SUPPORTING INFORMATION

Determinants and prediction of esterase substrate promiscuity patterns

Mónica Martínez-Martínez[†,○], Cristina Coscolín[†,○], Gerard Santiago[‡,○], Jennifer Chow[§], Peter J. Stogios[‖], Rafael Bargiela[†,▽], Christoph Gertler[⊥,Δ], José Navarro-Fernández[†], Alexander Bollinger[#], Stephan Thies[#], Celia Méndez-García[⊤,▲], Ana Popovic[‖], Greg Brown[‖], Tatyana N. Chernikova[⊥], Antonio García-Moyano[¥], Gro E.K. Bjerga[¥], Pablo Pérez-García[§], Tran Hai[⊥], Mercedes V. Del Pozo[†], Runar Stokke[ψ], Ida H. Steen[ψ], Hong Cui[‖], Xiaohui Xu[‖], Boguslaw P. Nocek[ξ], María Alcaide[†], Marco Distaso[⊥], Victoria Mesa[⊤], Ana I. Peláez[⊤], Jesús Sánchez[⊤], Patrick C. F. Buchholz[ϕ], Jürgen Pleiss[ϕ], Antonio Fernández-Guerra[¶,J,ʃ], Frank O. Glöckner[¶,J], Olga V. Golyshina[⊥], Michail M. Yakimov[♯,₸], Alexei Savchenko[‖], Karl-Erich Jaeger[#,ƨ], Alexander F. Yakunin[‖,*], Wolfgang R. Streit[§,*], Peter N. Golyshin[⊥,*], Víctor Guallar[‡,ғ,*], Manuel Ferrer[†,*]. The INMARE Consortium

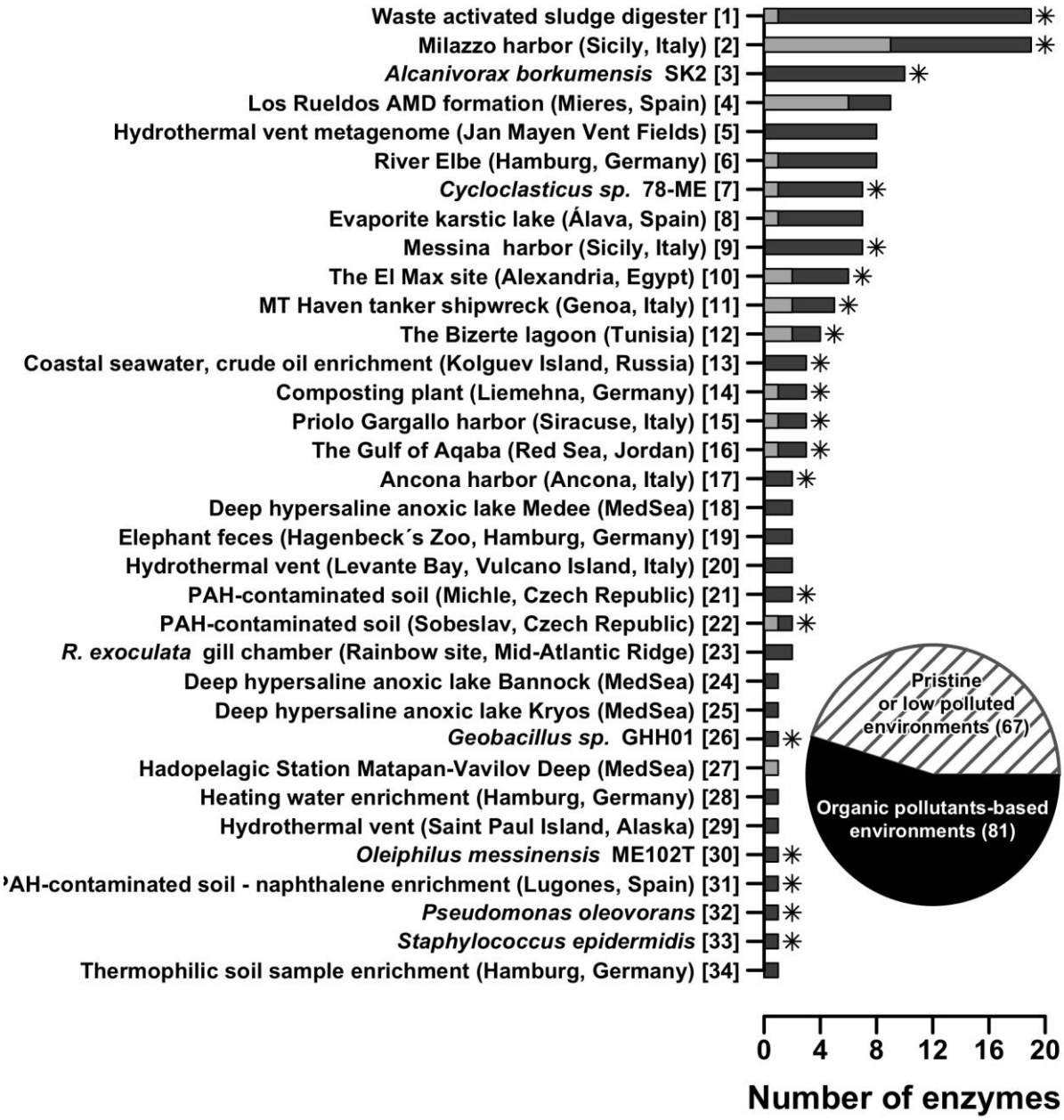

Figure S1. Bio-resources (habitats or microbes) of all 145 EHs in this study. This figure is generated from data in Supporting Information, Tables S1. The bars summarized the number of EHs characterized per each of the bio-resources screened. The number of highly promiscuous hits (i.e., using 30 or more esters) is shown by gray color in the bars. Two major bio-sources of enzymes were covered, namely organically polluted environments or bacteria inhabiting them, and pristine or low-pollution environments. The number of each of each of these bio-sources is summarized in the circular inset. The bio-sources defined as polluted are specifically marked with an asterisk in the bars. Each of the bio-resources is assigned an arbitrary number in square brackets, which is further used as ID in Fig. 2.

