Advances in Enzyme Research, 2016, 4, 68-86
Published Online June 2016 in SciRes. http://www.scirp.org/journal/aer
http://dx.doi.org/10.4236/aer.2016.42008

Scientific Research Publishing

Discovery and Characterization of a Thermostable Esterase from an Oil Reservoir Metagenome

Anna Lewin[1], Trine Aakvik Strand[1,2], Tone Haugen[1], Geir Klinkenberg[1], Hans Kristian Kotlar[3], Svein Valla[4], Finn Drabløs[2], Alexander Wentzel[1*]

[1]SINTEF Materials and Chemistry, Department of Biotechnology and Nanomedicine, Trondheim, Norway
[2]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[3]Former at Statoil ASA, Ranheim, Norway
[4]Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway
Email: *alexander.wentzel@sintef.no

Received 11 April 2016; accepted 24 June 2016; published 28 June 2016

Abstract

With the aim of identifying novel thermostable esterases, comprehensive sequence databases and cloned fosmid libraries of metagenomes derived from an offshore oil reservoir on the Norwegian Continental Shelf were screened for enzyme candidates using both sequence-and function-based screening. From several candidates identified in both approaches, one enzyme discovered by the functional approach was verified as a novel esterase and subjected to a deeper characterization. The enzyme was successfully over-produced in *Escherichia coli* and was shown to be thermostable up to 90˚C, with the highest esterase activity on short-chain ester substrates and with tolerance to solvents and metal ions. The fact that the thermostable enzyme was solely found by functional screening of the oil reservoir metagenomes illustrates the importance of this approach as a complement to purely sequence-based screening, in which the enzyme candidate was not detected. In addition, this example indicates the large potential of deep-sub-surface oil reservoir metagenomes as a source of novel, thermostable enzymes of potential relevance for industrial applications.

Keywords

Metagenomics, Enzyme Discovery, Thermostable, Esterase

*Corresponding author.

How to cite this paper: Lewin, A., Strand, T.A., Haugen, T., Klinkenberg, G., Kotlar, H.K., Valla, S., Drabløs, F. and Wentzel, A. (2016) Discovery and Characterization of a Thermostable Esterase from an Oil Reservoir Metagenome. *Advances in Enzyme Research*, **4**, 68-86. HTTP://DX.DOI.ORG/10.4236/AER.2016.42008

1. Introduction

The globe provides a large spectrum of environments with very diverse physical and chemical conditions for life to exist [1]. Some of these are considered extreme, providing for example high or low temperatures, pH or salt concentrations, or high pressure, radiation, as well as different combinations thereof. Deep sub-surface offshore oil reservoirs are poly-extreme environments due to a combination of high temperature, high pressure, high salinity, nutrient limitations, as well as the presence of toxic compounds like heavy metals and organic solvents [2]. Microorganisms populating these environments are generally considered (poly) extremophiles, attributing several unusual and interesting traits [3]. However, studies of these microbes are limited by difficulties in accessing relevant oil reservoirs, as well as technical challenges in representative oil reservoir sampling [4] [5]. Due to their nature, oil reservoir microorganisms with their special adaptations are potential high value targets for the discovery of useful biocatalysts for industrial processes [6]-[8]. By also covering the genetic information of the vast uncultivable fraction of microorganisms in nature [1] which is particularly pronounced in such special ecological niches, metagenomic libraries are an invaluable source for new discoveries by bioprospecting, using both metagenomic sequence data and cloned metagenomic DNA [9] [10]. Sequence-based bioprospecting using metagenomic DNA sequence databases has benefits in being fast and inexpensive to perform, and is independent of functional expression in a heterologous host and functional assays. However, its dependence on previously described sequence information limits its potential to discover entirely new enzymes. In contrast, function-based screening for activities of interest using cloned metagenomic libraries is capable of discovering novel enzymes based on the function, independent of sequence similarity to previously known genes. It represents therefore a valuable complement to sequence data mining, despite of its dependence on assay methods and vector-host systems for functional gene expression.

Esterases are enzymes with several applications in industrial processes, and novel variants of these providing special properties are still in demand [11]. They are today commercially used e.g. in the food industry, for animal feed and biofuel production, as well as detergents. New candidates with high tolerance to elevated temperature and pressure, stability and activity in organic solvents, and/or special substrate spectra may prove useful in these, or other applications [12]. In the current study, we have screened oil reservoir metagenomes [13] [14] for esterase enzyme candidates using both sequence-based and functional screening, and have subjected one candidate from the functional screening approach to further characterization. Our results document oil reservoir metagenomes as a valuable source of novel thermostable enzymes of potential industrial relevance.

2. Materials and Methods

2.1. Sample Collection, DNA Isolation and Fosmid Library Generation

Oil reservoir samples were collected and processed as described earlier [13] [14]. Metagenomic DNA was isolated [14] from the sample material and used for direct high throughput sequencing (water phases of Wells I and II of [14]), as well as whole genome amplification (WGA) using Qiagen REPLi-g kit according to the manufacturer's protocol (water phase of Well II of [14]) and both water and oil phase of a third, nearby production well (Well III) that, unlike Wells I and II, has experienced seawater breakthrough. Amplified DNA was cloned as fosmid libraries using the pCC1fos fosmid library kit from Epicentre according to the manufacturer's instructions [15]. Library clones of each sample were selected for on square (215 ml volume) LB agar plates containing 12.5 µg/ml chloramphenicol, pooled and stored in 25% glycerol at −80˚C. Based on colony counts, the total library harbors approximately 77,000 independent fosmid clones. Prior to functional screening, library pools from the different samples were again plated onto the same type of selective agar plates (LB-agar + 12.5 µg/ml chloramphenicol), with 350 - 1500 CFU per plate. Proportional to the number of total clones represented in each library, the corresponding number of agar plates and 384 multi-well plates were used for the different samples. In total 20 agar plates were processed; 14 with Well II water phase library clones, 4 with Well III water phase library clones and 2 with Well III oil phase library clones. Correspondingly, 21 individual 384 well plates were inoculated with Well II water phase library clones, 6 plates with Well III water phase clones and 3 plates with Well III oil phase clones. Clones within the well plates were grown in 80 µl reduced Hi-YE medium per well containing 12.5 µg/ml chloramphenicol (37˚C, 900 rpm, 85% humidity, 20 hours), plates were duplicated and then stored at −80˚C with 25% final concentration glycerol.

2.2. Sequence Mining

Twenty Pfam [16] profiles for known lipases and esterases were used as input to tBLASTn searches [17] against ORFs of the metagenomic DNA libraries, using an E-value cutoff of 0.00001 and a minimal overlap of 50% on the Pfam motif. In addition, ORFs of the metagenomic libraries were classified according to EC numbers using tBLASTn with exemplar sequences derived from the BRENDA database [18].

2.3. Functional Screening (HTS)

Approximately 11,520 fosmid clones carrying metagenomic DNA sequences originating from oil reservoir Well II (water phase sample) [14] and Well III (oil and water phase) were screened for lipase and/or esterase activities. Functional screening (schematically described in supplementary **Figure S1**) was performed using liquid culture extracts in 30 individual 384 well plates. Each of the 30 384 well plates of one stored replica set from library construction was used to inoculate 4 individual 96 well plates, in which the clones were grown for 20 hours under the same conditions as described above and subsequently the increase of fosmid copy number induced for 6 hours by addition of arabinose from a 10% (w/v) stock solution (0.01% final conc.). Cultivation plates were then stored at −80˚C until extraction. After thawing, crude cell extracts were prepared by adding 30 µl B-PER II (Thermo Scientific, prod# 89822) to each well, followed by an incubation at 25˚C, 700 rpm, 85% humidity for 1 hour. Cell debris was pelleted by centrifugation (4000 rpm, 30 minutes, 4˚C). Supernatants (extracts) were transferred in duplicates to 384 well plates and used in enzyme assays (**Figure S1**).

High throughput screening for short-chain esterase activity was performed in 384 well plates by addition of 5 µl crude cell extract to each well containing 40 µl reaction mixture (0.17 mM 4-nitrophenyl acetate substrate (Sigma-Aldrich, no. N8130) in 0.1 M phosphate buffer, pH 8, containing 4.4% EtOH), shaking of the plates (1600 rpm, 30 s) and absorption measurement at 410 nm immediately and after 100 min incubation at 37˚C. For screening for long-chain esterase activity, the same procedure was applied, with the exception of the reaction mixture containing 0.17 mM 4-nitrophenyl palmitate (Sigma-Aldrich, no. N2752), 4.4% EtOH and 1% Triton X-100 in 0.1 M phosphate buffer, pH 8, and performing absorption measurements at time points 0, 1 and 3 hours after 50˚C incubation.

2.4. Analysis of Positive Clones

Esterase positive clones identified in functional screening were subjected to fosmid isolation using the Promega Wizard Miniprep kit (arabinose induced cultures), and the genes of interest were identified by sequencing of respective fosmid clones and sub-cloning of candidate genes. Gene sequences were analyzed using BLASTx against the NCBI database. Amino acid sequences were analyzed on the Phylogeny.fr web server [19], using T-Coffee [20] for multiple alignment, PhyML [21] for generating the phylogenetic tree (after removing all columns with gaps), and TreeDyn [22] for drawing the tree. Branch support in the phylogenetic tree was estimated by the approximate likelihood ratio test (aLRT) [23]. The multiple alignment was visualized using Jalview [24]. Pairwise comparison of protein sequences was done with pairwise BLAST.

2.5. Enzyme Production

The esterase candidate genes were amplified using PCR from the relevant fosmid (primer sequences GGGGCCATGGGAGATAAGGAGGAGGG and GGGGATCCAAAGATAGAGGAGCAGATC for His-tagged enzyme, and sequences ATAAGGAGGAGGGCATATGGCTGA and GGGGATCCAAAGATAGAGGAGCAGATC for non-tagged enzyme), or synthesized as *E. coli* codon optimized versions (GenScript, Piscataway, NJ, USA) and sub-cloned in expression vector pET16b (with and without hexahistidyl-tag; however, only untagged protein was analyzed as enzyme isolation by heat incubation proved to be feasible). Enzymes were produced in *E. coli* BL21 (DE3) REPL (codon+) grown in shake flasks in 50 ml LB containing 100 µg/ml ampicillin. Gene expression was induced at OD_{600} = 1.0 using IPTG (0.5 mM final conc.). After continued incubation at 37˚C until observed OD_{600} decrease after approx. 7 h, the cultures were harvested by centrifugation (4000 rpm, 15 minutes, 4 ˚C) and cell pellets stored at −20˚C until extraction. Crude cell extracts were prepared by sonication in 2 ml 0.1 M Tris-HCl buffer, pH 8, per 1 g wet weight biomass (flat tip, duty cycle 50%, output control 4; 7 × 1 min). Cell debris was pelleted by centrifugation (20 000 x g, 20 min, 4˚C) and supernatant stored at −20˚C.

