

SCIENTIFIC REPORTS



OPEN

Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases

Dimitra Zarafeta^{1,2}, Danai Moschidi², Efthymios Ladoukakis², Sergey Gavrillov³, Evangelia D. Chrysina¹, Aristotelis Chatziioannou¹, Ilya Kublanov³, Georgios Skretas¹ & Fragiskos N. Kolisis²

Received: 20 June 2016

Accepted: 14 November 2016

Published: 19 December 2016

Biocatalysts exerting activity against ester bonds have a broad range of applications in modern biotechnology. Here, we have identified a new esterolytic enzyme by screening a metagenomic sample collected from a hot spring in Kamchatka, Russia. Biochemical characterization of the new esterase, termed EstDZ2, revealed that it is highly active against medium chain fatty acid esters at temperatures between 25 and 60 °C and at pH values 7–8. The new enzyme is moderately thermostable with a half-life of more than six hours at 60 °C, but exhibits exquisite stability against high concentrations of organic solvents. Phylogenetic analysis indicated that EstDZ2 is likely an *Acetothermia* enzyme that belongs to a new family of bacterial esterases, for which we propose the index XV. One distinctive feature of this new family, is the presence of a conserved GHSAG catalytic motif. Multiple sequence alignment, coupled with computational modelling of the three-dimensional structure of EstDZ2, revealed that the enzyme lacks the largest part of the “cap” domain, whose extended structure is characteristic for the closely related Family IV esterases. Thus, EstDZ2 appears to be distinct from known related esterolytic enzymes, both in terms of sequence characteristics, as well as in terms of three-dimensional structure.

Lipolytic enzymes (EC 3.1.1.x) hydrolyze the ester bonds in lipids and have received great attention due to their potential use in a broad range of biotechnological applications, such as the production of biofuels and fine chemicals^{1,2}. Depending on whether they hydrolyze small water-soluble esters or ester bonds of long-chain acylglycerols, they are classified as carboxylesterases (EC 3.1.1.1) or lipases (EC 3.1.1.3)³. Both types of enzymes are serine hydrolases that contain a functional serine at their active site, which is usually located within the conserved sequence motif G-X-S-X-G⁴. This serine, along with an acidic residue -aspartic acid or glutamic acid- and a histidine, to which the acidic residue is hydrogen-bonded, constitute a highly conserved catalytic triad⁵. In addition, ester hydrolases share common structural features, such as a characteristic α/β hydrolase fold, as well as biochemical characteristics, like co-factor independence^{3,6}.

Based on amino acid sequence homology and fundamental biochemical properties, bacterial lipolytic enzymes were initially classified into eight distinct families in the landmark paper by Arpingy and Jaeger⁵. Family I consists of lipases and includes six subfamilies⁵. The remaining seven families were denoted as Families II–VIII and include carboxyl esterases or “true esterases”. Despite the fact that this classification was proposed for bacterial esterases, it should be noted that many families in the original analysis included not only bacterial sequences, but also archaeal and eukaryotic representatives⁵. The more recent discovery of the lipase-like poly[(R)-3-hydroxybutyrate] depolymerase PhaZ7⁷, the hyperthermophilic esterase EstD⁸, and a large number of additional metagenome-derived esterolytic hydrolases^{9–12}, expanded this classification to include six new families, denoted as Families IX–XIV. More recently conducted metagenomic studies have revealed a plethora of new bacterial lipolytic enzymes that seem to define additional new families, which do not follow the ordered

¹Institute of Biology, Medicinal Chemistry & Biotechnology, National Hellenic Research Foundation, Athens, Greece.

²Laboratory of Biotechnology, School of Chemical Engineering, National Technical University of Athens, Athens, Greece. ³Winogradsky Institute of Microbiology, Research Center for Biotechnology Russian Academy of Sciences, Moscow, Russian Federation. Correspondence and requests for materials should be addressed to G.S. (email: gskretas@eie.gr) or F.N.K. (email: kolisis@chemeng.ntua.gr)

numerical indices of the original Arpingy and Jaeger classification. Examples of such recently discovered families of bacterial esterolytic enzymes are the ones defined originally by the esterases Est10¹³, EstWSD¹⁴, EstD2¹⁵, and EstLiu¹⁶, among others.

Carboxyl esterases and lipases are valuable industrial enzymes. Their native hydrolytic nature is complemented by a broad range of substrate specificities, as well as high stereo- and regio-selectivity¹⁷. Besides their physiological role, lipolytic enzymes possess the ability to perform synthesis of ester bonds or transesterification reactions in non-aqueous media, following the thermodynamic reversion of the hydrolysis reaction route¹⁸. These features render them protagonist biocatalysts in the flavor and aroma, food, biodiesel, fine chemical, cosmetics, detergent, and pharmaceutical industries^{19–21}. Their biotechnological significance is reflected by the fact that several esterases and lipases have been commercialized by leading biotech companies such as Fluka, Novozymes, Amano, Diversa, Roche Diagnostics, Thermogen and others⁴.

Nowadays, an important bottleneck to the wide industrial use of lipolytic enzymes is that they are often required to perform under harsh conditions. These include temperatures above 40 °C, significant concentrations of organic solvents, metal ions, surfactants, and other agents known to promote enzyme denaturation and inactivation^{22–24}. As a result, stability against elevated temperatures and overall tolerance to protein-destabilizing conditions is a crucial prerequisite before the broad industrial use of this type of enzymes is realized.

Lipolytic enzymes are ubiquitous in nature and they can be found in all three domains of life (bacteria, archaea, eukarya) residing in all types of environmental niches, including high-temperature ecosystems. Their abundance has led to the discovery and characterization of a number of thermostable esterases and lipases, the majority of which originate from cultivable thermophilic microorganisms⁶. While the need for novel enzymes is ever-growing, the biocatalytic capabilities of the vast majority of microbial organisms remain unexplored due to the fact that more than 99% of them cannot be cultured using standard laboratory procedures^{25,26}. Metagenomic approaches aim to address this issue and bypass the limitations imposed by current culturing techniques. This is achieved by providing direct access to the genomic material of microorganisms residing in extreme environments, where biocatalysts resistant to harsh “industrial conditions” are likely to be encountered. Furthermore, besides a discovery tool for biotechnology-relevant enzymes, metagenomic analysis enhances our understanding of the physiology of microbial communities residing in different types of environments, their complex relationships, and the related evolutionary paths²⁷.

In this study, we aimed to identify new lipolytic enzymes with novel sequence and functional characteristics, and with properties suitable for industrial applications. To achieve this, we investigated the putative lipolytic enzymes encoded in the genomes of microorganisms residing in a hot spring located in Kamchatka, Russia. This site possesses characteristics, which are expected to favor the presence of organisms with various hydrolytic activities, including lipolytic ones. Environmental sampling from this location, followed by high-throughput sequencing of the isolated metagenomic DNA and bioinformatics screening for the identification of putative esterolytic sequences, revealed an open reading frame, which presented a low – but notable – similarity to previously characterized esterases. The identified sequence was amplified, cloned, expressed in *Escherichia coli*, and purified in soluble and functional form. Detailed biochemical characterization of this recombinant protein demonstrated that it constitutes a new carboxyl esterase with preference towards esters of fatty acids with medium chain length. The new enzyme, termed EstDZ2, was found to exhibit moderate thermostability, as well as significant tolerance against a variety of organic solvents, metal ions and other denaturing agents. Notably, phylogenetic analysis revealed that the sequence of EstDZ2 does not cluster with that of any other characterized bacterial esterase, but instead exhibits characteristics that deviate from those encountered in related families of bacterial esterolytic enzymes. In accordance with this, protein structural modelling suggested that some of the structural features that characterize Family IV esterases, the most closely related family of EstDZ2 and its homologs, appear to be absent from the structure of EstDZ2. EstDZ2, thus, constitutes a new candidate enzyme for industrial biotransformations, which appears to belong to a new family of esterolytic enzymes. For this new family of bacterial esterolytic enzymes, we propose the index XV.

Results

Environmental sampling. An environmental sample was collected from the Solnechny (Sun) hot spring located in Uzon Caldera, Kronotsky reserve, Kamchatka, Russia (coordinates: 54° 29′ 56″ N, 159° 59′ 31″ E) in September 2011. At the moment when the sample was taken, the temperature of the site was 61–64 °C and the pH 5.8–6.0. Detailed description of the site is given by Menzel *et al.*²⁸. The hot spring is located in a geothermal swamp and constitutes a small pool that emanates CO₂. Its bottom is covered by a thick layer of plant residues buried in sludge and iron oxide depositions, and it is surrounded by stromatolites. It is, thus, characterized by a high content of organic carbon and active processes of organic matter degradation, conditions that should favor the presence of organisms with various hydrolytic activities, including lipolytic ones. A recent metagenomic study of the microbial diversity of the Solnechny hot spring revealed the prevalence of bacterial species, 40% of which could not be assigned to any phylum²⁸. These characteristics make the Solnechny hot spring a highly promising source for identifying new enzymes with a broad range of activities and high levels of stability against denaturing conditions. Metagenomic analysis in this work was carried out using a sample collected from the mud, as this was found to be the richest part of the hot spring in terms of microbial diversity²⁸.

