RESEARCH ARTICLE

Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment

Yongfei Hu[1,2], Chengzhang Fu[1,2], Yunpeng Huang[1], Yeshi Yin[1,2], Gong Cheng[1,2], Fang Lei[1,2], Na Lu[1], Jing Li[1], Elizabeth Jane Ashforth[1], Lixin Zhang[1,3] & Baoli Zhu[1]

[1]Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; [2]Graduate University of Chinese Academy of Sciences, Beijing, China; and [3]South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

Correspondence: Baoli Zhu, Institute of Microbiology, Chinese Academy of Sciences, no.1 West Beichen Road, Chaoyang District, Beijing 100101, China. Tel.: +86 10 64 80 7362; fax: +86 10 64 80 7358; e-mail: zhubaoli@im.ac.cn

Received 20 July 2009; revised 1 February 2010; accepted 1 February 2010.
Final version published online 16 March 2010.

DOI:10.1111/j.1574-6941.2010.00851.x

Editor: Michael Wagner

Keywords
metagenomic library; marine sediment; lipolytic; lipase/esterase; fosmid clone.

Abstract

Metagenomic cloning is a powerful tool for the discovery of novel genes and biocatalysts from environmental microorganisms. Based on activity screening of a marine sediment microbial metagenomic library, a total of 19 fosmid clones showing lipolytic activity were identified. After subcloning, 15 different lipolytic genes were obtained; their encoded proteins showed 32–68% amino acid identity with proteins in the database. Multiple sequence alignment and phylogenetic tree analysis demonstrated that most of these predicted proteins are new members of known families of bacterial lipolytic enzymes. However, two proteins, FLS18C and FLS18D, could not be assigned to any known family, thus probably representing a novel family of the bacterial lipolytic enzyme. The activity assay results indicated that most of these lipolytic enzymes showed optimum temperature for hydrolysis at 40–50 °C with *p*-nitrophenol butyrate as a substrate. The lipolytic gene *fls18D* was overexpressed, and the resulting protein FLS18D was characterized as an alkaline esterase. Furthermore, the whole sequence of fosmid pFL18 containing FLS18C and FLS18D was shotgun sequenced, and a total of 26 ORFs on it were analyzed and annotated.

Introduction

A broad range of living organisms, including animals, plants and microorganisms, produce lipolytic enzymes that catalyze the cleavage of ester bonds, releasing fatty acids and glycerols. These enzymes such as esterase (EC 3.1.1.1) and lipase (EC 3.1.1.3) are currently receiving considerable attention because of their potential industrial applications in cosmetic production, laundry, food flavoring, oil chemistry, fine chemistry, pharmaceutical and paper industries, as well as in biodiesel production (Bornscheuer, 2002; Jaeger & Eggert, 2002).

A common characteristic of these enzymes is that they contain a catalytic triad formed by Ser, His and Asp residues. The Ser residue usually appears in the conserved motif G-Xaa-S-Xaa-G. Lipolytic enzymes from prokaryotes have been classified into eight families (I–VIII) according to their amino acid sequences (Arpigny & Jaeger, 1999). Recently identified lipases and esterases such as LipG (Lee *et al.*, 2006), EstA (Chu *et al.*, 2008) and LipEH166 (Kim *et al.*, 2009) have further enriched these lipolytic enzyme families. The search for new esterases and lipases will increase the diversity of lipolytic enzymes and therefore aid the selection of suitable biocatalysts for challenging reactions (Ranjan *et al.*, 2005).

Culturable microorganisms have been a major source of novel enzymes and diverse biological active compounds; however, culturable microorganisms account for only a small fraction of the microbial diversity on Earth, which is limiting the spectrum of the discovery of new enzymes. By contrast, the vast majority are unculturable microorganisms (> 99%), which represent a large gene pool for biotechnological exploitation (Amann *et al.*, 1995; Ferrer *et al.*, 2005). Metagenomics, a cultivation-independent method, is regarded as one of the best approaches to access and investigate this potential without the need for culturing particular microorganisms (Lorenz & Eck, 2005). Various lipolytic enzymes have been identified using metagenomic libraries prepared from different environmental samples, including soils (Lämmle *et al.*, 2007), tidal flat sediments (Lee *et al.*,

FEMS MICROBIOLOGY ECOLOGY

2006), activated sludge (Roh & Villatte, 2008), surface seawater (Chu *et al.*, 2008) and pond and lake water (Rees *et al.*, 2003; Ranjan *et al.*, 2005). However, the origins of most of the lipolytic enzymes in these metagenomic libraries were from microorganisms living in terrestrial environments; very few of them were from deep-sea marine environments, especially from the marine sediment.

Deep-sea marine sediment, covering more than two-thirds of the Earth's surface, is considered to be an extreme environment. The prokaryotes in subseafloor sediments have been estimated to constitute as much as one-third of the Earth's total living biomass (D'Hondt *et al.*, 2004). Microorganisms inhabiting this environment are continuously exposed to extremes in pressure, salinity, temperature and nutrient availability (Kennedy *et al.*, 2008), and are therefore expected to have diverse biochemical and physiological characteristics, providing us with a vast resource for mining novel genes and biocatalysts.

In our previous study, we constructed a fosmid metagenomic library (named IMCAS-F003) from South China Sea marine sediment microbial DNA (Hu *et al.*, in press), and a novel low-temperature-active alkaline esterase was obtained in a preliminary screening step (unpublished data). Here, we focused on the large-scale screening of lipolytic genes from this library, with a total of 15 new lipolytic enzymes identified. Our results demonstrate that the metagenomic strategy is a very powerful tool for the discovery of new enzymes, and microorganisms in the marine sediment environments are a valuable resource for lipolytic enzymes.

Materials and methods

Lipolytic activity clones' screening and subcloning

The metagenomic library used for screening was constructed from deep-sea sediment microbial DNA in our previous study (Hu *et al.*, in press). The sampling site, 110°28′47.231″E and 17°34′11.603″N according to the global positioning system location, was located in the Qiongdongnan basin, South China Sea, water depth 778.5 m. In this study, the fosmid clones were plated on Luria–Bertani (LB) agar medium supplemented with 1% tributyrin (Park *et al.*, 2007) and 12.5 μg mL^{-1} chloramphenicol. Clones showing lipolytic activity were identified by the formation of clear zones around colonies after incubation for 24–48 h at 30 °C. For subcloning, the fosmid DNA of 19 lipolytic-positive clones named pFL1–pFL19 were extracted using the Omega Plasmid Mini Kit I and partially digested with Sau3AI. The DNA fragments of 1.5–4 kb were recovered from an agarose gel (BioSpin Gel Extraction Kit, Bioflux) and ligated into BamHI-digested pUC119 (TaKaRa). The ligation products were transformed into *Escherichia coli* DH5α, and the transformants were spread onto LB supplemented with 100 μg mL^{-1} ampicillin and 1% tributyrin to select the subclones that showed clear zones around the colonies. The positive subclones were named after their respective fosmid clones i.e. pFLS1–pFLS19.

