UNIVERSITY of the WESTERN CAPE

Efficient molecular cloning of environmental DNA from geothermal sediments

Dianna E. Wilkinson, Thomas Jeanicke & Don A. Cowan*

Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E6BT, UK

** Current address and author for correspondence: Department of Microbiology, University of the Western Cape, Bellville 7535, Cape Town, South Africa (Fax: +27 219593505; E-mail: dcowan@uwc.ac.za)*

Key words: culture-independent cloning, environmental DNA library, geothermal sediment, metagenome

Abstract

An efficient and simple method for constructing an environmental library using mechanically sheared DNA obtained directly from geothermal sediments is presented. The method is based on blunt-end modification of DNA fragments followed by 3'-adenylation using Vent DNA polymerase and *Taq* DNA polymerase, respectively. The prepared DNA fragments are then ligated into a TA cloning vector and used in the transformation of *Escherichia coli.* This method has been successfully applied to the cloning of ORFs derived from uncultivated prokaryotes present in geothermal sediment

Introduction

Recent progress in genomic analyses of uncultured microorganisms demonstrates that heterologous expression of environmental DNA can be achieved in *E. coli* hosts (Cottrell *et al.* 1999, Henne *et al.* 1999, 2000, Rondon *et al.* 2000, Brady & Clardy 2000, Brady *et al.* 2001). With this approach, DNA was directly isolated from an environmental sample, digested with restriction enzymes and cloned into either A-based, plasmid-based, or cosmid-based vectors or into a bacterial artificial chromosome. The resulting clones were screened for expression of enzymes. This approach provides a unique tool for enzyme discovery because, unlike other methods for accessing enzymes from the environment, it requires neither isolation of the source microorganism nor prior DNA sequence information for isolating the gene. One technical drawback to this approach, however, is that cohesive DNA ends must be generated for cloning into the respective vectors. This requires, as starting material, high molecular weight DNA. DNA extracted from environmental matrices samples such as soil and sediment can be very sheered (e.g., Ogram *et al.* 1987, Leff *et al.* 1995, Frostegard *et al.* 1999) making the restriction approach to environmental DNA construction nearly impracticable. Here we present a method for constructing environmental libraries that does not require restriction digests of the purified DNA. It is based on blunt-end modification followed by 3'-adenylation of the DNA fragments using Vent DNA polymerase and *Taq* DNA polymerase, respectively. The prepared DNA fragments are then ligated into a TA vector and transformed in *Escherichia coli.* With this technique, an environmental DNA library was generated using DNA isolated from Icelandic geothermal sediment. Here we report the sequences of two dark clones, p5ice16 p5ICE16 and p6ice16p6ICE16, identified during routine screening of the environmental DNA library.

Materials and methods

Sample collection

Wet biomass-containing sediment (ICE16), 200 ml, was aseptically collected from a piped run-off stream

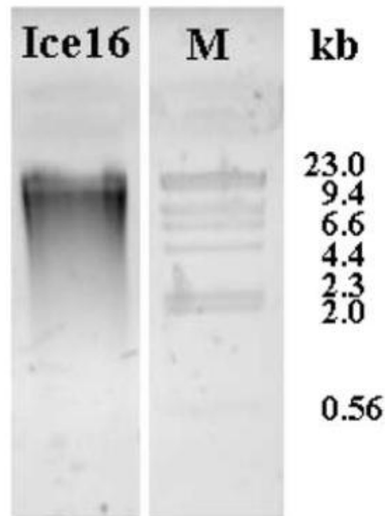

Fig. 1. DNA isolated from ICE16 geothermal sediment *via* bead beating plus spin column purification. One-tenth volume of purified DNA extracted from a 0.5 g aliquot of wet sediment was electrophoresed on 1% agarose gel. (M) *k-Hind* III DNA marker.

(69-79 °C, pH 9.5) that was emptying into the Hvere-girdi River, Iceland. In the laboratory, ICE16 sediment was distributed into 50 ml polypropylene tubes and stored at -80 °C.

