Extremophiles (2003) 7:415–421
DOI 10.1007/s00792-003-0339-2

ORIGINAL PAPER

**Helen C. Rees · Susan Grant · Brian Jones
William D. Grant · Shaun Heaphy**

Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries

Received: 16 April 2003 / Accepted: 19 June 2003 / Published online: 5 July 2003

Abstract A genomic DNA library was made from the alkaliphilic cellulase-producing *Bacillus agaradhaerans* in order to prove our technologies for gene isolation prior to using them with samples of DNA isolated directly from environmental samples. Clones expressing a cellulase activity were identified and sequenced. A new cellulase gene was identified. Genomic DNA libraries were then made from DNA isolated directly from the Kenyan soda lakes, Lake Elmenteita and Crater Lake. Crater Lake clones expressing a cellulase activity and Lake Elmenteita clones expressing a lipase/esterase activity were identified and sequenced. These were encoded by novel genes as judged by DNA sequence comparisons. Genomic DNA libraries were also made from laboratory enrichment cultures of Lake Nakuru and Lake Elmenteita samples. Selective enrichment cultures were grown in the presence of carboxymethylcellulose (CMC) and olive oil. A number of new cellulase and lipase/esterase genes were discovered in these libraries. Cellulase-positive clones from Lake Nakuru were isolated at a frequency of 1 in 15,000 from a library made from a CMC enrichment as compared to 1 in 60,000 from a minimal medium enrichment. Esterase/lipase-positive clones from Lake Elmenteita were isolated with a frequency of 1 in 30,000 from a library made from an olive-oil enrichment as compared to 1 in 100,000 from an environmental library.

Communicated by K. Horikoshi

H. C. Rees · S. Grant · W. D. Grant · S. Heaphy (✉)
Department of Microbiology and Immunology, University of Leicester, University Road, Leicester, LE1 9HN, UK
E-mail: sh1@le.ac.uk
Tel.: +44-116-2522973
Fax: +44-116-2525030

B. Jones
Genencor International B.V., Archimedesweg 30, 2333 CN Leiden, The Netherlands

Present address: H. C. Rees
Department of Civil and Structural Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Keywords Alkaliphile · Cellulase · Cloning · Environmental DNA · Esterase/lipase · Extremophile · Halophile · Libraries

Introduction

Industrially useful enzymes with novel applications, or which improve upon the activities of ones being currently used, are frequently sought (Marrs et al. 1999). The requirement for enzymes that are active under more extreme conditions of temperature, pH or salinity has meant that microorganisms isolated from 'extreme' environments, so-called extremophiles, have proved a valuable resource for the isolation of novel biotechnological products (Aguilar et al. 1998). A perceived benefit of enzymes from extremophiles for industrial applications is their high stability. The genes from these organisms are usually cloned into conventional host organisms in order to maximise enzyme production under less expensive growth conditions. This latter advance has in recent years lead to the realisation that direct cloning of environmental genomic DNA into such hosts circumvents the need for cultivation and isolation of the enzyme-producing organism itself (Marrs et al. 1999). Since it is generally accepted that a large proportion of microorganisms are not yet cultivable (Amann et al. 1995), this approach vastly increases the pool of genetic biodiversity from which novel enzyme activities may be isolated.

Alkaliphilic and halophilic microbial species have the potential to yield valuable new products for biotechnology and industry (Horikoshi 1999; Margesin and Schinner 2001). Alkaliphilic polymer-degrading enzymes such as proteases, lipases and cellulases are most frequently isolated from *Bacillus* or related species. Cellulases and lipases are not only already important components of many washing detergents, but also find roles in the paper, pulp, pharmaceutical, food, leather, chemical or waste treatment industries.

In this study we have constructed lambda libraries from genomic DNA isolated directly from environmental samples collected from the extremely alkaline and saline soda lakes of the East African Rift Valley in Kenya and the Wadi Natrun in Egypt. The East African lakes have proved to be a rich source of novel cultivable haloalkaliphilic organisms (Duckworth et al. 1996; Jones et al. 1998) and of as yet uncultivated novel phylotypes (Grant et al. 1999). These libraries were then screened for cellulase and lipolytic activities using plate assays. Activity screening directly for functional enzymes is likely to reveal greater enzyme diversity than DNA probe methods based on the sequence of known enzymes (Henne et al. 2000). In addition, enrichment libraries were screened for the same types of enzyme activity in order to determine whether this affected the proportion of clones showing the desired enzyme activity. Finally the similarity of the sequences of these newly isolated enzymes was compared to those already in the databases.