Lipolytic gene analysis and phylogenetic tree construction

The positive subclones were sequenced from two directions using the standard primers M13F and M13R in the pUC119 vector. Sequence assembly and analysis was performed using LASERGENE package, version 7.10 (DNA Star). ORFs in each assembled sequence were identified via the OPEN READING FRAME FINDER program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The ORFs identified were analyzed using the BLAST program (Ye *et al.*, 2006) of the National Center for Biotechnology Information (NCBI). The BLAST algorithm was also used for homology searches against the lipase engineering database. Conserved domains and patterns were analyzed using a conserved protein domain database (Marchler-Bauer & Bryant, 2004) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Multiple sequence alignments were carried out using CLUSTAL W, version 1.83 (Thompson *et al.*, 1994), in combination with GENEDOC (http://www.psc.edu/biomed/genedoc). Phylogenetic and molecular evolutionary analysis was conducted with the neighbor-joining method using MEGA version 4 (Tamura *et al.*, 2007). Bootstrapping (1000 replicates) was used to estimate the reliability of phylogenetic reconstructions (Felsenstein, 1985).

Optimum temperature assay

Culture supernatants of *E. coli* carrying the subclones were directly used for the determination of the optimum temperature of the lipolytic enzymes. *Escherichia coli* carrying only pUC119 was used as a control. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE, 12%) was performed to detect the lipolytic enzyme proteins encoded by individual subclones. Catalytic activity was examined using *p*-nitrophenol (*p*-NP) butyrate as a standard substrate. The substrate mixture consisted of 10 mM *p*-NP butyrate in acetonitrile, 20 mM sodium phosphate buffer (pH 7.4), 0.1 M NaCl and 0.2% Triton X-100. The standard assay mixture contained 900 μL of substrate mixture and 100 μL of individual culture supernatants. The activity of enzymes in the temperature range of 15–55 °C (with 5 °C interval) was determined at 37 °C by measuring the $A_{410\,nm}$ of *p*-NP released, and the measurement was carried out at least three times. One unit (U) of enzyme activity was defined as the amount of enzyme required to release 1 μM *p*-NP min^{-1} under the assay conditions.

Overexpression and characterization of FLS18D

The *fls18D* gene was amplified using the fosmid pFL18 as a template with the primers: 5′-GGAATTCCATATGCAGTCT GAGGCTTCCTCG-3′, 5′-CCCAAGCTTCTAGAATCCTTC GCCCGCGACG-3′ (NdeI and HindIII restriction enzyme sites are underlined). The PCR product was digested by NdeI and HindIII and the product was purified and cloned into the expression vector pET-28a (+), and the recombinant plasmid was transformed into *E. coli* BL21 (DE3). After 4 h of induction with 0.5 mM isopropyl-β-D-thiogalactoside, the target protein was first purified by Ni-NTA affinity chromatography, and then by gel filtration chromatography on Sephadex G-200 (GE Healthcare). The protein concentration was measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE).

The substrate specificity of the purified enzyme was investigated using various chain lengths of *p*-NP esters, C2–C16 (Sigma). The standard reaction mixture contained 0.01 mL of 10 mM *p*-NP butyrate in acetonitrile, 0.98 mL of 50 mM Tris-HCl (pH 8.0) buffer and 0.01 mL enzyme in a final volume of 1 mL (Park *et al.*, 2007). The activity of the enzyme was determined at 37 °C by measuring the $A_{410\,nm}$ of *p*-NP released. The background hydrolysis of the substrate was deduced using a reference sample with a composition identical to the reaction mixture, but with no enzyme added. The effects of pH on hydrolysis activity were examined over a pH range of 5.0–11.0. The buffers used were citrate buffer (pH 5.0–6.0), sodium phosphate buffer (pH 6.0–8.0), Tris-HCl buffer (pH 8.0–9.0) and sodium bicarbonate buffer (pH 9.0–11.0). The effects of NaCl on esterase activity were determined in the salt concentration range 0.2–1.4 M. The effects of detergents on esterase activity were examined using Tween 20, Tween 80, Triton X-100, SDS and CTAB at final concentrations of 0.2–5%. The effects of organic solvents on esterase activity were examined using dimethyl sulfoxide (DMSO), dimethylformamide (DMF), methanol, methanal, ethanol and acetone at final concentrations of 20% and 40%. All the tests were performed using the standard assay, and the activity of the enzyme without additives added in the reaction mixture was defined as 100%.

Fosmid sequencing and sequence analysis

Fosmid pFL18 was purified using the Plasmid Mini Kit I (Omega, Doraville, GA), and mechanically sheared DNA of 1.5–3 kb was ligated to the pUC119 vector for shotgun sequencing. The sequence coverage was above 10-fold, and the resulting DNA sequences were assembled using the LASERGENE package as mentioned above. ORF analysis was performed using the ORF FINDER program as mentioned above. The predicted function of ORFs was annotated using the BLASTX search of the NCBI nonredundant protein database and finalized by manual inspection. For clusters of orthologous groups of proteins (COG) assignments (Tatusov *et al.*, 2000), sequences were compared with the COG database using RPSBLAST (-p F) under a cutoff value of 1e − 5.

Nucleotide sequence accession numbers

The nucleotide sequence of the lipolytic gene of FLS1, FLS2, FLS3, FLS4, FLS5, FLS6, FLS9A, FLS9B, FLS10, FLS13, FLS14, FLS15 and FLS17 and the whole sequence of fosmid pFL18 have been submitted to the GenBank with accession numbers of FJ483455, FJ483456, FJ483457, FJ483458, FJ483459, FJ483460, FJ483461, FJ483462, FJ483463, FJ483465, FJ483466, FJ483467, FJ483468 and FJ483469, respectively.

Results

Screening of lipolytic activity clones

About one-fifth of the clones of the IMCAS-F003 library (total clone number, 200 000) were screened for lipolytic activity. Among the 40 000 clones screened, 19 clones (named pFL1–pFL19) showed a clear zone around the colonies, representing a hit rate of approximately one positive per 2100 tested clones (corresponding to one lipolytic gene per 75 6 Mb DNA, assuming an average insert size of 36 kb).