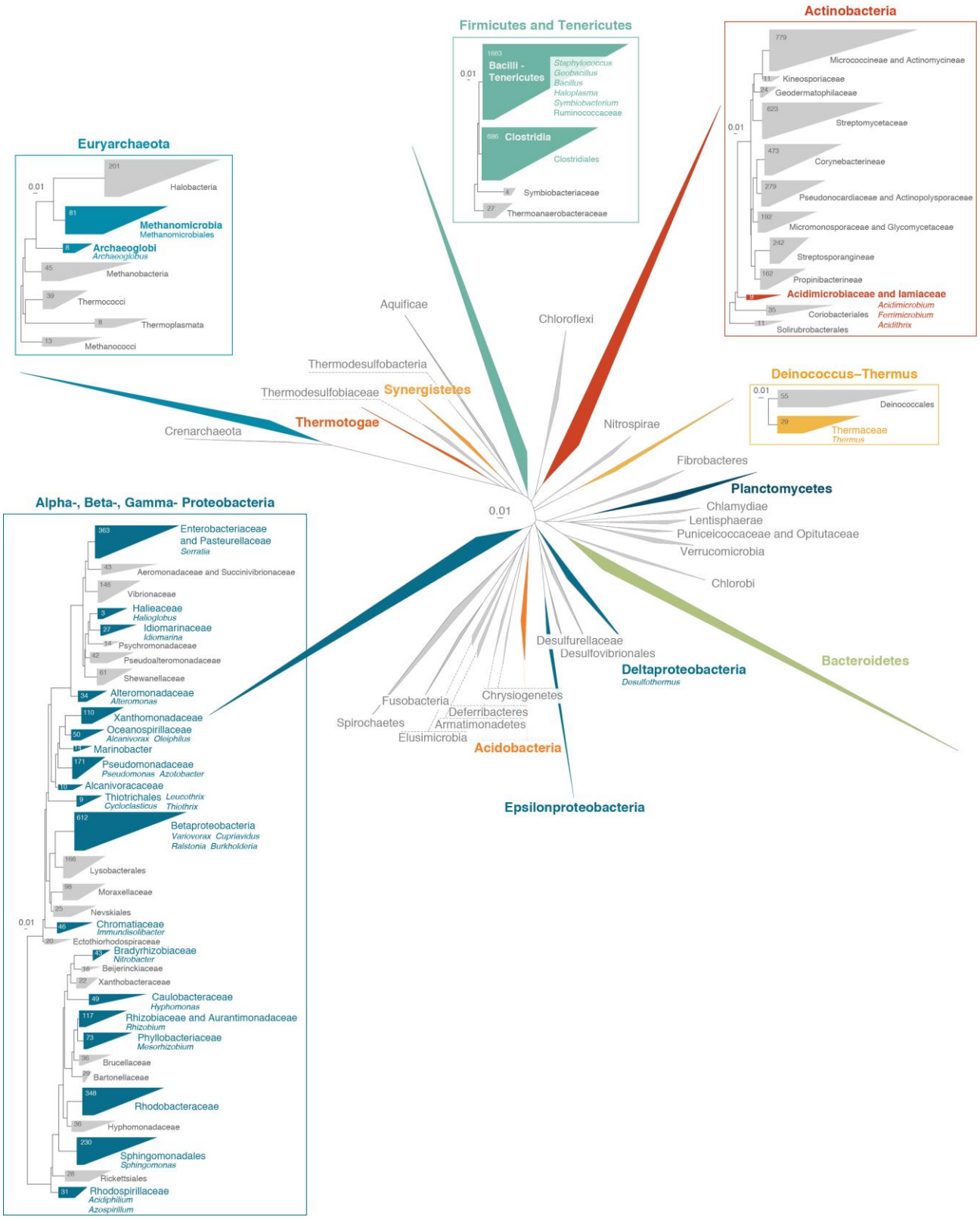

Figure S2. The associations of EHs to cultured and uncultured bacteria from multiple phylogenetic lineages. This figure is generated from data in Supporting Information, Tables S1. Maximum likelihood phylogenetic analysis of the 16S rRNA genes of all species present in the most recent

release of the Living Tree Project database. The groups in which enzymes have been characterized are presented with a color code. Microbial clades not covered in this study are colored in gray. The subtrees represent in detail the groups that contain EHs with broad substrate spectra (i.e., using 30 esters or more) described at the family and genus levels. The numbers in each collapsed clade from the subtrees designate the number of 16S rRNA gene sequences included in each group. Only major taxonomic groups with more than 5 type species present in the database were included in the phylogeny. The tree scales represent the nucleotide substitutions per site. Note that the phylogenetic binning of sequences encoding EHs (for details see Supporting Information, Table S1) was performed using a genome linguistics approach. Briefly, metagenomic fragments were searched for oligonucleotide compositional similarity (frequencies of tetranucleotides) against all sequenced bacterial chromosomes, plasmids and phages using the GOHTAM web tool.[1] For short DNA fragments, compositional analysis could not be performed, and a comprehensive analysis of the TBLASTX results was conducted. Both methods have been proven successful for suggesting the origin of metagenomics sequences.[1,2] However, we are aware that deep assignations (including at the species level) cannot be obtained for short DNA fragments and that this tool may not be appropriate for phylogenetic analysis of sequences with no homology in databases.

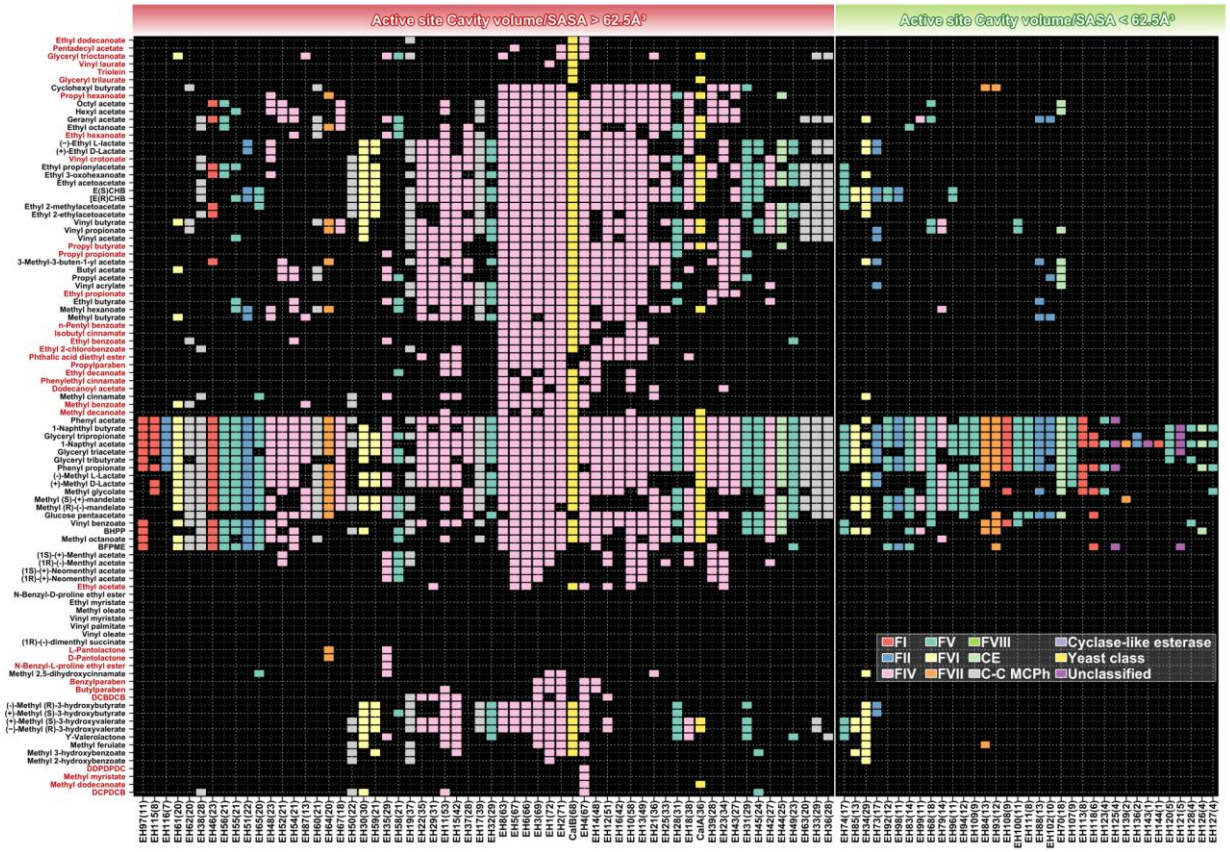

Figure S3. Active site effective volume and phylogeny may be also indicative of specific conversions. This figure is generated from data in Supporting Information, Tables S1 and S3. The figure is constructed as described in Fig. 2, although distance matrixes are not shown. Note that in Fig. 2 only EHs capable of converting 10 or more esters are given whatever their active site effective volume. In this case, the figure summarizes the substrate spectra of EHs from which active site effective volume could be unambiguously identified (96 in total) independently of the number of esters they can hydrolyze. The figure summarizes the substrate spectra of EHs with effective volumes ≥62.5 $Å^3$ (on the left side) or < 62.5 $Å^3$ (on the right side). The 34 esters found to be hydrolyzed under our assay conditions by EHs with effective volumes ≥62.5 $Å^3$ are indicated in the left side with a red color. Note that in all cases the threshold of 62.5 $Å^3$ was defined when the active site cavity volume was computed in $Å^3$ and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html).