Materials and methods

Sample collection

Samples were collected from the following lakes in the East African Rift system in Kenya in January 1999: Lake Nakuru at 0° 23′ S, 36° 05′ E; Crater Lake (Lake Sonachi) at 0° 49′ S, 36°16′ E; and Lake Elmenteita at 0° 25′ S, 36° 15′ E. Samples were also collected from the soda Lake Zugm of the Wadi Natrun, Eygpt (30° 15′ N, 30° 30′ E) in October 2000. Details of the environmental conditions in the lakes are described in Jones et al. (1994) and Duckworth et al. (1996).

DNA extraction

Most of the genomic DNA in the Lake Nakuru and Crater Lake libraries was partially extracted on-site from the lake samples in January 1999. Brines were sampled in a 250-ml stainless steel beaker and filtered on-site through a sequence of sterile 47-mm membrane filters in a 250-ml capacity polycarbonate filter unit (Sartorius) using a Nalgene hand pump, which produced a vacuum of 40–50 cm Hg under field conditions. A Whatman GF/A glass fibre prefilter was used first to remove suspended inorganic matter, which was discarded. The filtrate, 750 ml from Crater Lake and 350 ml from Lake Nakuru, was then passed through a Sartorius 8-μm cellulose nitrate filter until the flow stopped. The collected filtrate was then passed through a 3-μm cellulose nitrate filter, again until the flow stopped. Finally in the case of the Crater Lake brine, the collected filtrate was passed through a 0.22-μm cellulose acetate filter. The individual membranes were placed immediately into cold sterile saline buffer, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 2 M NaCl, and kept on ice in a refrigerated cool box until they could be processed further, usually within 4 h of sampling. The material on the filters was dispersed by vortexing and the cells pelleted by centrifugation. The cells were lysed using the GenomicPrep cell and tissue DNA isolation kit (Pharmacia Biotech) following the manufacturer's protocol. The samples are stable in this lysis solution at ambient temperature, and were transported back to the United Kingdom where extraction was completed.

The Nakuru and Crater Lake (Lake Sonachi) libraries were made by pooling the DNA from the on-site samples described above. In addition, both samples were supplemented with 30–50% extra DNA prepared from the stored refrigerated lake waters. In this case the microbial mass was pelletted by centrifugation. DNA extraction was carried out as described above, except that incubation with 50 μl lysozyme solution (50 mg/ml in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA) for 30 min at 37°C preceded the addition of the lysis solution, in order to degrade Gram-positive cell walls.

The Lake Zugm samples were transported to the United Kingdom and stored at 4°C. The library was constructed by pelleting the biomass. DNA was extracted following lysozyme treatment, using the Pharmacia kit as described above.

In addition to the material processed on-site, brines, sediments and microbial mats were collected from the same sampling sites. These were transported back to the United Kingdom and stored at 4°C. The Lake Elmenteita environmental library was made after return to the United Kingdom by pooling brine and sediment samples. DNA was extracted using the method of Chen and Kuo (1993), but with the addition of a prior lysozyme treatment as described above.

Enrichment cultures from Lakes Elmenteita and Nakuru were prepared by adding 1 ml pooled samples to 200 ml sterile alkaline minimal medium, pH 10.5, containing (per litre) 10 g glucose, 1 g yeast extract (Difco), 1 g K_2HPO_4, 0.2 g $MgSO_4.7H_2O$, 10 g Na_2CO_3, 40 g NaCl. For a minimal medium containing carboxymethylcellulose (CMC), glucose was exchanged for an equal weight of CMC. For minimal olive-oil medium, glucose was exchanged for 10% v/v olive oil. Cultures were grown at 37°C for 3–5 days until bacterial growth was visible. DNA was extracted according to the method of Pitcher et al. (1989).

Bacillus agaradhaerans (DSM 8721) was grown for 2 days, until good growth was visible, in an alkaline broth at pH 10.5, containing (per litre) 10 g glucose, 5 g yeast extract (Difco), 5 g bactopeptone (Difco), 1 g K_2HPO_4, 0.2 g $MgSO_4.7H_2O$, 10 g Na_2CO_3. Pelletted cells were lysozyme-treated prior to DNA extraction using the same DNA extraction kit.