Subcloning and lipolytic gene sequence analysis

Out of the 19 positive fosmid clones identified, 14 were successfully subcloned and screened for lipolytic activity in the secondary libraries. Five fosmid clones could not be obtained for lipolytic activity subclones, even after several trials at subcloning. The subclones showing lipolytic activity were sequenced from both ends and analyzed using various computational tools (Table 1). Both pFLS9 and pFLS18 harbor another incomplete ORF encoding an enzyme belonging to lipase/esterase superfamily proteins based on the result of the BLAST search. The full length of these two proteins named FLS9B and FLS18C was obtained by primer reverse walking sequencing using pFL9 and pFL18 as templates. One clone, pFLS19, had the same nucleotide acid sequences as pFLS4 and was thus excluded. In total, 15 different putative lipolytic genes were obtained on 13 subclones. The organization of the inserted fragments carried by these lipolytic subclones is shown in Fig. 1.

The ORF size of the putative lipolytic genes varied from 670 to 1608 bp, and the corresponding proteins translated were from 259 to 535 amino acids. BLASTX of the translated protein sequences showed 32–68% identity at the amino acid level with the proteins from various bacteria (Table 1). Seven matched sequences were derived from uncultured bacterium based on metagenome screening, and eight

Table 1. Lipolytic enzymes from metagenomic library IMCAS-F003 and similar enzymes in GenBank

Gene name	ORF range*	ORF (bp)†	G+C (%)‡	ORF (aa)§	Best match (accession no.)	Organism	Identity (%)	e-value
fls1	373–1395	1023	55.33	340	Esterase (ACB11219)	Uncultured bacterium	62	1e − 123
fls2	1247–2854	1608	58.77	535	Carboxylesterase (ZP_01385717)	*Chlorobium ferrooxidans* DSM 13031	36	1e − 82
fls3	1724–612	1113	59.21	370	Lipolytic enzyme (ZP_01387773)	*Geobacter* sp. FRC-32	36	2e − 30
fls4	141–1025	885	59.1	294	Putative esterase (ZP_01615262)	Marine gammaproteobacterium HTCC2143	63	5e − 93
fls5	1151–246	906	54.09	301	Lipase/esterase (ABQ11268)	Uncultured bacterium	55	3e − 92
fls6	698–1810	1113	57.77	370	∝/β hydrolase fold-3 domain protein (YP_001753536)	*Methylobacterium radiotolerans* JCM 2831	32	2e − 39
fls9A	1243–2178	936	61.75	311	Lipase/esterase (ACF04196)	Uncultured bacterium	59	2e − 105
fls9B	256–1203	948	65.19	315	∝/β hydrolase fold domain protein (EDX75541)	*Microcoleus chthonoplastes* PCC 7420	51	1e − 88
fls10	1505–582	924	71.65	307	Lipase/esterase (AAS77236)	Uncultured bacterium	68	8e − 110
fls13	360–1412	1113	57.95	370	∝/β hydrolase fold-3 domain protein (YP_001753536)	*Methylobacterium radiotolerans* JCM 2831	32	2e − 39
fls14	202–1137	936	65.81	311	Lipase/esterase (ACF04196)	Uncultured bacterium	60	2e − 102
fls15	147–1070	924	71.21	307	Lipase/esterase (AAS77236)	Uncultured bacterium	67	1e − 109
fls17	1155–250	906	53.09	301	Lipase/esterase (ABQ11268)	Uncultured bacterium	56	8e − 92
fls18C	1616–837	780	60.64	259	Putative poly(3-hydroxybutyrate) depolymerase (ZP_01876939)	*Lentisphaera araneosa* HTCC2155	38	2e − 44
fls18D	2567–1677	891	59.15	296	Putative poly(3-hydroxybutyrate) depolymerase (ZP_01876939)	*Lentisphaera araneosa* HTCC2155	49	4e − 70

*The nucleotide range of the ORFs most probably responsible for the lipolytic phenotype within the inserted DNA of individual subclones.
†Length of the predicted lipolytic enzyme in nucleotide acids.
‡GC content of ORFs of the predicted lipolytic enzyme.
§Length of the predicted lipolytic enzyme in amino acids.

sequences were annotated from whole-genome sequences. Comparison of the amino acid sequences of the 15 putative lipolytic enzymes indicated that FLS3, FLS6 and FLS13 had high similarity to each other, with amino acid identity from 83.8% to 95.7%. Similarly, FLS10 and FLS15 shared 97.1% amino acid identity, and FLS5 and FLS17 shared 98%. Thus, these seven proteins constituted three groups. The other enzymes from the library could not be grouped due to low identity to each other.

Phylogenetic tree and multiple-sequence alignment

A phylogenetic tree was constructed using 74 bacterial lipolytic enzymes representing 12 different families including family I–VIII (Arpigny & Jaeger, 1999), LipG (Lee *et al*., 2006), EstA (Chu *et al*., 2008), LipEH166 (Kim *et al*., 2009) and EstA (authors' unpublished data). As shown in Fig. 2, the 15 enzyme sequences grouped into four different families. Multiple-sequence alignment of the predicted amino acid sequences of these enzymes is shown in Fig. 3.

Lipolytic proteins of FLS5 and FLS17 were grouped into family V in the phylogenetic tree. Enzymes grouped in this family originate from mesophilic bacteria [*Acetobacter pasteurianus* (AB013096) as well as from cold-adapted *Moraxella* sp. (X53869), *Psychrobacter immobilis* (X67712) or heat-adapted (*Sulfolobus acidocaldarius*) (AF071233) organisms. However, no marine microorganism origin lipolytic enzymes belonging to this family have been identified previously.

FLS2 was assigned to family VII, a family comprising rather large bacterial esterases (55 kDa). Members of this family, such as esterase of *B. subtilis*, showing high hydrolysis activity to *p*-nitrobenzyl ester, can thus be used to remove the *p*-nitrobenzyl ester used as a protecting group in the synthesis of β-lactam antibiotics. Therefore, it would be interesting to investigate the possibility of FLS2 having a similar catalytic function.

Interestingly, most of the enzymes, 66.7% (10 sequences), grouped into family IV, a family displaying a striking amino acid sequence similarity to the mammalian hormone-sensitive lipase (HSL). Multiple-sequence alignment revealed that all of them contained the lipase-conserved catalytic triad residues Asp, His and the catalytic nucleophile Ser in the consensus pentapeptide G-X-S-X-G. For FLS3, FLS4, FLS6, FLS10, FLS13 and FLS15, the active site Asp was replaced by Glu, which is common as in other members of the family, such as the lipolytic enzyme of uncultured bacterium (AAX37296) and esterase of *Ralstonia solanacearum* (ZP_00943646). These results suggest that all of these enzymes are new members of the HSL family.

