VERSITA

Biologia **68**/6: 1079—1086, 2013
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-013-0246-7

Review

Bioprospecting microbial metagenome for natural products

Jana Nováková & Marián Farkašovský*

Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, SK-84551 Bratislava, Slovakia; e-mail: marian.farkasovsky@savba.sk

Abstract: Mining of natural sources for new secondary metabolites has a successful history, which is reflected by the fact that over 50% of all drugs, currently on the market, are derived from natural products. Bacteria are one of the most important sources of bioactive natural products destined for drug discovery. However, less than 1% of the microorganisms observed in different habitats have been cultivated and characterized. To explore the genomic and functional diversity of the vast majority of the microbial world, novel methods were introduced, which are based on analysis of a DNA isolated from environmental communities. Metagenomics represents a strategy offering access to the genetic information present in uncultured bacteria by screening of libraries constructed from DNA isolated from different habitats. Functional- and sequence-driven screens are the major approaches employed to mine metagenomic libraries. This review aims to highlight discoveries in this area and discusses the possible future directions of the field.

Key words: metagenomics; metagenomic library; natural products; secondary metabolites; unculturable bacteria.

Abbreviations: BAC, bacterial artificial chromosome; eDNA, environmental DNA; NGS, next generation sequencing; NRPS, non-ribosomal peptide synthase; PKS, polyketide synthase.

Introduction

Natural products, generated by bacteria, fungi, marine organisms and plants, are widely used in medicine and agriculture. Although synthetic molecules are rapidly created by combinatorial chemistry, they show lower chemical diversity than natural compounds. Prokaryotic cells have been identified as the largest source of biologically active compounds with antiviral, antibacterial, antifungal and anticancer activities. The total number of bacteria has been estimated to $4–6\times10^{30}$ (Whitman et al. 1998). Studies of different types of habitats indicate that more than 99% of microorganisms (Amann et al. 1995), which may include up to 10^9 different species (Schloss & Handelsman 2004), cannot be cultivated by routine techniques. Soil is probably the largest reservoir of all natural ecosystems with respect of diversity of microbial species. One gram of soil contains an estimated 10^7–10^{10} prokaryotic cells (Rossello-Mora & Amann 2001), which correspond up to the 10^6 different genomes (Torsvik et al. 1996, 2002; Gans et al. 2005). The enormous diversity of uncultured microorganisms in soil and other environmental sources provides a rich repository of novel natural products. Resistance against new antibiotics is increasing dramatically, leading to a return of diseases, such as tuberculosis that have almost vanished in industrialized countries. In addition, currently used drugs (e.g. anticancer) often lack high specificity for their targets. In view of these facts, it is evident that the new drug screening technologies applied to uncultured bacteria became extraordinarily important.

Screening strategies for new natural products in cultured microorganisms

Since the discovery of streptothricine and streptomycin in 1940's, roughly 70,000 new natural products of microbial origin have been described (Bérdy 2012). In the last two decades screens have been intensified due to high-throughput techniques. Different academic labs and companies were primarily focused on screening of large collection of actinomycetes, isolated from different geographical origins, to evaluate the patterns of the production of secondary metabolites under alternative conditions. By including of the dereplication tools into the early stages of the high-throughput screens, success in detection of novel compounds has been increased by exclusion of already known natural products. The discovery of platensimycin, a previously unknown class of antibiotics produced by *Streptomyces platensis*, was the most important result of these screens. This compound with no observed toxicity exhibits a broad-spectrum Gram-positive antibacterial effect through the selective targeting of FabF/B condensing enzymes in the synthetic pathway of fatty acids.

* Corresponding author

Springer

Because of its novel mode of action, platensimycin shows no cross-resistance to other antibiotic-resistant strains tested, including methicillin-resistant *Staphylococcus aureus*, vancomycin-intermediate *S. aureus*, vancomycin-resistant enterococci and highly pathogenic *Mycobacterium tuberculosis* (Basilio et al 2003; Wang et al. 2006). Platensimycin and the similar compound platencin have some problems with regard to their pharmacokinetic properties (high rate of clearance), which emphasizes the need for modification of chemical structure to enhance the *in vivo* stability without affecting of activity. Another examples of the novel molecules discovered in the actinomycete screens are thiazolyl peptides thiazomycin from *Amycolatopsis fastidiosa* and philipimycin from *Actinoplanes phillipinensis*, or the anthelmintic macrolactams 6-desmethyl-N methylfluvirucin A1 and N-methylfluvirucin A1 found in *Nonomuraea turkmeniaca* (Ayers et al. 2007; Jayasuriya et al. 2007; Zhang et al. 2008). The new protein synthesis inhibitors lucensimycins and okilactomycins were isolated from *Streptomyces* spp. using an antisense strategy (Singh et al. 2009; Zhang et al. 2009). Recently, the screening of actinomycetes cultured from marine sponges revealed novel compound JBIR-23 which inhibits tumour growth *in vivo* and promote tubulin polymerization analogous to the taxanes and epothilones, but possesses no structural resemblance to these compounds (Takagi & Shin-Ya 2012).