Batch cultivation for the production of esterase FS02 was performed in two 3 l Applicon fermentor vessels

using 1.8 l 3xLB medium (Tryptone 30 g/l, Yeast extract 15 g/l, NaCl 10 g/l) each, containing 100 µg/ml ampicillin, with constant airation at 0.25 VVM (0.45 l/min) and 1000 rpm fixed stirrer speed. The cultivation medium was inoculated with 1.5% (27 ml) overnight culture (LB, 100 µg/ml ampicillin). Gene expression was induced at OD_{600} = 4 using (3 hours cultivation) 0.5 mM IPTG. Cultures were thereafter harvested after 6 hours of induction and thereafter treated as described above to obtain crude cell extracts. Crude cell extracts were heat treated at 65˚C for 30 minutes, centrifuged at top speed (13 000 x g) in a table top centrifuge for 5 minutes to precipitate denatured host proteins, aliquoted and stored at −20˚C.

2.6. Enzyme Characterization

Verification of lipase activity. Crude cell extracts were used to assess lipase enzyme function using the same assay procedure as applied for esterase screening in multi-well plates, however, up-scaled to 1 ml volume for cuvette measurements. To 880 µl master mix (4.4% EtOH and 1% Triton X-100 in 0.1 M phosphate buffer, pH 8), 20 µl substrate solution (20 mM 4-nitrophenyl palmitate in acetonitrile) was added, incubated at 37˚C for 30 s prior to addition of 100 µl cell extract and absorbance measurement at 420 nm every 2 seconds for 4 minutes. End point measurements were performed using the same setup but with measurements after 30, 60 and 120 minutes, with assay and incubations done at 37˚C.

Heat stability. Temperature stability of enzymes was determined by incubating crude extracts at elevated temperatures (65˚C - 100˚C) for different time periods (15 min - overnight), followed by SDS-PAGE and/or activity assay as described above. Thermal unfolding transition midpoint Tm of the enzyme of interest was analyzed using the Thermal Shift assay from Applied Biosystems on a 7500 rtPCR machine according to the manufacturer's instructions.

pH optimum. The pH optimum of the enzyme of interest was analyzed by performing the esterase activity assay at 37˚C as described above but using buffer with different pH (0.1 M glycine buffer pH 2 - 3 and pH 9 - 12; 0.05 M citrate buffer pH 5, 0.1 M phosphate buffer pH 6 - 8).

Substrate spectrum. Enzyme activity on a spectrum of commercial ester substrates of different chain lengths (C2-C18; p-nitrophenyl acetate, p-nitrophenyl-butyrate, p-nitrophenyl-decanoate, p-nitrophenyl-myristate, p-nitrophenyl-palmitate and p-nitrophenyl-stearate) was analyzed using the assay procedure describe above.

Effect of selected additives on enzyme activity. Esterase activity in the presence of different additives was assayed as described above using p-nitrophenyl-butyrate as the substrate. Chemicals added were; metal ions (1 mM final concentrations from $CaCl_2{\cdot}2H_2O$, $CoSO_4{\cdot}7H_2O$, $CuCl_2{\cdot}2H_2O$, $FeCl_3$, $MgCl_2{\cdot}6H_2O$, $MnCl_2{\cdot}4H_2O$, RbCl, $ZnCl_2$, KCl, NaCl and Li-acetate*$2H_2O$ salt solutions, respectively), isopropanol, ethanol, methanol, acetone, acetonitrile, DMF and DMSO (1%, 10% and 30% each), Triton X-100, Tween 80 and SDS 5 (1% and 5% each) as well as DTT and EDTA (1, 5 and 10 mM each).

3. Results and Discussion

3.1. Homology-Based Sequence Data Mining for Lipase/Esterase Enzyme Candidates

The established metagenomic sequence databases [13] [14] were screened for lipase enzyme candidates using 1) BLAST searches with Pfam profiles for lipases and esterases and 2) screening by a local ORF database classified by EC-number. Using Pfam profiles against the Well I and Well II metagenomics databases, 72 and 55 candidates, respectively, were identified and in total 24 candidates were found associated with E.C. number 3.1.1.3 after matching ORFs against BRENDA exemplar sequences and requiring >60% sequence identity and >80% coverage (Supplementary **Table S1**). The list of candidates from sequence mining using the Pfam strategy contained candidates with very low sequence identity to Pfam motif (annotated lipases/esterases/hydrolases). In total only 19 candidates had sequence identity higher than 30% (37.5% as highest). increasing the probability of false positive predictions. However, the top five candidates (37.5% - 35.4% id) were used in a BLASTx search, and were indeed matching esterases, hydrolases and lipolytic proteins with fairly high sequence identity (>80%) for most, but with some exceptions having sequence identity as low as 40% - 50%. As examples, two gene candidates from sequence mining had sequence id of 78% against alpha/beta hydrolase fold (*Campylobacterales bacterium* GD) and 62% against biotin biosynthesis protein (*Pelobacter carbinolicus* DSM 2380), respectively (in addition verified using latest version of Pfam, resulting in abhydrolase 6, alpha/beta hydrolase family (e-value 2e-19), and abhydrolase 1, alpha/beta hydrolase fold (e-value 7e-25), respectively). These candidates

were synthesized as codon optimized versions (GenScript) and produced in *E. coli*, however showing non-conclusive data in activity analysis and only moderate thermostability (data not shown). These results illustrate a key challenge of sequence mining to reliably identify candidates with the predicted enzyme activity. In the light of the obtained results, these candidates were not analyzed further. Therefore, with the purpose of finding esterase enzymes with high novelty and functionality, focus was turned to functional screening of the constructed metagenome libraries.

3.2. Functional Screening for Lipases and Esterases

Functional screening for long-and short-chain esterases encoded in the 11,520 arrayed fosmid clones resulted in 13 positive hits from the initial screen. For identification of the genes of interest within the respective fosmids, fosmid DNA was isolated and subjected to a combination of end-sequencing and primer walking, supplemented by 454 pyrosequencing of pooled fosmid clones. For the end-sequencing and primer walking approach, technical difficulties appeared, and sequencing results could not be correlated with fosmid size determinations by gel electrophoresis (data not shown). After thorough evaluation of the fosmid library construction procedure, it was concluded that the encountered problems likely were associated with the amplification of the metagenomic DNA. Due to the low amount of DNA that could be isolated from the oil reservoir samples, target DNA was amplified using WGA based on Multiple Strand Displacement Amplification (MDA) [25] prior to fosmid cloning. MDA has been reported earlier to result in chimeric sequences [26], and for long insert cloning (such as fosmids of around 35 kilobases in size), repetitive and chimeric DNA fragments can potentially be expected. The obtained sequencing results suggest frequent occurrence of chimeric repeats with multiple primer binding sites within the cloned inserts during end-sequencing/primer walking, hampering elucidation of the complete sequence information contained in the cloned DNA. Nevertheless, some genes potentially associated with the detected esterase phenotype could still be identified. One of these genes was found to encode a previously described lipase of *Archaeoglobus fulgidus* [27] which was not chosen for further analysis due to lack of novelty, as well as two other candidate genes (FS01 and FS02, with annotations as hypothetical proteins). These could be associated with strong esterase and lipase phenotypes and were therefore chosen for sub-cloning, expression from pET16b in *E. coli* and then subjected to further characterization. Respective crude cell extracts were produced and analyzed by esterase activity assays. Sub-cloned candidate FS01 did not show any esterase activity, whereas the FS02 clone, despite of low enzyme production levels, clearly exhibited such an activity (**Figure S2**). FS02 was in addition found to be highly thermostable, as incubation at 85˚C for 30 minutes led only to a small decrease in activity compared to the not heat treated control (**Figure S2**).

3.3. Analysis of the FS02 Gene

The nucleotide sequence of the FS02 gene was used in BLASTx searches against the NCBI protein database, resulting in hits against hypothetical proteins or hydrolases of the alpha/beta superfamily (**Table S2**). The closest match was against a hypothetical protein encoded in the *Archaeoglobus fulgidus* genome [27], showing 99% sequence identity (1 amino acid mismatch) by 100% coverage. Further close homologues were a hypothetical protein from *Ferroglobus placidus* (60% identity, 97% coverage), a putative hydrolase of the alpha/beta superfamily from *Archaeoglobus sulfaticallidus* (55% identity, 96% coverage) and a hypothetical protein from *Archaeoglobus veneficus* SNP6 (52% identity and 99% coverage). For further evaluation of novelty, the FS02 amino acid sequence was combined with known lipase/esterase sequences of several Archaea and Bacteria (**Table S3**) in a multiple sequence alignment (**Figure S3**) in order to build a phylogenetic tree (**Figure 1**). This tree confirms the initial picture that the FS02 enzyme is closely related to predicted proteins from Archaea, but similar proteins are also found in Bacteria. Based on the functional confirmation of FS02 as an esterase, it is now possible to assign putative esterase function to these hypothetical proteins, demonstrating the strength of functional screening of metagenomes in elucidating function of the high number of genes of yet unknown function.

