 2007, **35**, D169–172.
10 T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu and G. L. Andersen, *Appl. Environ. Microbiol.*, 2006, **72**, 5069–5072.
11 G. Webster, L. Yarram, E. Freese, J. Koster, H. Sass, R. J. Parkes and A. J. Weightman, *FEMS Microbiol. Ecol.*, 2007, **62**, 78–89.
12 J. Piel, *Nat. Prod. Rep.*, 2004, **21**, 519–538.
13 J. Piel, *Nat. Prod. Rep.*, 2009, **26**, 338–362.
14 E. W. Schmidt, *Nat. Chem. Biol.*, 2008, **4**, 466–473.
15 G. Y. Wang, E. Graziani, B. Waters, W. Pan, X. Li, J. McDermott, G. Meurer, G. Saxena, R. J. Anderson and J. Davies, *Org. Lett.*, 2000, **2**, 2401–2404.
16 M. R. Rondon, P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman and R. M. Goodman, *Appl. Environ. Microbiol.*, 2000, **66**, 2541–2547.
17 A. Henne, R. Daniel, R. A. Schmitz and G. Gottschalk, *Appl. Environ. Microbiol.*, 1999, **65**, 3901–3907.
18 W. H. Yap, X. Li, T. W. Soong and J. E. Davies, *J. Ind. Microbiol. Biotechnol.*, 1996, **17**, 179–184.
19 C. DeBoer and A. Dietz, *J. Antibiot.*, 1976, **29**, 1182–1188.
20 P. Lorenz and J. Eck, *Nat. Rev. Microbiol.*, 2005, **3**, 510–516.
21 S. F. Brady, *Nat. Protoc.*, 2007, **2**, 1297–1305.
22 S. F. Brady and J. Clardy, *J. Am. Chem. Soc.*, 2000, **122**, 12903–12904.
23 S. F. Brady, C. J. Chao and J. Clardy, *Appl. Environ. Microbiol.*, 2004, **70**, 6865–6870.
24 S. F. Brady and J. Clardy, *J. Nat. Prod.*, 2004, **67**, 1283–1286.
25 S. F. Brady and J. Clardy, *Org. Lett.*, 2005, **7**, 3613–3616.
26 S. F. Brady, C. J. Chao and J. Clardy, *J. Am. Chem. Soc.*, 2002, **124**, 9968–9969.
27 P. R. August, T. H. Grossman, C. Minor, M. P. Draper, I. A. MacNeil, J. M. Pemberton, K. M. Call, D. Holt and M. S. Osburne, *J. Mol. Microbiol. Biotechnol.*, 2000, **2**, 513–519.
28 S. F. Brady, C. J. Chao, J. Handelsman and J. Clardy, *Org. Lett.*, 2001, **3**, 1981–1984.
29 D. E. Gillespie, S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman and J. Handelsman, *Appl. Environ. Microbiol.*, 2002, **68**, 4301–4306.
30 H. Budzikiewicz, H. Eckau and M. Ehrenberg, *Tetrahedron Lett.*, 1972, **13**, 3807–3810.
31 Z. L. Hegedus, *Toxicology*, 2000, **145**, 85–101.
32 S. F. Brady and J. Clardy, *Angewandte Chem.*, 44, pp. 7063–7065.
33 M. J. Garson and J. S. Simpson, *Nat. Prod. Rep.*, 2004, **21**, 164–179.
34 M. S. Edenborough and R. B. Herbert, *Nat. Prod. Rep.*, 1988, **5**, 229–245.
35 S. F. Brady and J. Clardy, *Angew. Chem., Int. Ed.*, 2005, **44**, 7045–7048.
36 S. F. Brady, J. D. Bauer, M. F. Clarke-Pearson and R. Daniels, *J. Am. Chem. Soc.*, 2007, **129**, 12102–12103.
37 I. A. MacNeil, C. L. Tiong, C. Minor, P. R. August, T. H. Grossman, K. A. Loiacono, B. A. Lynch, T. Phillips, S. Narula, R. Sundaramoorthi, A. Tyler, T. Aldredge, H. Long, M. Gilman, D. Holt and M. S. Osburne, *J. Mol. Microbiol. Biotechnol.*, 2001, **3**, 301–308.
38 H. K. Lim, E. J. Chung, J. C. Kim, G. J. Choi, K. S. Jang, Y. R. Chung, K. Y. Cho and S. W. Lee, *Appl. Environ. Microbiol.*, 2005, **71**, 7768–7777.
39 C. Guan, J. Ju, B. R. Borlee, L. L. Williamson, B. Shen, K. F. Raffa and J. Handelsman, *Appl. Environ. Microbiol.*, 2007, **73**, 3669–3676.
40 L. L. Williamson, B. R. Borlee, P. D. Schloss, C. Guan, H. K. Allen and J. Handelsman, *Appl. Environ. Microbiol.*, 2005, **71**, 6335–6344.
41 J. W. Craig, F. Y. Chang and S. F. Brady, *ACS Chem. Biol.*, 2009, **4**, 23–28.