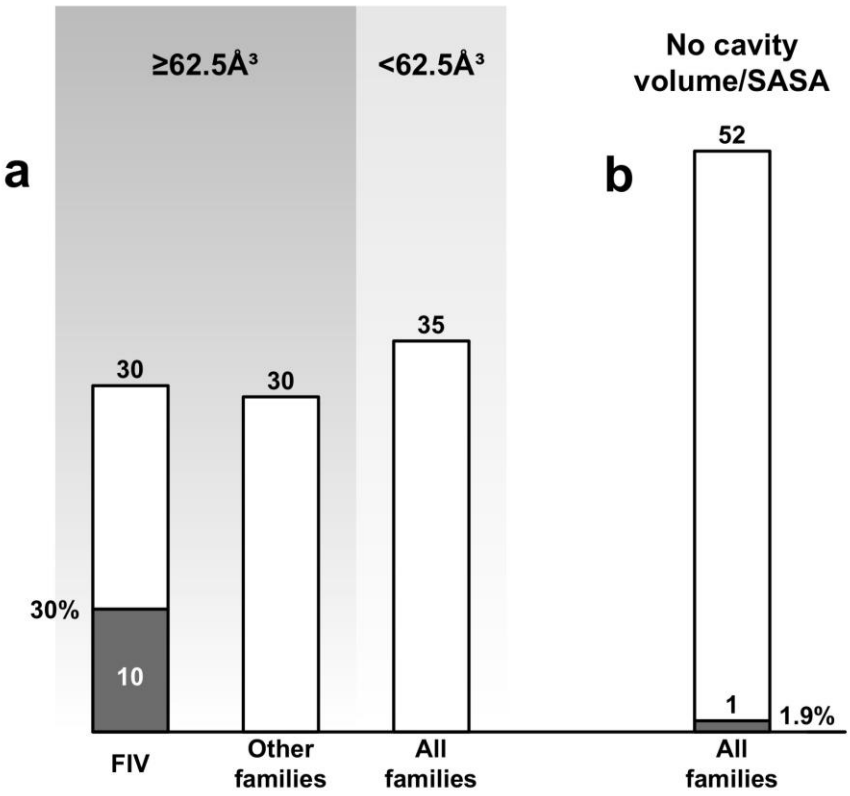

Figure S4. The active site effective volume and phylogeny are predictive markers of the capacity to convert benzyl-, butyl- and propyl-paraben esters. This figure is generated from data in Supporting Information, Tables S3 and S6. (a) Numbers (shown on the top of the bars) of EHs, assigned to FIV or other families, conforming to the ≥62.5 $Å^3$-threshold, or to all families and <62.5 $Å^3$-threshold. (b) Numbers of EHs (shown on the top of the bars) for which effective volume could not be unambiguously measured because sequence identities of less than 25% (to an existing crystal) or because not suitable alignments. In all cases, the number and percentage of EHs capable of converting paraben esters is shown by gray color in the bars. The threshold of 62.5 $Å^3$ was defined when the active site cavity volume was computed in $Å^3$ and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html).

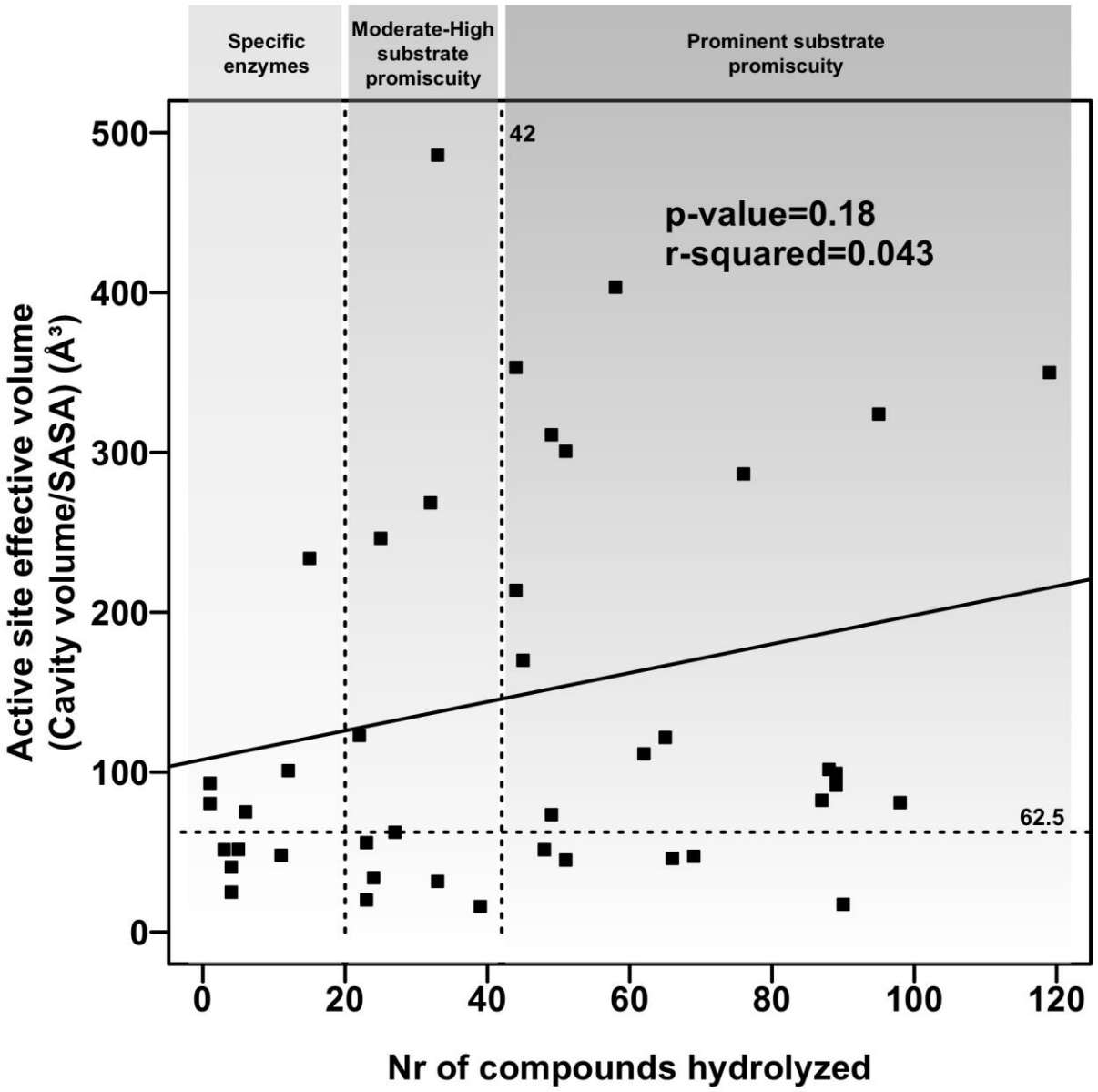

Figure S5. Relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed) of C2 members of HAD phosphatases. The number of substrates converted by each HAD phosphatase was obtained from Huang *et al*.[6] and is summarized in Supporting Information, Table S5. The panels contain information for HAD phosphatases for both crystal structures and homology models were available or could be unambiguously established (sequence identity ≥25%) and the catalytic residues identified. The threshold of substrates being converted, exemplifying the level of promiscuity as for EHs, is indicated. The active site effective volume (in Å^3) was calculated as described in Fig. 5.