Discovery of EstDZ2. Metagenomic DNA was isolated from the aforementioned environmental sample and sequenced using a Roche/454 Titanium FLX high-throughput sequencing platform. Reads were assembled into contigs, and the predicted open reading frames (ORFs) were examined against NCBI’s conserved domain database²⁹ for the presence of domains related to esterolytic activities, as described in the Methods section. From this analysis, we selected for further investigation a putative protein sequence, termed EstDZ2 α , which was assigned to the esterase/lipase superfamily by NCBI’s conserved domain database and contained multiple esterolytic-like

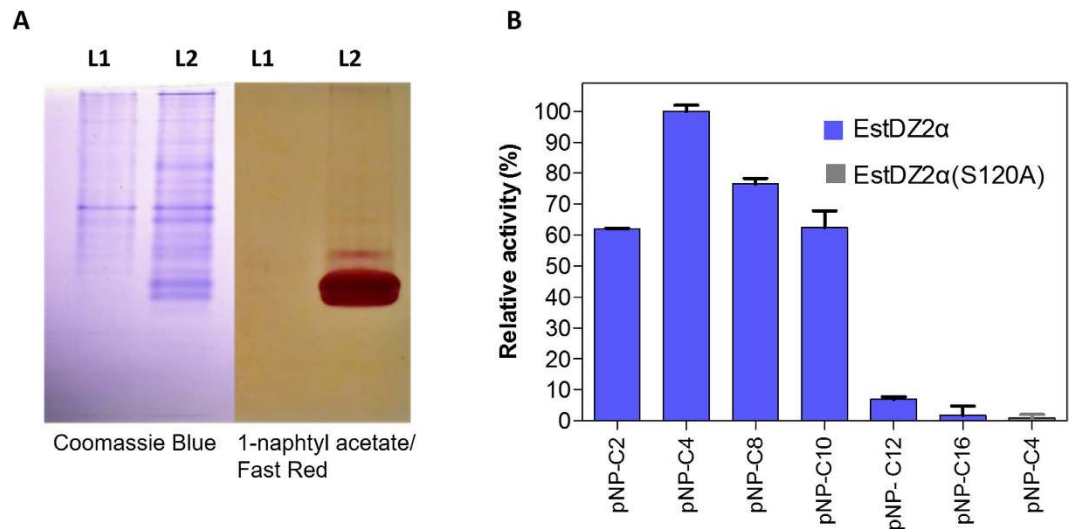


Figure 1. Discovery of the new esterolytic enzyme EstDZ2 and initial substrate specificity analysis.

(A) Detection of EstDZ2 α esterolytic activity via native PAGE analysis and Fast Red staining using 1-naphthyl acetate as a substrate. L1: Lysate of cells carrying an empty vector, L2: Lysate of cells overexpressing *estDZ2 α* . (B) Initial substrate profiling of the esterolytic activity of EstDZ2 α and the variant EstDZ2 α (S120A). The relative enzymic activity was measured as released pNP at 410 nm (pH 7, 40 °C). The reported values correspond to the mean value of three independent experiments performed in triplicate and the error bars to one standard deviation from the mean value.

domains. The EstDZ2 α sequence comprises 272 amino acids and has a predicted molecular mass of 29.4 kDa. Two sequence motifs, which are highly conserved among different families of esterolytic enzymes were identified: (i) the GX SXG motif, which is expected to be part of the active site, with the catalytic serine at the center, and (ii) the tetrapeptide sequence HG GG, which is known to contribute to the formation of the oxyanion hole required for catalysis in this class of enzymes³⁰. Analysis using Delta-BlastP revealed 23% amino acid sequence identity (90% coverage) to a previously characterized isoprenylcysteine methylesterase from *Arabidopsis thaliana*³¹. Analysis of the EstDZ2 α sequence using the SignalP 4.1³² and Signal-Blast³³ servers, predicted that the first 21 amino acids of the protein serve as a signal peptide, thus suggesting that EstDZ2 α is naturally a secreted enzyme.

The *estDZ2a* gene was amplified by PCR from the metagenomic DNA sample and was cloned into the plasmid pET-28a(+) to form the vector pET-EstDZ2 α . *E. coli* BL21(DE3) cells were transformed with pET-EstDZ2 α , grown in 5 ml of Luria-Bertani (LB) medium at 37 °C with shaking until the culture reached an optical density at 600 nm (OD₆₀₀) of 0.5, at which point 0.2 mM isopropyl thio- β -D-galactoside (IPTG) were added to induce *estDZ2a* overexpression. These cells were collected, lysed by brief sonication and their lysates were assayed for their ability to hydrolyze *p*-nitrophenyl butyrate colorimetrically. We observed that the characteristic yellow color of *p*-nitrophenol (pNP), which is indicative of ester bond cleavage, appeared only when *estDZ2 α* was expressed (see below), thus indicating that EstDZ2 α possesses esterolytic activity against pNP-butyrate. Furthermore, when the proteins contained in the same cell lysates were separated by native polyacrylamide gel electrophoresis (PAGE), the gel was exposed to 1-naphthyl acetate as a potential substrate for hydrolysis, and stained with Fast Red, a band of red-brown color appeared rapidly after the addition of the substrate/indicator mixture only for the EstDZ2 α -producing sample (Fig. 1A). Taken together, these results demonstrate that EstDZ2 α possesses esterolytic activity.

An initial substrate preference test, using soluble lysates from EstDZ2 α -producing cells and pNP-esters derived from fatty acids with a range of carbon chain lengths, demonstrated that this enzyme exhibits esterolytic activity against medium chain lengths (C2-C10), while its activity is barely detectable for C12 or larger substrates (Fig. 1B). To investigate the esterolytic activity of EstDZ2 α against natural substrates, a variety of naturally occurring esters, such as ethyl-acetate, butyl-acetate, cinnamyl-acetate, tributyrin and triolein, as well as natural oils, such as olive oil, palm oil, and sunflower oil, were tested. Esterolytic activity against these substrates was assayed using the pH indicator bromothymol blue, which detects the drop in the pH of the reaction that occurs due to the release of H⁺ ions upon ester bond hydrolysis³⁴. Clarified lysates of *estDZ2 α* -overexpressing cells caused a rapid color change in this assay only when tributyrin was added as a substrate (Supplementary Fig. S1). After long incubation times, a slight color change was also detected in the presence of cinnamyl-acetate, thus indicating that EstDZ2 α might have very low activity against this substrate (data not shown). No activity against triolein and the tested natural oils could be detected even at long incubation times (Supplementary Fig. S1 and data not shown). These results are in agreement with the initial substrate specificity test using pNP-esters, where long-chained pNP esters could be not hydrolyzed by EstDZ2 α and where the enzyme exhibited the highest levels of esterolytic activity against pNP-butyrate. Collectively, these results demonstrate that EstDZ2 α is a new esterolytic enzyme that acts as an esterase rather than a lipase.

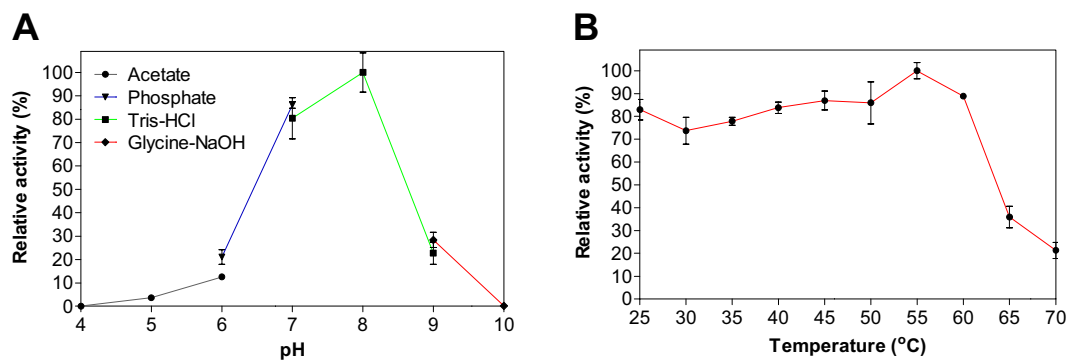


Figure 2. Effect of pH and temperature on EstDZ2 activity. (A) Effect of pH on EstDZ2 activity. Enzymic activity was measured in the standard assay reaction at 40 °C for 5 min, at pH values ranging from 4 to 10, using the indicated buffers. (B) Effect of temperature on EstDZ2 activity. Enzymic activity was measured at temperatures ranging from 25 to 70 °C and pH 8. The reported values correspond to the mean value from three independent experiments performed in triplicate and the error bars to one standard deviation from the mean value.

Substrate (pNP ester)	K_m (mM)	V_{max} ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1}\text{mM}^{-1}$)
Acetate (C2)	0.67 ± 0.15	36.8 ± 5.2	1,226	1,757
Butyrate (C4)	0.32 ± 0.09	139.6 ± 13.3	4,653	14,375
Caprylate (C8)	0.20 ± 0.02	61.8 ± 3.3	2,060	10,437
Caprate (C10)	0.12 ± 0.01	28.1 ± 0.5	936	7,551

Table 1. Kinetic parameters of EstDZ2 hydrolysis of various pNP-esters.

Biochemical characterization. In order to study the biochemical properties of the newly identified enzyme, we produced it recombinantly in *E. coli* and purified it in soluble form. For this purpose, we cloned a truncated version of the *estDZ2a* gene, in which the predicted N-terminal signaling sequence of EstDZ2 α (amino acids 2–21) was replaced with a hexa-histidine tag, and the resulting gene, referred to as *estDZ2*, was inserted again into pET-28a(+) to form plasmid pET-EstDZ2. Overexpression of *estDZ2* in *E. coli* BL21(DE3) resulted in the accumulation of primarily soluble enzyme with an apparent molecular mass of about 29 kDa under denaturing conditions, and the recombinant enzyme was purified via immobilized metal affinity chromatography to near homogeneity (Supplementary Fig. S2). Native PAGE, size exclusion chromatography and dynamic light scattering (DLS) analyses of EstDZ2 revealed that it functions as a monomer (Supplementary Fig. S3–S5).

Detailed biochemical characterization of EstDZ2 was carried out as described in the Materials and Methods section, using pNP-butyrate as a substrate. First, we determined the optimal pH for EstDZ2 esterolytic activity, which was assayed within the pH range of 4–10 at 40 °C. Optimal activity for EstDZ2 was recorded at pH 7–8 (Fig. 2A). Below pH 7 and above pH 8, the esterolytic activity of EstDZ2 was rapidly diminished (Fig. 2A). Measurements of its relative catalytic activity at different temperatures, on the other hand, revealed that EstDZ2 has a broad temperature range of action, as its levels of esterolytic activity were practically unaltered at temperatures between 25 and 60 °C (Fig. 2B).