Metagenomics as a strategy for discovery of natural products

The term "unculturable" does not necessarily mean the microbes cannot be cultured at all; therefore during the past two decades significant efforts have been made to develop new approaches to cultivate the uncultured majority of the microbial world. These strategies include the use of modified media, changes of growth conditions, community culture and co-culture, use of transwell plates, simulated natural environments using diffusion chambers, high-throughput microbioreactor and laser microdissection (reviewed in Pham & Kim 2012; Stewart 2012). Despite of the achievements in this field, the majority of environmental bacteria still remain recalcitrant to culturing. Considering the absence of cultivation techniques for the most of microorganisms, a new method was developed to cover the entire microbial diversity from different biotopes. This cultivation-independent approach is based on construction of complex libraries from environmental DNA (eDNA) in heterologous host, e.g. *Escherichia coli* (Fig. 1). Isolated microbial DNA from a particular habitat constitutes collective genome of all present microorganisms, known as metagenome (Handelsman et al. 1998). The tremendous diversity of uncultured microorganisms in soil and other environments constitutes a rich reservoir of undiscovered natural products, most prominent of them are represented by linear or cyclic polyketides and peptides. These compounds are synthesized by large multifunctional enzymes, known as polyketide synthases (PKSs) and non-ribosomal peptide synthases (NRPSs) using a simple building blocks of carboxylic acid and amino acid. Bacteria produce also other types of chemical compounds, e.g. carbohydrates and their derivatives, terpenoids, derivatives of amino acids, etc.

eDNA isolation

The construction of complex libraries derived from environmental samples is dependent on the high-quality of extracted DNA, as the enzymatic modifications required in the cloning procedure are sensitive to contamination by diverse compounds, such as humic and fulvic acids. The methods described for eDNA isolation from soil and sediment samples can be generally classified into direct and indirect isolation procedures. Direct DNA isolation is based on physical and/or chemical lysis of the cells within the sample matrix and subsequent separation of DNA from the matrix and cell debris (Zhou et al. 1996; Brady 2007). The indirect approach involves the separation of cells from the soil matrix using different centrifugation methods followed by cell lysis and DNA extraction. Alternatively, separated cells are embedded in agarose blocks, treated with enzymes and detergents and the DNA is finally purified with pulsed field gel electrophoresis (Liles et al. 2008). eDNA from aquatic ecosystems (sea, river, hot spring, etc.) can also be isolated by direct or indirect approach. Different filtration devices have been employed to concentrate microorganisms from the liquid samples and at the same time to separate them from large particles. However, this procedure may also remove subsets of the existing microbial populations living in attached fractions of the water (Zinger et al. 2012).

Small-insert and large-insert libraries

Two types of libraries can be constructed with respect to average insert size. Small-insert libraries in plasmid vectors, bearing less than 10 kb insert, are useful for the isolation of single genes or small operons encoding enzymes or small enzyme complexes. Overview on new biocatalysts that originate from metagenomic libraries can be found elsewhere (Fernández-Arrojo et al. 2010; Kennedy et al. 2011). The eDNA is generally isolated by the direct lysis for this type of libraries. Genes encoding enzymes required for biosynthesis of secondary metabolites are clustered on a coherent piece of DNA, which ranges in size from approximately 20 to 100 kb (Hopwood 1999; Zazopoulos et al. 2003). Clustering of biosynthetic pathways of natural products facilitates to clone an entire pathway into different vectors like fosmids, for the cloning of environmental genomic DNA of around 40 kb in size, or into bacterial artificial chromosome (BAC) vectors for the cloning of DNA fragments larger than 100 kb (Fig. 1). Both types of vectors, which are based on F factor, are stably maintained in *E. coli* at one or two copies per cell (Shizuya et al. 1992).