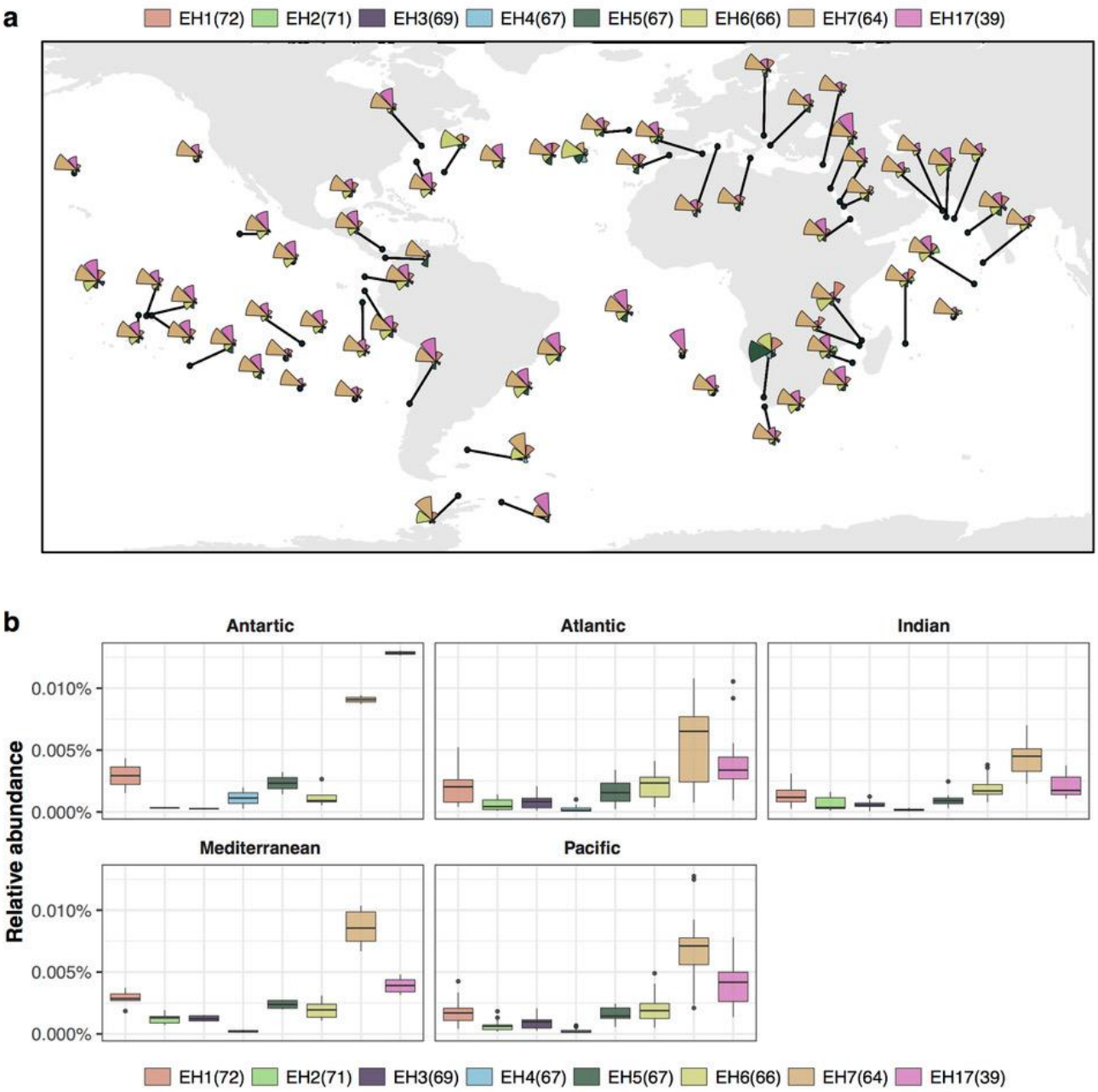

Figure S6. Exploring the sequence space for sequences that encode the most promiscuous EHs. The number in brackets close to the enzyme ID indicates the number of esters hydrolyzed by each enzyme, as described in the legend of Fig. 2. (a) Geographic distribution of the 8 selected promiscuous EHs in the 242 TARA Oceans samples. (b) Relative abundance of the 8 promiscuous EHs in five regions.

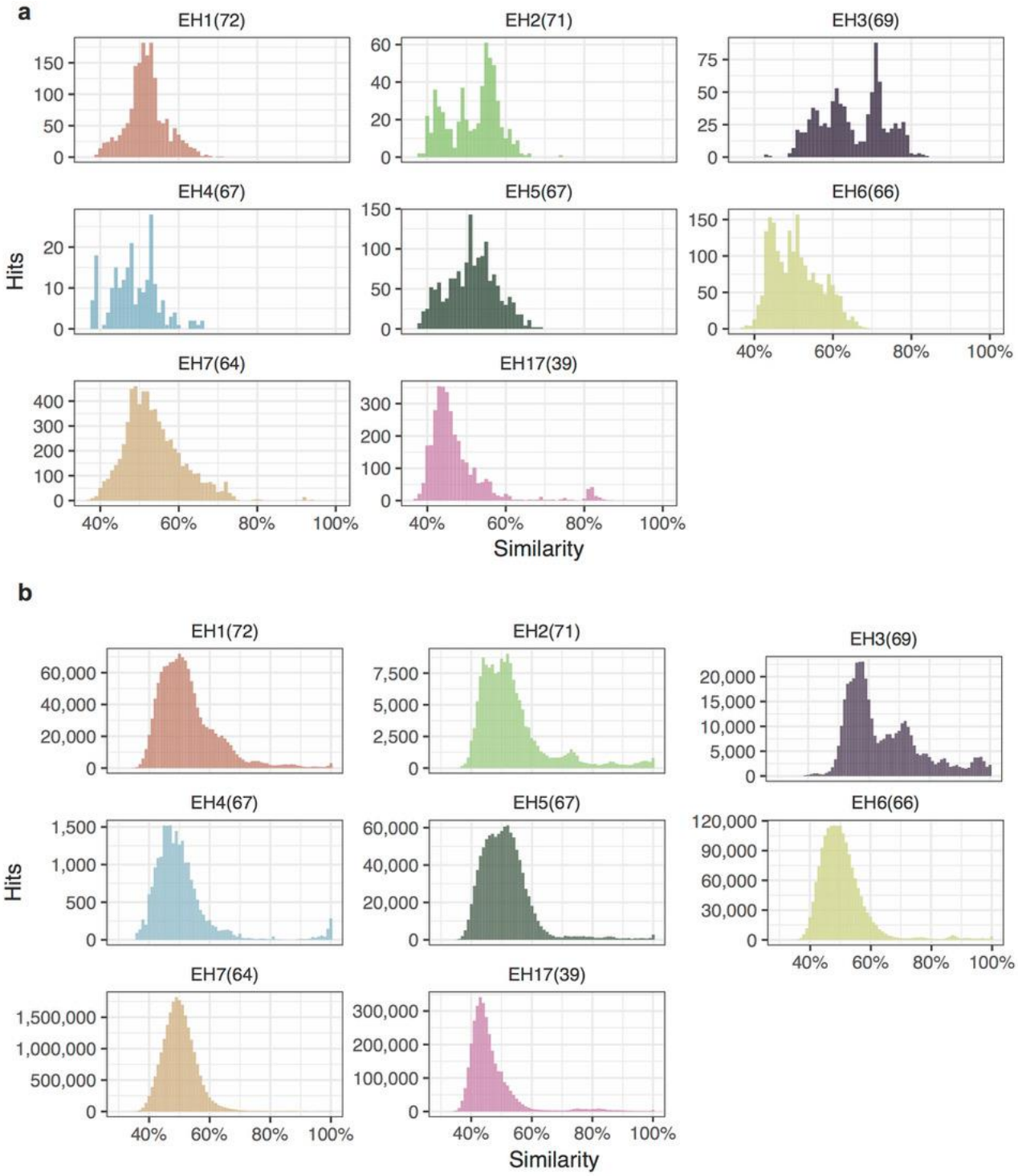

Figure S7. Distribution of the similarity values. The number in the bracket close to the enzyme ID indicates the number of esters hydrolyzed by each enzyme, as described in the legend of Fig. 2. (a) Blastp similarity distribution for every hit against each of the 8 selected promiscuous EHs. (b) All-vs-all blastp similarity distributions for all EH homologs found in the TARA Oceans samples.

Table S1. General information about all 145 EHs and the 2 commercial preparations investigated in this study. The database is an Excel table that provides the following information: ID code used in this study, accession number, PDB code if crystals were available, family to which each sequence was assigned, number of esters hydrolyzed (as found in Supporting Information, Table S3), active site cavity volume/SASA, average maximum specific activity (in U g^{-1} wet cells) (see Supporting Information, Table S3), sequence information, screening technique (sequence or naïve screen), enzyme source (with a full description and short description as in Figure 1), habitat type (as described in Figure 1 inset), cloning vector and expression host, expression conditions (i.e., antibiotic, inductor and temperature), a reference describing the identification, cloning and expression, the top hit in NCBI and the sequence identity, the theoretical molecular weight (ranging from 22 to 103 kDa) and the isoelectric point (ranging from 3.81 to 11.02). *Note 1*: To unambiguously assign sequences to families, a phylogenetic tree was generated using the FastTree v2.1.7 algorithm[3] implemented with the Shimodaira-Hasegawa test. Reference sequences that were unambiguously assigned to each of the esterase/lipase families[4,5] were used to help classify each of the 145 EHs. The final sequence alignments and the tree are available from the authors upon request. *Note 2*: The source organisms of selected polypeptides were identified by a search of oligonucleotide patterns against the GOHTAM database and TBLASTX to reveal compositional similarities between the gene sequences and/or DNA fragments containing the genes that encoded EHs and all analyzed sequenced bacterial chromosomes, plasmids and phages.[1,2] Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S2. Pairwise sequence similarities for all 147 EHs (including CalA/B) as calculated using Needleman-Wunsch alignments performed against all the other candidates ("all-vs.-all"). Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S3. The specific activity is given as U g^{-1} wet cells expressing esterases tested against a set of 96 structurally different esters. The table is an Excel table in which the following information is provided: name of ester, compound chemical class, log P value with standard deviation (calculated using the ACD/ChemSketch 2015.2.5 software), SMILES code, source or brand (with link to the source), molecular mass (g/mol), reaction conditions (including assay method, substrate and pH indicator concentration, buffer, shaking, temperature [30 °C] and pH [8.0], reaction time, additives and concentrations, cell amount), and average specific activity values for all 145 EH preparations in whole-cell assays given in U g^{-1} wet cell pellet. For the Cal A and Cal B preparations, the specific activity values are given in U g^{-1} total protein. The assays were performed as replicates, with the

average value given, and the standard deviation was less than 1% in all cases. Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S4. X-ray diffraction data collection and refinement statistics.