Library construction

DNA samples were used to construct genomic DNA lambda libraries following partial digestion with *Sau*3A1, optimised to maximise fragments in the 2–10 kb size range. Restricted DNA was run out on either 0.5% or 0.75% w/v Tris/acetate/EDTA agarose gels using Pharmacia N/A agarose. Fragments of 2–10 kb were excised and concentrated by reversed current electrophoresis, then extracted from the agarose gel using the QIAEX II system (QIAGEN) according to the manufacturer's protocol. The restricted DNA fragments were ligated into the ZAP Express vector (Stratagene), which had been predigested with *Bam*H1 and phosphatase-treated. It was then packaged with the Gigapack III Gold packaging kit (Stratagene). The primary libraries were titred with *Escherichia coli* XLI blue MRF′. Blue/white screening in the presence of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside and isopropyl beta-D-thiogalactopyranoside (IPTG) was used to determine the cloning efficiency. Libraries were amplified and stored at −80°C following the manufacturer's instructions. Library quality was also assessed by sizing the inserts in about 24 of the clones. Phagemid pBK-CMV was excised from the ZAP Express (Stratagene) libraries using ExAssist helper phage, and used to infect *E. coli* strain XLOLR, according to the Stratagene protocol. Phagemid-containing clones were recovered after plating on Luria-Bertoni medium (LB) containing kanamycin (50 μg/ml). Phagemid DNA was extracted using the Wizard Plus SV Minipreps DNA purification system (Promega). The insert size was determined by restriction digestion with *Hind*III or *Hind*III plus *Pst*I followed by gel electrophoresis.

Plate screening

E. coli clones containing the pBK-CMV phagemid were screened for cellulase or lipase/esterase activities by plating the phagemid-infected XLOLR strain onto substrate-containing media, to give

approximately 1,000 colonies per 7-cm-diameter plate. For cellulase activity this consisted of LB adjusted to pH 7.5 containing 50 μg ml^{-1} kanamycin and 0.5% w/v CMC cellulase substrate. IPTG (15 μl of 0.5 M solution) was spread on the plate prior to use. Detection of cellulolytic activity was based on the method of Teather and Wood (1982). After overnight growth the colonies were overlaid with 2.5 ml of 0.7% w/v water agarose. The plates were then flooded with an aqueous solution of 1 mg/ml Congo red and incubated at room temperature for 30 min with slow rocking. They were then washed twice with 1 M NaCl for 30 min. Positive clones were visible by a yellow zone of hydrolysis against a red background. For lipase/esterase activity, colonies were first grown on LB containing kanamycin, then replica-plated onto tributyrin agar pH 7.5 (Oxoid), containing IPTG as described above. Positive clones were identified by the faint zone of clearing around them.

Characterisation of cellulase and lipase positive clones

Phagemid DNA was isolated from positive clones and the insert size determined as described above. DNA sequencing was carried out at the Protein and Nucleic Acid Chemistry Laboratory at the University of Leicester, using Perkin Elmer BigDye terminator chemistry and a model 377 ABI automated DNA sequencer. Complete coverage of the sequence was obtained by primer walking from both the 5′ and 3′ ends, and was edited using the multisequence editor Seqed version 1.0.3 (Applied Biosystems). Sequences were assembled with programmes in the GCG Wisconsin Package, Version 10.2-UNIX, available at the University of Leicester. Comparison of sequences to those in the databases was made using BLASTX 2.1.3 at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). The open reading frame (ORF) finder at NCBI was used to identify possible open reading frames.

Results

General features of the libraries

Primary titres were generally $\geq 10^6$ pfu, and >90% of the clones contained inserts in the 2–10 kb size range. Analysis of the sequence of the first 500 bases in ten randomly selected clones from each of two different environmental libraries gave only low identity to known sequences in the databases (data not shown), confirming that we were cloning DNA from uncharacterised genomes. Table 1 summarises the libraries from which enzyme activities described in this paper have been discovered, including the names of the library, incidence of positive clones, enzyme activity type, name of clone, size of cloned insert, size of predicted protein and identity to proteins present in the data bases.

Screening the *B. agaradhaerans* genome for cellulases

To test the validity of enzyme activity screening of these libraries, and to optimise the conditions for the cellulase screen, we first made a genomic library from *B. agaradhaerans* (DSM 8721). This organism produces an alkaline cellulase, endo-1,4-beta-glucanase, belonging to cellulase family 5 of the glycosyl hydrolases (endoglucanase 5A; EC 3.2.1.4; Swiss-Prot 085465, entry name GUN5_BACAG; EBI accession number AF067428) the gene for which is 1,203 bp in length (Davies et al. 1998). Cellulase positive clones were detected with an incidence of 1/3,000 in the plate assay (Table 1, Fig. 1). One of the positive clones (BAGCEL) was retained for further study. Unexpectedly however, no polymerase chain

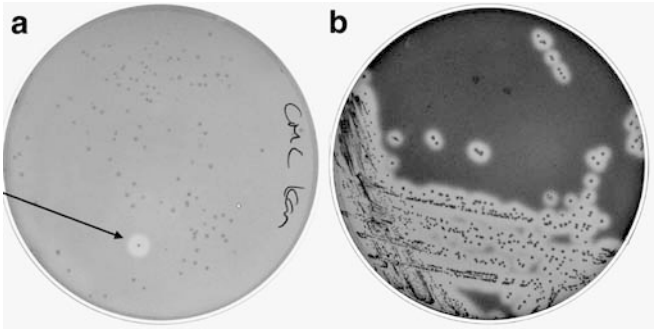

Fig. 1a, b Plate screening for cellulase activity using a Congo red assay. **a** Part of a library with a single cellulase-expressing clone indicated by the *arrow*. **b** The same clone after colony purification