In order to investigate the esterolytic activity of EstDZ2 in more detail, we determined the catalytic parameters of EstDZ2-mediated hydrolysis for a range of pNP-esters derived from fatty acids with variable carbon chain lengths. Michaelis-Menten kinetics were observed for the hydrolysis of all substrates, which could be rapidly hydrolyzed by EstDZ2. In accordance with the initial substrate profiling study described above, EstDZ2 exhibited high catalytic rates (k_{cat}) against middle- to long-chain fatty acid esters (C4–C10; Table 1). For C12 or larger substrates, pNP-ester hydrolysis was barely detectable. Catalytic efficiency was optimal for pNP-butyrate (C4) with a k_{cat}/K_m value of $14,375 \text{ s}^{-1}\text{mM}^{-1}$ and decreased with increasing chain length for C8 and C10 (Table 1).

Thermostability and effect of metals, detergents and organic solvents on the activity of EstDZ2. When exposed to elevated temperatures for prolonged time periods, EstDZ2 presented moderate stability, as determined by measurements of residual activity. Up to 55 °C, no significant changes in enzymic activity could be detected after exposure up to 24 h (Fig. 3A). When exposed to 60 °C, the enzyme retained more than 50% of its initial activity for at least six hours, while at 65 °C the enzyme was practically inactivated within the first hour of the high-temperature incubation (Fig. 3A).

EstDZ2 was found to be well tolerant against various organic solvents. It exhibited high tolerance to the presence of methanol, ethanol, and acetone, since it retained high levels of catalytic activity when these solvents were added at concentrations as high as 30% (v/v) or more (Fig. 3B). In the presence of acetonitrile and isopropanol, EstDZ2 exhibited intermediate levels of tolerance, while isooctane, n-hexane, and 1-butanol, started to affect its activity at concentrations as low as 10% (v/v). Importantly, EstDZ2 was found to exhibit exquisite stability in

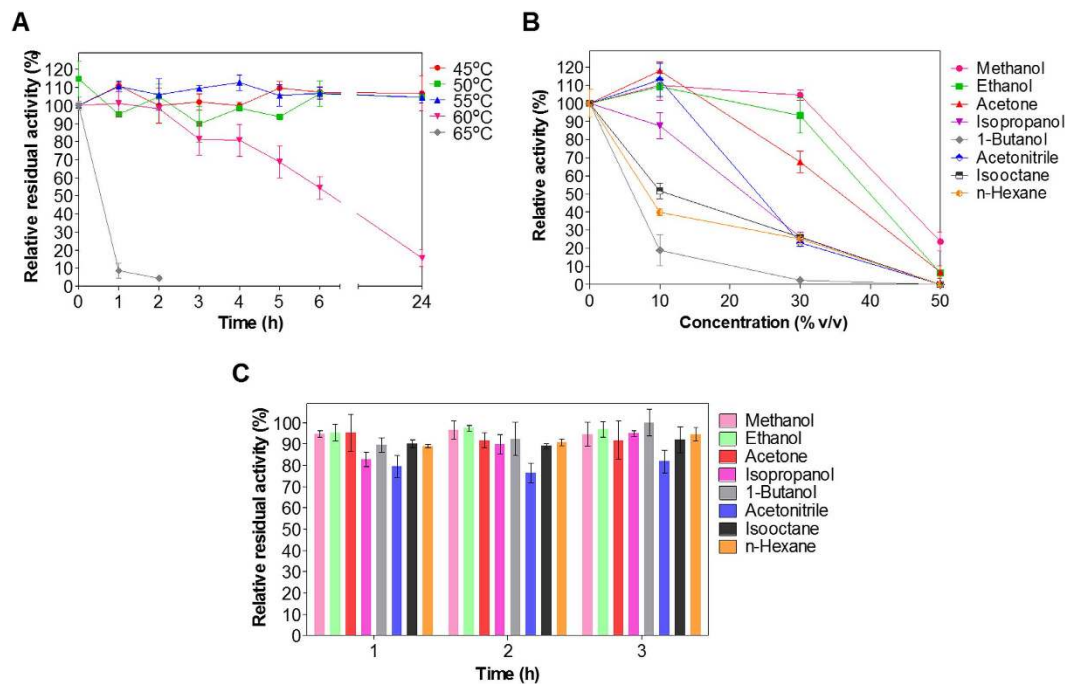


Figure 3. EstDZ2 thermostability and organic solvent tolerance and stability. (A) EstDZ2 thermostability was evaluated by measurements of residual esterolytic activity after exposure to 45, 50, 55, 60 and 65 °C for up to 24 h. (B) EstDZ2 tolerance against various organic solvents was evaluated by measuring its catalytic activity in the presence of 10, 30 and 50% (v/v) of the corresponding solvent. (C) EstDZ2 stability against various organic solvents was evaluated by measuring its residual catalytic activity in the presence of 50% (v/v) of the corresponding solvent for a period of 1–3 h. In all panels, the reported values correspond to the mean value from three independent experiments performed in triplicate and the error bars correspond to one standard deviation from the mean value.

the presence of high concentration of organic solvents, as it was found capable of retaining almost full residual catalytic activity after incubation in media containing up to 50% of various solvents for several hours (Fig. 3C).

Subsequently, we went on to test the effects of a range of metal ions, reducing agents and surfactants on the esterolytic efficiency of EstDZ2. We found that the addition of a variety of mono- and divalent metals (Na^+ , K^+ , Ca^{2+} , Fe^{2+} , Li^{2+} , Cu^{2+} , Mn^{2+} , and Mg^{2+}) at 1 mM did not significantly affect EstDZ2 activity (Table 2). Similar effects were observed when the chelating agent ethylenediaminetetraacetic acid (EDTA) was added to the reaction mixture at 1 mM, indicating that the activity of this esterase probably does not depend on a metal cofactor. The presence of nonionic surfactants, such as Tween 20 and Tween 80, did not have a significant impact on EstDZ2 enzymic activity when added at 1% (v/v). On the contrary, 1% Triton X-100 caused significant EstDZ2 inactivation, while the addition of the anionic detergent sodium dodecyl sulfate (SDS) at the same concentration, resulted in complete loss of EstDZ2 activity (Table 2). The addition of 1 mM phenylmethylsulfonyl fluoride (PMSF) resulted in significant inhibition of the enzyme's esterolytic activity, supporting the *in silico* prediction that EstDZ2 is a serine hydrolase. PMSF inhibits serine hydrolases by covalent linkage to the hydroxyl group of the catalytic Ser residue, a phenomenon that resembles the first transition state in ester hydrolysis, and prevents the formation of the hydrogen-bonding network that is crucial for the catalytic performance of this type of enzymes^{4,35}.

Phylogenetic analysis. In order to perform phylogenetic analysis of EstDZ2 in a manner that is as unbiased as possible, the Uniprot and Swissprot databases were first searched to identify the protein sequences with the highest degree of sequence similarity to EstDZ2. From this analysis, the top >100 hits corresponding to non-redundant protein sequences with evidence at the protein/transcript levels were selected in order to perform phylogenetic comparison with EstDZ2 (see the Methods section for more details). This analysis revealed that EstDZ2 and its close homologs form a deep branch (Fig. 4). Among the sequences that cluster with EstDZ2, those of bacterial origin have no assigned functions, and correspond to either hypothetical proteins, predicted α/β hydrolases, or sequences with putative carboxyl-esterase/lipase activity. Only a few enzymes within this branch have been characterized functionally, all of which are of plant or animal origin: an isoprenylcysteine methylesterase from *A. thaliana* [UniProtKB/Swiss-Prot: Q94AS5.1]³¹, and two kynurenine formamidases from the fruit fly *Drosophila melanogaster* [UniProtKB/Swiss-Prot: Q9VMC9.1]³⁶ and the mouse *Mus musculus* [UniProtKB/Swiss-Prot: Q8K4H1]³⁷. The appearance of animal kynurenine formamidases in this cluster is not surprising, as many enzymes of this type are known to resemble carboxyl-esterases more than typical amidases³⁸. This similarity is evident both in terms of sequence, since they contain the GX SXG signature catalytic motif of esterolytic enzymes, as well as in terms of folding, as they are known to adopt the typical α/β hydrolase fold, which

Metal ion or chemical agent	Concentration	Relative activity (%)
None	—	100.0 ± 7.8
K ⁺	1 mM	102.5 ± 9.2
Mn ²⁺	1 mM	106.7 ± 5.0
Ca ²⁺	1 mM	93.9 ± 5.4
Zn ²⁺	1 mM	87.8 ± 1.3
Li ²⁺	1 mM	112.9 ± 9.2
Mg ²⁺	1 mM	104.6 ± 4.0
Na ⁺	1 mM	99.3 ± 5.9
Fe ³⁺	1 mM	95.9 ± 9.6
Cu ²⁺	1 mM	109.8 ± 8.8
EDTA	1 mM	113.4 ± 7.0
PMSF	1 mM	31.0 ± 5.6
Triton X-100	1% (v/v)	55.0 ± 8.0
Tween 20	1% (v/v)	104.1 ± 9.1
Tween 80	1% (v/v)	95.9 ± 5.3
SDS	1% (w/v)	Undetectable

Table 2. Effect of metal ions, surfactants and other denaturing agents on EstDZ2 activity.