42 M. Weinstein, R. C. Roberts and D. R. Helinski, *J. Bacteriol.*, 1992, **174**, 7486–7489.

43 K. T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson and J. Davies, *J. Bacteriol.*, 1997, **179**, 7360–7368.

44 S. Courtois, C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet and J. L. Pernodet, *Appl. Environ. Microbiol.*, 2003, **69**, 49–55.

45 J. J. Banik and S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17273–17277.

46 O. Pylypenko, F. Vitali, K. Zerbe, J. A. Robinson and I. Schlichting, *J. Biol. Chem.*, 2003, **278**, 46727–46733.

47 N. Lopanik, N. Lindquist and N. Targett, *Oecologia*, 2004, **139**, 131–139.

48 R. M. Woollacott, *Mar. Biol.*, 1981, **65**, 155–158.

49 S. K. Davidson and M. G. Haygood, *Biol. Bull.*, 1999, **196**, 273–280.

50 S. K. Davidson, S. W. Allen, G. E. Lim, C. M. Anderson and M. G. Haygood, *Appl. Environ. Microbiol.*, 2001, **67**, 4531–4537.

51 S. Sudek, N. B. Lopanik, L. E. Waggoner, M. Hildebrand, C. Anderson, H. Liu, A. Patel, D. H. Sherman and M. G. Haygood, *J. Nat. Prod.*, 2007, **70**, 67–74.

52 N. B. Lopanik, J. A. Shields, T. J. Buchholz, C. M. Rath, J. Hothersall, M. G. Haygood, K. Hakansson, C. M. Thomas and D. H. Sherman, *Chem. Biol.*, 2008, **15**, 1175–1186.

53 R. L. Kellner, *Insect Biochem. Mol. Biol.*, 2002, **32**, 389–395.

54 R. L. Kellner, *J. Insect Physiol.*, 2001, **47**, 475–483.

55 R. L. Kellner, *J. Hered.*, 2000, **91**, 158–162.

56 J. Piel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14002–14007.

57 K. Zimmermann, M. Engeser, J. W. Blunt, M. H. Munro and J. Piel, *J. Am. Chem. Soc.*, 2009, **131**, 2780–2781.

58 R. A. Lewin and L. Cheng, *Prochloron: A Microbial Enigma*, Chapman & Hall, 1989.

59 E. W. Schmidt and M. S. Donia, *Methods Enzymol.*, 2009, **458**, 575–596.

60 C. M. Ireland and P. J. Scheuer, *J. Am. Chem. Soc.*, 1980, **102**, 5688–5691.

61 C. E. Salomon and D. J. Faulkner, *J. Nat. Prod.*, 2002, **65**, 689–692.

62 B. M. Degnan, C. J. Hawkins, M. F. Lavin, E. J. McCaffrey, D. L. Parry, A. L. van den Brenk and D. J. Watters, *J. Med. Chem.*, 1989, **32**, 1349–1354.

63 B. M. Degnan, C. J. Hawkins, M. F. Lavin, E. J. McCaffrey, D. L. Parry and D. J. Watters, *J. Med. Chem.*, 1989, **32**, 1354–1359.

64 P. F. Long, W. C. Dunlap, C. N. Battershill and M. Jaspars, *ChemBioChem*, 2005, **6**, 1760–1765.

65 E. W. Schmidt, S. Sudek and M. G. Haygood, *J. Nat. Prod.*, 2004, **67**, 1341–1345.

66 E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood and J. Ravel, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7315–7320.

67 L. P. Partida-Martinez and C. Hertweck, *Nature*, 2005, **437**, 884–888.

68 L. P. Partida-Martinez, C. F. de Looss, K. Ishida, M. Ishida, M. Roth, K. Buder and C. Hertweck, *Appl. Environ. Microbiol.*, 2007, **73**, 793–797.

69 L. P. Partida-Martinez, S. Monajembashi, K. O. Greulich and C. Hertweck, *Curr. Biol.*, 2007, **17**, 773–777.

70 G. Lackner, N. Mobius, K. Scherlach, L. P. Partida-Martinez, R. Winkler, I. Schmitt and C. Hertweck, *Appl. Environ. Microbiol.*, 2009, **75**, 2982–2986.

71 C. R. Currie, M. Poulsen, J. Mendenhall, J. J. Boomsma and J. Billen, *Science*, 2006, **311**, 81–83.

72 S. Haeder, R. Wirth, H. Herz and D. Spiteller, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4742–4746.

73 A. Schirmer, R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong and C. R. Hutchinson, *Appl. Environ. Microbiol.*, 2005, **71**, 4840–4849.