DNA extraction

Environmental DNA was extracted directly from 0.5 g aliquots of wet sediment using the FastDNA spin kit for soil (Bio101 Inc., Vista, CA) and Mini Bead Beater-8 (BioSpect Products Ltd., Bartlesville, OK). The procedure was conducted as described in the manufacturers' instructions, with the following modifications: bead-beating was set at ~1500 rpm for 1 min and the DNA was eluted from each purification column with 200 µl H_2O. The concentration of purified environmental DNA was determined fluorometrically using the PicoGreen dsDNA quantitation reagent and kit (Molecular Probes, Eugene, OR) as described in the manufacturer's instructions. Known concentrations of kDNA were used to produce standard curves. Samples were excited at ~480 nm and the emission was recorded at ~570 nm using a TD-700 Laboratory Fluorometer (Turner Designs, Sunnyvale,
CA).

TA cloning

Purified DNA fragments greater than 1 kb were recovered *via* preparative gel electrophoresis (Sambrook *et al.* 1989). Single-stranded 3' and 5'-termini were removed from the size-fractionated DNA using Vent DNA polymerase (New England Biolabs, Beverly, MA). Blunt-end polishing was carried out in 100 /zl reactions in the presence of ~5 µg input DNA and 50 /µM each of dATP, dCTP, dGTP and dTTP plus 1x

ThermoPol buffer which was supplied with the enzyme. One U Vent DNA polymerase was added and the mixture incubated at 55 °C for 30 min. The reaction was quenched on ice before purifying the DNA using the QIAquick PCR purification kit (Qiagen Ltd., Crawley, West Sussex, UK) according to the manufacturer's instructions except that 95 µl H2O was used to elute the DNA from the silica matrix.
For the addition of deoxyadenosine to the 3' termini of DNA, 11 µl of 10x ThermoPol buffer, 5 µl 2 mM dATP and 1 U *Taq* DNA polymerase (Gibco Life Technologies, Gaithersburg, MD) were added to the DNA solution and incubated at 72 °C for 20 min. The reaction mixture was then purified using the QIAquick PCR purification kit as above except that 100 µl TE buffer (10 mM Tris/HCl (pH 8), 1 mM EDTA) was used to elute the DNA from the purification matrix. 3'-Adenylated (3'A)-DNA was ethanol-precipitated, dried under vacuum and re-suspended in 10 µl of 1 x shrimp alkaline phosphatase (SAP) buffer (50 mM Tris/HCl (pH 8.5, at 20 °C), 5 mM $MgCl_2$). One U SAP (Gibco Life Technologies) was added and the mixture incubated at 37 °C for 30 min. Before proeceeding with the ligation reaction, SAP was inactivated by incubating the reaction mixture at 65 °C for 15 min.

The pCR-XL-TOPO cloning kit (Invitrogen Corp., Carlsbad, CA) was employed for cloning the prepared environmental DNA. Ligation was achieved by adding 10 ng pCR-XL-TOPO to 4 µl of 5'dephosphorylated, 3'A-DNA. The reaction mixture was incubated at 22 °C for exactly 5 min. The sample was quickly placed on ice and used immediately to transform elec-trocompetent *E. coli* TOP10 cells as described in the manufacturer's instructions. The amplified library was maintained as a cell suspension in 25% (v/v) glycerol at -80 °C (Sambrook *et al.* 1998).

Sequencing

DNA sequences were determined by Oswel DNA services (Northampton, UK). Nucleotide and deduced amino acid sequences obtained from environmental clones were compared with entries in the nonredundant nucleic acid and protein databases of the National Center for Biotechnology Information (NCBI) server using BLASTN or BLASTX (Altschul *et al.* 1990, Gish & States 1993). Open reading frames (ORFs) were located within DNA sequences using the ORF finder, also located on the NCBI server. The sequence data of p5ICE16 and p6ICE16 have been submitted to the GenBank nucleotide sequence database under accession numbers AF429956 and AF429957, respectively.