Table 1 Libraries, genes and encoded enzymes described in this paper

Library	Incidence of positive clones	Activity	Clone name	Insert bp	Open reading frame information
Bacillus agaradhaerans genomic	1/3,000	CEL	BAGCEL	4,205	1,713 bp, 570 aa, 68% identity *B. halodurans* AP001509
Crater Lake environmental	1/36,000	CEL	CRATCEL	3,410	1,746 bp, 581 aa, 29% identity *Fusobacterium mortiferum* AAB49340
Lake Nakuru, minimal alkaline enrichment	1/60,000	CEL		Approx. 2,500	
Lake Nakuru, minimal carboxymethylcellulose enrichment	1/15,000	CEL	HKCEL	3,796	1,716 bp, 571 aa, 67% identity *B. halodurans* AP001509
Lake Elmenteita environmental	1/100,000	LIP	ELIP	4,313	1,209 bp, 402 aa, 67% identity *Salmonella typhymurium* NP_460582.1
Lake Elmenteita, minimal olive-oil enrichment	1/30,000	LIP	LIP1	2,285	792 bp, 263 aa, 42% identity *Vibrio cholerae* NP_232345
			LIP2	3,112	645 bp, 214 aa 43% identity *Escherichia coli* U82664

BAGCEL
MGYTKAKCTL KKTVLFGLIL CLSVSMFVPM TSAEDVTSSQ LDIHSYVADM QPGWNLGNTF
DAVGDDETAW GNPRVTRELI KTIADEGYKS IRIPVTWQNQ MGGSPDYTIN EDYINRVEQA
IDWALEEDLY VMLNVHHDSW LWMYDMEHNY DEVMARYTAI WEQLSEKFKS HSHKLMFESV
NEPRFTQEWG EIQENHHAYL EDLNKTFYYI VRESGGNNVE RPLVLPTIET ATSQDLLDRL
YQTMEDLDDP YLIATVHYYG FWPFSVNIAG YTHFEQETQQ DIIDTFDRVH NTFTARGVPV
VLGEFGLLGF DKSTDVIQQG EKLKFFEFLI HHLNERDITH MLWDNGQHLN RETYAWYDQE
FHDILKASWE GRSATAESNL IHVKDGKPIR DQDIQLYLNG NELTALQAGE ESLVLGEDYE
LAGGVLTLKA DTLTRLITPG QLGTNAVITA QFNSGADWRF QLQNVDVPTV ENTDGSTWHF
AIPTHFNGDS LATMEAVYAN GEYAGPQDWT SFKEFGEAFS PNYATGEIII SEAFFNAVRD
DDIHLTFHFW SGETVEYTLR KNGNYVQGRR *

CRATCEL
MNFSLPIRFF GIGPTAGFFL LLMGSSALVA STHTDKIRID QFGYPADATK VAVIADPQIG
WNSAESYSPG ATLEVRRVND GVVVFSGSPV PWNGGAIHIQ SGDRVWWFDF TVVAEPGHYR
IHDPANNTHS DSFAIGADVY DVVLREAVRM FFYQRSGFAK EVPYAHANWA DAASHPQDVA
SRPIWDMGNA SLERDLSGGW FDAGDFNKYS EWTGRVILEL LLAYQGRPDV FTDDFGIPES
GNGVPDLLDE VKWGMDWLLR MQEPSGAILG KVSVTGHQSA SPPSTDTHPR YYGPVSTEAT
AMAAAAFALG ATVFESVGMS DYAVTLESAA IAAWNWTMVH PHVPFDNTGF ASVSPSRNAH
DTLANRVMAA AMLFERTGGA VYRDFFDVRY LDMEPVQWWY FFPFQGELQK ALAHYTTLPG
ATPSVSADIR NRMAASINGG EFLGAWNNQT DAYRAYLKDQ DYTWGSNKTK SQAGFFFEGV
RRLGLNPADA AAHRDAAMGY LHYLHGVNPM GMVYLSNMYA SGADRAANEI YHHWFRDGRT
GTMPSLHSTV PLLVFFRAGP NAQIQRKHSG DPRPTRAKSL P*