Protein	EH1	EH102
PDB code	5JD4	5DJ3
Data collection		
Space group	P 1	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	90.28, 90.10, 110.76	66.59, 129.45,
α, β, γ (°)	68.02, 79.60, 67.57	129.85
		90, 97.49, 90
Resolution, Å	25.00 – 2.05	29.73 – 2.30
R_{merge}[a]	0.072 (0.606)[b]	0.125 (0.543)
$I / \sigma(I)$	21.66 (2.52)	9.86 (2.50)
Completeness, %	94.9 (91.2)	96.9 (93.4)
Redundancy	3.0 (2.9)	3.6 (3.6)
Refinement		
Resolution, Å	24.95 – 2.05	29.73 – 2.30
No. of unique reflections:	16392, 1875	86908, 4375
working, test		
R-factor/free *R*-factor[c]	16.9/21.1 (27.9/30.0)	18.4/23.6
		(22.4/28.1)
No. of refined atoms,		
molecules	19 262, 8	13 721, 8
Protein	324	23
Solvent	2463	1572
Water		
B-factors		
Protein	37.1	37.1
Solvent	63.1	60.5

Water	48.8	47.0
r.m.s.d.		
Bond lengths, Å	0.003	0.004
Bond angles, °	0.730	0.684

[a]$R_{sym} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle/\Sigma_h\Sigma_i I_i(h)$, in which $I_i(h)$ and $\langle I(h)\rangle$ are the *i*th and mean measurement of the intensity of reflection *h*.

[b]The values for the outer shells of the data are indicated in parentheses.

[c]$R = \Sigma|F_p^{obs} - F_p^{calc}|/\Sigma F_p^{obs}$, in which F_p^{obs} and F_p^{calc} are the observed and calculated structure factor amplitudes, respectively.

* = molecules in the active site cleft.

Table S5. General information about all HAD phosphatases analyzed in this study. The database is an Excel table that provides the following information: ID code, Uniprot ID and number of substrates being converted as reported,[6] and active site cavity volume/SASA. In case of protein for which a crystal structure was not available, the homology to the PDB template (in %) is given. Due to the extensive size, the table is submitted as a separate file in Excel format.

Table S6. The specific activity is given as U g^{-1} pure protein tested against a set of 96 structurally different esters. The data for the Cal A and Cal B commercial preparations are also included. The table is an Excel table in which the following information is provided: name of ester, compound chemical class, log P value with standard deviation (calculated using the ACD/ChemSketch 2015.2.5 software), SMILES code, source or brand (with link to the source), molecular mass (g/mol), reaction conditions (including assay method, substrate and pH indicator concentration, buffer, shaking, temperature [30°C] and pH [8.0], reaction time, additives and concentrations, enzyme amount), and average specific activity values for all tested EHs in pure protein assays given in U g^{-1} protein. The assays were performed as replicates, with the average value given, and the standard deviation was less than 0.5% in all cases. It should be noted that all preparations were tested in pure form as His-tagged proteins, but for the sake of simplicity, the specific activities of only the most promiscuous EHs (i.e., those that reacted with 42 or more esters) are shown. All datasets are available upon request from the authors. Due to the extensive size, the table is submitted as a separate file in Excel format.

RESULTS

Source of EH. A search of oligonucleotide patterns against the GOHTAM database[1,2] and TBLASTX analysis indicated that all 145 EHs were distributed across the entire phylogenetic tree (i.e., in at least 10 phyla and 40 genera), although a bias (70%) toward proteobacterial EH was noted (Supporting Information, Table S1, Fig. S2). No clear affiliation other than *Bacteria* was observed for 9 EHs, and an additional set of 43 EHs could not be assigned at the genus level. Within the 31 highly promiscuous EHs (i.e., using 30 or more esters; excluding CalA/B), an unambiguous affiliation was found for 21 EHs that were associated with at least 2 phyla (the most abundant was Proteobacteria (94%)) and 15 genera. These genera included *Sphingomonas*, *Rhizobium*, *Pseudomonas*, *Alteromonas* and *Acidiphilium*, which have been well-explored with respect to their enzymatic contents, as well as genera that have been largely neglected, such as *Acidithrix*, *Acidimicrobium*/*Ferrimicrobium*, *Alcanivorax*, *Cycloclasticus*, *Immundisolibacter*, *Idiomarina*, *Hyphomonas*, and *Halioglobus* (Supporting Information, Fig. S2). Of the known 24 bacterial and 5 archaeal phyla with cultured representatives,[7] the present study covered EHs from 10 and 1 phyla, respectively (Supporting Information, Fig. S2); much diversity remains to be uncovered in the coming years. The reason that enzyme diversity is so biased is intriguing, even though a very broad diversity was sampled in this study. Note that because many microbial lineages others than those herein analyzed may contain promiscuous EHs a statistical significance of the enrichment of substrate-promiscuous EHs in the above bacterial species cannot be assessed, which is out of the scope of the present study. Whatever the case, we have provided first preliminary evidences that a number of underexplored microbial phylogenetic lineages contained EHs with prominent substrate range. Similarly, we noticed a high percentage of EHs with prominent substrate promiscuity in the chronically polluted seashore area of Milazzo harbor in Sicily (Italy) compared to other sites (Supporting Information, Fig. S1), but again the statistical significance cannot be obtained because the number of enzymes examined was over-represented in this site and because habitats others than those herein examined remains to be explored.

Substrate profiling: general considerations and dynamic range of the assay. Identifying EHs with a broad substrate profile remains a major bottleneck for biocatalysis and biotechnological processes in general.[5] Rapid screening methods will facilitate the identification of such enzymes. Medium-to-high throughput colorimetric protocols have been previously developed that enable EH hydrolytic activity testing with pure proteins.[1,8-11] However, protein purification on a large scale is a very time-consuming and expensive process.[12,13] Consequently, the methods in the present study were adapted to use whole cells, which allowed broad sampling and reduced the time, effort and

cost of identifying enzymes with ample or restricted substrate spectra. Prior to substrate profiling, all clones were tested for activity on agar plates using the three model esters 1-naphthyl acetate, glyceryl tri-acetate and tri-propionate, and their activity was confirmed (see Supporting Information, Methods).

The activity protocol established and used in this study included growing *E. coli* cultures expressing the enzyme of interest overnight in the presence of the appropriate antibiotics and an inducer, followed by activity analysis of the pellet fractions against 96 esters via a pH indicator assay in 384-well plates, at pH 8.0 and 30 °C. Acid is produced after ester bond cleavage by the action of the hydrolytic enzyme contained in the cell pellet,[8] which induces a color change of the pH indicator that can be measured spectrophotometrically at 550 nm. Following the recommendation of Janes *et al*.[8] the concentrations of the pH indicator (0.45 mM) and each of the esters (from 1.3 to 13.2 mM; see concentrations in Methods) were chosen to maximize the accuracy and sensitivity. The number of cells should be as high as possible to maximize sensitivity and to ensure accurate "activity data." In our assay, activity data refers to experimental "time course data" of substrate conversion using 0.4 mg of wet cells per ester, in which time course data might indicate specific activity (in U per g wet cells) from time-course conversion after a maximum of 24 h. This amount of cells (0.4 mg per ester) and time frame (up to 24 h) were found to ensure sensitivity and detection of all esters hydrolyzed by any given EH, as determined by studying the frequency of each ester considered as a hit using different numbers of cells and different time frames and comparing the data from whole-cell assays with pure proteins. Thus, the substrate utilization for all pure proteins supported the validity of the whole-cell screening assay in this study. Under our assay conditions, all esters hydrolyzed by whole cells (Supporting Information, Table S3) were also hydrolyzed when using pure proteins and *vice versa* (Supporting Information, Table S6), which demonstrated that substrate limitation problems using whole cells were not anticipated and that the rapid assay used in this study can be applied to screen substrate ambiguity for any type of cells containing EHs (i.e., fosmid clones, expression clones, microbial cells, etc.). However, we cannot rule out the possibility that unambiguous detection of the conversion of a substrate at a significantly low rate may not be possible using whole cells.

The rapid whole-cell assay used in this study to analyze the substrate profile may be of interest in future genomic and metagenomics screening programs to help identify and prioritize clones with a broad substrate range, whether they contain one or more sequences that encode EHs in a cloned DNA fragment. To screen for promiscuous enzymes, cells from pure cultures, enrichments or clones that express a particular DNA fragment can be screened with standard substrates. Those found to be active against a common standard ester (e.g., 1-napthyl-acetate or glyceryl tri-butyrate)

but do not show a restricted promiscuous profile in a subsequent substrate profiling test will not merit *a priori* sequencing and/or cloning efforts if the objective of the screening program is to find substrate-promiscuous enzymes. By contrast, clones active against a broad range of esters at a high rate merit further sequencing, cloning, expression and characterization. This will substantially reduce the reagent and labor cost while using modest resources. If the activity tests, which can be extended to a broader set of chemical blocks of industrial interest, are also directly performed under conditions resembling industrial requirements, for example in the presence of solvents or at high temperature, efforts can be made toward identifying new, versatile and robust biocatalysts fulfilling industry criteria. Indeed, preliminary tests indicated that the assay method used in this study can be performed in the presence of solvents such as methanol, ethanol, acetonitrile and DMSO at final concentrations of up to 30% (*v*/*v*).