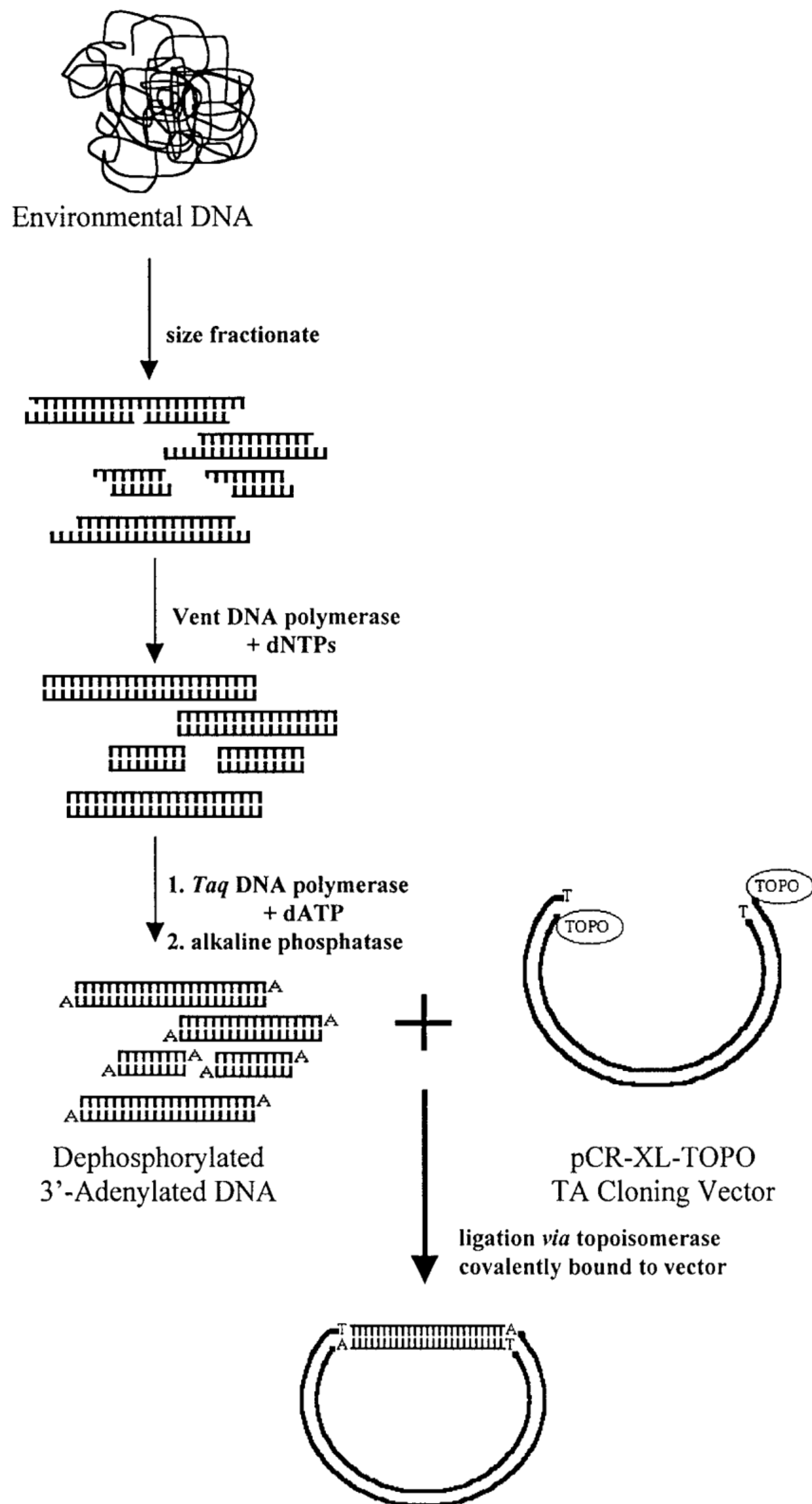

Fig. 2. Overview of environmental library construction using pCR-XL-TOPO. TOPO is topoisomerase covalently bound to vector ends.