74 M. G. Haygood, E. W. Schmidt, S. K. Davidson and D. J. Faulkner, *J. Mol. Microbiol. Biotechnol.*, 1999, **1**, 33–43.

75 C. A. Bewley, N. D. Holland and D. J. Faulkner, *Experientia*, 1996, **52**, 716–722.

76 E. W. Schmidt, A. Y. Obraztsova, S. K. Davidson, D. J. Faulkner and M. G. Haygood, *Mar. Biol.*, 2000, **136**, 969–977.

77 W. M. Bruck, S. H. Sennett, S. A. Pomponi, P. Willenz and P. J. McCarthy, *ISME J.*, 2008, **2**, 335–339.

78 E. Esquenazi, C. Coates, L. Simmons, D. Gonzalez, W. H. Gerwick and P. C. Dorrestein, *Mol. BioSyst.*, 2008, **4**, 562–570.

79 O. Gillor, S. Carmeli, Y. Rahamim, Z. Fishelson and M. Ilan, *Mar. Biotechnol.*, 2000, **2**, 213–223.

80 R. Sakai, K. Yoshida, A. Kimura, K. Koike, M. Jimbo, A. Kobiyama and H. Kamiya, *ChemBioChem*, 2008, **9**, 543–551.

81 K. H. Sharp, S. K. Davidson and M. G. Haygood, *ISME J.*, 2007, **1**, 693–702.

82 J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani and S. Matsunaga, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16222–16227.

83 S. Hrvatin and J. Piel, *J. Microbiol. Methods*, 2007, **68**, 434–436.

84 T. Diyabalanage, C. D. Amsler, J. B. McClintock and B. J. Baker, *J. Am. Chem. Soc.*, 2006, **128**, 5630–5631.

85 C. S. Riesenfeld, A. E. Murray and B. J. Baker, *J. Nat. Prod.*, 2008, **71**, 1812–1818.

86 M. D. Unson and D. J. Faulkner, *Experientia*, 1993, **49**, 349–353.

87 Z. Chang, P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis and D. H. Sherman, *Gene*, 2002, **296**, 235–247.

88 C. P. Ridley, P. R. Bergquist, M. K. Harper, D. J. Faulkner, J. N. A. Hooper and M. G. Haygood, *Chem. Biol.*, 2005, **12**, 397–406.

89 C. P. Ridley, D. J. Faulkner and M. G. Haygood, *Appl. Environ. Microbiol.*, 2005, **71**, 7366–7375.

90 P. Flatt, J. Gautschi, R. Thacker, M. Musafija-Girt, P. Crews and W. Gerwick, *Mar. Biol.*, 2005, **147**, 761–774.

91 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.

92 K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews and J. Piel, *Nat. Chem. Biol.*, 2009, **5**, 494–501.

93 V. Miao, M. F. Coeffet-LeGal, P. Brown, S. Sinnemann, G. Donaldson and J. Davies, *Trends Biotechnol.*, 2001, **19**, 349–355.

94 J. Piel, D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen and S. Matsunaga, *J. Nat. Prod.*, 2005, **68**, 472–479.

95 C. E. Salomon, N. A. Magarvey and D. H. Sherman, *Nat. Prod. Rep.*, 2004, **21**, 105–121.

96 M. W. Taylor, R. Radax, D. Steger and M. Wagner, *Microbiol. Mol. Biol. Rev.*, 2007, **71**, 295–347.

97 T. K. Kim and J. A. Fuerst, *Environ. Microbiol.*, 2006, **8**, 1460–1470.

98 T. Hochmuth and J. Piel, *Phytochemistry*, 2009, DOI: 10.1016/j.phytochem.2009.04.010.

99 L. Fieseler, U. Hentschel, L. Grozdanov, A. Schirmer, G. Wen, M. Platzer, S. Hrvatin, D. Butzke, K. Zimmermann and J. Piel, *Appl. Environ. Microbiol.*, 2007, **73**, 2144–2155.

100 J. Kennedy, C. E. Codling, B. V. Jones, A. D. Dobson and J. R. Marchesi, *Environ. Microbiol.*, 2008, **10**, 1888–1902.

101 M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2006, **2**, 729–735.

102 S. Sudek, M. G. Haygood, D. T. Youssef and E. W. Schmidt, *Appl. Environ. Microbiol.*, 2006, **72**, 4382–4387.

103 M. S. Donia, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2008, **4**, 341–343.

104 J. Lee, J. McIntosh, B. J. Hathaway and E. W. Schmidt, *J. Am. Chem. Soc.*, 2009, **131**, 2122–2124.

105 S. J. Pidot, H. Hong, T. Seemann, J. L. Porter, M. J. Yip, A. Men, M. Johnson, P. Wilson, J. K. Davies, P. F. Leadlay and T. P. Stinear, *BMC Genomics*, 2008, **9**, 462.