HKCEL
MGYTQAKCMV KKTVLFGLIL CLGVSMFVPV TSAEDRVSSS QVDIQSYVAD MQPGWNLGNT
FDAIGDDETA WGNPRVTREL IEMIADEGYK SIRIPVTWQN QMGGSPDYTI NEDYIKRVEQ
VIDWALEEDL YVMLNVHHDS WLWMYDMEHN YDEVMARYTA IWEQLSEKFK NHSHKLMFES
VNEPRFTQEW GEIQENHHAY LEDLNKTFYY IVRESGGNNV ERPLVLPTIE TATSQDLLDR
LYQTMEDLDD PHLIATVHYY GFWPFSVNIA GYTRFEQETQ QDIIDTFDRV HNTFTANGIP
VVLGEFGLLG FDKSTDVIQQ GEKLKFFEFL IHHLNERDIT HMLWDNGQHL KRETYSWYDQ
EFHDILKASW EGRSATAESN FIHVKDGEPI RDQHIQLYLN GNELTALQAG DESLVLGEDY
ELAGDVLTLK AGILTRLITP GQLGTNAVIT AQFNSGADWR FQLQNVDVPT VENTDGSIWH
FAIPTHFNGD SLATMEAVYA NGEYAGPQDW TSFKEFGEAF SPNYATGEII ITEAFFNAVR
DDDIHLTFHY WSGETVEYTL RKNGNYVQGR R*

ELIP
MVWLHGGGYT IGAGSLPPYD GAAFASRDVV LVTVNYRLGH LGFFAHPALD EENPDGPVHN
FALLDQIAAL KWVQENIAAF GGDAGNVTLF GESAGARSVL SLLASPLAKN LFHKGIIQSA
YTLPDVDRKK ALKRGVALAG HYGLQNATAD ELRALPADGL WALEGPLNIG PTPISGDVVL
PEPMLDIFFA GRQHRMPLMV GSNSDEASVL SYFGIDPAGQ VELLRRGAAF PDWGLIKLLY
SRSEXGMPEL GRQVCRDMAF XXLGFVVMQA QQRVNQPCWR YYFDYVGEAE RKIYANGTWH
GNEVPYVFDT LSLTPPASEY VNQNDLTFAG QICDYWTRFA RSAGPHSKAI PGPLSWPACV
RGKDRTMRLG VHSRARFKVE NRFMRMRMQL FKRVMKHHVS LD*

LIP1
MPQNTGNLSI NVQGSGPALV LLHGWGLNQA IWHPIQAKLA AHFTVYCVDL PGFGDSSWLP
GNESFHHACT RVAEQLIRRI GHSFALAGWS MGGLIATQIA LDFPTYVQRL ITIASSPCFI
AHEGEQWPGI KPTTLETFRY QLTTDYRKTL ERFLAVQALG SPAARDEIKA MRVLLSERAE
PQPEALVAGL RWLAQVDMRK QLMALEVPLL RLYGKRDSLV PIATQEAIRP LLPHHLQDTV
VTFQGSAHAP FFTEPEQFVA ALL*

LIP2
MFKKVLNHSL LISFGFICLL LIRPASANVS LVVLGDSLSA GYGISQADTW VAEIERRWQR
DYPEFSIINA SISGDTTQGG LNRLADVVER HQPDAVFIEL GGNDGLRGFR VDTIRDNLSQ
MIDYLHGQGI YVALSQIEIP PNMGRRYTSQ FSGLFAEVAE AHEIALVPFF MIDIATDSTL
MQSDGIHPNL AAQPVIADIM EPKLREILER VADL*

Fig. 2 Predicted amino acid sequences encoded by the putative open reading frames described in this paper

reaction (PCR) product was obtained from this clone using primers known to amplify the known *B. agaradhaerans* cellulase, despite the fact that we could amplify this gene from the library. The complete sequence of the 4,205-bp insert coding for the cellulase was therefore determined by primer walking. BLASTX and ORF-finder analysis identified a potential gene of 1,713 base pairs encoding a protein of 570 amino acids (see Fig. 2), and having 68% identity over 1,575 base pairs with the endo-β-1,4-glucanase (cellulase B) of the facultative alkaliphile *B. halodurans* C125 (EBI accession AP001509; protein ID BAB04322) (Takami et al. 2000). The sequence of the clone containing BAGCEL has been deposited at the EBI and can be found under accession number AJ537597.

This activity screen of the *B. agaradhaerans* library for cellulase-producing clones was therefore successful, and fortuitously produced a novel enzyme.