Results and discussion

DNA extraction

For construction of an environmental DNA library, DNA was isolated by direct lysis of microorganisms present within the geothermal sediment. This was achieved *via* ballistic disintegration in a bead beater followed by spin-column purification. The yield was 17.7 µg DNA per g dry sediment (SD 0.96, $n = 3$). The size distribution of recovered ICE16-DNA ranged between ~1 kb and 20 kb (Figure 1). While high-molecular-weight DNA is usually required for general cloning procedures, DNA fragmented to between 1 kb and 10 kb was preferred as starting material for TA cloning. This is because prior cleavage of the DNA fragments with restriction endonuclease(s) is not required for ligation to pCR-XL-TOPO. This is an important consideration since DNA isolated from soil and sediment can be fragmented to 10 kb or less especially when recovered *via* bead-beating (Ogram *et al.* 1987, Leff *etal.* 1995).

TA cloning

A schematic describing the construction of environmental DNA libraries *via* TA cloning is shown in Figure 2. TA cloning was selected for constructing environmental DNA libraries because after initial cloning experiments using Sau3A1 partial digests of environmental DNA were unsuccessful (results not shown). The reason for this failure was attributed to the molecular weight of the input DNA. Complete or even partial digestion of relatively low-molecular-weight DNA was thought to have generated a mixed population of DNA termini with only some fragments possessing cohesive ends.

Purified ICE16-DNA was size fractionated and DNA fragments > 1 kb were retained (Materials and methods section). Vent DNA polymerase was selected for blunt-end polishing of DNA fragments because the termini generated by this enzyme are >95% blunt-ended (Imai *et al.* 1991). After a clean-up step, *Taq* DNA polymerase was then utilized to produce 3'A-DNA fragments. Ligation of dephosphorylated input DNA and plasmid DNA was achieved by exploiting the ligase activity of the enzyme topoisomerase I, which is supplied covalently bound to the linearised vector (Shuman 1994). Ligation products were then used to transform *E. coli* TOP10 cells (Materials and methods). A transformation efficiency of 1.4×10^6 colonies μg^{-1} vector was obtained. Agarose gel electrophoresis of restriction digests of randomly selected clones identified 84% recombinant plastids containing inserts ranging between ~1 and 10 kb. The unamplified library contained 37 000 independent clones. With an average insert size of ~5 kb, this represents approximately 185Mbp of cloned environmental DNA. Given an average gene size of ~1 kb, the environmental DNA library is estimated to contain 185 000 genes.

Screening

During routine screening of the amplified environmental library for amylase activity (Gerhardt *et al.* 1981), recombinant transformants were identified fortuitously by their dark coloured phenotype. Although flooding the starch plates with I_2 solution revealed no zones of starch hydrolysis, five dark transformants were noted amongst the general population of creamy-white transformants typical of the library. When *E. coli* was re-transformed with purified plasmid DNA isolated from these five dark clones, three re-transformants conferring the dark phenotype were recovered. Restriction digests of plasmids from these three transformants (p5ICE16, p6ICE16 and p12ICE16, respectively), indicated that p6ICE16 and p12ICE22 are derived from the same clone (Figure 3).

This was confirmed by comparing the DNA-insert sequence of p6ICE16 with the partial DNA sequence of the insert from p12ICE16 (results not shown).