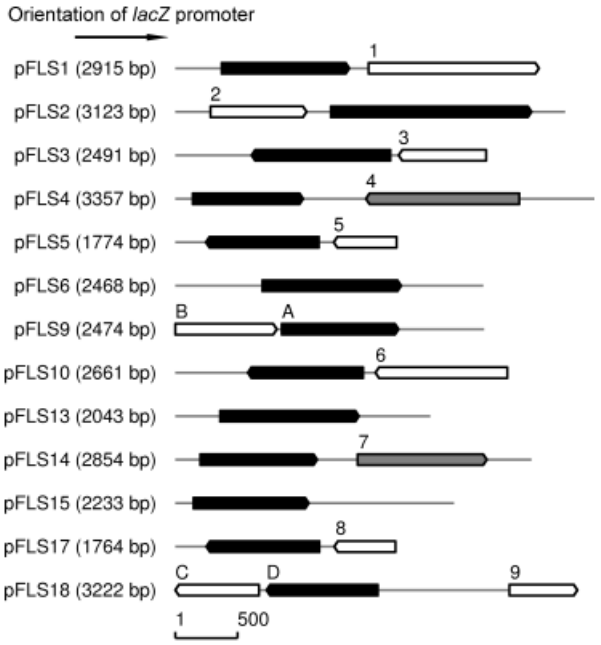

Fig. 1. Schematic diagram of the organization of the inserted fragments carried by the lipolytic subclones. The arrows represent the length, location and orientation of the ORFs. Lipolytic ORFs are indicated by shaded arrows, and other complete and partial ORFs adjacent to lipolytic ORFs are numbered and indicted by gray arrows and blank arrows, respectively. The numbered genes are: 1, sulfatase; 2, hypothetical protein; 3, hypothetical protein VspiD_05225; 4, short-chain dehydrogenase/reductase SDR; 5, pyruvate formate-lyase activating enzyme; 6, sulfatase; 7, short-chain dehydrogenase/reductase SDR; 8, pyruvate formate-lyase activating enzyme; and 9, nuclease inhibitor. Note that both pFLS9 and pFLS18 harbor two tandem lipolytic genes. A, lipolytic gene *fls9A*; B, lipolytic gene *fls9B*; C, lipolytic gene *fls18C*; and D, lipolytic gene *fls18D*.

The homology search revealed that both FLS18C and FLS18D showed maximum identity to putative poly(3-hydroxybutyrate) depolymerase from *Lentisphaera araneosa* HTCC2155. Another close match of these two proteins was lipase/esterase (AAX37300) from uncultured bacterium, showing 39% and 43% identity to FLS18C and FLS18D, respectively. Phylogenetic analysis indicated that FLS18C, FLS18D and their closest related members did not belong to any known family. Hence, we assumed that this group may constitute a new family of bacterial lipolytic enzymes.

Effects of temperature on enzyme activity

SDS-PAGE analysis of crude lysates of *E. coli* carrying individual subclones showed that proteins of the expected molecular weight were expressed compared with *E. coli* carrying only pUC119. Five examples are shown in Fig. 4. Because the lipolytic enzymes can be secreted into the culture broth, the culture supernatants of *E. coli* carrying the subclones were directly used for the activity assay. Most of these enzymes showed optimum temperature for hydrolysis at 4050 °C; two of them (namely

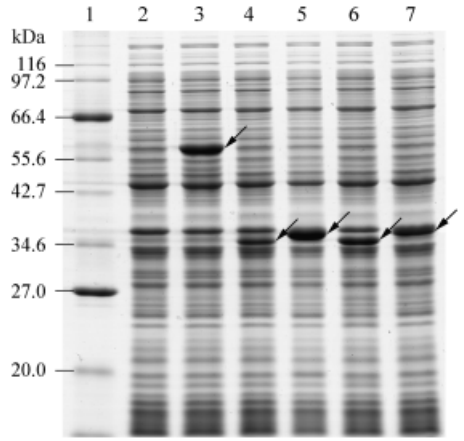

Fig. 2. SDS-PAGE of crude lysates of *Escherichia coli* carrying individual subclones. Lane 1, molecular weight standards; lane 2, *E. coli* cells with pUC119 only; lane 3, FLS2; lane 4, FLS4; lane 5, FLS10; lane 6, FLS17; and lane 7, FLS14. Arrows, expected proteins.

FLS10 and FLS15) showed optimum temperature at 30 °C (Table 2). Members of the three groups of enzymes, FLS3, FLS6 and FLS13, FLS5 and FLS17, FLS10 and FLS15, as mentioned above, showed the same optimum temperature.

Characterization of the FLS18D protein

The lipolytic gene *fls18D* was overexpressed and the resulting protein FLS18D was purified (Fig. 5a). This enzyme exhibited the highest hydrolysis activity toward *p*-NP butyrate (345.9 ± 38.3 U mg^{-1}). High hydrolysis activity was also obtained with other short- and middle-chain *p*-NP esters, such as *p*-NP acetate, *p*-NP captylate and *p*-NP caprylate, but very low activity was observed with longer-chain *p*-NP esters (Fig. 5b). The optimum activity of FLS18D occurred at an alkaline pH of 8, which was about 1.5 times higher than that at pH 7.0. Approximately 40% activity remained even at pH 10, compared with that at pH 8.0 (Fig. 5c). The hydrolysis activity of FLS18D increased to about 170% in the presence of 0.6 M NaCl, but the addition of 1.4 M NaCl decreased the enzyme activity to 28% (Fig. 5d). Addition of 1% w/v Tween 20, Tween 80 and Triton X-100 significantly increased enzyme activity to 264%, 213% and 225%, respectively. The hydrolysis activity, however, was almost completely inactivated by the same concentration of SDS and CTAB (Table 3). FLS18D exhibited more tolerance to DMSO and DMF than the other organic solvents examined, and 20% concentration of DMSO even increased the activity to about 127%. FLS18D was sensitive to methanal; 40% concentration of methanal inactivated the enzyme completely.