3.4. Production of FS02 Enzyme and Enzyme Characterization

Due to the apparently low production levels of FS02 enzyme in *E. coli*, the FS02 gene was synthesized in an *E. coli* codon usage optimized version (GenScript), cloned in the T7 expression vector pET16b and expressed in *E. coli*. Enzyme production based on the new gene version was found to be significantly higher compared to the

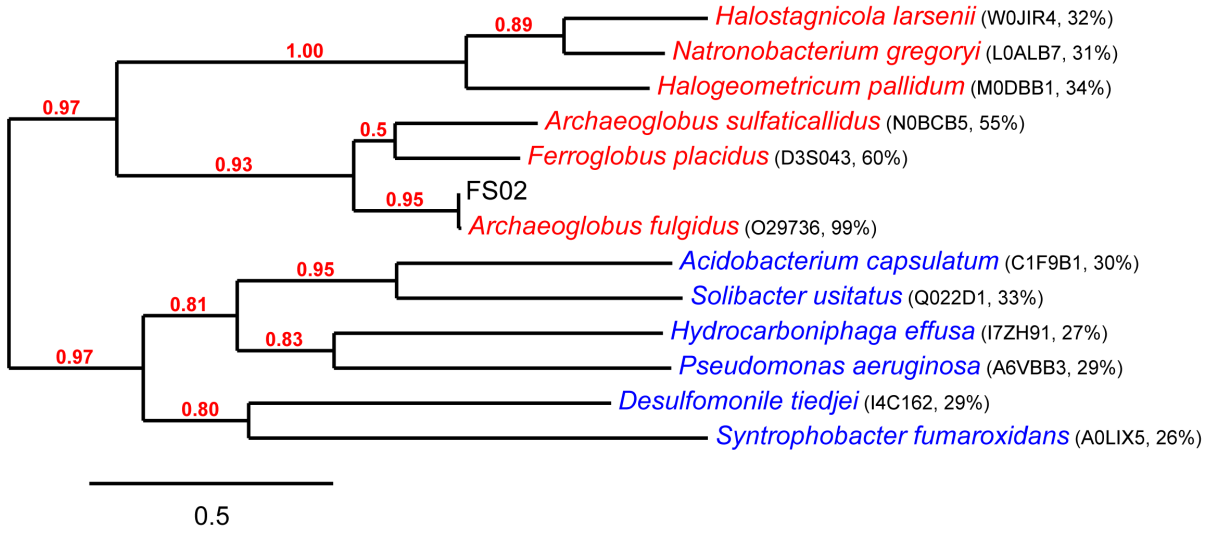


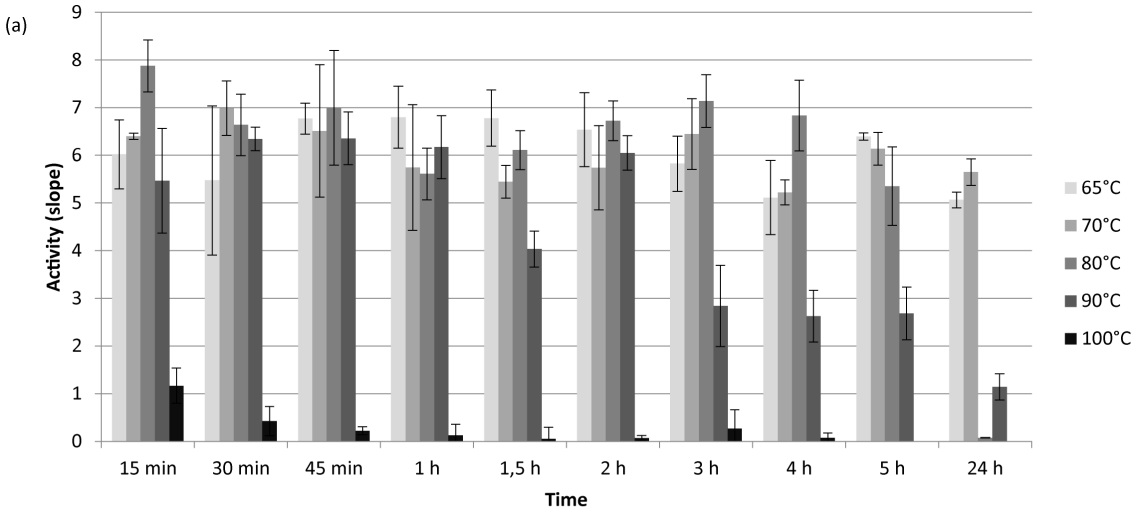
Figure 1. Phylogenetic analysis of the amino acid sequence of lipase FS02 with a selection of similar alpha/beta hydrolases (**Table S2**). Protein inputs are annotated as hypothetical proteins or alpha/beta hydrolases, and their respective UniProt IDs are indicated. The tree shows a clear separation into sequences from Archaea (red) and Bacteria (blue). Numbers show branch support by the approximate likelihood ratio test (aLRT).

native gene sequence (data not shown). FS02 thermostability (**Figure S4**) and activity were verified, and a larger batch of enzyme was produced to allow multiple analysis of the enzyme from the same batch (3 - 4 g wet cell mass/fermentor).

Batch-produced FS02 was successfully tested for heat stability by incubation at 65˚C for 30 minutes, thus leading to an FS02 enriched crude extract for subsequent studies. The enzyme's temperature stability was analyzed in detail by incubation at different temperatures for various periods of time, demonstrating that FS02 is thermostable up to 90˚C for at least 1 hour without any notable decrease in activity, whereas a 100˚C incubation resulted in rapid inactivation of the enzyme (**Figure 2(a)**). Longer incubation at 90˚C resulted in lowered residual activity, however, high enzyme activity could still be detected after incubations at up to 80˚C for as long as 5 hours (**Figure 2(a)**), indicating that FS02 is a highly thermostable enzyme. Tm values were determined by Thermal Shift, resulting in values of Boltzman Tm (TmB; **Figure 2(b)**) that were consistent with observed residual enzyme activities at high temperature. When comparing Tm values determined under different pH conditions, it could be concluded that the chosen pH range did not affect heat stability at notable levels (**Figure 2(b)**). In attempts to analyze the pH optimum of FS02, it was found that enzyme activity was high at pH 11, however, assay background was high as well at this pH (data not shown), and therefore pH 9 was chosen as standard pH in all subsequent additional enzyme characterizations. The substrate spectrum analysis of FS02 revealed a higher enzymatic activity on short-chain substrates compared to long-chain esters (**Figure 3**). Enzyme activity was highest using p-nitrophenyl acetate as substrate.

Despite relatively high standard deviations in replicate measurements (**Figure 4**) it was found that enzyme activity on this substrate was little affected by a number of different additives, such as Ca^{2+}, Cu^{2+}, Na^{+}, Li^{+}, Co^{2+}, Fe^{3+} and Mg^{2+} ions, as well as by addition of 5 to 10 mM DTT (**Figure 4**). Ions like Mn^{2+}, Rb^{2+}, Zn^{2+} and K^{+} however, led to a notable increase in activity up to 5-fold. Stimulation and inhibitory effects of different metal ions on esterase activity have been observed before [28], including an increased activity in the presence of Zn^{2+} and a decreased activity in the presence of e.g. Ca^{2+}, Mn^{2+} and Na^{+}. Addition of low concentrations of detergent (1% Triton X-100 or SDS) was found to increase activity, whereas a higher concentration of the same compounds (5%) resulted in loss of activity, which is consistent with observations with other esterases (e.g. [29]). Presence of solvents such as ethanol, methanol, DMF, acetonitrile, acetone and isopropanol apparently led to significant increases in activity (**Figure 4**). Lipase activity in the presence of different solvents has been analyzed in several previous studies, mainly showing no or negative effects; e.g. unaffected activity with methanol, ethanol and acetone, and a decreased lipase activity in the presence of isopropanol [30], or reduced lipase activity in the presence of isopropanol, methanol, acetone, acetonitrile and/or ethanol [27] [29]. On the other hand,

(a)



(b)

Melting temperature FS02		
pH	Tm B	Stdev
8	95.90	± 0.268
9	96.14	± 0.323
11	96.01	± 0.488

Figure 2. Temperature stability of FS02. a) Enzyme activity assay (37˚C) after heat series incubation of crude cell extracts (CCE) containing FS02. Incubation at given temperatures was performed ranging from 15 min to overnight, and applying three replica measurements each. From each replicate measurement, a respective background value (using extract of cells not harboring the FS02 encoding plasmid) was subtracted (background values not exceeding 7% of the experiment data), the resulting initial slope values of all three parallels determined and the average and standard deviation values plotted, giving the relative activity in units presented in the graph. b) Thermal shift assay of FS02 enzyme to determine melting temperatures (TmB) at pH 8, 9 and 11 from the protein melting curve when in different pH buffers. Experiments were performed twice, including four replicas each. Averages and standard deviations were calculated collectively from all eight data points.

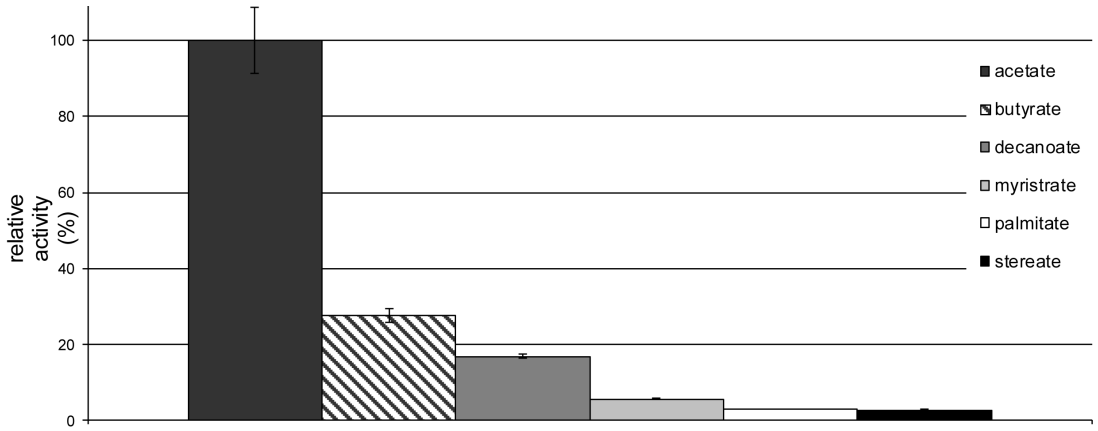


Figure 3. Determination of the substrate spectrum of FS02. Activity assays were performed for different ester substrates ranging from C2-C16. Four replica measurements were performed. Background values (based on extracts of cells not harboring the FS02 encoding plasmid) were subtracted from each measurement (background values not exceeding 16% of the experiment data). Resulting initial slope values were determined and their average and standard deviation values plotted. The average activity value on acetate was set to 100%.