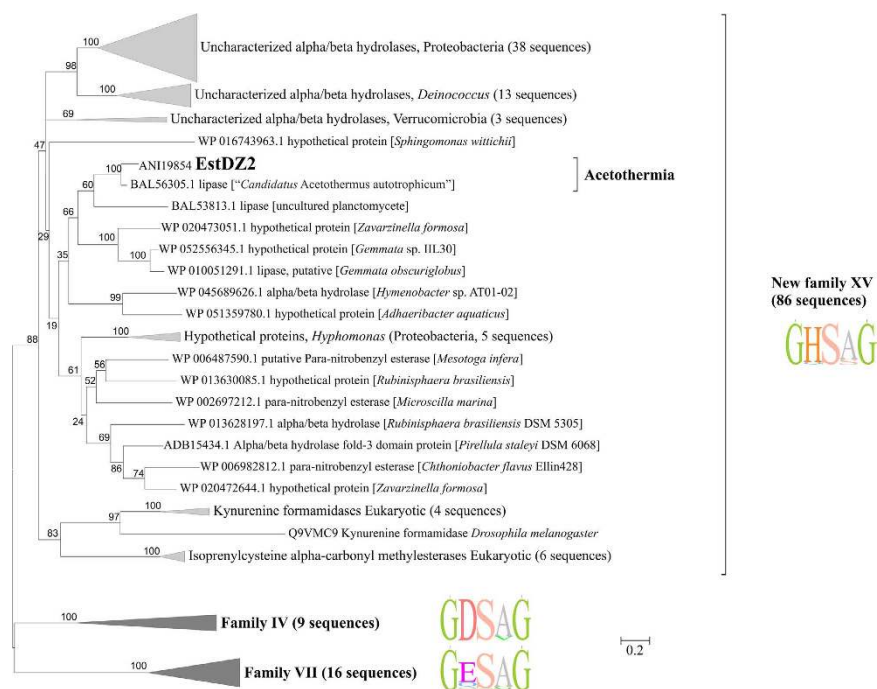


Figure 4. Phylogenetic analysis of EstDZ2 and related sequences. Unrooted Maximum-Likelihood tree of the esterase families XV (this study), IV and VII, which resulted from the phylogenetic analysis described in the Results and Methods sections. The tree with the highest log likelihood is shown. The bootstrap values (100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Sequence logos indicating the preferred amino acids for each family in the highly conserved catalytic motif GXSXG motif are shown to the right of each family cluster. The tree was generated in MEGA6⁸⁴.

is characteristic for carboxyl-esterases and lipases³⁹. Furthermore, certain kynurenine formamidases have been reported to exhibit esterolytic activity in addition to their ability to hydrolyze amides^{38,39}.

Strikingly, the nearest homolog of EstDZ2 was an uncharacterized lipase (identity 82%, query coverage 99%) from “*Candidatus Acetothermus autotrophicum*”⁴⁰, an uncultivated representative of *Acetothermia* (OP1) candidate division⁴¹, which is regarded as one of the deepest bacterial lineages^{40,42,43}. EstDZ2 and this uncharacterized *Acetothermia* lipase are located within a clade of mostly *Planctomycetes* representatives, thus indicating horizontal transfer from *Planctomycetes* to *Acetothermia*. Notably, all five predicted proteins encoded by the genes localized in the same contig as *estDZ2* have their nearest homologs in “*Candidatus Acetothermus autotrophicum*”.

Family of bacterial esterolytic enzyme	Conserved catalytic motif
Family I	GHSQG
Family II	GDSL
Family III	GXSMMG
Family IV	GDSAGG
Family V	GXSMMGG
Family VI	GFSQG
Family VII	GESAG
Family VIII	GGSVG
PhaZ7 (Family IX) ⁸⁹	AHSMG
EstD (Family X) ⁸	GHSLG
LipG (Family XI) ⁹	GHSLGG
LipEH166 (Family XII) ¹⁰	GHSLG
Est30 (Family XIII) ¹¹	GLSLGG
EstA3 (Family XIV) ⁹⁰	CHSMG
EstA ⁹¹	GHSMG
VLip509 ⁹²	GHSLGG
EstD2 ¹⁵	GHSQG
EstZ3, EstGK1 ⁹³	XHSQX
FLS18 ⁹⁴	AHGMG
EM3L4 ⁹⁵	GHSQG
EstWSD ¹⁴	GHSQG
EstLiu ¹⁶	GFSAG
Family XV (this study)	GHSAG

Table 3. List of the amino acid sequences corresponding to the highly conserved oligopeptide GXSXG catalytic motif that appears as the most dominant one in the previously described bacterial esterase families following the Arpingy and Jaeger classification⁵ and numbering (Families I–XIV)¹² and of additional families of bacterial esterolytic enzymes that do not follow the Arpingy and Jaeger numbering. **Note:** The list of additional families of bacterial esterolytic enzymes that do not follow the Arpingy and Jaeger numbering is indicative and may not be exhaustive.

One distinctive feature of EstDZ2 and its homologs is that they share a highly conserved GHSAG catalytic motif. The GHSAG motif is not encountered frequently in any other known family of bacterial esterolytic enzymes^{12,44} and its high level of conservation appears to be characteristic for the herein described group of EstDZ2 and its homologous enzymes. Taken together, our results indicate that EstDZ2 belongs to a new family of bacterial esterolytic enzymes, for which we propose the index XV. Based on sequence analyses, enzymes belonging to this new family are most closely related to Family IV esterases (also referred to as hormone-sensitive lipases, HSLs) and –to a smaller extent– to Family VII esterases, but they are sufficiently distinct in terms of sequence characteristics. The most notable among these differences is the presence of a dominant GHSAG catalytic motif, whereas Family IV and Family VII esterases are characterized by the presence of dominant GDSAG and GESAG motifs, respectively (Fig. 4; Table 3).

Structural model of EstDZ2. Modelling studies to predict the three-dimensional structure of EstDZ2 were performed by selected automated web-servers for protein structure homology modelling included in “Protein Model Portal”⁴⁵, such as SWISS-MODEL⁴⁶, M4T⁴⁷, InFOLD2⁴⁸, Phyre2⁴⁹ and I-TASSER⁵⁰, by applying standard protocols. Comparison of the overall structures generated by the different models, showed that the architecture of the catalytic domain was highly conserved, while differences were mainly located in loop regions (data not shown). Therefore, we focused on the results produced by the Iterative Threading ASSEMBLY Refinement program (I-TASSER)⁵⁰, since it is the most recently developed suite that utilizes iterative threading assembly simulations, coupled with secondary structure enhanced Profile-Profile threading alignment and *ab initio* Monte Carlo simulations for unaligned regions. The top ten threading templates selected by I-TASSER included microbial esterases and the *D. melanogaster* kynurenine formamidase, with sequence identity ranging from 34 to 23% (alignment coverage 96–92%). This finding is in perfect agreement with our phylogenetic analysis, which classified EstDZ2 among putative esterases and indicated that the fruit fly kynurenine formamidase is the most closely related protein sequence of known structure to EstDZ2 (Fig. 4). The final predicted model structure of EstDZ2 produced by I-TASSER was selected according to the confidence and structural similarity scores (see Methods for details) and is shown in Fig. 5.

The predicted EstDZ2 structure is characterized by the α/β hydrolase fold, which is typical for esterolytic enzymes, forming a twisted β -sheet, sandwiched between two layers of helices (Fig. 5). Residues Ser120, Asp200 and His232 are predicted to constitute the catalytic triad. These residues are found to be in close proximity in the model structure, ready to participate in catalysis. The position of their backbone atoms was found to be absolutely conserved in all model structures derived with other utilized structure prediction tools (data not shown). Ser120 is part of the highly conserved GXSXG catalytic motif³⁰, as also demonstrated in the sequence alignment

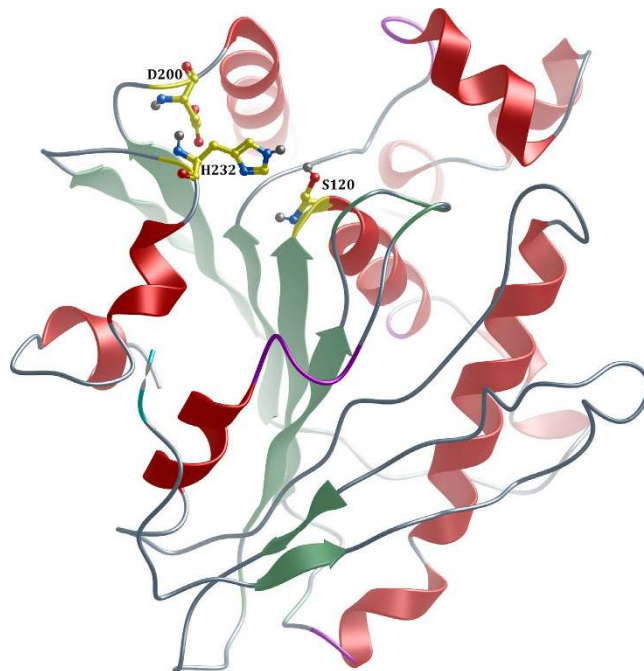


Figure 5. The modelled three-dimensional structure of EstDZ2. Residues Ser120, Asp200 and His232 forming the catalytic triad are indicated in ball-and-stick representation. The figure was prepared using *MolSoft*⁸⁶.

of EstDZ2 and other members of the closely related Family IV esterases with known 3D structure (Fig. 6). The participation of Ser120 to the catalytic mechanism of EstDZ2 was verified by mutational analysis, as the amino acid substitution S120A abolished esterolytic activity (Fig. 1B). The tetrapeptide sequence HGGG (residues 54–57 in EstDZ2), which contributes to the formation of the oxyanion hole³⁰ is also absolutely conserved (Fig. 6). Inactivation of the enzyme's esterolytic activity by the serine-specific inhibitor PMSF⁵¹ (Table 2) supports further the prediction that EstDZ2 is a serine hydrolase.