Assays were performed at pH 8.0 and 30 °C in the absence of any chemical other than the ester and a small amount (4.5% v/v) of solvent (DMSO or acetonitrile) needed to dissolve the ester. Therefore, no surfactant was added. Although this may compromise the solubility of very hydrophobic molecules, the majority of large hydrophobic molecules, such as triolein, 2,4-dichlorophenyl 2,4-dichlorobenzoate, 2,4-dichlorobenzyl 2,4-dichlorobenzoate, and diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro pyridine-3,5-dicarboxylate, were hydrolyzed by some EHs under our assay conditions, suggesting that our assay conditions did not introduce any bias in the detection of substrate-enzyme pairs.

Reactions were performed at pH 8.0 and 30 °C. Because we sampled a very broad diversity of habitats from moderately cold to thermophilic, we decided to use an assay temperature of 30 °C, the temperature at which we anticipated most EHs would show appreciable activity. A pH of 8.0 was selected because it was the pH required for the Phenol Red pH indicator used, although pH 7.0 with *p*-nitrophenol as the pH indicator may also be used.[8] We used pH 8.0 because most EHs have been reported to show neutral-to-slightly alkaline pH optima. We are aware that we also included samples from environments ranging from sea-water like habitats to acid mine drainage and that the enzymes from these environments may show different pH optima. Consequently, in many cases (not shown), pH and temperature values for optimal activity were compiled (from previous studies or after the evaluation of optimal parameters for the enzymes investigated herein for the first time) to ensure that the EHs were most active at 30 °C and pH 8.0. Having said that the assay herein used can be extended to pH 7.0 (using *p*-nitrophenol as pH indicator) and at any temperature, preferably below 70ºC to avoid evaporation; a minimum concentration of 0.5 mM substrate is recommended to ensure detection of pH shift.[8] The concentration of buffer (5 mM EPPS) and pH indicator

(Phenol red 0.45 mM) were empirically determined as being optimal to ensure that changes in pH shift during reaction gives linear changes in absorbance.[8]

The predictive capacity of cavity volume/SASA is not influenced by the presence of flexible elements in the structure. The predictive capacity of the active site effective volume may also sensitively depend on the conformational state of flexible elements such as the lid.[5] Indeed, some EH proteins contain a lid domain that covers the active site, preventing substrate binding and requiring structural rearrangement to attain an open conformation. To evaluate whether the homology model-derived active site effective volume value of an EH differs if the template is in the open or closed state, we used the crystal structures of CalA and CalB in closed and open conformations (3GUU/2VEO and 4K6G/5A71). The variations were minimal. Since CalA and CalB do not have the classic lid, which consists of one or two helices that move considerably as rigid bodies, we extended the applicability of this measure to sequences of typical-lid fungal-like lipases such as lipases from *Candida rugosa* (open/close PDB: 1CRL/1THR) and *Thermomyces lanuginosa* (1DTE/1DT5). In all cases, the effective volume ($\geq$76.9 Å^3) was above the 62.5 Å^3-threshold, which is indicative of high substrate promiscuity, with minimal variations between the open and closed conformations. Therefore, the presence of lid domains does not have an observable effect on the substrate promiscuity level, which is mostly defined by the sequence-defined topology of the active site environment, herein exemplified by the active site effective volume.

An example of the potential of the effective volume measure for bioprospecting of EH with prominent substrate promiscuity. This study also produced a subset of EHs with large substrate ranges, competitive with the best industrial prototypes. Their sequences can be used as targets for bioprospecting similar sequences in large-scale datasets. As example, we used MMSEQS2[14] to screen the occurrence of promiscuous EHs against the predicted open reading frames from the TARA Oceans project assemblies,[15] which was used as a study case. All hits with an E-value threshold of 10^{-10} and sequence coverage $\geq$0.6 were selected (see Supporting Information, Methods). For simplicity, the search was restricted to sequences encoding EH1 to EH6 (assigned to FIV), EH7 (assigned to FVIII), and EH17 (assigned to MCP hydrolase family). All of these EHs were among the most substrate-promiscuous EHs and encompassed phylogenetically and environmentally diverse sequences (see Clusters 1 and 2 in Fig. 2). The data presented in Supporting Information, Fig. S6a reveal that not all sequences encoding promiscuous EHs were equally abundant. The homologues to FVIII serine beta-lactamase EH7 and MCP hydrolase EH17 were the most abundant in the majority of the sites, which is most noticeably when examining their accumulative abundances in all 5 oceanic regions (Supporting Information, Fig. S6b). At this stage we would like to notice that obtaining statistically significant differences to environmental

metadata and to global distributions, albeit of interest, is out the scope of the present study. Rather, we studied the degree of novelty of the EH homologs found in the TARA Oceans samples, by performing an all-vs-all blastp search.[16] Supporting Information, Fig. S7a shows the distribution of the similarity values of every hit against our reference sequences, and Supporting Information, Fig. S7b shows the values for all comparisons. The results revealed that the sequences of EH herein reported are distantly related to those in the TARA Oceans Project, which are also quite diverse. The active site effective volume calculation for all homologs (sequence identity >40%) in Tara Ocean samples indicated that all have volumes ranging from approx. 500 to 66.7 $Å^3$, which are above the 62.5 $Å^3$-throshold, and indicative of a broad substrate range, yet to be elucidated.

METHODS

Protein samples. Of the 145 EHs, the screening, cloning and expression of 69 were previously reported. The expression systems and purification conditions used in this study were as reported previously, in order to proceed with their characterization. These 69 EHs were identified by screening meta-genome clone libraries with the short esters α-naphthyl acetate and/or glyceryl tri-butyrate or, in one case, *p*-nitrophenyl-octanoate. Full details of these enzymes are extensively provided in Supporting Information, Table S1. The remaining 77 sequences encoding EHs are reported in this study for the first time and include 15 from Milazzo harbor (Sicily, Italy),[11,17] 9 from the Los Rueldos acid mine (Mieres, Spain),[18] 8 from a hydrothermal vent metagenome (Jan Mayen Vent Fields) (the meta-sequences are available from the National Center for Biotechnology Information (NCBI) database under the ID PRJNA296938, SAMN04111445), 8 from the River Elbe (Hamburg, Germany),[19] 6 from the El Max site (Alexandria, Egypt),[17] 4 from the Bizerte lagoon (Tunisia),[17] 3 from Messina harbor (Sicily, Italy),[17] 3 from the Gulf of Aqaba (Red Sea, Jordan),[17] 2 from Ancona harbor (Ancona, Italy),[20] 2 from elephant feces (Hagenbeck's Zoo, Hamburg, Germany),[21] 1 from Priolo Gargallo harbor (Syracuse, Italy),[17] 6 from *A. borkumensis* SK2,[22] 5 from *Cycloclasticus* sp. 78-ME,[23] 1 from *Oleiphilus messinensis* ME102T,[24] 1 from *Staphylococcus epidermidis*, 1 from *Pseudomonas oleovorans* DSM-1045 (genome not yet published), and 1 from *Geobacillus* sp. GHH01 (the genome sequence has been deposited in GenBank under accession no. CP004008).[25] In all cases, DNA extraction from the corresponding material,[11,17-25] preparation of pCCFOS1 libraries and naïve screening using α-naphthyl acetate and glyceryl tri-butyrate were performed as described elsewhere.[26] Positive clones containing presumptive EHs were selected, and their DNA inserts were sequenced using a MiSeq Sequencing System (Illumina, San Diego, USA) with a 2 × 150-bp sequencing kit. After sequencing, the reads

were quality-filtered and assembled to generate non-redundant meta-sequences, and genes were predicted and annotated as described previously.[27] The meta-sequences used for *in silico* screening are available from the NCBI database.[17,18,20] Protein-coding genes from metagenomes (sequence-based mining) and from the DNA inserts from positive clones (naïve screening) were screened (score > 45; e-value < $10e^{-3}$) using BLASTP and PSI-BLAST searches[28] for enzymes of interest against the ESTHER (*ESTerases and alpha/beta-Hydrolase Enzymes and Relatives*) and LED (*Lipase Engineering*) databases.[29,30]