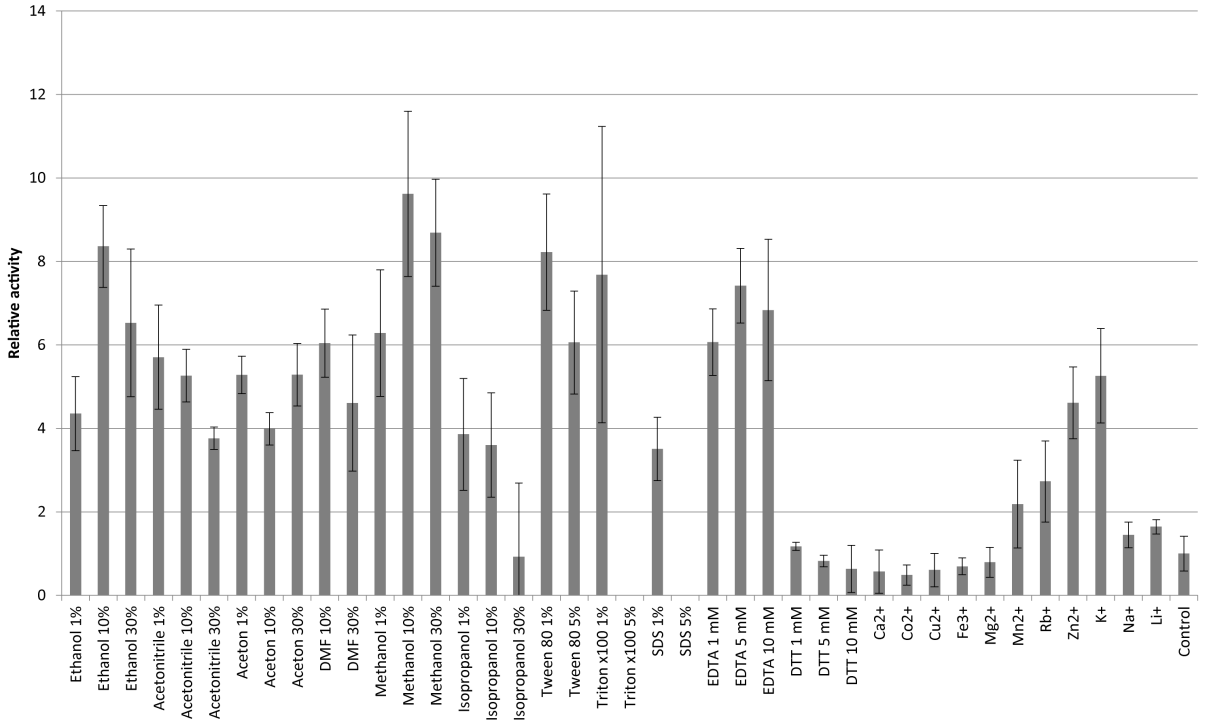


Figure 4. Lipase activity of FS02 in the presence of different additives. All experiments were performed in four replicates and are presented as mean value and standard deviations of measurements with average background (same conditions but without cell extracts) subtracted. For conditions ethanol 1% - 30%, acetonitril 10% - 30%, acetone 10% - 30%, DMF 30%, methanol 30%, isopropanol 10%, Tween80 1%, Triton X-100 1%, EDTA 5 mM, Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Rb^{+} and Li^{+}, each one of the four parallels was excluded as obvious outliers.

similar to the observed activity increase of FS02 in the current work, enzymatic activity of an alkaline lipase was shown to be increased in the presence of e.g. methanol [31], and high solvent tolerance of a *Pseudomonas* lipase has been shown [30]. Hence enzyme activity increase in the presence of solvents observed for FS02 is not unique; however, despite of some uncertainties concerning the absolute activity levels from high standard deviations in the analyses, the increases in activity observed for enzymes in previous studies are rarely as high as observed for FS02, rendering the newly discovered enzyme with distinct features compared to other similar enzymes.

4. Conclusions

Esterases are used in several industrial processes [1] [32], and many of these enzymes are well characterized and applied in different processes [33] [34]. However, the discovery of novel esterases with beneficial properties, e.g. high thermo-and solvent-stability, might still serve current and potential new uses in e.g. the food, polymer, pharmaceutical and fine chemical industries [32] [34]. Metagenome-based bioprospecting is a powerful tool for biodiscovery of novel enzymes for various applications, and metagenomic libraries from (poly) extreme environments might be a particularly valuable source of enzymes with unique features. Several lipase/esterase enzyme candidates were identified from the oil reservoir-originating metagenomes, both using homology-based sequence mining and functional screening. The latter helps assigning function to the vast pool of hypothetical proteins encoded and annotated in genomes and metagenomes.

One lipase/esterase candidate from functional screening, denoted FS02, was found to be active, highly thermostable, as well as showed high activity in the presence of various compounds, generating noteworthy results (**Figure 4**) on the enzyme's tolerance to e.g. solvents, potentially relevant for some industrial applications. Solvent tolerance is often observed for lipases (e.g. [31]), but the large stimulation of enzyme activity appears rare. However, dependent on a potentially envisioned industrial application, a more profound investigation of these

properties will be necessary.

The discovery of an active and thermostable esterase in a metagenomic library originating from a deep subsurface oil reservoir illustrates the powerful approach of functional metagenomics in search for novel enzymes with desired properties from poly-extreme environments. It also demonstrates the potential for biodiscovery from such environments as deep subsurface oil reservoirs. The finding of the described thermostable esterase represents only the top of the ice berg of what potentially can be found by harvesting and screening metagenomes from this special ecological niche.

Acknowledgements

This work is supported by the Research Council of Norway (grant numbers 187317/S30, and 208541/O10), Statoil ASA and SINTEF.

References

[1] Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: The Unseen Majority. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6578-6583. http://dx.doi.org/10.1073/pnas.95.12.6578

[2] Wentzel, A., *et al*. (2013) Deep Sub-Surface Oil Reservoirs as Poly-Extreme Habitats for Microbial Life. A Current Review. In: *Polyextremophiles*, Volume 27 of the series Cellular Origin, Life in Extreme Habitats and Astrobiology, Springer, 439-466. http://dx.doi.org/10.1007/978-94-007-6488-0_19

[3] Lewin, A., Wentzel, A. and Valla, S. (2013) Metagenomics of Microbial Life in Extreme Temperature Environments. *Current Opinion in Biotechnology*, **24**, 516-525. http://dx.doi.org/10.1016/j.copbio.2012.10.012

[4] Magot, M., Ollivier, B. and Patel, B.K. (2000) Microbiology of Petroleum Reservoirs. *Antonie Van Leeuwenhoek*, **77**, 103-116. http://dx.doi.org/10.1023/A:1002434330514

[5] Orphan, V.J., *et al.* (2000) Culture-Dependent and Culture-Independent Characterization of Microbial Assemblages Associated with High-Temperature Petroleum Reservoirs. *Applied and Environmental Microbiology*, **66**, 700-711. http://dx.doi.org/10.1128/AEM.66.2.700-711.2000

[6] Ramesha, B., *et al.* (2010) Biodiversity and Chemodiversity: Future Perspectives in Bioprospecting. *Current Drug Targets*, **12**, 1515-1530.

[7] Lee, M.H. and Lee, S.-W. (2013) Bioprospecting Potential of the Soil Metagenome: Novel Enzymes and Bioactivities. *Genomics & Informatics*, **11**, 114-120. http://dx.doi.org/10.5808/GI.2013.11.3.114

[8] Koskinen, P., *et al.* (2008) Bioprospecting Thermophilic Microorganisms from Icelandic Hot Springs for Hydrogen and Ethanol Production. *Energy & Fuels*, **22**, 134-140. http://dx.doi.org/10.1021/ef700275w

[9] Novakova, J. and Farkasovsky, M. (2013) Bioprospecting Microbial Metagenome for Natural Products. *Biologia*, **68**, 1079-1086. http://dx.doi.org/10.2478/s11756-013-0246-7

[10] Vester, J., Glaring, M. and Stougaard, P. (2015) Improved Cultivation and Metagenomics as New Tools for Bioprospecting in Cold Environments. *Extremophiles*, **19**, 17-29. http://dx.doi.org/10.1007/s00792-014-0704-3

[11] Xu, Q., *et al.* (2007) Cellulases for Biomass Conversion. In: Polaina, J. and MacCabe, A.P., Eds., *Industrial Enzymes*, Springer, Valencia, 65-82. http://dx.doi.org/10.1007/1-4020-5377-0_3

[12] Acharya, S. and Chaudhary, A. (2012) Bioprospecting Thermophiles for Cellulase Production: A Review. *Brazilian Journal of Microbiology*, **43**, 844-856. http://dx.doi.org/10.1590/S1517-83822012000300001

[13] Kotlar, H.K., *et al.* (2011) High Coverage Sequencing of DNA from Microorganisms Living in an Oil Reservoir 2.5 Kilometres Subsurface. *Environmental Microbiology Reports*, **3**, 674-681. http://dx.doi.org/10.1111/j.1758-2229.2011.00279.x

[14] Lewin, A., *et al.* (2014) The Microbial Communities in Two Apparently Physically Separated Deep Subsurface Oil Reservoirs Show Extensive DNA Sequence Similarities. *Environmental Microbiology*, **16**, 545-558.

[15] Aakvik, T., *et al.* (2009) A Plasmid RK2-Based Broad-Host-Range Cloning Vector Useful for Transfer of Metagenomic Libraries to a Variety of Bacterial Species. FEMS Microbiology Letters, **296**, 149-158. http://dx.doi.org/10.1111/j.1574-6968.2009.01639.x

[16] Finn, R.D., *et al*. (2014) Pfam: The Protein Families Database. *Nucleic Acids Research*, **42**, D222-D230. http://dx.doi.org/10.1093/nar/gkt1223

[17] Altschul, S.F., *et al*. (1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, **25**, 3389-3402. http://dx.doi.org/10.1093/nar/25.17.3389

[18] Chang, A., *et al*. (2015) BRENDA in 2015: Exciting Developments in Its 25th Year of Existence. *Nucleic Acids Re-*

search, **43**, D439-D446. http://dx.doi.org/10.1093/nar/gku1068

[19] Dereeper, A., *et al*. (2008) Phylogeny.fr: Robust Phylogenetic Analysis for the Non-Specialist. *Nucleic Acids Research*, **36**, W465-W469.