Screening environmental libraries for cellulase genes

Genomic DNA libraries made from the Wadi Natrun, Lake Nakuru and Crater Lake samples were screened for cellulases. In the first two cases approximately 100,000 colonies were screened for cellulase activity without isolating any positives. The Crater Lake library, however, gave four positive clones in approximately 110,000 colonies (see Table 1). Three of these were cellulase-positive on repeat screening. Sequencing with the M13 forward primer (contained in the pBK-CMV phagemid) and restriction analysis indicated that these clones contained identical inserts about 3.5 kb in size. The complete sequence of the 3,410-bp insert in one of these (CRATCEL) was again obtained by primer walking. This sequence has been deposited at the EBI under accession number AJ537595. Analysis reveals a putative ORF of 1,746 bp (581 amino acids) (Fig. 2). Results of a BLASTX search show significant identity to endo-β-1, 4-glucanases from a number of species such as *Clostridium thermocellum, C. cellulovorans, Fusobacterium mortiferum,* and *Fibrobacter succinogenes*, with identities in the range 26–29%. Another putative ORF in this sequence is 516 amino acids in length and shows homology to chitinases such as those from *Emericella nidulans* and *Aspergillus nidulans*, with 21% identity. Enzymes of a particular specificity can therefore be isolated from genomic DNA libraries made from environmental samples.

Do enrichment cultures increase the frequency of enzyme isolation?

Microbes present in sediment from Lake Nakuru were grown either in a minimal alkaline medium or in a minimal alkaline medium supplemented with CMC. Screening of the library made from genomic DNA extracted after growth of the Lake Nakuru samples in the minimal alkaline medium for cellulase activity gave one positive in 60,000 clones (see Table 1). This clone contains an insert of 2.5 kb, which has not been further characterised. A similar library made following enrichment by growth on CMC increased the frequency of cellulase positive clones to one in 15,000 (Table 1). One of these clones was further characterised. The 3,796-bp insert in this clone (HKCEL) was sequenced and shown to contain a candidate ORF of 1,716 bp, encoding a protein of 571 amino acids (Fig. 2). The sequence can be found at EBI under accession number AJ537596. Database searches showed 67% identity to the cellulase B of *B. halodurans* C125 (EBI accession AP001509; protein ID BAB04322). Comparison of this putative ORF to that for the new activity obtained from *B. agaradhaerans* (encoded in the BAGCEL clone, described above) gave 94.8% identity for the nucleotide sequence and 94.7% identity (95.6% similarity) for the translated amino acid sequence. Comparison of nucleotide sequence for the two complete inserts gives 94.5% identity over 3,608 bp. It

is therefore likely derived from a *Bacillus* species closely related to *B. agaradhaerans*.

Clones HKCEL and BAGCEL also contained an incomplete ORF of about 200 amino acids showing 47% identity to the putative acetylesterase NP_465956 from *Listeria monocytogenes*.

Esterase activities from environmental libraries

We then looked to see if other enzyme activities could be isolated from these libraries. Screening of the Lake Elmenteita environmental library for lipase/esterase activity gave one positive clone (ELIP) in 100,000 screened (Table 1). The insert size was estimated at 4.5 kb. Sequencing by primer walking showed that it comprised 4,313 bp, encoding two major putative ORFs. One encoding 402 amino acids (see Fig. 2) gave highest homology of the translated protein sequence to a putative carboxylesterase from *Salmonella typhymurium* LT2 (NCBI entrez NP_460582.1), having 67% identity over 402 amino acids. It is therefore a strong candidate for the lipase/esterase activity. The sequence can be found at the EBI under accession number AJ537554. The second was 326 amino acids long, and had highest identity (53% over 305 amino acids) to a probable pyridoxal phosphate aminotransferase protein (NP_520132.1) from *Ralstonia solanacearum*.

Esterase screening from enrichment libraries

The sample from which the Lake Elmenteita environmental genomic library was made was enriched for microorganisms able to grow in the presence of olive oil as the major carbon source. A library of genomic DNA was made from this minimal olive-oil enrichment culture and yielded two positive clones after screening of 60,000 colonies for esterase/lipase activity on tributyrin agar (Table 1). Agarose gel electrophoresis after restriction enzyme digestion showed that one, LIP1, was contained in a cloned insert of about 2.5 kb in size and the other, LIP2, in a cloned insert of about 3.5 kb.

Sequencing by primer walking gave an insert size of 2,285 bp for LIP1; the sequence can be found at the EBI under accession number AJ537555. Three putative ORFs were encoded. The most likely candidate for the lipase gene was a 792-bp region (263 amino acids) (Fig. 2) with the highest identity of 42% for the translated amino acid sequence over 252 amino acids to the hypothetical bioH protein of *Vibrio cholerae* (accession NP_232345, PID g15642712). A second putative ORF of 260 amino acids, 783 bp, had closest identity of 34% to competence protein F of *Xylella fastidiosa* (NP_297356, pir C82852). The third, of 98 amino acids, 297 bp, had 48% identity over 97 amino acids to a conserved hypothetical protein from *R. solanacearum* (NP_519040.1) and 44% over 92 amino acids to the acetylspermidine deacetylase from *Brucella melitensis* (NP_540421.1).