Sequence analysis of fosmid pFL18

In order to explore more genetic information on the FLS18C- and FLS18D-containing fosmid, pFL18 was

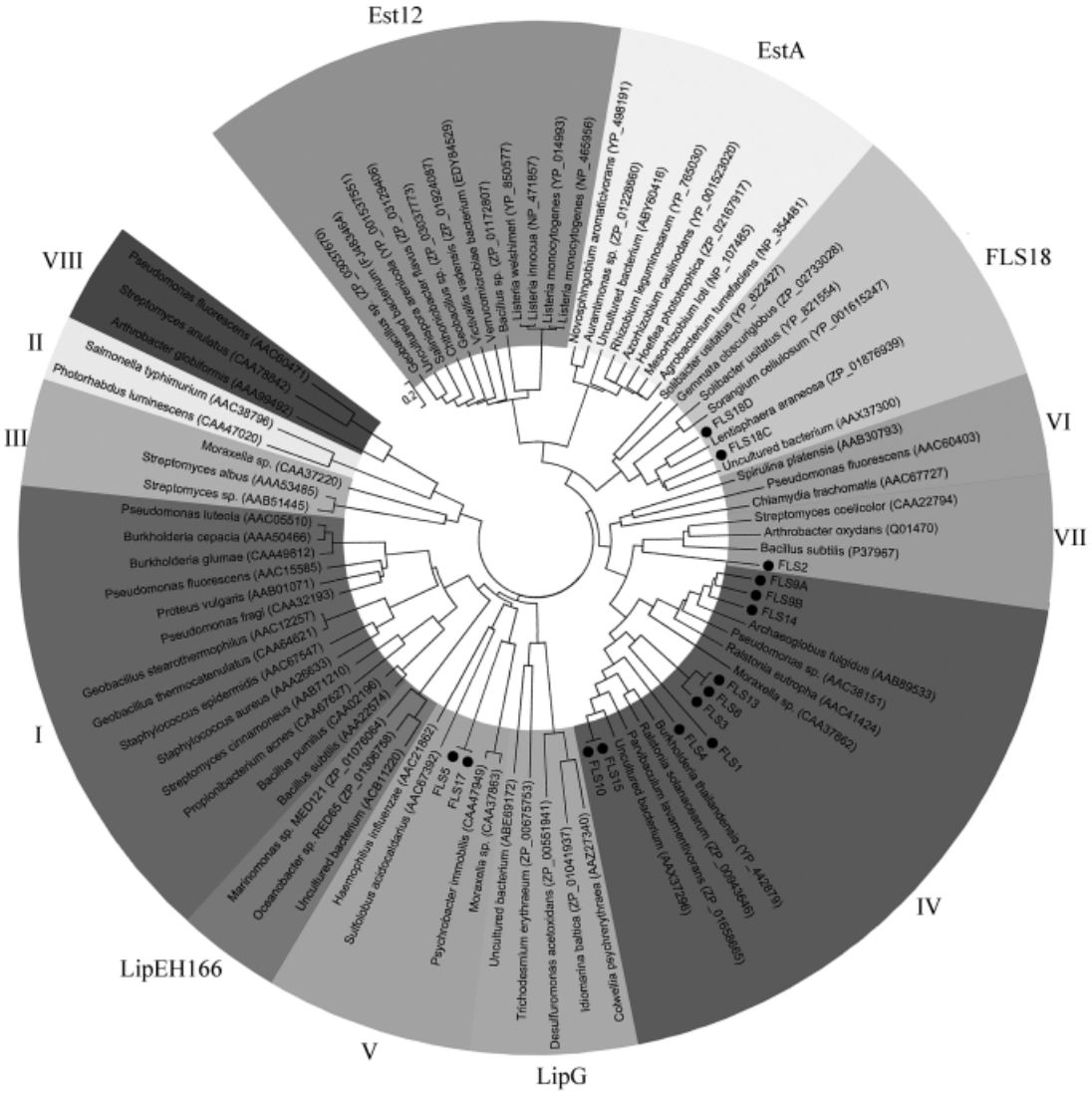

Fig. 3. Phylogenetic tree of lipolytic enzymes in this study and representative members of other previously identified families. Except for the 15 lipolytic enzymes, all the protein sequences were retrieved from GenBank. The phylogenetic positions of the enzymes in this study are indicated by solid circles (•). The length of the branches indicates the divergence among the amino acid sequences. The scale bar corresponds to 20% estimated sequence divergence.

randomly shotgun sequenced. This fosmid contained a 37282-bp foreign DNA fragment with a G+C content of 58.06%. A total of 25 complete ORFs (orf1–orf25) and one incomplete ORF (orf26) were found on the insert DNA (for details, see Supporting Information, Table S1). Most of the genes (17 orfs) exhibited a low identity of $< 50\%$ with database hits below the cutoff value of $1e-5$, and 13 genes were annotated as hypothetical proteins or conserved hypothetical proteins according to their BLASTX results. Among the matching genes in the database, six genes were of *Planctomyces* origin and six genes were of *Proteobacteria* origin. COG category analysis indicated that only half of the genes had significant hits below the e-value of $1e-5$ (for details, see Table S2). No phylogenetic marker genes or gene clusters related to a specific biological process were found.

Discussion

The library screening procedure for lipolytic activity was based on functional expression. This activity-based screening method allows for the discovery of hitherto entirely novel genes. In this study, a total of 15 different lipolytic genes were identified, and these genes were completely novel with a low identity (32–68%) to those in the database. Former studies on lipolytic genes have been focused mainly on terrestrial and aquatic environments. To our knowledge, only five lipolytic enzymes, h1Lip1 (Hardeman & Sjöling,