[20] Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *Journal of Molecular Biology*, **302**, 205-217. http://dx.doi.org/10.1006/jmbi.2000.4042

[21] Guindon, S., *et al*. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321. http://dx.doi.org/10.1093/sysbio/syq010

[22] Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. and Christen, R. (2006) TreeDyn: Towards Dynamic Graphics and Annotations for Analyses of Trees. *BMC Bioinformatics*, **7**, 439. http://dx.doi.org/10.1186/1471-2105-7-439

[23] Anisimova, M. and Gascuel, O. (2006) Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Systematic Biology*, **55**, 539-552. http://dx.doi.org/10.1080/10635150600755453

[24] Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—A Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics*, **25**, 1189-1191. http://dx.doi.org/10.1093/bioinformatics/btp033

[25] Rodrigue, S., *et al*. (2009) Whole Genome Amplification and *De novo* Assembly of Single Bacterial Cells. *PLoS ONE*, **4**, e6864. http://dx.doi.org/10.1371/journal.pone.0006864

[26] Lasken, R.S. and Stockwell, T.B. (2007) Mechanism of Chimera Formation during the Multiple Displacement Amplification Reaction. *BMC Biotechnology*, **7**, 19. http://dx.doi.org/10.1186/1472-6750-7-19

[27] Klenk, H.P., *et al*. (1997) The Complete Genome Sequence of the Hyperthermophilic, Sulphate-Reducing Archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364-370. http://dx.doi.org/10.1038/37052

[28] Ugur, A., *et al*. (2014) New Lipase for Biodiesel Production: Partial Purification and Characterization of LipSB 25-4. *ISRN Biochemistry*, **2014**, Article ID: 289749.

[29] Tian, J.W., Lei, Z.C., Qiu, P., Wang, L. and Tian, Y.Q. (2014) Purification and Characterization of a Cold-Adapted Lipase from *Oceanobacillus* Strain PT-11. *PLoS ONE*, **9**, e101343. http://dx.doi.org/10.1371/journal.pone.0101343

[30] Borkar, P.S., Bodade, R.G., Rao, S.R. and Khobragade, C.N. (2009) Purification and Characterization of Extracellular Lipase from a New Strain: *Pseudomonas aeruginosa* SRT 9. *Brazilian Journal of Microbiology*, **40**, 358-366. http://dx.doi.org/10.1590/S1517-83822009000200028

[31] Bisht, D., Yadav, S.K. and Darmwal, N.S. (2013) An Oxidant and Organic Solvent Tolerant Alkaline Lipase by *P. aeruginosa* Mutant: Downstream Processing and Biochemical Characterization. *Brazilian Journal of Microbiology*, **44**, 1305-1314. http://dx.doi.org/10.1590/S1517-83822013000400040

[32] Hasan, F., Shah, A.A. and Hameed, A. (2006) Industrial applications of Microbial Lipases. *Enzyme and Microbial Technology*, **39**, 235-251. http://dx.doi.org/10.1016/j.enzmictec.2005.10.016

[33] Balan, A., Ibrahim, D., Abdul, R.R. and Ahmad, R.F.A. (2012) Purification and Characterization of a Thermostable Lipase from *Geobacillus thermodenitrificans* IBRL-nra. *Enzyme Research*, **2012**, Article ID: 987523.

[34] Zamost, B.L., Nielsen, H.K. and Starnes, R.L. (1991) Thermostable Enzymes for Industrial Applications. *Journal of Industrial Microbiology*, **8**, 71-81. http://dx.doi.org/10.1007/BF01578757

Appendix

Table S1. Candidates found by sequence mining for lipases and esterases using BLAST searches against a database of Pfam profiles and a collection of exemplar sequences from BRENDA classified by EC number. The most significant hit for each contig is shown, listing Pfam or BRENDA annotation, as well as % identity, % coverage and e-value. For Pfam hits the % identity is relative to the consensus sequence.

Search/candidate	Candidate contig/read ID	Pfam ID/Annotation	Identity (%)	Coverage (%)	e-value
Pfam1	contig10634_532_4	pfam00561-Abhydrolase_1	23.45	62.67	5.00E−12
Pfam2	contig03006_2979_63	pfam12695-Abhydrolase_5	19.62	100	4.00E−18
Pfam3	contig03006_2979_63	pfam12695-Abhydrolase_5	25.51	57.24	4.00E−11
Pfam4	contig06306_3078_37	pfam12697-Abhydrolase_6	24.39	69.64	1.00E−05
Pfam5	contig03391_8603_101	pfam13472-Lipase_GDSL_2	30.64	97.69	3.00E−30
Pfam6	contig04922_8688_114	pfam12695-Abhydrolase_5	27.37	59.31	5.00E−08
Pfam7	contig05190_528_3	pfam12695-Abhydrolase_5	28.75	52.41	3.00E−14
Pfam8	contig07066_7149_180	pfam12697-Abhydrolase_6	33.06	54.02	1.00E−20
Pfam9	contig04904_2432_80	pfam13472-Lipase_GDSL_2	16.98	52.60	8.00E−09
Pfam10	FYXZCPA02FPU0B_473_1	pfam07859-Abhydrolase_3	21.56	74.88	3.00E−07
Pfam11	contig06801_3003_27	pfam13472-Lipase_GDSL_2	32.95	100	3.00E−28
Pfam12	contig03009_1909_60	pfam06821-Ser_hydrolase	37.50	97.08	1.00E−46
Pfam13	contig08344_800_5	pfam12697-Abhydrolase_6	18.59	78.13	2.00E−06
Pfam14	contig03736_4913_82	pfam12695-Abhydrolase_5	23.40	63.4483	1.00E−05
Pfam15	contig06405_970_8	pfam12697-Abhydrolase_6	30.00	99.1071	1.00E−50
Pfam16	FYXZCPA02G140B_424_1	pfam12695-Abhydrolase_5	33.33	62.7586	6.00E−18
Pfam16	contig01573_4582_113	pfam12695-Abhydrolase_5	25.29	100	1.00E−12
Pfam17	contig05511_3393_28	pfam12697-Abhydrolase_6	19.26	96.4286	1.00E−13
Pfam18	FYXZCPA02IP2FT_387_1	pfam12697-Abhydrolase_6	22.83	55.3571	6.00E−10
Pfam19	contig02150_5033_123	pfam12697-Abhydrolase_6	19.40	88.8393	7.00E−07
Pfam2	contig03063_7553_90	pfam12697-Abhydrolase_6	19.83	96.875	2.00E−19
Pfam21	contig08668_499_3	pfam12697-Abhydrolase_6	36.09	56.6964	2.00E−17
Pfam22	FYXZCPA01BCVY3_478_1	pfam12695-Abhydrolase_5	30.59	53.7931	4.00E−12
Pfam23	contig01280_2950_36	pfam12697-Abhydrolase_6	21.30	95.9821	9.00E−23
Pfam24	contig07165_797_5	pfam12695-Abhydrolase_5	16.50	100	1.00E−10
Pfam25	contig00489_2553_26	pfam00561-Abhydrolase_1	22.34	97.3333	1.00E−20
Pfam26	contig06642_507_3	pfam12697-Abhydrolase_6	24.18	61.6071	4.00E−16
Pfam27	contig07477_1639_15	pfam00561-Abhydrolase_1	18.12	62.2222	4.00E−12
Pfam28	contig02638_1148_7	pfam12695-Abhydrolase_5	27.91	50.3448	3.00E−08
Pfam29	contig09772_490_3	pfam12695-Abhydrolase_5	20.75	79.3103	4.00E−13
Pfam30	contig01826_5617_105	pfam12697-Abhydrolase_6	19.40	95.5357	1.00E−08
Pfam31	contig11358_484_4	pfam12697-Abhydrolase_6	35.40	50.4464	6.00E−33
Pfam32	contig02573_1545_14	pfam12697-Abhydrolase_6	21.43	66.5179	2.00E−08
Pfam33	FYXZCPA02INS3R_479_1	pfam13472-Lipase_GDSL_2	22.66	69.9422	4.00E−09
Pfam34	FYXZCPA02I0GHD_502_1	pfam12697-Abhydrolase_6	20.83	50.8929	5.00E−13
Pfam35	FYXZCPA01DINQ9_368_1	pfam12695-Abhydrolase_5	27.27	56.5517	4.00E−09
Pfam36	FYXZCPA01CXWGZ_504_1	pfam13472-Lipase_GDSL_2	21.35	50.8671	1.00E−12
Pfam37	contig02497_4406_116	pfam00561-Abhydrolase_1	18.70	89.7778	2.00E−10
Pfam38	contig05216_3751_38	pfam12697-Abhydrolase_6	25.58	92.8571	3.00E−22
Pfam39	contig02933_3670_42	pfam12697-Abhydrolase_6	19.70	80.3571	6.00E−13
Pfam40	contig01893_2307_67	pfam13472-Lipase_GDSL_2	36.72	98.2659	4.00E−28
Pfam41	contig10277_498_3	pfam12697-Abhydrolase_6	22.01	66.0714	1.00E−20
Pfam42	FYXZCPA02IIP42_477_1	pfam12697-Abhydrolase_6	36.59	54.4643	1.00E−29