The LIP2 insert sequence was 3,112 bp in length; the sequence can be found at the EBI under accession number AJ537556. Three main ORFs were identified, with the most probable lipase/esterase encoding region being 645 bp, encoding a predicted protein 214 amino acids long (Fig. 2). It had highest homology of 43% over 197 amino acids in a BLASTP search to an acyl-coA thioesterase I precursor of *Escherichia coli,* U82664. A second putative ORF of 242 amino acids has 34% identity over 230 amino acids to a probable short-chain dehydrogenase from *Pseudomonas aeruginosa* (NP_252127.1). The third ORF of 187 amino acids has 37% identity to a 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase from *Xylella fastidiosa* (NP_298745.1, pir A82680) over 94 amino acids.

Discussion

In this work we have made environmental genomic libraries with insert sizes ranging from 2–10 kb. Analysis via 16 S rRNA sequencing of the DNA from which these libraries were made shows that they contain a wide degree of species diversity from both the archaeal and bacterial domains (Rees 2002). Sequence analysis of 20 clones chosen at random from the libraries confirms this. FASTA analysis of the nucleotide sequences showed about 60–70% identity to sequences present in the database. BlastX analysis, translating the nucleotide sequences in all six reading frames, showed that 15 out of 20 clones had 60–70% identity to known proteins and 5 out of 20 had no identity to known proteins. Highest identities were shown with proteins from a number of genera, members of which are known to be found in these environments including *Halobacterium, Methanobacterium, Synechocystis, Aquifex, Klebsiella, Proteus, Serratia, Escherichia, Enterobacter* and *Aeropyrum* (data not shown).

With our screening approach we will only detect enzyme activity if the gene is transcribed, translated and sufficient protein is made and folded correctly. Expression detection also depends on the enzyme either being secreted from the cell or cell lysis allowing sufficient to accumulate extracellularly and the enzyme having activity with the substrates used in the screen. Possible transcriptional promoters include the vector T3, T7 and Lac Z promoters. We have made no attempt to optimise transcription by these promoters. It is also possible that the endogenous gene promoters in the cloned inserts may be responsible for transcription. This seems particularly probable when the predicted ORF is far removed from the vector-derived promoter sequences. It also seems intuitively likely that ORFs derived from species distantly related to *E. coli* are unlikely to be expressed well or even at all using *E. coli* as a host.

A clear example of the problems involved in expression screening was our failure to detect the previously known cellulase gene AF067428 from our *Bacillus*

agaradhaerans library even when a PCR analysis showed that it was present. Despite these limitations we readily detected several lipase and cellulase enzyme activities from clones in several libraries. During the course of sequencing these clones we also discovered other putative ORFs with recognisable identity to a wide variety of other proteins. We have also been able to identify mannanases present in other libraries using appropriate screening assays. An advantage of the expression screening strategy is that we have isolated enzymes with very low identity to known sequences. CRATCEL (see below) had only 29% identity at the protein level to data base entries and no significant identity at all at the nucleotide level. An identification strategy based on sequence homology would never identify such a divergent sequence. Even with nucleic acid identities of about 40–70%, typical of the other ORFS identified in this study, a nucleic acid probe or PCR-based identification approach would be difficult. These libraries are clearly a rich resource for the future identification of novel enzymes.

We also looked at the effects of biasing a culture to grow on either a lipid or a cellulose derivative as the sole carbon source. Enrichment of a Lake Nakuru culture by growth on CMC resulted in 1/15,000 clones being cellulase positive compared to 1/60,000 positives in the unenriched library (Table 1). The frequency of lipase activity in a Lake Elmenteita library clones was 1 in 100,000. After enrichment it was 1 in 30,000 clones (Table 1). These two experiments suggest that biasing an environmental sample to express a particular type of enzyme activity does result in an increased frequency for isolating that enzyme activity. Clearly this needs to be repeated with other libraries biased in alternative ways, before generalised rules can be made. It is also important to establish if such an enrichment process results in the loss of gene diversity recovered.

Turning to the enzymes described in this paper residues 57–347 in the new *B. agaradhaerans* cellulase that we unexpectedly discovered, BAGCEL (Fig. 2) contain a motif from the cellulase family 5 pfam00150 group, LMFESVNEPR (see below). One of the cellulases we discovered from an environmental gene library made from Lake Nakuru shows about 95% identity to this cellulase at both the nucleotide and protein level. HKCEL is also a family 5 glycoside hydrolase containing the amino acid motif described above. This motif fits into the degenerate sequence motif characteristic of family 5 as indicated in bold (**L**IV)-(LIV**M**FYWGA)-(LIVM**F**YWGA)-(DN**E**QG)-(LIVMG**S**T)-**x**-**N**-**E**-(**P**V)-(**R**HDNSTLIVFY). Neither BAGCEL or HKCEL appears to contain a known cellulose binding domain.