In order to perform a structural comparison of EstDZ2 with closely related bacterial esterolytic enzymes, we superimposed the modelled EstDZ2 structure with five representative carboxyl esterases of known structure, which have been classified in the literature as Family IV members, the most closely related family to EstDZ2 and other Family XV enzymes. In accordance with conclusions from previous studies, all five Family IV enzymes were characterized by the presence of multiple N- and C-terminal residues belonging to structurally conserved α -helices of various lengths that form an extended “cap” domain (Fig. 7), a structural feature which is highly conserved among members of this family^{30,52,53}. This domain acts as a shield for the catalytic site of many esterolytic enzymes and appears to play a key role in several functional aspects of Family IV esterases, such as activity, substrate specificity, and thermostability⁵². EstDZ2, on the other hand, seems to completely lack these N-terminal helices, while only part of the C-terminal ones are present, resulting in a structure which provides a different type of shielding of the EstDZ2 active site compared to Family IV enzymes (Figs 6 and 7). This structural difference in the environment of the catalytic site is illustrated by the superposition of the EstDZ2 predicted structure on an ensemble of representative Family IV esterases (Fig. 7), and indicates that EstDZ2 may be distinct from related enzymes not only in terms of sequence, but also in terms of structure.

Discussion

In this study, we have identified a new thermostable carboxyl esterase by subjecting the metagenomic DNA isolated from a hot spring located in Kamchatka-Russia, to bioinformatic screening for new lipolytic enzymes. The identified esterase, termed EstDZ2, was cloned, purified from *E. coli* and characterized biochemically. While the physiological role of this new enzyme is unknown, EstDZ2 presented a typical esterolytic profile and was found to be highly active against tributyrin and pNP-esters of fatty acids of medium chain length. For pNP-butyrate (C4), the catalytic efficiency (k_{cat}/K_m) of EstDZ2-mediated hydrolysis was found to be $14,375 \text{ s}^{-1} \text{ mM}^{-1}$. Compared to esterolytic enzymes deposited in the BRENDA database⁵⁴, a collection of enzyme and metabolic information based on primary literature, this level of catalytic efficiency is markedly higher than that of most characterized esterases. Out of the twenty esterases in the database that have been assayed against pNP-butyrate, sixteen exhibit a catalytic efficiency that is one or two orders of magnitude lower than that of EstDZ2. On the contrary, comparison with the rest of the four esterases, which are present in this group and have been found to exhibit higher catalytic efficiency, EstDZ2 exhibits a k_{cat}/K_m value that is only two- to three-fold lower. Among this set, the enzyme with the closest catalytic efficiency value for pNP-butyrate ($k_{\text{cat}}/K_m = 10,000 \text{ s}^{-1} \text{ mM}^{-1}$) is the esterase Est from the hyperthermophilic archaeon *Pyrobaculum calidifontis*⁵⁵.

EstDZ2 was found to be a moderately thermostable enzyme, with a half-life of more than six hours at 60 °C. Importantly, it exhibited good tolerance and exquisite stability when challenged against high concentrations of

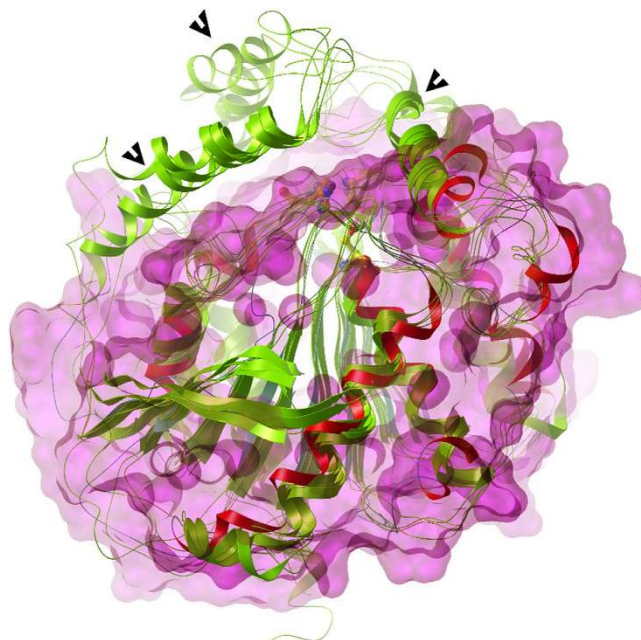


Figure 7. Superposition of the modelled EstDZ2 structure onto the structures of five representative Family IV esterases. The modelled EstDZ2 structure is shown in red, while the structures of the five representative Family IV esterases described in Fig. 6 are shown in green. The molecular surface of EstDZ2 is depicted in magenta and demonstrates that the conserved “cap” domain for Family IV esterases (denoted by black arrows) formed by the N-terminal helices is absent from EstDZ2, while only part of the C-terminal helices contributing to the formation of the “cap” domain is present.

organic solvents. Compared to the recently discovered esterase EstOF4⁵⁶, EstDZ2 demonstrated greater tolerance to the presence of all organic solvents tested. When compared in terms of organic solvent stability with other recently discovered thermostable and biotechnologically promising esterolytic enzymes, such the esterases Pf_Est from *Pyrococcus furiosus*⁵⁷ and EstLiu from the marine bacterium *Zunongwangia profunda*¹⁶, EstDZ2 exhibited superior stability against commonly utilized solvents, such as methanol, ethanol and isopropanol. In particular, while EstDZ2 activity remained practically unaffected in media containing 50% of these solvents after incubation for up to 3 h, exposure of Pf_Est to the same conditions resulted in loss of half of its maximal activity only after 30 min of incubation⁵⁷. Similarly, when EstLiu was incubated under similar conditions, 40–80% of its maximal activity was lost within the first 40 min of solvent exposure¹⁶.

When compared to other known industrially relevant enzymes, the biochemical profile of EstDZ2 resembles that of the landmark lipolytic enzyme Novozym 435, as described in the BRENDA database⁵⁴. Similarly to EstDZ2, Novozym 435 exhibits optimal activity at 50 °C and functions in the temperature range of 10–70 °C and at pH values between 7 and 9. Furthermore, tributyrin has been reported to be the natural substrate of Novozym 435, a preferred substrate also for EstDZ2. Finally, EstDZ2 retains high levels of enzymic activity in media containing organic solvents, a characteristic which is also shared with Novozym 435, and which has been exploited several times for performing biocatalytic reactions in non-aqueous media^{58–60}.

Phylogenetic analysis indicated that EstDZ2 belongs to a putative new family of bacterial esterolytic enzymes, for which we have proposed the index XV. The amino acid sequence of EstDZ2 did not cluster with that of any previously characterized bacterial esterolytic enzyme. Furthermore, EstDZ2 and EstDZ2-resembling sequences were characterized by a distinctive GHSAG pentapeptide catalytic motif, which is not frequently encountered in any described family of bacterial esterases/lipases. In the most closely related Family IV enzymes, the His residue of the GHSAG motif in EstDZ2 and homologous enzymes, is replaced by an Asp residue, yielding the characteristic GDSAG pentapeptide catalytic motif for this family of esterolytic enzymes¹².

Multiple sequence alignment, coupled with homology modelling for the construction of a three-dimensional structure of the new enzyme, supported further the notion that EstDZ2 belongs to a new family of bacterial esterolytic enzymes. EstDZ2 seems to lack a significant part of the so-called “cap” domain, whose extended structure is characteristic for Family IV esterases^{30,52,61,62}. In accordance with the above prediction, EstDZ2 activity was found to be inhibited by PMSF. Esterases with a full-size “cap” domain are typically found to be unaffected by the addition of this serine hydrolysis-inactivating agent, as this domain acts as a “shield” over the active site and protects the catalytic Ser^{66–68}. Other known esterolytic enzymes with minimal “cap” domains include a *Streptomyces exfoliates* lipase⁶³, and two carboxyl esterases from *Lactobacillus plantarum*⁶⁴ and *Thermogutta terrifontis*⁶⁵. Determination of the EstDZ2 3D structure will allow a direct comparison to Family IV enzymes and other esterase families, thus providing insights on the mechanism of action of EstDZ2, both in its native form and in complex with substrate analogues. EstDZ2 structural studies using X-ray crystallography are currently underway in our laboratories.

According to our phylogenetic analysis, horizontal gene transfer was likely the main route of evolution of the representatives of the XV family of esterases and their homologues among eukaryotes and various bacterial phyla, including the deepest lineages, such as the *Acetothermia* (OP1) candidate division. Analysis of the “*Candidatus Acetothermus autotrophicum*” draft genome (four contigs) characterized its putative physiology as facultative anaerobic, acetogenic, chemolithoautotrophic or chemoorganoheterotrophic. Although the natural substrates of EstDZ2 are currently unknown, our findings provide support to the claim made previously by Takami *et al.*⁴⁰ for a possible heterotrophic lifestyle of *Acetothermia* representatives, and indicate that the spectrum of their possible substrates may include esters and lipids. The discovery of EstDZ2 suggests that OP1 may play an important role in the hydrolysis of lipids, providing more easily utilized substances, such as glycerol, fatty acids, and/or the products of their degradation to other members of the microbial consortium.

The ability of EstDZ2 to efficiently catalyze the hydrolysis of short to medium-length acyl chains, combined with its good thermostability and optimum activity at moderately high temperatures (50–60 °C), suggest that EstDZ2 could be utilized in food-processing applications, such as dairy product modifications and synthesis of flavor compounds. Its tolerance against non-ionic surfactants and metal ions renders EstDZ2 a candidate enzyme for use in detergent mixtures and emulsifiers, while its good tolerance and excellent stability against organic solvents, could signify its potential use towards the production of biodiesel and biopolymers^{5,19,69}.

In conclusion, EstDZ2 constitutes a new candidate enzyme for industrial biotransformations, with novel sequence characteristics that appear to define a new family of esterolytic enzymes.

Methods

Reagents and chemicals. All chemical reagents used in this study were purchased from Sigma-Aldrich (St. Louis, USA). All molecular biology-related products (restriction enzymes, protein markers, etc.) were purchased from New England BioLabs (Ipswich, USA).