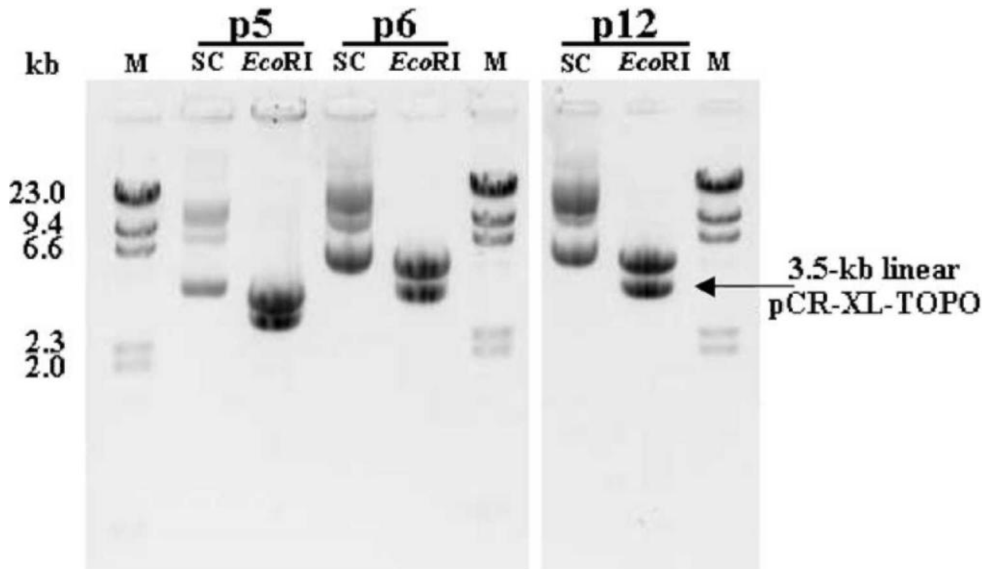

Fig. 3. One percent agarose gel of plasmids (p5, p6 and p12) isolated from *E. coli* TOP10/pCR-XL-TOPO clones 5ICE16, 6ICE16 and 12ICE16, respectively. These recombinant transformants displayed a dark-coloured phenotype when assayed on starch indicator plates. M, *k-HindIII* DNA marker; SC, supercoiled plasmid DNA; EcoRI, plasmid DNA digested with EcoRI.