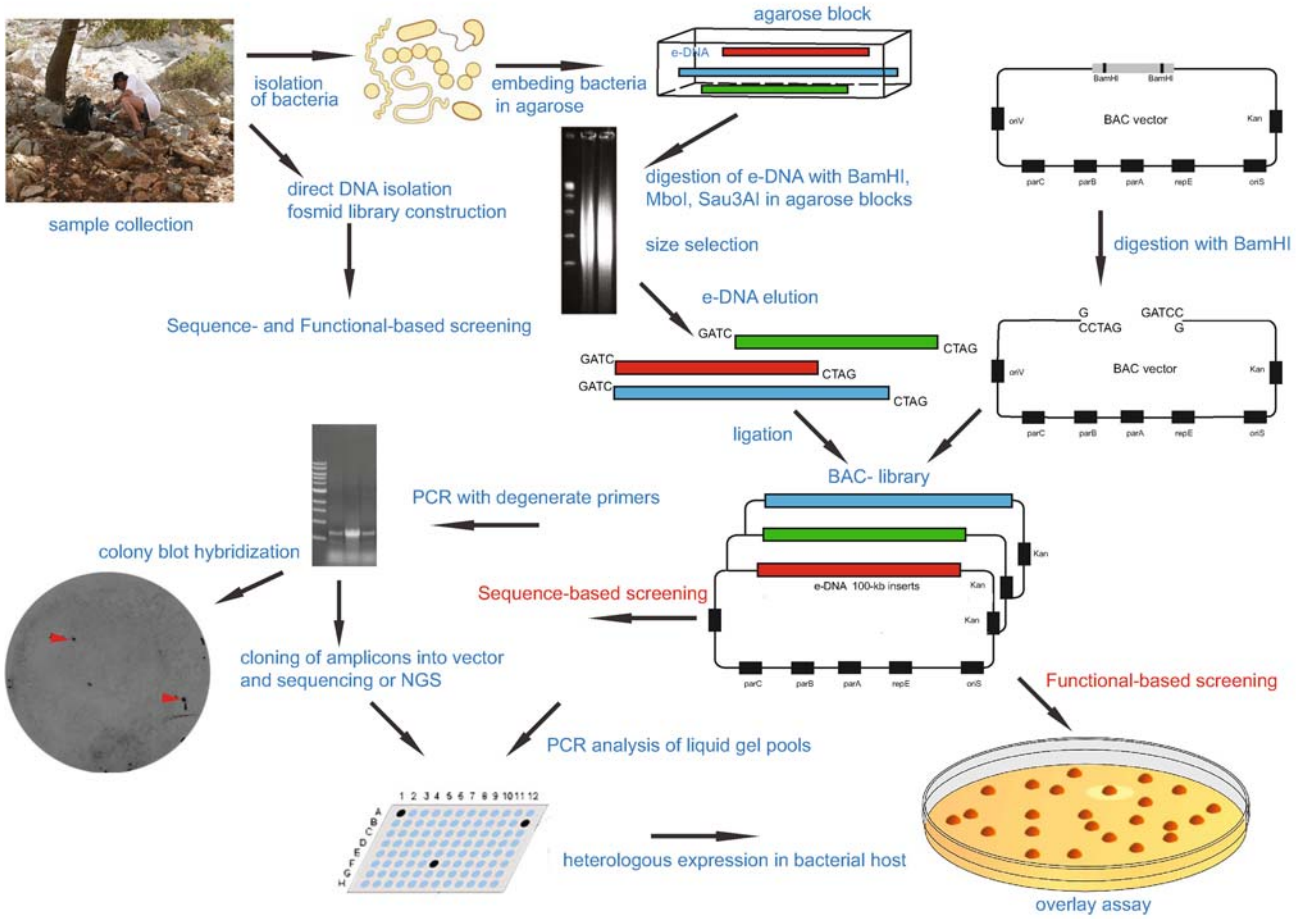

Fig. 1. Overview of metagenomic library construction and screening of natural products (for details, see text).

Functional-based approach of metagenomic library screening

Currently, two different strategies are used to identify biocatalysts and secondary metabolites from metagenomic libraries: activity-based and sequence-based screening. In the first approach, the functional gene products are detected upon biological activity and further characterized by molecular-biological and biochemical methods. This strategy allows quick identification of new compounds and enzymes. One drawback is requirement of sensitive analytical method and dependence on expression of cloned gene in the foreign host.

Overlay assay represents the most straightforward screen based on the detection of clones exhibiting inhibition zones around clones growing on bacterial (Rondon et al. 2000; Wang et al. 2000) or fungal (Chung et al. 2008) test lawns (Fig. 1). Pigmentation observed by a visual screening of libraries can also indicate the presence of secondary metabolite as observed in the case of the blue violaceins (Brady et al. 2001), indigo and indirubin (Lim et al. 2005) or brown melanin (Craig et al. 2010). One of the first novel natural products discovered by hemolytic screens of soil metagenomic library in BAC vector was the red broad-spectrum antibiotic turbomycin B (Fig. 2) (Gillespie et al. 2002). The *N*-acyl amino acid derivatives (Brady & Clardy 2000; Brady et al. 2002), palmitoylputrescine (Brady & Clardy 2004) and isocyanide (Brady & Clardy 2005) are further examples of natural product identified by inhibition zone-based screens of environmental libraries. The more laborious approach of the functional screen employed detection of natural products in extracts by high-performance liquid chromatography-mass spectrometry (HPLC-MS) analysis. This strategy was successfully used for the identification of biosynthetic genes for patellamide D (Fig. 2) and ascidiacyclamide in a BAC library constructed from the DNA of uncultivated *Lissoclinum patella* photosynthetic prokaryotic symbiont, *Prochloron* sp. (Long et al. 2005).

The use of standard *E. coli* strains as the host might, however, limit the ability to express DNA from environmental microorganisms. Requirement for post-translational modifications of biosynthetic protein complexes (e.g. the addition of phosphopantetheine groups), absence of biosynthetic precursors in *E. coli* (methylmalonyl-CoA, a building block for polyketide biosynthesis), compatibility with transcription (absence of particular sigma factors of RNA polymerase) and translation apparatus (different codon usage) are only a few examples of possible hurdles for successful heterologous expression in *E. coli*. To overcome these expression-related limitations, several non-*E. coli*