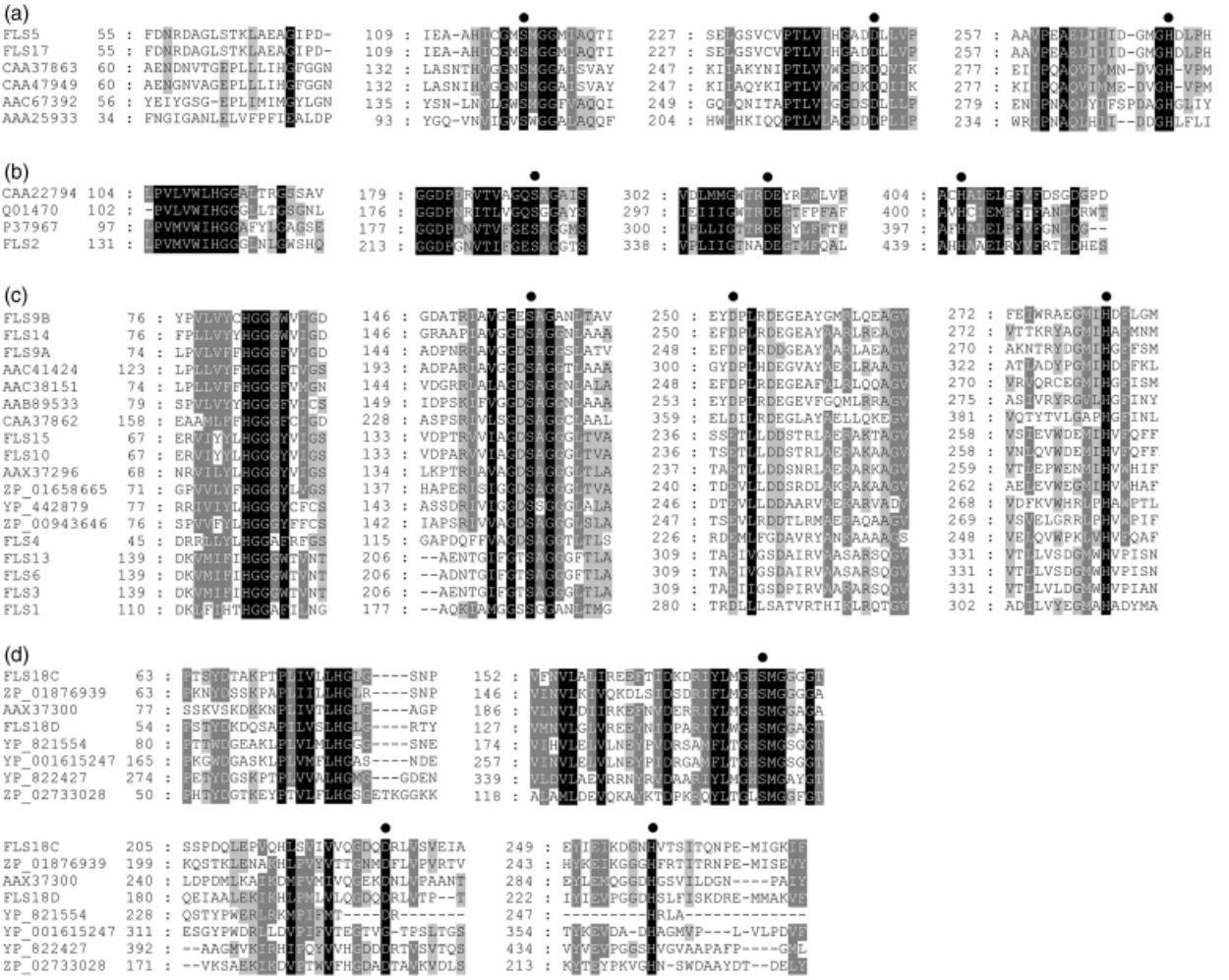

Fig. 4. Multiple sequence alignment of the partial amino acid sequences containing the conserved blocks of the lipolytic enzymes. The numbers before each sequence indicate the amino acid position of the displayed region within the complete ORF. Residues identical to the consensus are shaded. The first block in each family contains putative oxyanion hole residues, and the last three blocks contain putative active sites. •, putative amino acid residues belonging to the catalytic triad. (a) Sequence alignment of FLS5 and FLS17 with representative members of family V. (b) Sequence alignment of FLS2 with representative members of family VII. (c) Sequence alignment of FLS1, FLS3, FLS4, FLS6, FLS9A, FLS9B, FLS10, FLS13, FLS14 and FLS15 with representative members of family IV. (d) Sequence alignment of FLS18C and FLS18D with their closest members.

2007), EM2L8 (Park *et al.*, 2007), estAT1 and estAT11 (Jeon *et al.*, 2008) and EML1 (Jeon *et al.*, 2009), have been reported from the deep-sea sediment metagenome before this study.

Considering the number of positive lipolytic fosmid clones and the genes obtained, our frequency of hits was in fact quite high compared with studies that have screened metagenomic libraries for the same activities using the *E. coli* as host (see the review by Steele *et al.*, 2009). This difference can be partly attributed to the procedure of cloning and screening, such as for example the source of the environmental sample, the composition of the inherent microorganisms, representative of the extracted DNA, the efficiency of ligation and transformation or the screening methods (Lämmle *et al.*, 2007). Besides these, we suggest that screening of a large insert fosmid library containing more information is more efficient than a small insert DNA library, which requires screening of a large number of clones in the search for enzymes. Furthermore, we propose that the lipolytic genes are prevalent in marine sediments, at least in our site. It was shown in our previous study that sulfate-reducing bacteria (SRB) were very abundant in this sampling site (Hu *et al.*, in press), and the richness of SRB in marine sediments has also been reported elsewhere (Purdy *et al.*, 2003; Kjeldsen *et al.*, 2007; Leloup *et al.*, 2009). These microorganisms, responsible for up to 50% of the organic carbon mineralization in marine sediments, may use volatile fatty acids together with hydrogen as electron donors (Sorensen *et al.*, 1981; Jørgensen, 1982). Therefore, we consider that microorganisms producing lipolytic enzymes,

releasing fatty acids, could be important in generating electron donors for the sulfate-reduction process implemented by these SRB.

Table 2. Optimum temperatures of lipolytic activity of the culture supernatants from *Escherichia coli* carrying individual subclones

Name	Optimum temperature (°C)	Activity (U mL^{-1})
pFLS1	45	13.5 ± 1.2
pFLS2	50	2.3 ± 0.4
pFLS3	45	3.1 ± 0.5
pFLS4	40	2.9 ± 0.6
pFLS5	50	20.6 ± 1.6
pFLS6	45	4.7 ± 0.6
pFLS9	50	66.0 ± 3.2
pFLS10	30	46.8 ± 5.5
pFLS13	40	8.3 ± 2.4
pFLS14	40	29.3 ± 1.9
pFLS15	30	34.6 ± 3.8
pFLS17	50	18.8 ± 2.0
pFLS18	45	14.1 ± 2.7

Given that most of the enzymes in this study belonged to family IV (mammalian HSL), we infer that members of this family are very common in a marine sediment environment, but the reason for this particular abundance was not clear. The sequence conservation of genes belonging to this family was once thought to be linked to cold adaption. However, esterases from psychrophilic, mesophilic and thermophilic microorganisms share common motifs with HSL, indicating that temperature adaptation is not responsible for such extensive sequence conservation (Arpigny & Jaeger, 1999).