As summarized in Supporting Information, Table S1, 128 of 145 genes encoding EHs were available or cloned in this study in common expression vectors (Ek/LIC 46, pET21a, pET22b, pVLT31, pCR-XL-TOPO, and p15Tv-L, among others) using a PCR-based approach and appropriate DNA samples as templates. The remaining set of 18 genes (accession codes KY483640-KY483649 and KY203030-KY203037) were synthesized by GenScript (Hong Kong) Limited in expression vectors. Of these 18 genes, the genes KY483640–KY483649 were synthesized in the pCDFDuet expression vector. The genes KY203030 to KY203037 were cloned using a recently developed vector suite that facilitates sub-cloning based on fragment exchange (FX) into multiple expression vectors.[31,32] Proteins were His-tagged at the C-terminus and purified as previously described (see Supporting Information, Table S1) with slight modifications. Briefly, selected *E. coli* clones that expressed each protein were grown at 37 °C on solid Luria Bertani (LB) agar medium supplemented with the appropriate antibiotics (Table S1), and one colony was picked and used to inoculate 10 mL of LB broth plus antibiotic in a 0.25-L flask. The cultures were then incubated at 37 °C and 200 rpm overnight. Afterward, 10 mL of this culture was used to inoculate 0.5 L of LB medium, which was then incubated to an OD_{600nm} of approximately 0.7 (ranging from 0.55 to 0.75) at 37 °C. Protein expression was induced by adding the appropriate inducer to a final concentration of approx. 1 mM (see Supporting Information, Table S1), followed by incubation for 16 h at 16°C. The cells were harvested by centrifugation at 5000 × *g* for 15 min to yield a pellet of 2-3 g/L pellet (wet weight). The wet cell pellet was frozen at -86 °C overnight, thawed and resuspended in 15 mL of 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0. Lysonase Bioprocessing Reagent (Novagen, Darmstadt, Germany) was added (4 μL/g wet cells) and incubated for 60 min on ice with rotating mixing. The cell suspension was sonicated for a total of 5 min and centrifuged at 15000 × *g* for 15 min at 4°C, and the supernatant was retained. The His-tagged proteins were purified at 4 °C after binding to a Ni-NTA His·Bind resin (Sigma Chemical Co. (St Louis, MO, USA)), followed by extensive dialysis of the protein solutions against 20 mM HEPES buffer (pH 7.0) by ultra-filtration through low-adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) and storage at -86°C.

Purity was assessed as >98% using SDS-PAGE analysis[33] in a Mini PROTEAN electrophoresis system (Bio-Rad). The protein concentration was determined according to Bradford with bovine serum albumin as the standard.[34]

Quick tests for the production of active proteins and preparation of protein samples for hydrolysis activity assessment. Prior to substrate profiling, an activity test was performed to verify that each protein was active when expressed in *E. coli*. Protein activity was followed by monitoring the hydrolytic activity directly in the induced cells. Selected *E. coli* clones that expressed each protein were grown at 37°C on solid LB agar media supplemented with the appropriate antibiotics (Supporting Information, Table S1), and one colony was picked and used to inoculate 500 µL of LB broth plus antibiotic in a 2-mL Eppendorf tube. The culture was then incubated at 37 °C and 700 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 7 h. To obtain uniform colonial growth, 5 µL of each culture was spotted on the surface of an LB agar plate (90 cm ø) supplemented with the appropriate antibiotic and expression inducer (Supporting Information, Table S1). The plates were incubated overnight at 37°C, and the agar surface was then covered with three different esters, 1-naphthyl acetate, glyceryl tri-acetate and tri-propionate, which are common EH substrates. Because EHs are commonly active at neutral or slightly alkaline pH, the activity tests were performed at pH 7.0. Briefly, the agar surface was covered with a layer of 20 mL of 1-naphthyl acetate/Fast Blue RR salt in HEPES buffer, pH 7.0, containing 0.4% (w/v) agar.[26] The hydrolysis of 1-naphthyl acetate was monitored by following the formation of an intense brown precipitate because of the release of naphthol, which was further oxidized. In parallel, a second plate was covered with a layer of 20 mL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*'-(3-propanesulfonic acid (EPPS) buffer supplemented with 0.45 mM Phenol Red as a pH indicator, 1 mL of a glyceryl tri-acetate and tri-propionate stock solution (200 mg/mL in acetonitrile each), and 0.4% (w/v) agar. If EH activity occurred, a yellow halo was evident around the colony due to acid formation. Clones containing all 145 EHs investigated in this study tested positive in in at least one of the two screening methods, indicating the production of soluble active proteins, as further confirmed by SDS-PAGE analysis in a Mini PROTEAN electrophoresis system (Bio-Rad).[33]

Selected *E. coli* clones that expressed active and soluble proteins were grown at 37 °C on solid LB agar supplemented with the appropriate antibiotics (Supporting Information, Table S1), and one colony was picked and used to inoculate 500 µL of LB broth plus antibiotic in a 2-mL Eppendorf tube. The culture was then incubated at 37 °C and 700 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 7 h, after which 300 µL of each culture was used to seed LB agar petri dishes (90 cm ø) supplemented with the appropriate antibiotic and expression inducer (i.e., isopropyl β-D-1-thiogalactopyranoside or arabinose; see Table S1). The 300-µL culture was spread

on the plate to obtain uniform growth that covered the entire surface of the plate. The plates were incubated overnight at 37 °C. After incubation, 5 mL of 40 mM HEPES buffer, pH 7.0, was added to the surface of each plate. The bacterial cells were detached using sterile disposable Drigalsky spatulas, and the cell suspensions were transferred to a 5-mL Eppendorf tube and pelleted by centrifugation at 8000 rpm for 10 min at 4 °C. The supernatant was discarded, and the pellets were washed and centrifuged twice in the same buffer. The washed, wet pellets were weighed. For the activity tests, 100 mg of wet intact cells was re-suspended in 0.5 mL of 5 mM EPPS buffer, pH 8.0. This suspension was mixed by vortexing for 1 min and 2 μL of these suspensions were used immediately for the activity test as described in Methods. The activity was further confirmed using purified His-tagged proteins. Prior to use, stock solutions of 1-5 mg mL^{-1} protein in 5 mM EPPS buffer, pH 8.0, were prepared after extensive dialysis against this buffer by ultra-filtration through low-adsorption hydrophilic 10,000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) and 2 μL of protein solutions (from 1.0-17 μg of total protein depending on the specific activity of the particular EH) were immediately used for the activity tests.

Structural determinations. The proteins EH1 and EH102 were expressed and purified according to previously described procedures.[10] The His_6 tags were removed by TEV protease cleavage. The proteins were crystallized using the sitting-drop method in Intelliplate 96-well plates and a Mosquito liquid-handling robot (TTP LabTech), which mixed 0.5 μL of 20-25 mg mL^{-1} protein with 0.5 μL of the following reservoir solutions: EH1 - 0.1 M Tris pH 8.5, 0.2 M ammonium sulfate, 25% (w/v) PEG3350, EH102 - 0.1 M Tris pH 8.5, 0.2 M ammonium sulfate, 25% (w/v) PEG3350, 1/70 units of chymotrypsin. The crystal was cryoprotected with the reservoir solution supplemented with Paratone-N oil prior to flash freezing in an Oxford Cryosystems Cryostream. For the EH1 crystal, diffraction data were collected at 100 K and the Cu K° emission wavelength using a Rigaku HF-0007 rotating anode with a Rigaku R-AXIS IV++ detector. For the EH102 crystal, diffraction data were collected at 100 K and the Se absorption edge wavelength at the Structural Biology Center, Advanced Photon Source, beamline 19-ID using an ADSC Quantum 315R CCD detector. All diffraction data were reduced with HKL3000.[35] The structures were determined by molecular replacement using the structures from the Est2 Protein Databank (PDB) codes 1EVQ and 3RJT for EH1 and EH102, respectively, and the software Phenix.phaser.[36] Refinement was completed with Phenix.refine.[37] All B-factors were refined as isotropic with TLS parameterization. All geometries were verified using Phenix and the wwPDB server.