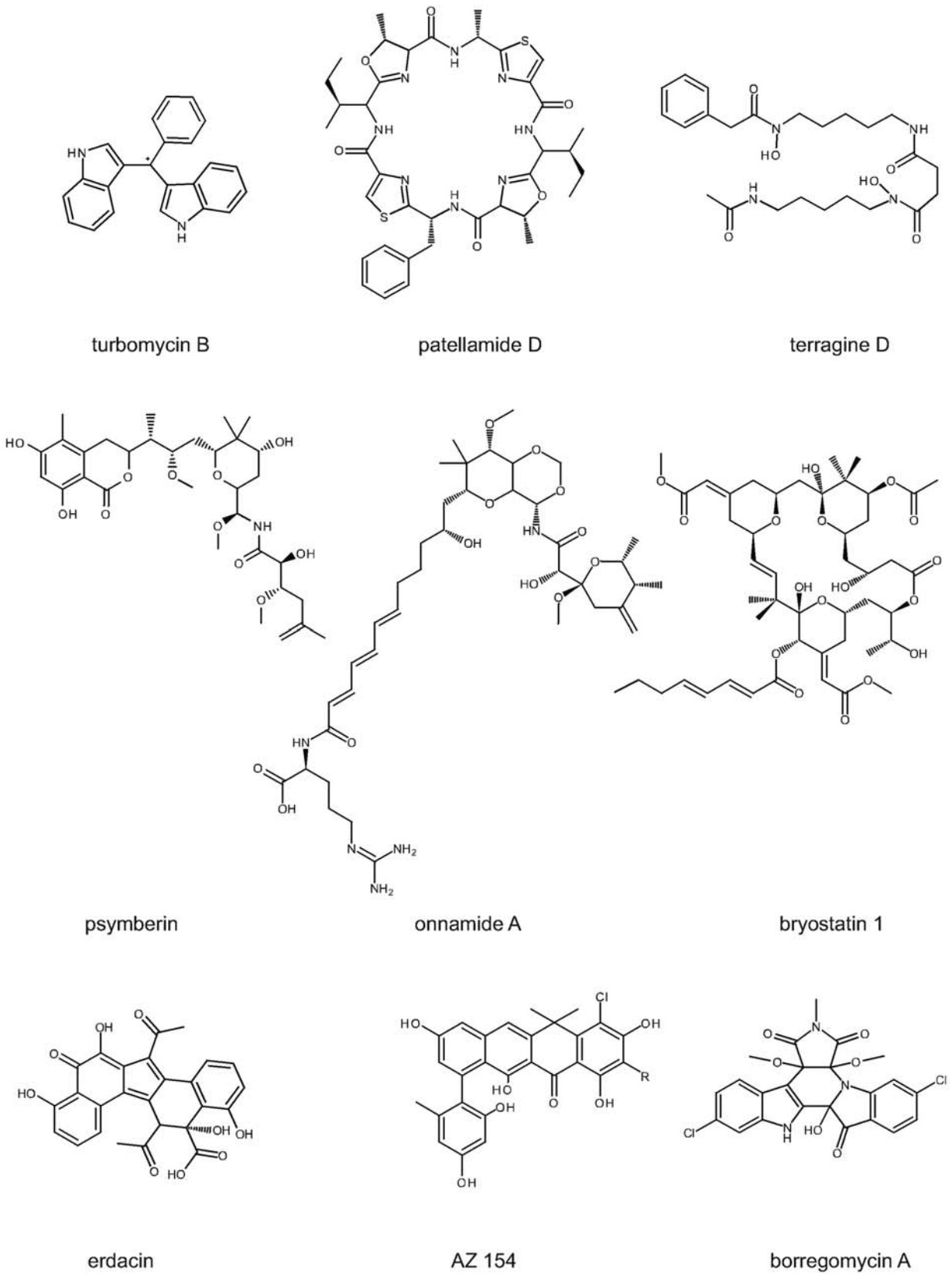

Fig. 2. Representative secondary metabolites discovered by screens of metagenomic libraries.

host expression systems have been developed, including *Streptomyces lividans* (Wang et al. 2000; Courtois et al. 2003; Martinez et al. 2004), *Streptomyces albus* (King et al. 2009; Bauer et al. 2010; Feng et al. 2010), *Pseudomonas putida* (Martinez et al. 2004; Brady et al, 2007; Craig et al. 2010), *Agrobacterium tumefaciens* (Craig et al. 2010), *Burkholderia graminis* (Craig et al. 2010), *Caulobacter vibrioides* (Craig et al. 2010) and

Ralstonia metallidurans (Craig et al. 2009, 2010). Functional screen of these strains yielded terragines A-E (Fig. 2), fatty dienic alcohol isomers (Wang et al. 2000; Courtois et al. 2003), N-acyl amino acids and several polyketides (Craig et al. 2009, 2010).

Sequence-based approach of metagenomic library screening

Sequence-based approach employs PCR or colony blot hybridization with DNA probes or primers specific to sequences of genes coding target proteins. These sequences are derived from conserved regions of already known genes, which represents at the same time limitation of this method, because only gene variants of known protein families can be identified. Nevertheless, this approach allows finding clusters of genes coding for new secondary metabolites regarding the high degree of structural variability of polyketides and NRPS. Considerable advantage of sequence-based approach is the independence on expression of cloned genes. Although homology-based screens are more labour intensive than activity-based screens, efficient methods have been developed allowing for the rapid isolation of clones from libraries. One possibility is to use the library DNA as a template for PCR with degenerate primers designed to recognize the most conserved DNA regions found in the PKS and NRPS genes. Amplicons are then cloned and sequenced or directly used for next generation sequencing (NGS) methods. PCR product can also be labelled and employed in colony hybridization to identify individual clones in the library. In another approach, unbiased clone pools are generated in the liquid agarose medium and screened by whole-cell PCR by using the same degenerated primers or by primers derived from sequenced amplicons. Individual positive clones are then isolated by several rounds of dilution and PCR screening (Hrvatin & Piel 2007). Highly viscose soft gel allows amplifying plasmid libraries in a strictly representative fashion, decreasing the possibility that less abundant clones disappear during amplification of the library pools due to differential rates of replication (Fig. 1).