Continued

Pfam43	contig00476_8726_271	pfam12697-Abhydrolase_6	24.03	98.6607	5.00E−36
Pfam44	FYXZCPA01E5UXP_460_1	pfam12695-Abhydrolase_5	31.00	63.4483	3.00E−15
Pfam45	contig04216_513_5	pfam12695-Abhydrolase_5	30.00	64.1379	8.00E−20
Pfam46	contig02413_18858_411	pfam12695-Abhydrolase_5	25.70	100	2.00E−22
Pfam47	contig00197_534_3	pfam12695-Abhydrolase_5	29.00	64.1379	2.00E−19
Pfam48	FYXZCPA02GZ6XY_538_1	pfam12695-Abhydrolase_5	21.29	77.2414	3.00E−09
Pfam49	FYXZCPA02GN27Q_460_1	pfam12695-Abhydrolase_5	20.83	52.4138	1.00E−06
Pfam50	FYXZCPA01DX4R7_474_1	pfam12697-Abhydrolase_6	27.27	54.0179	3.00E−20
Pfam51	contig00250_28097_740	pfam12695-Abhydrolase_5	27.03	74.4828	1.00E−15
Pfam52	FYXZCPA01A44W5_490_1	pfam12695-Abhydrolase_5	24.49	81.3793	8.00E−10
Pfam53	contig04095_6514_173	pfam12697-Abhydrolase_6	18.71	67.4107	2.00E−06
Pfam54	contig00394_1735_8	pfam12695-Abhydrolase_5	20.60	100	3.00E−15
Pfam55	contig07114_639_5	pfam12695-Abhydrolase_5	24.77	65.5172	1.00E−17
Pfam56	contig04783_1443_31	pfam12697-Abhydrolase_6	24.00	51.3393	1.00E−11
Pfam57	contig00882_9723_289	pfam12697-Abhydrolase_6	22.07	93.3036	2.00E−16
Pfam58	contig01290_10856_428	pfam00561-Abhydrolase_1	16.30	97.3333	1.00E−11
Pfam59	FYXZCPA02HSGF6_484_1	pfam12697-Abhydrolase_6	33.06	53.5714	6.00E−20
Pfam60	FYXZCPA02JVPXC_526_1	pfam00561-Abhydrolase_1	16.55	62.6667	4.00E−07
Pfam61	contig01615_9318_182	pfam13472-Lipase_GDSL_2	34.66	100	3.00E−28
Pfam62	FYXZCPA02H13EN_510_1	pfam12695-Abhydrolase_5	30.00	60.6897	6.00E−15
Pfam63	contig08515_1275_11	pfam12697-Abhydrolase_6	20.32	80.8036	2.00E−18
Pfam64	contig06241_7320_88	pfam12697-Abhydrolase_6	25.31	91.9643	5.00E−26
Pfam65	contig01479_20270_446	pfam12697-Abhydrolase_6	30.33	100	1.00E−33
Pfam66	contig05933_1877_36	pfam00756-Esterase	32.33	100	1.00E−58
Pfam67	contig06201_5298_75	pfam12695-Abhydrolase_5	24.00	64.1379	3.00E−10
Pfam68	FYXZCPA01DJ12Y_449_1	pfam12695-Abhydrolase_5	25.47	64.1379	4.00E−17
Pfam69	contig01611_15601_375	pfam12695-Abhydrolase_5	26.03	97.2414	2.00E−12
Pfam70	contig00834_8473_222	pfam00561-Abhydrolase_1	22.32	94.2222	9.00E−16
Pfam71	contig01303_2032_11	pfam12695-Abhydrolase_5	25.77	52.4138	2.00E−07
Pfam72	GJ054VR02ISR5J_494_1	pfam13472-Lipase_GDSL_2	15.79	57.2254	7.00E−08
Pfam73	GJ054VR02HA4RF_492_1	pfam12697-Abhydrolase_6	20.96	65.625	5.00E−13
Pfam74	contig10048_381_3	pfam12697-Abhydrolase_6	16.41	51.3393	5.00E−06
Pfam75	contig07636_309_2	pfam12695-Abhydrolase_5	25.00	54.4828	1.00E−06
Pfam76	GJ054VR01C67B5_457_1	pfam13472-Lipase_GDSL_2	26.47	64.1618	1.00E−15
Pfam77	contig02952_2692_17	pfam12697-Abhydrolase_6	24.64	85.2679	9.00E−27
Pfam78	contig00620_19347_416	pfam12697-Abhydrolase_6	26.22	97.3214	6.00E−29
Pfam79	contig00889_16596_343	pfam12697-Abhydrolase_6	25.77	100	4.00E−27
Pfam80	contig03816_967_7	pfam12695-Abhydrolase_5	24.48	84.1379	5.00E−09
Pfam81	contig08847_841_6	pfam12695-Abhydrolase_5	24.83	97.2414	2.00E−10
Pfam82	contig05729_1202_11	pfam02230-Abhydrolase_2	27.10	93.4272	1.00E−28
Pfam83	contig10760_22051_6260	pfam12697-Abhydrolase_6	27.39	97.7679	1.00E−29
Pfam84	GJ054VR02FQQYK_497_1	pfam12695-Abhydrolase_5	21.56	99.3103	1.00E−14
Pfam85	contig00012_70893_20885	pfam12695-Abhydrolase_5	16.50	100	2.00E−07
Pfam86	contig02372_1061_6	pfam12697-Abhydrolase_6	26.88	66.0714	1.00E−05
Pfam87	contig00035_25497_6840	pfam12695-Abhydrolase_5	20.50	99.3103	1.00E−08
Pfam88	GJ054VR02IK5IL_506_1	pfam12695-Abhydrolase_5	31.68	66.2069	1.00E−19
Pfam89	contig03737_2033_16	pfam12695-Abhydrolase_5	27.97	66.2069	5.00E−11
Pfam90	contig01505_1164_6	pfam12697-Abhydrolase_6	23.33	87.9464	7.00E−19

Continued

Pfam91	contig00049_11245_240	pfam12697-Abhydrolase_6	21.43	66.5179	3.00E−07
Pfam92	GJ054VR01EWNW0_530_1	pfam12695-Abhydrolase_5	27.93	60.6897	2.00E−08
Pfam93	GJ054VR01BPBL5_460_1	pfam12697-Abhydrolase_6	29.51	50.4464	1.00E−24
Pfam94	contig00454_19280_412	pfam12697-Abhydrolase_6	30.17	100	1.00E−42
Pfam95	contig00119_19232_365	pfam12697-Abhydrolase_6	19.26	96.4286	1.00E−12
Pfam96	contig00119_19232_365	pfam12695-Abhydrolase_5	24.00	64.1379	1.00E−08
Pfam97	GJ054VR01CNOSD_447_1	pfam12695-Abhydrolase_5	25.25	57.931	4.00E−14
Pfam98	contig00960_13684_281	pfam12697-Abhydrolase_6	27.43	100	3.00E−39
Pfam99	contig07622_1530_12	pfam12697-Abhydrolase_6	25.00	82.5893	6.00E−23
Pfam100	GJ054VR02H7I8R_531_1	pfam00561-Abhydrolase_1	24.85	74.6667	2.00E−17
Pfam101	GJ054VR02G60MH_452_1	pfam12697-Abhydrolase_6	23.08	58.0357	5.00E−17
Pfam102	GJ054VR02JWL8C_458_1	pfam02230-Abhydrolase_2	20.98	58.216	8.00E−08
Pfam103	contig07870_496_3	pfam12697-Abhydrolase_6	15.20	72.3214	2.00E−06
Pfam104	contig10986_45864_11909	pfam12697-Abhydrolase_6	31.22	95.5357	2.00E−27
Pfam105	GJ054VR01BULKR_485_1	pfam13472-Lipase_GDSL_2	28.87	82.0809	2.00E−27
Pfam106	contig01919_852_7	pfam13472-Lipase_GDSL_2	30.99	94.7977	5.00E−25
Pfam107	contig03267_4254_30	pfam12695-Abhydrolase_5	25.00	80	2.00E−09
Pfam108	contig00064_9159_317	pfam00756-Esterase	32.33	100	4.00E−54
Pfam109	contig08275_501_3	pfam12695-Abhydrolase_5	22.12	60.6897	3.00E−06
Pfam111	GJ054VR01C7J7L_531_1	pfam13472-Lipase_GDSL_2	32.41	57.8035	3.00E−17
Pfam112	contig03549_307_2	pfam12695-Abhydrolase_5	23.91	57.931	3.00E−16
Pfam113	GJ054VR01DFKXI_458_1	pfam12697-Abhydrolase_6	29.55	58.4821	3.00E−25
Pfam114	GJ054VR02HLP9Q_442_1	pfam12697-Abhydrolase_6	19.05	60.7143	4.00E−10
Pfam115	contig00103_70784_19920	pfam12697-Abhydrolase_6	24.02	96.875	3.00E−19
Pfam116	contig01310_928_5	pfam12695-Abhydrolase_5	30.28	70.3448	5.00E−23
Pfam117	GJ054VR01B2IB1_498_1	pfam12697-Abhydrolase_6	17.05	57.1429	4.00E−11
Pfam118	contig00490_53254_1018	pfam12697-Abhydrolase_6	26.75	98.2143	6.00E−23
Pfam119	GJ054VR01D1B3V_512_1	pfam12695-Abhydrolase_5	27.36	63.4483	7.00E−10
Pfam120	GJ054VR01C6IR9_411_1	pfam12695-Abhydrolase_5	19.05	68.9655	2.00E−16
Pfam121	contig00981_111696_2437	pfam00561-Abhydrolase_1	22.34	97.3333	2.00E−18
Pfam122	contig00363_90062_1905	pfam12697-Abhydrolase_6	19.70	80.3571	2.00E−11
Pfam123	contig07913_1207_7	pfam12697-Abhydrolase_6	18.80	99.5536	4.00E−07
Pfam124	contig08413_774_5	pfam12697-Abhydrolase_6	23.91	97.3214	2.00E−40
Pfam125	contig00941_2929_50	pfam12695-Abhydrolase_5	28.24	56.5517	7.00E−10
Pfam126	contig00109_61254_1373	pfam13472-Lipase_GDSL_2	22.99	100	1.00E−24
Pfam127	contig00271_130367_2942	pfam12695-Abhydrolase_5	27.37	59.3103	8.00E−07
E.C. no. 1	contig01459_13828_351	triacylglycerol lipase	76.75	77.82	1.00E−90
E.C. no. 2	FYXZCPA01D1H7X_445_1	protein-glutamate methylesterase	53.15	88.80	5.00E−29
E.C. no. 3	FYXZCPA01EWMQU_502_1	protein-glutamate methylesterase	52.63	55.07	9.00E−06
E.C. no. 4	FYXZCPA01ATIY8_536_1	triacylglycerol lipase	50.96	61.33	2.00E−38
E.C. no. 5	contig01695_17539_381	protein-glutamate methylesterase	77.60	55.01	3.00E−84
E.C. no. 6	contig07322_3581_40	triacylglycerol lipase	60.40	99.60	2.00E−70
E.C. no. 7	FYXZCPA02J31IF_470_1	protein-glutamate methylesterase	94.55	69.62	3.00E−55
E.C. no. 8	contig07990_2621_29	protein-glutamate methylesterase	74.50	99.43	2.00E−157