Environmental sampling. A detailed description of the environmental sampling site and DNA extraction procedures is given by Menzel *et al.*²⁸. Briefly, 150 ml of wet mud sample was collected and processed on-site to fix the total DNA in RNAlater Stabilization Reagent (Qiagen - Hilden, Germany) before further processing in the lab. The microbial cells contained in the collected sample were disrupted by freeze-thaw cycles, lysozyme, and sodium lauroyl sarcosinate treatments. Total DNA was extracted by a standard phenol extraction method and purified from residual polysaccharides and humic substances. The final preparation contained 1.8 µg of metagenomic DNA as determined by dsDNA BR assay on a Qubit 2.0 Fluorometer (Invitrogen - Waltham, USA) according to the manufacturer's instructions.

Bioinformatic analyses. The isolated metagenomic DNA was sequenced on a Roche/454 Titanium FLX next-generation sequencing platform and assembled into contigs using three different assembly tools as described by Menzel *et al.*²⁸. For the *de novo* prediction of coding sequences in the generated contigs, three different types of software were utilized, each based on a different machine-learning model: MetaGeneAnnotator⁷⁰, MetaGeneMark⁷¹ and Prodigal⁷². The combined results of all three analyses consisted of 32,525 putative gene sequences. These were subsequently submitted to a similarity search using BLASTp and DELTA-BLAST⁷³ against sequences with lipolytic activity deposited in both NCBI-nr and UniProt/Swiss-Prot⁷⁴. The generated hits were imported into a local MySQL⁷⁵ database by running an in-house Python script, and comparative tables were created using appropriate search queries. These tables comprised the highest identity scoring results from both databases for every single sequence, including the corresponding EC numbers from the UniProt/Swiss-Prot database. The sequences were also examined for Pfam domains using HMMER (Hmmscan tool)⁷⁶ against the Pfam-A database. The generated results from the Hmmscan analysis were also imported in a local MySQL database via another in-house Python script and were queried in order to return sequences with domains related to esterolytic activity. From the BLAST hits, the ones with the highest similarity score to sequences with esterolytic activity (EC number: 3.1.1.x) were filtered out from the UniProt/Swiss-Prot database and were compared with the corresponding hits from the NCBI-nr database. The sequence subsequently nominated as EstDZ2 α was one of the hits considered of high interest as a candidate esterolytic enzyme, since HMMscan analysis returned a statistically significant Pfam-A match from the list of lipolytic-related domains, an α/β hydrolase_3 domain (PF07859). Further curation of the sequence included putative EC assignment by exploiting a machine-learning based tool; EFICAZ 2.5⁷⁷ returned a low-confidence (EFICAZ_components: CHIEFc_FDR, MTTSI_bin: 2) EC prediction of 3.1.1.x. Furthermore, EstDZ2 α showed 23% amino acid sequence identity (e-value 6×10^{-24} , query coverage 90%, positive percentage 42%) with a previously characterized isoprenylcysteine alpha-carbonyl methylesterase³¹ in UniProt (Accession number: Q94AS5.1). The corresponding hit in NCBI-nr had 82% identity (e-value $< 10^{-161}$, query coverage 99%, positive percentage 88%) with a sequence annotated as a putative lipase from an uncultured *Acetothermia* bacterium (Accession number: BAL56305.1).

Plasmid construction. pET-EstDZ2 α was constructed by amplifying *estDZ2a* from the isolated metagenomic DNA by PCR using primers containing XbaI (5'-AAAAATCTAGAAGGAGGAAACGATGCGACAATACTCTATGCAACTTAGC-3') and HindIII (5'-AAAAAAGCTTTTAAATCTCTTCGGTCGTGGGATCATTAGACGCTCCAATGGTGGCG-3') restriction sites (underlined) and was cloned into the expression vector pET-28a(+) (Novagen - Darmstadt, Germany). pET-EstDZ2 was constructed by replacing the sequence encoding for amino acids 2–21 of EstDZ2 α with a hexa-histidine tag. For this, *estDZ2* was amplified from pET-EstDZ2 α using the forward primer 5'-AAAAATCTAGAAGGAGGAAACGATGCACCACCACCACCACCACCACGCGCCGCCGTTCTCGTTCC-3' (XbaI restriction site underlined and the hexa-histidine tag doubly underlined) and the reverse primer utilized for the cloning of *estDZ2a*. For

the generation of the EstDZ2 variant S120A, overlap extension PCR was carried out using the forward primer (5'-GAAGATCTTCGTGACGGGGCACgCAGCTGGGGGACATCTGACGGCT-3') and the reverse primer (5'-AGCCGTCAGATGTCCCCCAGCTGcGTGCCCGTCACGAAGATCTTC-3'), in combination with the aforementioned external primers. The mutant gene was cloned into pET-28a(+) using the same cloning strategy as described above, thus leading to the generation of the recombinant plasmid pET-EstDZ2 α (S120A). The correct sequence of all constructs was verified by standard DNA sequencing.

Protein expression and purification. BL21 (DE3) cells carrying the pET-EstDZ2 plasmid were grown in LB broth containing 50 μ g/ml kanamycin at 37 °C under constant shaking until the culture reached an OD₆₀₀ of about 0.5. At that point, the expression of *estDZ2* was induced by the addition of 0.2 mM IPTG followed by overnight incubation at 25 °C with shaking. For EstDZ2 purification, the cells from a 500 mL culture grown in a 2 L shake flask were harvested, washed, re-suspended in 10 mL equilibration buffer NPI20 enhanced with 1% Triton X-100 (v/v), and lysed by brief sonication steps on ice. The cell extract was clarified by centrifugation at 10,000 \times g for 15 min at 4 °C and the supernatant was incubated for 1 h at 60 °C in order to denature *E. coli* proteins. After this heat-treatment step, the precipitated material was removed again by centrifugation at 10,000 \times g for 15 min at 4 °C. The supernatant was collected and mixed with 0.5 mL Ni-NTA agarose beads (Qiagen - Hilden, Germany) and shaken mildly for 1 h at 4 °C. The mixture was then loaded onto a 5 mL polypropylene column (Thermo Fisher Scientific - Waltham, USA), the flow-through was discarded, and the column was washed with 10 mL of NPI20 wash buffer containing 1% (v/v) Triton X-100. Next, Triton X-100 was washed away by passing 20 mL of standard NPI20 wash buffer. EstDZ2 was eluted using 1 mL of NPI200 elution buffer. All buffers used for purification were prepared according to the manufacturer's protocol (Qiagen - Hilden, Germany) unless stated otherwise. Imidazole was subsequently removed from this protein preparation using a Sephadex G-25 M PD10 column (GE Healthcare - Chicago, USA). Protein concentration was estimated according to the assay described by Bradford⁷⁸ using bovine serum albumin as a standard. The purified protein was visualized by SDS-PAGE analysis and staining with Coomassie blue or Western blotting using an anti-polyhistidine monoclonal antibody conjugated with horseradish peroxidase (Sigma - St. Louis, USA), and stored at -80 °C after the addition of 50% (v/v) glycerol.

Protein electrophoresis and zymography. *E. coli* BL21(DE3) cells were transformed with either pET-28a(+) or pET-EstDZ2 α , grown in LB medium at 37 °C with shaking until the culture reached an OD 600 nm of 0.5, at which point 0.2 mM IPTG was added to induce *estDZ2 α* overexpression. After additional incubation at 25 °C for 12 h, the cells were lysed by brief sonication and the proteins contained in 15 μ l of total cell lysate were separated by native PAGE. To avoid denaturation, protein separation was performed at 120 V/35 mA for two gels, and the tank was immersed in ice-cold water. Esterase activity was assayed by zymography according to Dherbecourt *et al.*⁷⁹ with minor modifications: the gel was rinsed in distilled water and incubated for 30 min at 40 °C in 0.1% Fast Red TR-salt (4-chloro-2-methylbenzenediazonium salt) in 0.1 M Tris-HCl buffer (pH 7.0) containing 2% of a 1% (v/v) 1-naphthyl acetate solution in acetone. Esterolytic activity was visualized by the appearance of a band of red-brown color.

Enzyme activity assays. For the initial specificity analysis using natural substrates, cell lysates were assayed against ethyl-acetate, butyl-acetate, cinnamyl-acetate, tributyrin and triolein. The substrates were used at a 5 mM concentration in 100 μ L reactions, containing a 25 mM Tris-HCl pH 7.3 buffer with 0.01% bromothymol blue (BTB) as a pH indicator. The reaction was carried out for 15 min under constant agitation. Natural oils, such as olive oil, sunflower oil and palm oil, were tested in the same reaction at 1% (v/v). For biochemical characterization, the catalytic activity of EstDZ2 was determined by quantification of the amount of *p*-nitrophenol released from the substrate by photometric measurement at 410 nm. Unless stated otherwise, the 100 μ L standard reaction mixture consisted of 25 mM Tris-HCl pH 8 buffer with 0.05% Triton X-100 (v/v), 2 mM pNP-butyrate and 1 μ g of pure enzyme, and was carried out for 10 min at 50 °C on a MJ Research thermal cycler, with a pre-incubation setting of the buffer to the target temperature, before the enzyme was added. Enzymic activity was recorded using a Safire II-Basic plate reader (Tecan, Austria) by measuring the absorbance of the released pNP at 410 nm, immediately after the reaction was completed. For the substrate specificity experiments, a range of different pNP-fatty acyl esters, such as acetate (C2), butyrate (C4), octanoate (C8), decanoate (C10), laurate (C12), and palmitate (C16) were used at concentrations ranging from 0.1 to 1 mM. For the determination of the enzyme's optimal pH, reactions were carried out at 40 °C in 25 mM acetate, phosphate, Tris-HCl and glycine-NaOH buffers for pH values 4–6, 6–7, 7–9 and 9–10, respectively. Activity was measured by recording absorbance at 348 nm, the isosbestic point of pNP, so as to exclude the pH effect on the readings, as described previously⁸⁰. Temperature profiling of EstDZ2 was performed by incubating the standard reaction at temperatures ranging from 25 to 70 °C. The assays for the determination of EstDZ2 tolerance to the presence of metal ions, detergents and organic solvents were also executed in the standard reaction with the only difference being the addition of these agents at the specified concentrations. The standard reaction without addition of the tested denaturant was recorded as maximal (100%) activity, while reactions that included the tested denaturing agents but no enzyme, were used as blanks. Organic solvent stability was evaluated after incubating the enzyme in 50% of the tested solvents for time periods of 1, 2 and 3 h. Subsequently, the incubation medium was diluted to remove the solvent, and the residual activity of EstDZ2 was measured in the standard reaction. All measurements were obtained from at least three independent experiments carried out in triplicates.