Many natural products were found in symbiotic microorganisms of plants (Firáková et al. 2007), insects (Bode 2009) and marine invertebrates (Leal et al. 2012). These systems are very convenient for application of the metagenomic approach because symbionts are reproducibly found in defined habitats represented by the hosts and display rather limited diversity. Symbionts have been co-evolved for millions of years with their eukaryotic hosts and developed very complicated relations. To date is not really known how many of the natural products originate from the host or from the microorganisms. However, results from last decade have indicated that symbionts could be the true source of many natural products isolated from these complex systems. Pederin was found in *Paederus* spp. beetles, which contain one dominant endosymbiont of the genus *Pseudomonas*. The limited metagenomic complexity allowed allocating of the isolated pederin biosynthetic genes to the *Pseudomonas* symbiont (Piel 2002). Other examples of natural products generated by symbiotic microorganisms are cytotoxic polyketides onnamide A (Fig. 2) from the sponge *Theonella swinhoei* (Piel et al. 2004), rhizoxin A from intracellular *Burkholderia* spp. symbionts of *Rhizopus* spp. fungi (Partida-Martinez & Hertweck 2007), psymberin (Fig. 2) from a symbiont of the sponge *Psammocinia* sp. aff. *bulbosa* (Fisch et al. 2009), and pateamine and mycalamide A from the sponge *Mycale hentscheli* (Fisch et al. 2009). In order to reduce the DNA complexity, different centrifugation and filtration techniques have been developed for presorting of particular cells. This approach resulted in the discovery of the barbamide from the free-living cyanobacterium *Lyngbya majuscule* (Chang et al. 2002), dysidin and dysidenin in symbiotic cyanobacterium *Oscillatoria spongeliae* of the sponge *Lamellodysidea herbacea* (Flatt et al. 2005) or anticancer drug candidate bryostatin 1 (Fig. 2) produced by the γ-proteobacterial symbiont "*Candidatus* Endobugula sertula" of the bryozoan *Bugula neritina* (Sudek et al. 2007). Non-*E. coli* hosts are frequently employed in heterologous expression of positive clones obtained by sequence-driven screens. PCR screening of a soil metagenomic library with degenerate primers (type II PKS), followed by the expression of biosynthetic genes in *S. lividans* and *S. albus* resulted in the novel natural products erdacin (Fig. 2) (King et al. 2009) and utahmycins A and B (Bauer et al. 2010), which exhibit novel carbon skeletons. Another soil metagenomic library was screened with primers specific for type I PKS genes (Courtois et al. 2003). The cosmids from positive clones were transferred into *S. lividans* and the resulting recombinant clones expressed the novel dienols. In the third example the PKS construct obtained by PCR screen from the multimillion soil DNA library was expressed in *S. albus*, which yielded in three novel antibiotics of the fluostatin group (Feng et al. 2010).

Feng et al. (2011) reported another successful sequence-based identification and heterologous expression of type II PKS. Three soil-derived PKS e-DNA cosmids in *S. albus* produced well-known antibiotic landomycin E, previously uncharacterized pentacyclic ring system, and unique KB-3346-5 derivatives (AZ 154; Fig. 2) which show activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecalis*. Very recently, a unique tryptophan dimerization biosynthetic gene cluster was identified within a large soil DNA library. When heterologously expressed in *S. albus*, this cluster produced an indolotryptoline antiproliferative agent borregomycin A with CaMKIIδ kinase inhibitory activity (Fig. 2), along with several dihydroxyindolocarbazole anticancer/antibiotics (borregomycins B-D) (Chang & Brady 2013).

Conclusions and future prospects

Although novel natural products, frequently synthe-

sized by new enzymology, have been revealed in activity-driven screens of metagenomic libraries, the frequency of the identified compounds is relatively low. In several reports (Brady & Clardy 2000; Brady et al. 2004) it was described that 1 in every 10,000-20,000 eDNA cosmid clones displays an antibacterial activity. One reason for the low detection rate is the huge size of the most clusters of polyketide and NRPS genes, which frequently exceeds the cloning capacity of fosmids (40 kb). Improved BAC cloning technologies resulting in libraries of millions of independent clones and insert size over 100 kb would substantially improve the discovery rate. Another point is that many gene clusters are not expressed in *E. coli* because of the reasons described above. The development of alternative library hosts is therefore crucial for better efficiency of activity-based natural product screening. *Streptomyces*, the genus exhibiting the ability to produce a wide variety of secondary metabolites, are the most common bacteria used for heterologous expression of natural products. However, the expression level varies substantially between different species or may even fail completely, reflecting the high diversity and complexity of secondary metabolism regulation (for a review, see van Wezel & McDowall 2011). Recently, cyclic AMP receptor protein (Crp) (Gao et al. 2012), chemical inhibition of fatty acid synthesis (Craney et al. 2012) and RNA polymerase β-subunit mutants (Tanaka et al. 2013) were identified as new global regulatory elements of secondary metabolism and they represent promising tools to better utilize the potential of actinomycetes in high-throughput screens. Sequence-based methods are a good way of overcoming the lack of heterologous expression in *E. coli*, but the size issue remains. As already mentioned, many fosmid clones do not comprise the entire cluster of biosynthetic genes. Nevertheless, functional genes are often distributed over several library clones. Transformation-assisted recombination represents the best method to join multiple overlapping linear DNA fragments in a single transformation step. Most of the homology-based screens have been performed using degenerate primers derived from conserved sequences found in PKS or NRPS adenylation domain. In order to reach a wider structural diversity of these compounds, designing of primers should also focus on tailoring genes involved in reactions, such as oxygenation, cyclization, glycosylation or halogenation (Banik & Brady 2008).