As the source of the genes was a deep-sea sediment sample, we assumed that the lipolytic enzymes presented here are cold-adapted enzymes. However, all the lipolytic enzymes have optimum temperature for activity in the mesophilic or the thermophilic range. A similar phenomenon has been reported in other studies, for example, the cold-adapted esterase EM2L8 from a deep-sea sediment metagenome showed an optimum temperature at 50–55 °C (Park *et al.*, 2007) and the low-temperature active lipase hiLip1 showed optimal activity at 35 °C (Hardeman & Sjöling, 2007). These studies, together with the study presented here,

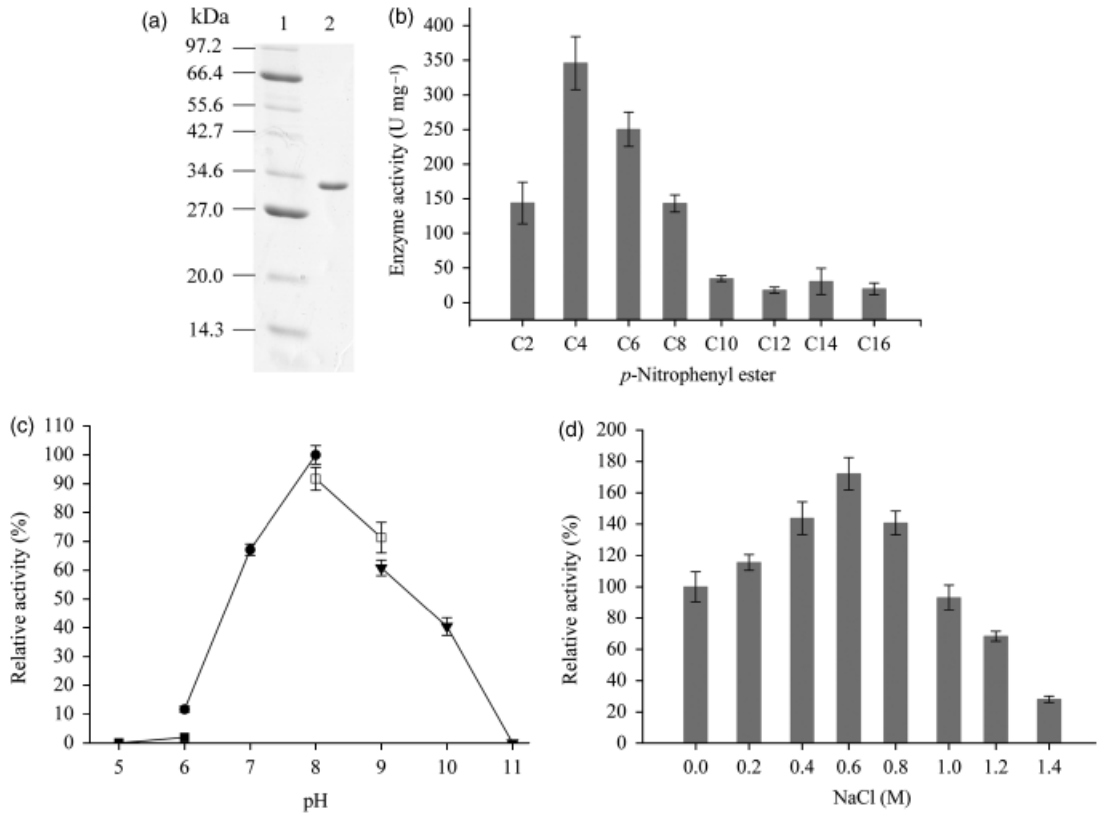

Fig. 5. The purified protein and characteristics of FLS18D. (a) SDS-PAGE analysis of the purified FLS18D protein. Lane 1, molecular weight standards; lane 2, the protein purified by gel filtration chromatography. (b) Substrate specificity of FLS18D. Hydrolytic activity was determined toward different chain lengths of *p*-NP esters: acetate (C2), butyrate (C4), caproate (C6), caprylate (C8), caprate (C10), laurate (C12), myristate (C14) and palmitate (C16). (c) Effects of pH on the enzyme activity of FLS18D. Buffers used were citrate buffer (■) for pH 5.0–6.0, sodium phosphate buffer (●) for pH 6.0–8.0, Tris-HCl buffer (□) for pH 8.0–9.0 and sodium bicarbonate buffer (▼) for pH 9.0–11.0. (d) Effects of NaCl concentration on the hydrolysis activity.

Table 3. Effects of detergents and organic solvents on the esterase activity of FLS18D

Detergents (w/v)	Relative activity (%)	Organic solvents (v/v)	Relative activity (%)
Tween 20 (0.2%)	215.2 ± 10.1	DMSO (20%)	127.5 ± 8.9
Tween 20 (1%)	264.5 ± 9.8	DMSO (40%)	40.3 ± 2.8
Tween 20 (5%)	176.9 ± 11.4	DMF (20%)	84.3 ± 5.3
Tween 80 (0.2%)	178.5 ± 6.5	DMF (40%)	38.7 ± 4.1
Tween 80 (1%)	213.0 ± 13.6	Methanol (20%)	54.9 ± 2.4
Tween 80 (5%)	139.4 ± 7.2	Methanol (40%)	17.2 ± 1.1
Triton X-100 (0.2%)	116.4 ± 9.9	Methanal (20%)	3.8 ± 0.8
Triton X-100 (1%)	225.0 ± 14.7	Methanal (40%)	0
Triton X-100 (5%)	127.2 ± 8.7	Ethanol (20%)	44.4 ± 2.1
SDS (0.2%)	11.2 ± 1.2	Ethanol (40%)	13.9 ± 0.7
SDS (1%)	3.7 ± 0.7	Acetone (20%)	33.4 ± 2.3
SDS (5%)	0	Acetone (40%)	9.6 ± 0.6
CTAB (0.2%)	12.6 ± 1.9		
CTAB (1%)	3.8 ± 0.3		
CTAB (5%)	1.9 ± 1.1		

support the fact that the optimal temperature of activity for an enzyme is usually higher than the optimal growth temperature for the organism that is the source of the enzyme (Sheridan *et al.*, 2000). In addition, the results that FLS3, FLS6 and FLS13, FLS5 and FLS17, FLS10 and FLS15 showed the same optimum temperature, together with the information of their sequence comparison, suggest that these three groups of enzymes were probably derived from three different groups of phylogenetically closely related marine microorganisms living in the same habitat.