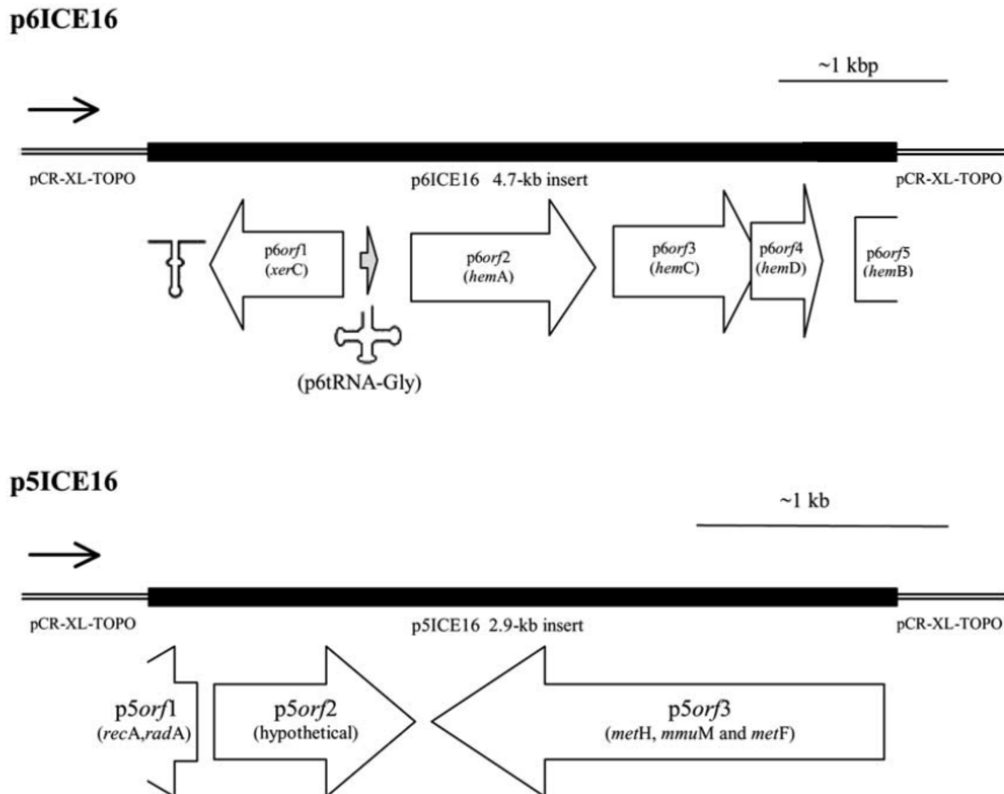

Fig. 4. Physical maps of DNA inserts (black bars) from p5ICE16 and p6ICE16. Block arrows show the relative localisation of each gene and the orientation of coding sequences found. Truncated ORFs are indicated by incomplete arrows. The genes having homologous sequences to the coding sequences identified are shown in parentheses and summarised in Table 1. The black arrow indicates the direction of transcription from the vector-encoded *lac* promoter. Vector sequences are indicated by double lines. Maps are at different scales.

Table 1. Protein- and tRNA-encoding genes identified in environmental DNA sequences recovered from dark-coloured transformants as described in Results. Trunc. indicates that the coding sequence was truncated at the cloning site. The GenBank sequence accession numbers of the homologues are given in parentheses. For most of the query sequences, BLAST local sequence alignments showed highest homologies either within the central region or across the entire sequence. *N* and *C*, respectively, refer to those cases in which highest homology was found within the *N*-terminal and *C*-terminal sequences of the putative gene product.

DNA insert/ coding sequence	Size of product	Putative gene	Organism (accession number)	% Amino acid or % nucleotide identity of gene product
p5ICE16/				
p5*orf*1	trunc. 37 aa	*RecA* recombinase	*Aquifex aeolicus* (O66827)	55% over 37 aa
p5*orf*2	251 aa	Hypothetical	*Bacillus holodurans* (BAB04808)	24% over 208 aa
p5*orf*3	612 aa	Methionine biosynthesis	*Bacillus holodurans* (BAB05348)	37% over 609 aa
			Thermotoga maritima (AAD35357, AAD35359)	*N* 27% over 401 aa *C* 32% over 278 aa
p6ICE16/				
p6*orf*1	312 aa	Site-specific recombinase	*Methanothermobacter thermautotrophicus* (AAB85391)	42% over 148 aa
p6*orf*2	377 aa	Glutamyl-tRNA reductase	*Aquifex aeolicus* (O67314)	29% over 335 aa
p6*orf*3	299 aa	Porphobilinogen deaminase	*Aquifex aeolicus* (AE000681)	38% over 214 aa
p6*orf*4	143 aa	Uroporphyrinogen synthase	*Deinococcus radiodurans* (AAF10363)	40% over 143 aa
p6*orf*5	trunc. 76 aa	Porphobilinogen synthase	*Chlamydophila pneumoniae* (Q9Z7G1)	58% over 39 *N*-terminal aa
p6*tRNA*	76 nt	tRNA-Gly	*Stigmatella aurantiaca* (X82820)	96% over 76 nt

Sequence studies

The DNA inserts of plasmids p5ICE16 and p6ICE16 were sequenced and compared to the sequences in the NCBI databases (Materials and methods section). The physical maps of the p5ICE16- and p6ICE16-DNA inserts are shown in Figure 4. The coding sequences identified for the cloned DNA fragments are summarised in Table 1.