Diverse eDNA enrichment strategies have been developed in order to simplify the screening of large libraries for genes of interest. The fluorescence *in situ* hybridization with cell sorting (Kalyuzhnaya et al. 2006) and subtractive hybridization (Chew & Holmes 2009) belong to these in particular. Another approach, single cell genomics, which has appeared in the foreground of interest of microbiologists in the last few years, may also find application in the discovery of natural products. In this strategy, flow cytometric sorting of bacteria or microfluidic device is used for single cell isolation, then the genome is amplified by multiple displacement amplification followed by the screening of the DNA samples with primers and NGS. Currently, the first report that applied the single cell genomics to the localization of natural product genes described the discovery of PKS in the unusual candidate phylum "Poribacteria" (Fieseler et al. 2004), which are ubiquitously associated with sponges (Siegl et al. 2011).

Accessibility to enormous amounts of sequencing data at reasonable cost makes NGS very attractive alternative for eDNA gene discovery (Shokralla et al. 2012). Although the sequence data reside on short contigs, it has also been possible to reconstruct near-complete genomes from low-diversity environments like acidophilic biofilm (Tyson et al. 2004). Despite the progress in sequencing methodologies, for the complex metagenomes, like those originating from soil, systematic discovery of entire sequences of large biosynthetic cluster by NGS sequencing is practically not feasible today (Mende et al. 2012). NGS can be very useful in sequence-based approach as it has been demonstrated recently by deep sequencing of amplicons obtained using degenerate primers and eDNA or library as a template. This strategy uncovered more than 50% of the unique PKS and NRPS domains that are accessible from metagenome (Reddy et al. 2012; Woodhouse et al. 2013).

Although until now only a small number of secondary metabolites have been identified and characterized using metagenomic approach, uncultured bacteria are very promising source of novel natural products as obvious from the studies of the last ten years. However, several technical and conceptual challenges remain unresolved and only systematic approach integrating researchers from different fields will allow us to address these issues effectively.

Acknowledgements

We gratefully acknowledge the financial support from the Alexander von Humboldt Foundation research group linkage program (3.3 DEU7/1152594).

References

Amann R.I., Ludwig W. & Schleifer K.H. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. **59:** 143–169.

Ayers S., Zink D.L., Mohn K., Powell J.S., Brown C.M., Murphy T., Grund A., Genilloud O., Salazar O., Thompson D. & Singh S.B. 2007. Anthelmintic macrolactams from *Nonomuraea turkmeniaca* MA7381. J. Nat. Prod. **70:** 1371–1373.

Banik J.J. & Brady S.F. 2008. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc. Natl. Acad. Sci. USA **105:** 17273–17277.

Basilio A., González I., Vicente M.F., Gorrochategui J., Cabello A., González A. & Genilloud O. 2003. Patterns of antimicrobial activities from soil actinomycetes isolated under different conditions of pH and salinity. J. Appl. Microbiol. **95:** 814–823.

Bauer J.D., King R.W. & Brady S.F. 2010. Utahmycins A and B, azaquinones produced by an environmental DNA clone. J. Nat. Prod. **73:** 976–979.

Bérdy J. 2012. Thoughts and facts about antibiotics: where we are now and where we are heading. J. Antibiot. **65:** 385–395.

Bode H. 2009. Insects: true pioneers in anti-infective therapy and what we can learn from them. Angew. Chem. Int. Ed. **48:** 2–5.

Brady S.F. 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nat. Protoc. **2:** 1297–1305.

Brady S.F., Bauer J.D., Clarke-Pearson M.F. & Daniels R. 2007. Natural products from *isnA* containing biosynthetic gene clusters recovered from the genomes of cultured and uncultured bacteria. J. Am.Chem. Soc. **129:** 12102–12103.

Brady S.F., Chao C.J. & Clardy J. 2002. New natural product families from an environmental DNA (eDNA) gene cluster. J. Am. Chem. Soc. **124:** 9968–9969.

Brady S.F., Chao C.J. & Clardy J. 2004. Long-chain *N*-acyltyrosine synthases from environmental DNA. Appl. Environ. Microbiol. **70:** 6865–6870.

Brady S.F., Chao C.J., Handelsman J. & Clardy J. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org. Lett. **3:** 1981–1984.

Brady S.F. & Clardy J. 2000. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. J. Am. Chem. Soc. **122:** 12903–12904.

Brady S.F. & Clardy J. 2004. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. J. Nat. Prod. **67:** 1283–1286.

Brady S.F. & Clardy J. 2005. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. Angew. Chem. Int. Ed. **44:** 7063–7065.

Chang F.Y. & Brady S.F. 2013. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. Proc. Natl. Acad. Sci. U S A. **110:** 2478–2483.

Chang Z.X., Flatt P., Gerwick W.H., Nguyen V.A.,Willis C.L. & Sherman D.H. 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-nonribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. Gene **296:** 235–247.

Chew Y.V. & Holmes A.J. 2009. Suppression subtractive hybridisation allows selective sampling of metagenomic subsets of interest. J. Microbiol. Methods **78:** 136–143.

Chung E.J., Lim H.K., Kim J.C., Choi G.J., Park E.J., Lee M.H., Kim J.C., Choi G.J., Cho K.Y. & Lee S.W. 2008. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. Appl. Environ. Microbiol. **74:** 723–730.