Protein Energy Landscape Exploration (PELE) sampling. We used Protein Energy Landscape Exploration (PELE) software to sample the binding modes of glyceryl tri-acetate with EH1 and EH102.[38,39] PELE is a Monte Carlo algorithm composed of a sequence of perturbation,

relaxation, and Metropolis acceptance tests. In the first step, the ligand is subjected to random rotations and translations, while the protein is perturbed based on the anisotropic network model (ANM).[40] The maximum allowed translation for ligand perturbation was 1.5 Å, and the maximum rotation was 20°. During the protein perturbation, all atoms were displaced by a maximum of 0.5 Å by moving the α-carbons following a random linear combination of the 6 lowest eigenvectors obtained in the ANM model. The relaxation step included the repositioning of all amino acid side chains within 6 Å of the ligand and the 5 side chains with the highest energy increase along the previous ANM step. The relaxation stage ended with a truncated Newton minimization using the OPLS all-atom force field and an implicit surface-generalized Born continuum solvent. The new proposed minima were then accepted or rejected based on a Metropolis test. The substrate binding plots contained all accepted conformations for three 12-h simulations using 200 processors.

Homology modeling. Homology models were developed using Prime software from Schrödinger. Prime uses BLAST (with BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW.

Cavity Volume and Solvent Accessible Surface Area (SASA) calculation. The relative Solvent Accessible Surface Area (SASA) for a residue was obtained using the GetArea web server.[41] This service allows a user to submit a PDB file and retrieve the relative SASA or the solvation energy in a variety of formats. This server has several options and allows the user to compute individual residue exposure under the "Select desired level of output" option. Thus, the "exposure" of the active site can be computed using the catalytic amino acids, e.g. the conserved catalytic Ser/Asp/His triad in esterases and the two conserved aspartic catalytic residues in HAD phosphatases. Cavity volumes were computed with Fpocket,[42] a very fast open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation. Fpocket includes two other programs (dpocket and tpocket) that allow the extraction of pocket descriptors and the testing of owned scoring functions, respectively.

SUPPORTING INFORMATION REFERENCES

(1) Ménigaud, S., Mallet, L., Picord, G., Churlaud, C., Borrel, A., Deschavanne, P. (2012) GOHTAM: a website for 'Genomic Origin of Horizontal Transfers, Alignment and Metagenomics'. *Bioinformatics 28*, 1270-1271.

(2) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, Mdel C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol.* 79, 3553-3562.

(3) Price, M.N., Dehal, P.S., Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490.

(4) Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J. 343*, 177-183.

(5) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235-249.

(6) Huang, H., Pandya, C., Liu, C., Al-Obaidi, N.F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N., and Farelli, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. USA 112*, E1974-1983.

(7) Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol. 12*, 635-645.

(8) Janes, L. E., Löwendahl, C., and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. *Chem. Eur. J. 4*, 2317-2324.

(9) Alcaide, M., Tornés, J., Stogios, P.J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., López-Cortés, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T.Y ., Plumeier, I., Pieper, D. H., Yakimov, M.M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J. 454*,157-166.

(10) Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C.C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., Yakimov, M. M., Yakunin, A. F., Golyshin, P. N., Golyshina, O. V., Savchenko, A., Ferrer, M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol. 17*, 332-345.

(11) Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A. N., Brown, G., Glinos, J., Flick, R., Skarina, T., Chernikova, T. N., Yim, V., Brüls, T., Paslier, D.L., Yakimov, M. M., Joachimiak, A., Ferrer, M., Golyshina, O. V., Savchenko, A., Golyshin, P. N., and Yakunin, A. F. (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep. 7*, 44103.

(12) Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J. 2*,171-179.

(13) Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol. 9*, 22-34

(14) Steinegger, M., Soeding, J. (2016) Sensitive protein sequence searching for the analysis of massive

data sets. *bioRxiv* 079681.

(15) Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S.; Tara Oceans coordinators, Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., and Bork P. (2015) Structure and function of the global ocean microbiome. *Science 348*, 1261359.

(16) Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L (2009) BLAST+: architecture and applications. *BMC Bioinformatics 10*, 421.

(17) Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., Richter, M., Del Pozo, M. V., Cappello, S., Gertler, C., Genovese, M., Denaro, R., Martínez-Martínez, M., Fodelianakis, S., Amer, R. A., Bigazzi, D., Han, X., Chen, J., Chernikova, T. N., Golyshina, O. V., Mahjoubi, M., Jaouanil, A., Benzha, F., Magagnini, M., Hussein, E., Al-Horani, F., Cherif, A., Blaghen, M., Abdel-Fattah, Y. R., Kalogerakis, N., Barbas, C., Malkawi, H. I., Golyshin, P. N., Yakimov, M. M., Daffonchio, D., and Ferrer, M. (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Sci Rep 5*, 11651.

(18) Méndez-García, C., Mesa, V., Sprenger, R. R., Richter, M., Diez, M. S., Solano, J., Bargiela, R., Golyshina, O. V., Manteca, Á., Ramos, J. L., Gallego, J. R., Llorente, I., Martins dos Santos, V. A., Jensen, O. N., Peláez, A. I., Sánchez, J., and Ferrer, M. (2014) Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J. 8*, 1259-1274.

(19) Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., Schulte, M., Streit, W. R. (2013) Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl. Environ. Microbiol. 79*, 4551-4563.

(20) Bargiela, R., Gertler, C., Magagnini, M., Mapelli, F., Chen, J., Daffonchio, D., Golyshin, P. N., and Ferrer, M. (2015) Degradation network reconstruction in uric acid and ammonium amendments in oil-degrading marine microcosms guided by metagenomic data. *Front. Microbiol*. *6*, 1270.

(21) Güllert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B., Alawi, M., Rattei, T., Daniel, R., Schmitz, R. A., Grundhoff, A., and Streit, W. R. (2016) Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol. Biofuels 9*, 121.

(22) Schneiker, S., Martins dos Santos, V. A., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., Chernikova, T. N., Denaro, R., Ferrer, M., Gertler, C., Goesmann, A., Golyshina, O. V., Kaminski, F., Khachane, A. N., Lang, S., Linke, B., McHardy, A. C., Meyer, F., Nechitaylo, T., Pühler, A., Regenhardt, D., Rupp, O., Sabirova, J. S., Selbitschka, W., Yakimov, M. M., Timmis, K. N., Vorhölter, F. J., Weidner, S.,

Kaiser, O., and Golyshin, P. N. (2006) Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat. Biotechnol. 24*, 997-1004.

(23) Messina, E., Denaro, R., Crisafi, F., Smedile, F., Cappello, S., and Genovese, M. (2016) Genome sequence of obligate marine polycyclic aromatic hydrocarbons-degrading bacterium *Cycloclasticus* sp. 78-ME, isolated from petroleum deposits of the sunken tanker Amoco Milford Haven, Mediterranean Sea. *Mar. Genomics* 25, 11-13.

(24) Golyshin, P. N., Chernikova, T. N., Abraham, W. R., Lünsdorf, H., Timmis, K. N., Yakimov, M. M. (2002) Oleiphilaceae fam. nov., to include *Oleiphilus messinensis* gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. *Int. J. Syst. Evol. Microbiol. 52*, 901-911.

(25) Wiegand, S., Rabausch, U., Chow, J., Daniel, R., Streit, W. R., Liesegang, H. (2013) Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announc.* 1, e0009213.

(26) Reyes-Duarte, D., Ferrer, M., and García-Arellano, H. (2012) Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods Mol Biol 861*, 101-113.

(27) Placido, A., Hai, T., Ferrer, M., Chernikova, T. N., Distaso, M., Armstrong, D., Yakunin, A. F., Toshchakov, S. V., Yakimov, M. M., Kublanov, I. V., Golyshina, O. V., Pesole, G., Ceci, L. R., Golyshin, P. N. (2015) Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Appl. Microbiol. Biotechnol. 99*, 10031-10046.

(28) Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389-3402.

(29) Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res. 31*, 319-321.

(30) Barth, S., Fischer, M., Schmid, R. D., and Pleiss J. (2004) The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. *Bioinformatics 20*, 2845-2847.

(31) Geertsma, E. R., and Dutzler, R. (2011) A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry 50*, 3272-3278.

(32) Bjerga, G. E., Lale, R., and Williamson, A. K. (2016) Engineering low-temperature expression systems for heterologous production of cold-adapted enzymes. *Bioengineered 7*, 33-38.

(33) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680-685.

(34) Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248-254.

(35) Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr. 62*, 859-866.

(36) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D: Biol. Crystallogr. 66*, 213-221.

(37) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr. 60*, 2126-2132.

(38) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474-6487.

(39) Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Chem. Theory Comput. 1*, 1304-1311.

(40) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Hall, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum. Chem. 113*, 2110-2142.

(41) Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319.

(42) Guilloux, V. L., Schmidtke, P., and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics 10*, 168.