CRATCEL, discovered from a genomic library made from Crater Lake is a member of the glycoside hydrolase family 9, pfam00759. This group comprises enzymes with several known activities, such as endoglucanase (EC 3.2.1.4) and cellobiohydrolase (EC 3.2.1.91). These enzymes were formerly known as cellulase family E. The organism it is derived from is of course unknown but is probably most closely related to members of the *Fusarium* or *Clostridium* genera. It does not appear to contain a known cellulose binding domain but does contain an *N*-terminal Ig-like region, pfam 02927, found in other cellulases from bacteria including the proteobacteria.

Of the esterase/lipase enzymes described here, ELIP is most closely related to the carboxyesterase type B family having a catalytic triad of serine, a glutamate or aspartate and a histidine. It contains the sequence FGGDAGNVTLFGESAG highlighted in the degenerate sequence motif characteristic of this family, **F**-(**G**R)-**G**-x-x-x-x-(**L**I**V**M)-x-(**L**IV)-x-**G**-x-**S**-(ST**A**G)-**G**(Prosite PDOC00112). LIP1 belongs most closely to pfam00975, a thioesterase family. Lip2 is most related to an acyl-coA thioesterase I precursor of *E. coli* which belongs to the GSDL family of esterases (Upton and Buckley 1995). LIP2 contains the sequence LVVLGDSLSAG which belongs to the degenerate sequence motif characteristic of this family, (**L**IVMFYAG)-(LI**V**MFYAG)-(LI**V**M-FYAG)-(**L**IVMFYAG)-**G**-**D**-**S**-(**L**IVM)-x(1,2)-(T**A**G)-**G.**

Other work not described here shows that the same gene can be isolated from our libraries in different *Sau*3A fragments.

To conclude, we have a library resource made from alkaliphilic and halophilic organisms containing literally billions of clones. These clones contain many uncharacterised enzymes. We have identified more than 1 dozen enzymes having cellulase activity, as well as lipases and mannanases. Future studies involve the expression and biochemical characterisation of protein sequences expressed by the genes isolated in this study. Preliminary work shows that the cellulases are alkalitolerant; HKCEL, for example, still shows enzyme activity above pH 11 (data not shown). Improvements in our expression and cloning strategies are also required. For instance it is possible to clone environmental genomic DNA in vectors allowing expression and secretion in different hosts. It may also be possible to clone cDNA sequences made from mRNAs isolated from uncultivated organisms.

References

Aguilar A, Ingemansson T, Magnien E (1998) Extremophile microorganisms as cell factories: support from the European Union. Extremophiles 2:367–373

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Chen WP, Kuo TT (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res 21:2260

Davies GJ, Dauter M, Brzozowski AM, Bjoernvad ME, Andersen KV, Schuelein M (1998) Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6-A and its cellobiose complex at 2.0-A resolution. Biochemistry 37:1926–1932

Duckworth AW, Grant WD, Jones BE, Steenbergen R (1996) Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiol Ecol 19:181–191

Grant S, Grant WD, Jones BE, Kato C, Li L (1999) Novel archaeal phylotypes from an East African saltern. Extremophiles 3:139–145

Henne A, Schmitz RA, Bömeke 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

Horikoshi K (1999) Alkaliphiles: some applications of their products for biotechnology. Microbiol Mol Biol Revs 63:735–750

Jones BE, Grant WD, Collins NC, Mwatha WE (1994) Alkaliphiles: diversity and identification. In: Priest FG, Ramos-Cormenzana A, Tindal BJ (eds) Bacterial diversity and systematics. Plenum, New York, pp195–230

Jones BE, Grant WD, Duckworth AW, Owenson GG (1998) Microbial diversity of soda lakes. Extremophiles 2:191–200

Margesin R, Schinner F (2001) Potential of halotolerant and halophilic microorganisms for biotechnology. Extremophiles 5:73–83

Marrs B, Delagrave S, Murphy D (1999) Novel appproaches for discovering industrial enzymes. Curr Opin Microbiol 2:241–245

Pitcher DG, Saunders NA, Owen RJ (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett Appl Microbiol 8:151–156

Rees H (2002) Environmental gene screening. PhD thesis, University of Leicester

Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res 28:4317–4331

Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterisation of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43:777–780

Upton C, Buckley JT (1995) A new family of lipolytic enzymes? Trends Biochem Sci 20:178–179