Courtois S., Cappellano C.M., Ball M., Francou F.X., Normand P., Helynck G., Martinez A., Kolvek S.J., Hopke J., Osburne M.S., August P.R., Nalin R., Guérineau M., Jeannin P., Simonet P. & Pernodet J.L. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. **69:** 49–55.

Craig J.W., Chang F.Y. & Brady S.F. 2009. Natural products from environmental DNA hosted in *Ralstonia metallidurans*. ACS Chem. Biol. **4:** 23–28.

Craig J.W., Chang F.Y., Kim J.H., Obiajulu S.C. & Brady S.F. 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl. Environ. Microbiol. **76:** 1633–1641.

Craney A., Ozimok C., Pimentel-Elardo S.M., Capretta A. & Nodwell J.R. 2012. Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. Chem. Biol. **19:** 1020–1027.

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

Feng Z.Y., Kim J.H. & Brady S.F. 2010. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. J. Am. Chem. Soc. **132:** 11902–11903.

Fernández-Arrojo L., Guazzaroni M.E., López-Cortés N., Beloqui A. & Ferrer M. 2010. Metagenomic era for biocatalyst identification. Curr. Opin. Biotechnol. **21:** 725–733.

Fieseler L., Horn M., Wagner M. & Hentschel U. 2004. Discovery of the novel candidate phylum "Poribacteria" in marine sponges. Appl. Environ. Microbiol. **70:** 3724–3732.

Firáková S., Šturdíková M. & Múčková M. 2007. Bioactive secondary metabolites produced by microorganisms associated with plants. Biologia **62:** 251–257.

Fisch K.M., Gurgui C., Heycke N., van der Sar S.A., Anderson S.A., Webb V.L., Taudien S., Platzer M., Rubio B.K., Robinson S.J., Crews P. & Piel J. 2009. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. **5:** 494–501.

Flatt P., Gautschi J., Thacker R., Musafija-Girt M., Crews P. & Gerwick W. 2005. Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea* (*Lamellodysidea*) *herbacea* and symbiotic cyanobacterium *Oscillatoria spongeliae* by CARD-FISH analysis. Mar. Biol. **147:** 761–774.

Gans J., Wolinsky M. & Dunbar J. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science **309:** 1387–1390.

Gao C., Hindra, Mulder D., Yin C. & Elliot M.A. 2012. Crp is a global regulator of antibiotic production in *Streptomyces*. mBio **3:** e00407–12.

Gillespie D.E., Brady S.F., Bettermann A.D., Cianciotto N.P., Liles M.R., Rondon M.R., Clardy J., Goodman R.M. & Handelsman J. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. **68:** 4301–4306.

Handelsman J., Rondon M.R., Brady S.F., Clardy J. & Goodman R.M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. **5:** R245–R249.

Hopwood D.A. 1999. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. Microbiology **145:** 2183–2202.

Hrvatin S. & Piel J. 2007. Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. J. Microbiol. Methods **68:** 434–436.

Jayasuriya H., Herath K., Ondeyka J.G., Zhang C., Zink D.L., Brower M., Gailliot F.P., Greene J., Birdsall G., Venugopal J., Ushio M., Burgess B., Russotti G., Walker A., Hesse M., Seeley A., Junker B., Connors N., Salazar O., Genilloud O., Liu K., Masurekar P., Barrett J.F. & Singh S.B. 2007. Isolation and structure elucidation of thiazomycin. A potent thiazolyl peptideantibiotic from *Amycolatopsis fastidiosa*. J. Antibiot. **60:** 554–564.

Kalyuzhnaya M.G., Zabinsky R., Bowerman S., Baker D.R., Lidstrom M.E. & Chistoserdova L. 2006. 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. **72:** 4293–4301.

Kennedy J., O'Leary N.D., Kiran G.S., Morrissey J.P., O'Gara F., Selvin J. & Dobson A.D. 2011. Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. J Appl. Microbiol. **111:** 787–799.

King R.W., Bauer J.D. & Brady S.F. 2009. An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. Angew. Chem. Int. Ed. **48:** 6257–6261.

Leal M.C., Puga J., Serôdio J., Gomes N.C.M., Calado R. 2012. Trends in the discovery of new marine natural products from invertebrates over the last two decades – where and what are we bioprospecting? PLoS ONE **7:** e30580.

Liles M.R., Williamson L.L., Rodbumrer J., Torsvik V., Goodman R.M. & Handelsman J. 2008. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. Appl. Environ. Microbiol. **74:** 3302–3305.

Lim H.K., Chung E.J., Kim J.C., Choi G.J., Jang K.S., Chung Y.R., Cho K.Y. & Lee S.W. 2005. Characterization of a forest

soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Appl. Environ. Microbiol. **71:** 7768–7777.

Long P.F., Dunlap W.C., Battershill C.N. & Jaspars M. 2005. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. ChemBioChem. **6:** 1760–1765.

Martinez A., Kolvek S.J., Yip C.L.T., Hopke J., Brown K.A., MacNeil I.A. & Osburne M.S. 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl. Environ. Microbiol. **70:** 2452–2463.

Mende D.R., Waller A.S., Sunagawa S., Järvelin A.I., Chan M.M., Arumugam M., Raes J. & Bork P. 2012. Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS ONE **7:** e31386.

Partida-Martinez L.P. & Hertweck C. 2007. A gene cluster encoding rhizoxin biosynthesis in "Burkholderia rhizoxina", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. ChemBioChem. **8:** 41–45.