Evidence indicates that the isolated environmental DNA fragments were derived from extremophilic prokaryotes. In all cases, the putative genes identified in this work possess highest homologies to bacterial or archaeal sequences. Except for the truncated p6orf5 and p6tRNA-Gly, high-scoring homo-logues were from thermophilic species. A putative *hem* operon encoding tetrapyrrole biosynthetic genes was identified in p6ICE16. Operons or similar gene clusters are features unique to prokaryotes. All ORFs identified in the cloned DNA-inserts of p5ICE16, p6ICE16 were preceded by putative ribosome binding sites indicating that the environmental DNA fragments were prokaryotic.

p5*orf*2, encoded by p5ICE16, was the likely candidate for conferring the dark phenotype to *E. coli* TOP10. It possessed a moderate SD sequence 13nt upstream to an ATG start codon and was located downstream of a bacterial-like promoter sequence. p5*orf*1 and p5*orf*3 were not considered likely candidates because the former was severely truncated at the cloning sight and the latter was positioned in the wrong orientation with respect to the vector-encoded *lac* promoter. Because p5*orf*3 appeared to lack its own promoter, heterologous expression was not likely. BLAST searches (Materials and
methods section) indicated that p5*orf*2 encoded a hypothetical protein, the function of which could not be inferred. The activity of the p5*orf*2 gene product, therefore, could not be investigated directly.

For p6ICE16, BLAST searches using the deduced protein sequences encoded by p6*orfs*2-5 revealed homologies to enzymes involved in the early steps of tetrapyrrole biosynthesis. This putative heme biosynthetic *(hem)* operon was considered the likely candidate for conferring the dark phenotype to *E. coli* TOP10 because bacterial-like promoter elements were identified upstream of the *hem* gene cluster and each *hem* gene possessed a putative ribosome binding site. The accumulation of tetrapyrrole pigments due to functional expression of one or more of the *hem* genes was thought to have been responsible for the dark phenotype observed with the transformant harbouring p6ICE16. The putative site-specific recombinase was not considered a likely candidate for conferring the dark phenotype because it was positioned in the wrong orientation with respect to the vector-encoded *lac* promoter. and lacked an *E. coli*-like promoter.

Our results have demonstrated that blunt-ended cloning of sheared DNA is a valid method for the generation of multigenomic libraries, which are viable targets for the recovery of novel activities from the metagenome in a culture-independent manner. Further work is underway to functionally characterise these heterologously expressed gene products.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local sequence alignment search tool. *J. Mol. Biol.* **215:** 403—410.

Brady SF, Clardy J (2000) Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J.Am. Chem. Soc.* **122:** 12903-12904. Brady SF, Chao CJ, Handelsman J,

Clardy J (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* **3:** 1981-1984.

Cottrell MT, Moore JA, Kirchman DL (1999) Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **65:** 2553-2557.

Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P (1999) Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* **65:** 5409-5420.

Gerhardt P, Murry RGE, Costilow RN, Nester EW, Wood WA, Krieg NR, Phillips GB (1981) *Manual of Methods for General Bacteriology.* Washington, DC: ASM Press. Gish W, States DJ (1993) Identification of protein regions by database similarity search. *Nat. Genet.* **3:** 266-272.

Henne A, Daniel R, Schmitz RA, Gottschalk G (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* **65:** 3901-3907.

Henne A, Schmitz RA, Bomeke M, Gottschalk G, Daniel R (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **66:** 3113-3116.

Imai Y, Matsushima Y, Sugimura T, Terada M (1991) A simple and rapid method for generating a deletion by PCR. *Nucl. Acids Res.* **19:** 2785.

Kawata Y, Yano S, Kojima H (1998) Efficient library construction with a TA vector and its application to cloning of the phytoene synthase gene from the cyanobacterium *Spirulina platensis*. *Curr. Microbiol.* **37:** 289-291.

Leff LG, Dana JR, McArthur JV, Shimkets LJ (1995) Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* **61:** 1141-1143.

Ogram A, Sayler GS, Barkay T (1987) The extraction and purification of microbial DNA from sediments. *J. Microbiol. Meth.* **7:** 57-66.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66:** 2541-2547.

Shuman S (1994) Novel approach to molecular cloning and polynu-cleotide synthesis using vaccinia DNA topoisomerase. *J. Biol.Chem.* **269:** 32678-32684.