Continued

E.C. no. 9	contig10134_349_3	protein-glutamate methylesterase	51.76	60.43	5.00E−17
E.C. no. 10	contig01574_9423_255	protein-glutamate methylesterase	54.29	50.72	4.00E−04
E.C. no. 11	contig04439_2235_47	protein-glutamate methylesterase	54.96	52.49	2.00E−72
E.C. no. 12	contig07769_983_5	protein-glutamate methylesterase	54.31	92.80	2.00E−31
E.C. no. 13	contig05129_1344_6	protein-glutamate methylesterase	59.58	97.90	3.00E−121
E.C. no. 14	contig01573_4582_113	carboxymethylenebutenolidase	54.38	85.38	6.00E−52
E.C. no. 15	contig09355_501_4	protein-glutamate methylesterase	62.50	84.55	1.00E−34
E.C. no. 16	FYXZCPA01C4V38_432_1	triacylglycerol lipase	51.41	56.71	9.00E−19
E.C. no. 17	contig05158_2403_20	protein-glutamate methylesterase	50.85	50.64	8.00E−25
E.C. no. 18	contig04980_2741_49	protein-glutamate methylesterase	53.38	97.79	2.00E−35
E.C. no. 19	contig00779_27300_655	protein-glutamate methylesterase	77.68	99.44	2.00E−160
E.C. no. 20	FYXZCPA02JFV78_481_1	protein-glutamate methylesterase	50.00	68.80	2.00E−20
E.C. no. 21	contig01548_1890_12	protein-glutamate methylesterase	54.29	50.72	2.00E−05
E.C. no. 22	FYXZCPA02HPIN0_465_1	6-phosphogluconolactonase	95.48	53.63	3.00E−74
E.C. no. 23	contig00680_7380_200	protein-glutamate methylesterase	62.16	97.76	2.00E−75
E.C. no. 24	FYXZCPA01AW52H_501_1	protein-glutamate methylesterase	64.38	58.40	4.00E−24
E.C. no. 25	contig04738_1406_28	protein-glutamate methylesterase	52.21	83.70	6.00E−27
E.C. no. 26	contig01007_1067_9	triacylglycerol lipase	75.79	82.25	1.00E−70
E.C. no. 27	contig01899_519_4	triacylglycerol lipase	50.29	50.88	3.00E−48
E.C. no. 28	contig04083_1665_9	protein-glutamate methylesterase	50.42	93.60	8.00E−28
E.C. no. 29	contig00864_10444_246	protein-glutamate methylesterase	54.79	99.47	2.00E−53
E.C. no. 30	contig01748_3935_54	triacylglycerol lipase	65.05	90.88	5.00E−117
E.C. no. 31	contig01312_6616_156	triacylglycerol lipase	78.60	98.28	9.00E−86
E.C. no. 32	contig00956_967_6	protein-glutamate methylesterase	60.07	79.17	4.00E−93
E.C. no. 33	contig03179_1774_53	protein-glutamate methylesterase	50.56	73.55	3.00E−22
E.C. no. 34	contig05175_614_5	protein-glutamate methylesterase	51.35	88.80	2.00E−29
E.C. no. 35	contig04148_2863_53	protein-glutamate methylesterase	51.11	59.42	4.00E−04
E.C. no. 36	contig01109_4523_139	protein-glutamate methylesterase	51.33	98.67	2.00E−41
E.C. no. 37	FYXZCPA02IIP42_477_1	3-oxoadipate enol-lactonase	75.84	55.19	2.00E−60
E.C. no. 38	FYXZCPA01CK4RL_454_1	protein-glutamate methylesterase	52.70	58.73	6.00E−16
E.C. no. 39	contig06284_1669_16	protein-glutamate methylesterase	51.26	95.20	9.00E−30
E.C. no. 40	FYXZCPA02GN27Q_460_1	acetylcholinesterase	56.49	57.09	5.00E−41
E.C. no. 41	contig01142_9836_320	protein-glutamate methylesterase	78.51	97.58	3.00E−52
E.C. no. 42	contig05497_1971_21	protein-glutamate methylesterase	50.00	92.11	2.00E−45
E.C. no. 43	contig00811_7133_147	protein-glutamate methylesterase	72.13	100.00	1.00E−141
E.C. no. 44	contig03571_4243_108	protein-glutamate methylesterase	51.22	58.91	2.00E−56
E.C. no. 45	contig00177_37689_2549	protein-glutamate methylesterase	64.58	99.18	7.00E−137
E.C. no. 46	contig01355_5503_132	protein-glutamate methylesterase	51.30	92.00	2.00E−26
E.C. no. 47	contig07567_976_5	triacylglycerol lipase	56.25	52.03	2.00E−61
E.C. no. 48	contig04073_10150_130	6-phosphogluconolactonase	70.09	92.24	7.00E−73
E.C. no. 49	contig04394_3033_83	protein-glutamate methylesterase	60.83	96.77	3.00E−37
E.C. no. 50	contig00873_13383_308	protein-glutamate methylesterase	55.26	55.07	5.00E−05
E.C. no. 51	contig00238_4315_99	protein-glutamate methylesterase	60.21	82.65	1.00E−101
E.C. no. 52	contig01644_12678_360	protein-glutamate methylesterase	51.79	52.58	5.00E−23

Continued

E.C. no. 53	contig05752_1086_9	protein-glutamate methylesterase	54.70	63.59	5.00E−87
E.C. no. 54	contig06520_2685_59	protein-glutamate methylesterase	51.20	98.81	2.00E−90
E.C. no. 55	FYXZCPA01D95YE_442_1	protein-glutamate methylesterase	50.00	54.03	4.00E−09
E.C. no. 56	contig12591_616_6	protein-glutamate methylesterase	50.43	53.99	1.00E−22
E.C. no. 57	contig01479_20270_446	protein-glutamate methylesterase	50.82	93.85	6.00E−32
E.C. no. 58	contig04528_7408_175	protein-glutamate methylesterase	62.18	96.75	5.00E−41
E.C. no. 59	contig01961_1964_14	protein-glutamate methylesterase	62.50	89.82	6.00E−120
E.C. no. 60	contig05933_1877_36	carboxylesterase	72.10	97.18	6.00E−119
E.C. no. 61	FYXZCPA02JFTDD_499_1	protein-glutamate methylesterase	50.00	56.97	5.00E−32
E.C. no. 62	contig02175_3096_107	protein-glutamate methylesterase	54.46	99.11	5.00E−49
E.C. no. 63	contig05606_1287_8	protein-glutamate methylesterase	65.03	55.04	2.00E−58
E.C. no. 64	contig08366_295_2	protein-glutamate methylesterase	51.81	65.35	2.00E−18
E.C. no. 65	contig03691_2959_22	protein-glutamate methylesterase	58.57	51.85	8.00E−20
E.C. no. 66	contig06683_514_3	protein-glutamate methylesterase	50.39	94.85	8.00E−33
E.C. no. 67	GJ054VR02IO0TL_534_1	triacylglycerol lipase	50.39	59.76	7.00E−53
E.C. no. 68	GJ054VR01CUP6J_454_1	protein-glutamate methylesterase	52.94	75.56	6.00E−23
E.C. no. 69	contig00456_28478_646	protein-glutamate methylesterase	50.23	96.09	7.00E−49
E.C. no. 70	contig03921_533_3	protein-glutamate methylesterase	64.22	88.62	7.00E−38
E.C. no. 71	contig10760_22051_6260	triacylglycerol lipase	81.57	99.22	2.00E−115
E.C. no. 72	contig04473_1195_7	protein-glutamate methylesterase	50.43	85.19	4.00E−26
E.C. no. 73	GJ054VR02GCI2F_479_1	protein-glutamate methylesterase	50.00	97.60	1.00E−33
E.C. no. 74	GJ054VR01EFZ0J_498_1	protein-glutamate methylesterase	62.50	59.02	3.00E−49
E.C. no. 75	contig07065_401_2	protein-glutamate methylesterase	50.00	60.87	6.00E−06
E.C. no. 76	contig02240_1506_10	protein-glutamate methylesterase	50.00	55.61	1.00E−27
E.C. no. 77	contig05971_806_4	protein-glutamate methylesterase	51.69	73.55	6.00E−23
E.C. no. 78	GJ054VR02JO9KC_470_1	protein-glutamate methylesterase	51.28	95.12	5.00E−27
E.C. no. 79	contig04043_1277_11	protein-glutamate methylesterase	50.44	83.70	1.00E−27
E.C. no. 80	contig00960_13684_281	protein-glutamate methylesterase	65.05	90.88	5.00E−117
E.C. no. 81	GJ054VR01B5TSY_488_1	protein-glutamate methylesterase	78.51	97.58	2.00E−52
E.C. no. 82	GJ054VR02II30F_510_1	protein-glutamate methylesterase	51.96	81.60	6.00E−27
E.C. no. 83	contig00732_3200_23	protein-glutamate methylesterase	57.14	50.72	8.00E−04
E.C. no. 84	GJ054VR02H7I8R_531_1	dihydrocoumarin hydrolase	60.23	63.77	2.00E−59
E.C. no. 85	contig00987_35243_1622	protein-glutamate methylesterase	64.58	99.18	7.00E−137
E.C. no. 86	contig03267_4254_30	Carboxymethylenebutenolidase	53.18	67.98	4.00E−42
E.C. no. 87	contig00064_9159_317	carboxylesterase	72.10	97.18	6.00E−119
E.C. no. 88	GJ054VR02JYT6B_479_1	protein-glutamate methylesterase	62.03	63.20	1.00E−24
E.C. no. 89	contig00955_1565_14	triacylglycerol lipase	78.09	60.75	6.00E−71
E.C. no. 90	contig07351_1200_9	protein-glutamate methylesterase	54.29	50.72	6.00E−04
E.C. no. 91	contig05469_1382_15	protein-glutamate methylesterase	54.76	59.42	8.00E−05
E.C. no. 92	contig04613_1156_8	protein-glutamate methylesterase	55.10	87.50	5.00E−10
E.C. no. 93	contig05233_960_7	protein-glutamate methylesterase	51.43	50.72	2.00E−04
E.C. no. 94	contig06045_1712_10	protein-glutamate methylesterase	62.61	72.78	6.00E−37
E.C. no. 95	contig05435_745_4	protein-glutamate methylesterase	50.00	81.63	2.00E−06

Continued

E.C. no. 96	contig00154_53031_1113	protein-glutamate methylesterase	50.85	50.64	6.00E−25
E.C. no. 97	GJ054VR02F0FR6_436_1	protein-glutamate methylesterase	61.74	51.11	6.00E−28
E.C. no. 98	GJ054VR01A7XTC_544_1	protein-glutamate methylesterase	80.21	51.28	5.00E−78
E.C. no. 99	contig06676_1954_11	protein-glutamate methylesterase	54.29	53.60	1.00E−15
E.C. no. 100	contig01396_3317_22	triacylglycerol lipase	51.28	80.29	3.00E−79
E.C. no. 101	contig09830_293_3	protein-glutamate methylesterase	58.76	59.88	5.00E−29
E.C. no. 102	contig00023_26624_552	6-phosphogluconolactonase	69.06	96.12	1.00E−76
E.C. no. 103	GJ054VR02GMPRP_548_1	triacylglycerol lipase	84.03	62.34	5.00E−58
E.C. no. 104	GJ054VR01EH9PQ_550_1	protein-glutamate methylesterase	50.00	67.86	1.00E−05
E.C. no. 105	GJ054VR01CJYLF_496_1	protein-glutamate methylesterase	50.00	50.38	3.00E−10
E.C. no. 106	contig00373_119927_2499	protein-glutamate methylesterase	74.36	100.00	2.00E−157
E.C. no. 107	GJ054VR01BUKCP_412_1	triacylglycerol lipase	59.12	53.52	2.00E−46

Table S2. Blastx search using FS02 sequence.