Pham V.H. & Kim J. 2012. Cultivation of unculturable soil bacteria. Trends Biotechnol. **30:** 475–484.

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

Piel J., Hui D.Q., Fusetani N. & Matsunaga S. 2004. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. Environ. Microbiol. **6:** 921–927.

Reddy B.V., Kallifidas D., Kim J.H., Charlop-Powers Z., Feng Z. & Brady S.F. 2012. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Appl. Environ. Microbiol. **78:** 3744–3752.

Rondon M.R., August P.R., Bettermann A.D., Brady S.F., Grossman T.H., Liles M.R., Loiacono K.A., Lynch B.A., MacNeil I.A., Minor C., Tiong C.L., Gilman M., Osburne M.S., Clardy J., Handelsman J. & Goodman R.M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. **66:** 2541–2547.

Rossello-Mora R. & Amann R. 2001. The species concept for prokaryotes. FEMS Microbiol. Rev. **25:** 39–67.

Schloss P.D. & Handelsman J. 2004. Status of the microbial census. Microbiol. Mol. Biol. Rev. **68:** 686–691.

Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y. & Simon M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA **89:** 8794–8797.

Shokralla S., Spall J.L., Gibson J.F. & Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. **21:** 1794–1805.

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

Singh S.B., Zink D.L., Dorso K.L., Motyl M.R., Salazar O., Basilio A., Vicente M.F., Byrne K.M., Ha S.N. & Genilloud O. 2009. Isolation, structure and antibacterial activities of lucensimycins D-G, discovered from *Streptomyces lucensis* MA7349 using an antisense strategy. J. Nat. Prod. **72:** 345–352.

Stewart E.J. 2012. Growing unculturable bacteria. J. Bacteriol. **194:** 4151–4160.

Sudek S., Lopanik N.B., Waggoner L.E., Hildebrand M., Anderson C., Liu H., Patel A., Sherman D.H. & Haygood M.G. 2007. Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus* Endobugula sertula", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. J. Nat. Prod. **70:** 67–74.

Takagi M. & Shin-Ya K. 2012. Construction of a natural product library containing secondary metabolites produced by actinomycetes. J. Antibiot. **65:** 443–447.

Tanaka Y., Kasahara K., Hirose Y., Murakami K., Kugimiya R. & Ochi K. 2013. Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (rpoB) mutations in actinomycetes. J. Bacteriol. **195:** 2959–2970.

Torsvik V., Daae F. L., Sandaa R.A. & Øvreås L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. **5:** 240–245.

Torsvik V., Sorheim R. & Goksoyr J. 1996. Total bacterial diversity in soil and sediment communities. J. Ind. Microbiol. **17:** 170–178.

Tyson G.W., Chapman J., Hugenholtz P., Allen E.E., Ram R.J., Richardson P.M., Solovyev V.V., Rubin E.M., Rokhsar D.S. & Banfield J.F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:** 37–43.

van Wezel G. P. & McDowall K. J. 2011. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. Nat. Prod. Rep. **28:** 1311–1333.

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

Wang J., Soisson S.M., Young K., Shoop W., Kodali S., Galgoci A., Painter R., Parthasarathy G., Tang Y.S., Cummings R., Ha S., Dorso K., Motyl M., Jayasuriya H., Ondeyka J., Herath K., Zhang C., Hernandez L., Allocco J., Basilio A., Tormo J.R., Genilloud O., Vicente F., Pelaez F., Colwell L., Lee S.H., Michael B., Felcetto T., Gill C., Silver L.L., Hermes J.D., Bartizal K., Barrett J., Schmatz D., Becker J.W., Cully D. & Singh S.B. 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature **441:** 358–361.

Whitman, W.B., Coleman D.C. & Wiebe W.J. 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA **95:** 6578–6583.

Woodhouse J. N., Fan L., Brown M.V., Thomas T. & Neilan B.A. 2013. Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. ISME J. (in press); doi:10.1038/ismej.2013.65.

Zazopoulos E., Huang K., Staffa A., Liu W., Bachmann B.O., Nonaka K., Ahlert J., Thorson J.S., Shen B. & Farnet, C.M. 2003. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nature Biotechnol. **21:** 187–190.

Zhang C., Ondeyka J.G., Zink D.L., Basilio A., Vicente M.F., Salazar O., Genilloud O., Dorso K.L., Motyl M.R., Byrne K.M. & Singh S.B. 2009. Discovery of okilactomycin and congeners from *Streptomyces scabrisporus* by antisense differential sensitivity assay targeting ribosomal protein S4. J. Antibiot. **62:** 55–61.

Zhang C.W., Occi J., Masurekar P., Barrett J.F., Zink D.L., Smith S., Onishi R., Ha S.H., Salazar O., Genilloud O., Basilio A., Vicente F., Gill C., Hickey E.J., Dorso K., Motyl M. & Singh S.B. 2008. Isolation, structure, and antibacterial activity of philipimycin, a thiazolyl peptide discovered from *Actinoplanes philippinensis* MA7347. J. Am. Chem. Soc. **130:** 12102–12110.

Zhou J., Bruns M.A. & Tiedje J.M. 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. **62:** 316–322.

Zinger L., Gobet A. & Pommier T. 2012. Two decades of describing the unseen majority of aquatic microbial diversity. Mol. Ecol. **21:** 1878–1896.

Received May 20, 2013
Accepted August 9, 2013