Phylogenetic analysis. The analysis dataset included the amino acid sequence of EstDZ2 α , along with its 100 top BLAST hits obtained via the NCBI BLASTp utility ([http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins; threshold: 0.01; word length: 2; non-redundant protein sequences \(nr\) database](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins; threshold: 0.01; word length: 2; non-redundant protein sequences (nr) database)), and the 31 top BLAST hits, obtained via the Uniprot BLASTp utility (<http://www.uniprot.org/blast/>; threshold: 0.01; UniProtKB-SwissProt

database). For the latter, only complete proteins with evidence at the protein/transcript levels were retrieved. The results were filtered through an 0.95% identity filter using the CD-hit utility (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=h-cd-hit)⁸¹ to eliminate duplicates and reduce redundancy. Preliminary investigations revealed that EstDZ2 α is mostly similar to Family IV and VII esterases. Based on this, three additional Family IV (Uniprot accession numbers: O52270, Q8NKS0, A0A068LG40) and three additional Family VII (P37967, Q8GCC7, Q9Z545) proteins were included in the analysis. The final dataset consisted of 111 sequences. The dataset was aligned in Geneious 8.1 using MAFFT v.7 (algorithm E-INS-i, default parameters)⁸². The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model⁸³. All positions with less than 95% site coverage were eliminated. There were a total of 196 positions in the final dataset. The tree was generated in MEGA6⁸⁴ and the tree with the highest log likelihood (-30586.7910) was selected. Sequence logos were constructed with Geneious 8.1.

Modelling studies of EstDZ2. Modelling studies on the three-dimensional structure of EstDZ2 were performed by employing a variety of automated web-servers for protein structure homology modelling and applying standard protocols as implemented in “Protein Model Portal”⁴⁵. These include SWISS-MODEL⁴⁶, M4T⁴⁷, InFOLD2⁴⁸, Phyre2⁴⁹ and I-TASSER⁵⁰. Model structures of EstDZ2, were created based on templates derived from previously determined structures that exhibited esteroytic and/or other activities, i.e. an engineered feruloyl esterase derived from a soil metagenomic library (PDB code 4ZRS), a putative thioesterase from *Silicibacter* sp. tm1040 (PDB code 2PBL), esterase B from *Lactobacillus rhamnosis* (PDB code 4N5I), the Ttest2 esterase from *Thermogutta terrifontis* (PDB code 5AO9), a kynurenine formamidase from *D. melanogaster* (PDB code 4E11), and a carboxyl esterase from *Ferroplasma* (PDB code 3WJ1). From the top five final models predicted by I-TASSER, the most recently developed suite, only one was created with reliable prediction scores (C-score 0.61 and TM-score 0.8 ± 0.09). The C-score represents the confidence of the generated model, with values ranging between zero and one, while the TM-score is a measure of the global structural similarity between the predicted EstDZ2 structure and that of template proteins. Superposition of the modelled structure with Family IV representatives was performed by COOT⁸⁵ using secondary structure elements. Visual representation of the modelled structure was performed with the molecular graphics software MolSoft⁸⁶.

References

- Yeoman, C. J. *et al.* Thermostable enzymes as biocatalysts in the biofuel industry. *Adv. Appl. Microbiol.* **70**, 1–55 (2010).
- Schulze, B. & Wubbolts, M. G. Biocatalysis for industrial production of fine chemicals. *Curr. Opin. Biotechnol.* **10**, 609–615 (1999).
- Hausmann, S. & Jaeger, K.-E. *Handbook of hydrocarbon and lipid microbiology* 1099–1126 (Springer, 2010).
- Bornscheuer, U. T. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol. Rev.* **26**, 73–81 (2002).
- Arpigny, J. & Jaeger, K. Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* **343**, 177–183 (1999).
- Levison, M., van der Oost, J. & Kengen, S. W. Carboxylic ester hydrolases from hyperthermophiles. *Extremophiles* **13**, 567–581 (2009).
- Handrick, R. *et al.* A new type of thermoalkalophilic hydrolase of *Paucimonas lemoignei* with high specificity for amorphous polyesters of short chain-length hydroxyalkanoic acids. *J. Biol. Chem.* **276**, 36215–36224 (2001).
- Levison, M., van der Oost, J. & Kengen, S. W. Characterization and structural modeling of a new type of thermostable esterase from *Thermotoga maritima*. *FEBS J.* **274**, 2832–2842 (2007).
- Lee, M. H., Oh, T.-K., Song, J. K. & Yoon, J. H. 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. Microbiol.* **72**, 7406–7409 (2006).
- Kim, E.-Y. *et al.* Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl. Environ. Microbiol.* **75**, 257–260 (2009).
- Montoro-García, S. *et al.* Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterase family. *J. Bacteriol.* **191**, 3076–3085 (2009).
- Rao, L. *et al.* A thermostable esterase from *Thermoanaerobacter tengcongensis* opening up a new family of bacterial lipolytic enzymes. *Biochim. Biophys. Acta.* **1814**, 1695–1702 (2011).
- Rodríguez, M. C. *et al.* Est10: A novel alkaline esterase isolated from bovine rumen belonging to the new family XV of lipolytic enzymes. *PLoS ONE* **10**, e0126651 (2015).
- Wang, S., Wang, K., Li, L. & Liu, Y. Isolation and characterization of a novel organic solvent-tolerant and halotolerant esterase from a soil metagenomic library. *J. Mol. Catal. B: Enzymatic* **95**, 1–8 (2013).
- Lee, M. H. *et al.* A new esterase EstD2 isolated from plant rhizosphere soil metagenome. *Appl. Microbiol. Biotechnol.* **88**, 1125–1134 (2010).
- Rahman, M. A. *et al.* Characterization of a novel cold active and salt tolerant esterase from *Zunongwangia profunda*. *Enzyme Microb. Tech.* **85**, 1–11 (2016).
- Brocknerhoff, H. *Lipolytic enzymes*. (Elsevier, 2012).
- Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409**, 241–246 (2001).
- Panda, T. & Gowrishankar, B. Production and applications of esterases. *Appl. Environ. Microbiol.* **67**, 160–169 (2005).
- Phythian, S. J. Esterases. *Biotechnology Set, Second Edition*, 193–241 (1998).
- Bornscheuer, U. T. & Kazlauskas, R. J. *Hydrolases in organic synthesis: regio- and stereoselective biotransformations*. (John Wiley & Sons, 2006).
- Gomes, J. & Steiner, W. The biocatalytic potential of extremophiles and extremozymes. *Food Technol. Biotech.* **42**, 223–235 (2004).
- Li, X. & Yu, H.-Y. Characterization of an organic solvent-tolerant lipase from *Haloarcula* sp. G41 and its application for biodiesel production. *Folia microbial.* **59**, 455–463 (2014).
- Rathi, P., Saxena, R. & Gupta, R. A novel alkaline lipase from *Burkholderia cepacia* for detergent formulation. *Process Biochem.* **37**, 187–192 (2001).
- Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* **20**, 616–622 (2009).
- Zarafeta, D. *et al.* Discovery and Characterization of a Thermostable and Highly Halotolerant GH5 Cellulase from an Icelandic Hot Spring Isolate. *PLoS ONE* **11.1** e0146454 (2016).
- Hugenholtz, P. & Tyson, G. W. Microbiology: metagenomics. *Nature* **455** (7212), 481–483 (2008).
- Menzel, P. *et al.* Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* **70.2** 411–424 (2015).
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y. & Lanczycki, C. J. CDD: NCBI’s conserved domain database. *Nucl. Acids Res.* (2014).