FLS18C, FLS18D and their closest BLAST match potentially constitute a new bacterial lipolytic enzyme family. The expression and characterization of FLS18D provided the first experimental data for a member of this potential new family. Lipases, by definition, have the ability to hydrolyze long-chain acylglycerols ($\geq$C10), whereas esterases hydrolyze ester substrate with short-chain fatty acids ($\leq$ C10) (Verger, 1997). The enzyme identified here preferred short- and middle-chain *p*-NP esters as a substrate, and the optimum hydrolysis activity of FLS18D occurred at an alkaline range, indicating that FLS18D was an alkaline esterase. In addition, characteristics such as activation by the appropriate concentration of NaCl, Tween 20, Tween 80 and Triton X-100, and tolerance to organic solvents DMSO and DMF suggested that esterase FLS18D is a very important candidate for industrial applications.

We expected to obtain more genetic information from the whole sequence of the fosmid containing FLS18C and FLS18D, to understand their biological function and identify their microorganism of origin. However, most of the genes on the fosmid had low identity to their hits in the database, and for many of them, the functions were unknown or were simply classified as hypothetical proteins. In addition, the microorganism of origin could not be identified because the insert DNA did not carry any phylogenetic marker genes. We speculate that this DNA fragment originated from an unknown marine microorganism belonging to *Proteobacteria* or *Planctomyces* based simply on the protein-based taxonomy.

In conclusion, a total of 19 lipolytic-active positive fosmid clones were screened out using an activity-based method from a deep-sea sediment microbial metagenomic library, and 15 lipolytic enzyme genes were obtained from the secondary subclone libraries. All of the enzymes are new members of different bacterial lipolytic families, and most of the enzymes were assigned to bacterial lipolytic family IV. The optimum temperature of these enzymes was higher than their cold environment origin. Characterization of the enzyme FLS18D indicated that it was an alkaline esterase and showed the best tolerance to DMSO. The study here broadens the diversity of lipolytic genes and demonstrates that marine sediments are a very important source of novel lipolytic genes, especially those from uncultured microorganisms. Further overexpression, purification and biochemical characterization of these lipolytic enzymes should provide clues of their potential biotechnological applications.

Acknowledgements

This work was supported in part by the National High Technology Research and Development Program of China (2007AA09Z443 and 2007AA021301) and the Knowledge Innovation Project of the Chinese Academy of Sciences (KSCX2-YW-G-022 and O62A131BB4). B.Z. and L.Z. are awardees for the Hundred Talents Program.

Authors' contribution

Y.H. and C.F. contributed equally to this work.

References

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

Arpigny JL & Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* **343**: 177–183.

Bornscheuer UT (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* **26**: 73–81.

Chu X, He H, Guo C & Sun B (2008) Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl Microbiol Biot* **80**: 615–625.

D'Hondt S, Jørgensen BB, Miller DJ *et al.* (2004) Distributions of microbial activities in deep subseafloor sediments. *Science* **306**: 2216–2221.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Ferrer M, Martínez-Abarca F & Golyshin P (2005) Mining genomes and 'metagenomes' for novel catalysts. *Curr Opin Biotech* **16**: 588–593.

Hardeman F & Sjöling S (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol Ecol* **59**: 524–534.

Hu YF, Fu CZ, Yin YS, Cheng G, Lei F, Yang X, Li J, Ashforth EJ, Zhang LX & Zhu BL (in press) Construction and preliminary analysis of a deep-sea sediment metagenomic fosmid library from Qiongdongnan Basin, South China Sea. *Mar Biotechnol*.

Jaeger KE & Eggert T (2002) Lipases for biotechnology. *Curr Opin Biotech* **13**: 390–397.

Jeon JH, Kim JT, Kang SG, Lee JH & Kim SG (2008) Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar Biotechnol* **11**: 307–316.

Jeon JH, Kim JT, Kim YJ *et al.* (2009) Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biot* **81**: 865–874.

Jørgensen BB (1982) Mineralization of organic matter in the sea bed – the role of sulphate reduction. *Nature* **296**: 643–645.

Kennedy J, Marchesi JR & Dobson ADW (2008) Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb Cell Fact* **7**: 27.

Kim EY, Oh KH, Lee MH, Kang CH, Oh TK & Yoon JH (2009) Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microb* **75**: 257–260.

Kjeldsen KU, Loy A, Jakobsen TF, Thomsen TR, Wagner M & Ingvorsen K (2007) Diversity of sulfate-reducing bacteria from an extreme hypersaline sediment, Great Salt Lake (Utah). *FEMS Microbiol Ecol* **60**: 287–298.

Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H & Rupp S (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* **127**: 575–592.

Lee MH, Lee CH, Oh TK, Song JK & Yoon JH (2006) Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases? *Appl Environ Microb* **72**: 7406–7409.

Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C & Jørgensen BB (2009) Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. *Environ Microbiol* **11**: 1278–1291.

Lorenz P & Eck J (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3**: 510–516.

Marchler-Bauer A & Bryant SH (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**: W327–W331.

Park HJ, Jeon JH, Kang SG, Lee JH, Lee SA & Kim HK (2007) Functional expression and refolding of new alkaline esterase, EM2L8 from deep-sea sediment metagenome. *Protein Expres Purif* **52**: 340–347.

Purdy KJ, Munson MA, Cresswell-Maynard T, Nedwell DB & Embley TM (2003) Use of 16S rRNA-targeted oligonucleotide probes to investigate function and phylogeny of sulphate-reducing bacteria and methanogenic archaea in a UK estuary. *FEMS Microbiol Ecol* **44**: 361–371.

Ranjan R, Grover A, Kapardar RK & Sharma R (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochem Bioph Res Co* **335**: 57–65.

Rees HC, Grant S, Jones B, Grant WD & Heaphy S (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **7**: 415–421.

Roh C & Villatte F (2008) Isolation of a low-temperature adapted lipolytic enzyme from uncultivated micro-organism. *J Appl Microbiol* **105**: 116–123.

Sheridan PP, Panasik N, Coombs JM & Brenchley JE (2000) Approaches for deciphering the structural basis of low temperature enzyme activity. *Biochim Biophys Acta* **1543**: 417–433.

Sorensen J, Christensen D & Jorgensen BB (1981) Volatile fatty acids and hydrogen as substrates for sulfate-reducing bacteria in anaerobic marine sediment. *Appl Environ Microb* **42**: 5–11.

Steele HL, Jaeger KE, Daniel R & Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microb Biotech* **16**: 25–37.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Tatusov RL, Galperin MY, Natale DA & Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Verger R (1997) 'Interfacial activation' of lipases: facts and artifacts. *Trends Biotechnol* **15**: 32–38.

Ye J, McGinnis S & Madden TL (2006) BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34**: W6–W9.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Molecular analysis of predicted protein-encoding genes on fosmid pFL18.

Table S2. COG assignments of the predicted ORFs on fosmid pFL18.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.