Protein hit	Score	Query coverage	E-value	Ident
Hypothetical protein AF0514 (Archaeoglobus fulgidus DSM 4304)	377	100%	2e−131	99%
Hypothetical protein Ferp 1969 (Ferroglobus placidus DSM10642)	227	97%	3e−72	60%
Putative hydrolase of alpha/beta superfamily (Archaeoglobus sulfaticallidus)	214	96%	4e−67	55%
Hypothetical protein Arcve 1663 (Archaeoglobus veneficus SNP6)	190	99%	1e−57	52%
Hydrolase of alpha/beta superfamily (Natronobacterium gregoryi SP2)	100	100%	8e−23	31%
Hypothetical protein HacjB3 (Halalkalicoccus jeotgali)	99.0	96%	3e−22	32%
Hypothetical protein (Haloferax larsenii)	98.6	93%	5e−22	32%
Hypothetical protein (Haloferax elongans)	98.2	93%	5e−22	32%
Alpha/beta hydrolase (Halostagnicola larsenii XH-48)	97.8	98%	7e−22	32%
Hydrolase of alpha/beta superfamily-like protein (Natrialba taiwanensis)	96.7	97%	2e−21	32%
Hydrolase of alpha/beta superfamily-like protein (Natroncoccus amylolyticus)	97.4	98%	2e−21	32%
Alpha/beta hydrolase (Halosarcina pallida)	94.4	91%	1e−20	34%

Table S3. Protein sequences used for phylogenetic analysis.

Uniprot ID	NCBI RefSeq	Archaea/Bacteria	Species	Annotation	Identity to FS02
W0JIR4	WP_049951831.1	A	*Halostagnicola larsenii*	alpha/beta hydrolase	32%
L0ALB7	WP_005580491.1	A	*Natronobacterium gregoryi*	alpha/beta hydrolase	31%
M0DBB1	WP_008384884.1	A	*Halogeometricum pallidum*	alpha/beta hydrolase	34%
N0BCB5	WP_015590857.1	A	*Archaeoglobus sulfaticallidus*	hydrolase of the alpha/beta superfamily	55%
D3S043	WP_012966445.1	A	*Ferroglobus placidus*	hypothetical protein	60%
O29736	WP_010878021.1	A	*Archaeoglobus fulgidus*	hypothetical protein	99%
C1F9B1	WP_012680666.1	B	*Acidobacterium capsulatum*	alpha/beta hydrolase	30%
Q022D1	WP_011684934.1	B	*Solibacter usitatus*	hypothetical protein	33%
I7ZH91	WP_007184188.1	B	*Hydrocarboniphaga effuse*	hypothetical protein	27%
A6VBB3	WP_003150434.1	B	*Pseudomonas aeruginosa*	alpha/beta hydrolase	29%
I4C162	WP_052315987.1	B	*Desulfomonile tiedjei*	hypothetical protein	29%
A0LIX5	WP_011698547.1	B	*Syntrophobacter fumaroxidans*	alpha/beta hydrolase family protein	26%

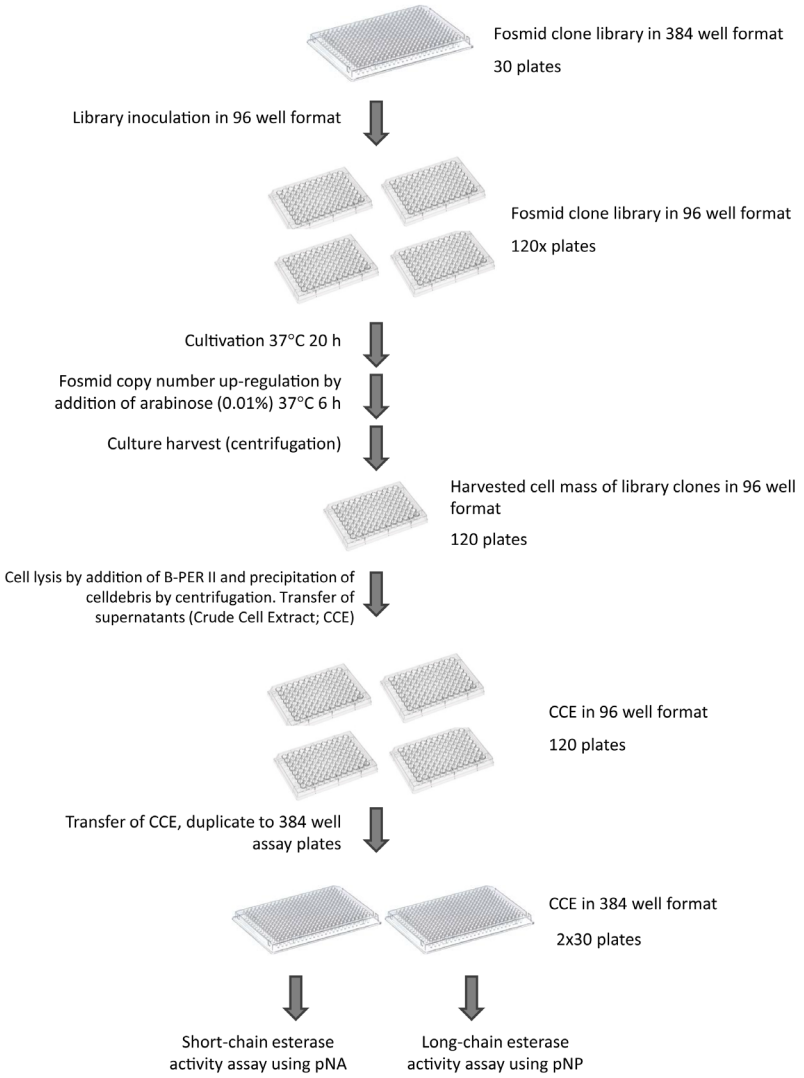


Figure S1. Schematic overview of functional screening experiment set-up. For details on experimental conditions it is referred to the Materials and Methods section.

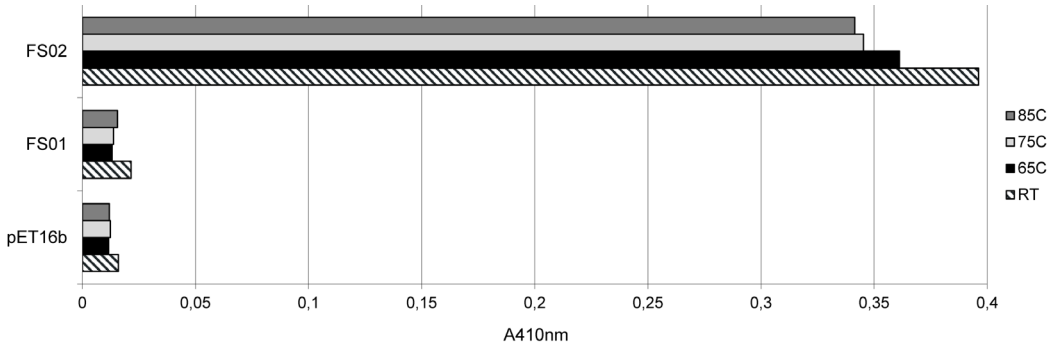


Figure S2. End-point (30 min) long-chain esterase assay (substrate; 4-Nitrophenyl palmitate) using crude cell extracts containing FS01 and FS02 from expression of the native sequence from pET16b in E. coli BL21 (DE3) RELP (codon+). Extracts were pre-incubated for 30 minutes at 20˚C, 65˚C, 75˚C or 85˚C, respectively, and the remaining soluble fraction assayed. FS02 activity (measured at 37˚C) was found well over background (empty pET16b vector control), and remained even after incubation at 85˚C, indicating high thermostability of FS02. In contrast, FS01 did not show any measureable esterase activity.

Figure S3. Multiple sequence alignment of the amino acid sequence of enzyme FS02 against selected alpha/beta hydrolase sequences from multiple archaea and bacteria (see **Table S3**). The corresponding amino acid numbers are given in right as well as left panel, and conserved positions are indicated by lower panel. This alignment was used for construction of the phylogenetic tree presented in **Figure 1**.

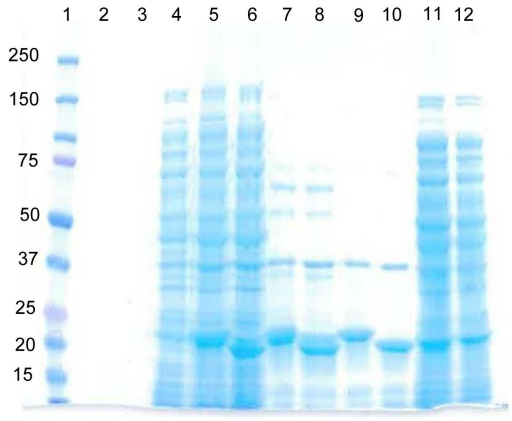


Figure S4. Verification of the production of enzyme FS02 as determined by (10%) SDS-PAGE using sample from induced cultivation of E. coli BL21 (DE3) RELP (codon+) containing plasmids for expression of gene variants with and without a His6-tag. Lane 1, BioRad Dual Color Precision Plus standard (161-0374) kDa indicated; 2, extracellular fraction FS02-His6; 3, extracellular fraction FS02; 4, crude cell extract (CCE) pET16b; 5, CCE FS02-His6; 6, CCE FS02; 7, CCE FS02-His6 incubated 65˚C 25 minutes; 8, CCE FS02 incubated 65˚C 25 minutes; 9, CCE FS02-His6 incubated 80˚C 30 minutes; 10, CCE FS02-His6 incubated 80˚C 30 minutes; 11, CCE FS02-His6 non-induced; 12, CCE FS02 non-induced. Molecular mass of FS02 with and without His6-tag is 23.4 and 21.2 kDa, respectively. SDS-PAGE analysis of samples after incubation at elevated temperatures show FS02 thermal stability, and visualizes the heat-based "purification" of the enzyme.