30. Wei, Y. *et al.* Crystal structure of brefeldin A esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nat. Struct. Mol. Biol.* **6**, 340–345 (1999).
31. Tabata, S. *et al.* Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* **408**, 823–826 (2000).
32. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
33. Frank, K. & Sippl, M. J. High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* **24**, 2172–2176 (2008).
34. Martínez-Martínez, I., Montoro-García, S., Lozada-Ramírez, J. D., Sánchez-Ferrer, Á. & García-Carmona, F. A colorimetric assay for the determination of acetyl xylan esterase or cephalosporin C acetyl esterase activities using 7-amino cephalosporanic acid, cephalosporin C, or acetylated xylan as substrate. *Anal. Biochem.* **369**, 210–217 (2007).
35. Rashamuse, K. *et al.* Discovery of a novel carboxylesterase through functional screening of a pre-enriched environmental library. *J. Appl. Microbiol.* **106**, 1532–1539 (2009).
36. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
37. Pabarcus, M. K. & Casida, J. E. Cloning, expression, and catalytic triad of recombinant arylformamidase. *Protein Expr. Purif.* **44**, 39–44 (2005).
38. Pabarcus, M. K. & Casida, J. E. Kynurenine formamidase: determination of primary structure and modeling-based prediction of tertiary structure and catalytic triad. *Biochem. Biophys. Acta* **1596**, 201–211 (2002).
39. Han, Q., Robinson, H. & Li, J. Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. *Biochem. J.* **446**, 253–260 (2012).
40. Takami, H. *et al.* A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. *PLoS ONE* **7**, e30559 (2012).
41. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
42. Jumas-Bilak, E., Roudiere, L. & Marchandin, H. Description of ‘Synergistetes’ phyl. nov. and emended description of the phylum ‘Deferribacteres’ and of the family Syntrophomonadaceae, phylum ‘Firmicutes’. *Int. J. Syst. Evol. Microbiol.* **59**, 1028–1035 (2009).
43. Yakimov, M. M. *et al.* Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci. Rep.* **3** (2013).
44. Li, P.-Y. *et al.* Structural Basis for Dimerization and Catalysis of a Novel Esterase from the GTSAG Motif Subfamily of Bacterial Hormone-Sensitive Lipase (HSL) Family. *J. Biol. Chem.* (2014).
45. Haas, J. *et al.* The Protein Model Portal—a comprehensive resource for protein structure and model information. *Database* **2013**, bat031 (2013).
46. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* gku340 (2014).
47. Fernandez-Fuentes, N., Madrid-Aliste, C. J., Rai, B. K., Fajardo, J. E. & Fiser, A. M4T: a comparative protein structure modeling server. *Nucleic Acids Res.* **35**, W363–W368 (2007).
48. Webb, B. & Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **5.6**, 1–5.6. 32 (2014).
49. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
50. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
51. Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M. & Baker, E. N. Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus Ak.1* protease at 1.8 Å resolution. *J. Mol. Biol.* **294**, 1027–1040 (1999).
52. Mandrich, L. *et al.* Role of the N terminus in enzyme activity, stability and specificity in thermophilic esterases belonging to the HSL family. *J. Mol. Biol.* **345**, 501–512 (2005).
53. De Simone, G. *et al.* A snapshot of a transition state analogue of a novel thermophilic esterase belonging to the subfamily of mammalian hormone-sensitive lipase. *J. Mol. Biol.* **303**, 761–771 (2000).
54. Schomburg, I. *et al.* BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* **32**, D431–D433 (2004).
55. Hotta, Y., Ezaki, S., Atomi, H. & Imanaka, T. Extremely stable and versatile carboxylesterase from a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* **68**, 3925–3931 (2002).
56. Rao, L., Xue, Y., Zheng, Y., Lu, J. R. & Ma, Y. A novel alkaliphilic *Bacillus* esterase belongs to the 13 th bacterial lipolytic enzyme family. *PLoS ONE* **8**, e60645 (2013).
57. Mandelli, F. *et al.* Characterization and Low-Resolution Structure of an Extremely Thermostable Esterase of Potential Biotechnological Interest from *Pyrococcus furiosus*. *Mol. Biotechnol.* 1–10 (2016).
58. Danieli, B., Luisetti, M., Sampognaro, G., Carrea, G. & Riva, S. Regioselective acylation of polyhydroxylated natural compounds catalyzed by *Candida antarctica* lipase B (Novozym 435) in organic solvents. *J. Mol. Catal. B: Enzym.* **3**, 193–201 (1997).
59. Morrone, R., Nicolosi, G., Patti, A. & Piattelli, M. Resolution of racemic flurbiprofen by lipase-mediated esterification in organic solvent. *Tetrahedron: Asymmetry* **6**, 1773–1778 (1995).
60. Du, W., Xu, Y. y., Zeng, J. & Liu, D. h. Novozym 435-catalysed transesterification of crude soya bean oils for biodiesel production in a solvent-free medium. *Biotechnol. Appl. Biochem.* **40**, 187–190 (2004).
61. Byun, J.-S. *et al.* Crystal structure of hyperthermophilic esterase EstE1 and the relationship between its dimerization and thermostability properties. *BMC Struct. Biol.* **7**, 1 (2007).
62. Yang, S., Qin, Z., Duan, X., Yan, Q. & Jiang, Z. Structural insights into the substrate specificity of two esterases from the thermophilic *Rhizomucor miehei*. *J. Lipid Res.* **56**, 1616–1624 (2015).
63. Wei, Y. *et al.* Structure of a microbial homologue of mammalian platelet-activating factor acetylhydrolases: *Streptomyces exfoliatus* lipase at 1.9 Å resolution. *Structure* **6**, 511–519 (1998).
64. Benavente, R. *et al.* Structure, biochemical characterization and analysis of the pleomorphism of carboxylesterase Cest-2923 from *Lactobacillus plantarum* WCFS1. *FEBS J.* **280**, 6658–6671 (2013).
65. Sayer, C., Szabo, Z., Isupov, M. N., Ingham, C. & Littlechild, J. A. The Structure of a Novel Thermophilic Esterase from the Planctomycetes Species, *Thermogutta terrifontis* Reveals an Open Active Site Due to a Minimal ‘Cap’ Domain. *Front. Microbiol.* **6** (2015).
66. Das, A. K. *et al.* Structural basis for the insensitivity of a serine enzyme (palmitoyl-protein thioesterase) to phenylmethylsulfonyl fluoride. *J. Biol. Chem.* **275**, 23847–23851 (2000).
67. Peng, Q. *et al.* A novel esterase gene cloned from a metagenomic library from neritic sediments of the South China Sea. *Microb. Cell Fact.* **10**, 1 (2011).
68. De Simone, G. *et al.* The crystal structure of a hyper-thermophilic carboxylesterase from the archaeon *Archaeoglobus fulgidus*. *J. Mol. Biol.* **314**, 507–518 (2001).
69. Lagarde, D. *et al.* High-throughput screening of thermostable esterases for industrial bioconversions. *Org. Process Res. Dev.* **6**, 441–445 (2002).
70. Noguchi, H., Taniguchi, T. & Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **15**, 387–396 (2008).

71. Wenhan, Zhu, Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**(12), e132 (2010)
72. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
73. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
74. Consortium, U. UniProt: a hub for protein information. *Nucleic Acids Res.* gku989 (2014).
75. MySQL, A. MySQL. <https://www.mysql.com/> (2001).
76. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* gkr367 (2011).
77. Kumar, N. & Skolnick, J. EFICAZ2. 5: application of a high-precision enzyme function predictor to 396 proteomes. *Bioinformatics* **28**, 2687–2688 (2012).
78. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
79. Dherbécourt, J. *et al.* Identification of a secreted lipolytic esterase in *Propionibacterium freudenreichii*, a ripening process bacterium involved in Emmental cheese lipolysis. *Appl. Environ. Microbiol.* **76**, 1181–1188 (2010).
80. Sayer, C. *et al.* Structural and biochemical characterisation of *Archaeoglobus fulgidus* esterase reveals a bound CoA molecule in the vicinity of the active site. *Sci. Rep.* **6** (2016).
81. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
82. Katoh, K., Misawa, K., Kuma, K. i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
83. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
84. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* mst197 (2013).
85. Emsley, P., Brunger, A. T. & Lütke, T. Tools to assist determination and validation of carbohydrate 3D structure data. *Glycoinformatics*. 229–240 (2015).
86. MolSoft, I. 2.8 Program Manual. MolSoft LLC. San Diego (2000).
87. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7** (2011).
88. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
89. Braaz, R., Handrick, R. & Jendrossek, D. Identification and characterisation of the catalytic triad of the alkaliphilic thermotolerant PHA depolymerase PhaZ7 of *Paucimonas lemoignei*. *FEMS Microbiol. Lett.* **224**, 107–112 (2003).
90. Elend, C. *et al.* Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl. Environ. Microbiol.* **72**, 3637–3645 (2006).
91. Chu, X., He, H., Guo, C. & Sun, B. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl. Microbiol. Biotech.* **80**, 615–625 (2008).
92. Park, S.-Y. *et al.* A new esterase showing similarity to putative diene lactone hydrolase from a strict marine bacterium, *Vibrio* sp. GMD509. *Appl. Microbiol. Biotech.* **77**, 107–115 (2007).
93. Bayer, S., Kunert, A., Ballschmiter, M. & Greiner-Stoeffele, T. Indication for a new lipolytic enzyme family: isolation and characterization of two esterases from a metagenomic library. *J. Mol. Microbiol. Biotechnol.* **18**, 181–187 (2010).
94. Hu, Y. *et al.* Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS Microbiol. Ecol.* **72**, 228–237 (2010).
95. Jeon, Jeong Ho *et al.* Novel lipolytic enzymes identified from metagenomic library of deep-sea sediment. *Evid. Based Complementary Altern. Med.* 2011 (2011).

Acknowledgements

This work was financed by the European Union 7th Framework Programme FP7/2007–2013 in the framework of the EU FP7 collaborative project HotZyme (<http://hotzyme.com>; Grant agreement no. 265933). The aim of this programme was the application of metagenomic approaches to identify new thermostable hydrolases from diverse hot environments with improved performances and/or novel functionalities suitable for different industrial processes. Additional support was also provided by the Synthetic Biology research infrastructure OMIC-ENGINE, financed by the Hellenic General Secretariat of Research and Technology and the National Strategic Reference Framework (NSRF). DZ is the recipient of a PhD fellowship from the Greek State Scholarships Foundation (Idryma Kratikon Ypotrofon-IKY) in the framework of the Research Grant Excellence IKY-Siemens Program, which is co-financed by the European Social Fund and the Greek Government. We would like to thank Vassiliki Papadimitriou and Aristotelis Xenakis for facilitating the dynamic light scattering analyses.

Author Contributions

D.Z., G.S. and F.K. conceived the research; D.Z., E.L., S.G., E.D.C., A.C., I.K., G.S. and F.K. designed the research; D.Z., D.M., E.L., S.G., E.D.C., A.C. and I.K. performed the research; D.Z., D.M., E.L., S.G., E.D.C., I.K. and G.S. analyzed the data; G.S. and F.K. supervised the research; D.Z. and G.S. wrote the paper with contributions from E.L., S.G., E.D.C., A.C. and I.K. All authors read and approved the final manuscript.

Additional Information

Accession codes: The nucleotide sequence of *estDZ2a* has been deposited in GenBank under accession code KX301277.

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zarafeta, D. *et al.* Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Sci. Rep.* **6**, 38886; doi: 10.1038/srep38886 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016