New insights into the function and global distribution of polyethylene terephthalate (PET) degrading bacteria and enzymes in marine and terrestrial metagenomes

Dominik Danso1, Christel Schmeisser1, Jennifer Chow1, Wolfgang Zimmermann2, Ren Wei2, Christian Leggewie3, Xiangzhen Li4, Terry Hazen5 and Wolfgang R. Streit1

1Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Hamburg, Germany
2Institute of Biochemistry, Department of Microbiology and Bioprocess Technology, Leipzig University
3evoxx Technologies GmbH, Monheim am Rhein, Germany
4Chengdu Institute of Biology, Chengdu, China
5The university of Tennessee, Knoxville, Tennessee, USA

Keywords: HMM, hydrolases, metagenome, metagenomic screening, PET degradation, Polyethylene terephthalate (PET), BHET, TPA

The authors declare no conflict of interest.

*Correspondence
E-Mail: wolfgang.streit@uni-hamburg.de
Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg
Ohnhorststr.18, D-22609 Hamburg, Germany
Tel. (+49) 40-42816-463, Fax. (+49) 40-42816-459
Polyethylene terephthalate (PET) is one of the most important synthetic polymers used nowadays. Unfortunately, the polymers accumulate in nature and until now, no highly active enzymes are known that can degrade it at high velocity. Enzymes involved in PET degradation are mainly α/β-hydrolases like cutinases and related enzymes (E.C. 3.1.-). Currently, only a small number of such enzymes are well characterized. Within this work, a search algorithm was developed that identified 504 possible PET hydrolase candidate genes from various databases. A further global search that comprised more than 16 GB of sequence information within 108 marine and 25 terrestrial metagenomes obtained from the IMG data base detected 349 putative PET hydrolases. Heterologous expression of four such candidate enzymes verified the function of these enzymes and confirmed the usefulness of the developed search algorithm. Thereby, two novel and thermostable enzymes with high potential for downstream application were in part characterized. Clustering of 504 novel enzyme candidates based on amino acid similarities indicated that PET hydrolases mainly occur in the phylum of Actinobacteria, Proteobacteria and Bacteroidetes. Within the Proteobacteria, the Beta-, Delta- and Gammaproteobacteria were the main hosts. Remarkably enough, in the marine environment, bacteria affiliated with the phylum of the Bacteroidetes appear to be the main host of PET hydrolase genes rather than Actinobacteria or Proteobacteria as observed for the terrestrial metagenomes. Our data further imply that PET hydrolases are truly rare enzymes. The highest occurrence of 1.5 hits/Mb was observed in a sample site containing crude oil.
Danso et al., 2018; PET hydrolases from metagenomes

**IMPORTANCE**

Polyethylene terephthalate (PET) accumulates in our environment without significant microbial conversion. Although few PET hydrolases are already known it is still unknown how frequent they appear and which main bacterial phyla they are affiliated with. In this study, deep sequence mining of protein databases and metagenomes demonstrated that PET hydrolases indeed are occurring at very low frequencies in the environment. Further it was possible to link them to phyla which were previously unknown to harbor such enzymes. This work contributes novel knowledge to the phylogenetic relationship, the recent evolution and the global distribution of PET hydrolases. Finally, we describe biochemical traits of four novel PET hydrolases.
INTRODUCTION

Since the discovery, its first synthesis and patenting in 1941 polyethylene terephthalate (PET) became a widely used material in several industrial branches (1). The worldwide PET resin production amounted to 27.8 million tons in 2015 (2).

Due to its massive use, PET is highly enriched in nature. Microplastic and bigger fragments are found worldwide in oceans and terrestrial environments. The most prominent example is the so called Pacific garbage patch. PET debris is often eaten by fish and other marine creatures (3, 4). Thereby PET degradation products and additives (i.e. solubilizers) are introduced into the food chain and have negative impact on human and animal health (5). Until now, only few bacteria and fungi have been described as capable to partially degrade PET to oligomers or even monomers (6). Within this framework, it is however noteworthy, that all known PET hydrolases have relatively low turnover rates, which makes their use for efficient bioremediation almost impossible (TABLE 1).

Intriguingly, the trait for PET degradation appears to be limited to a few bacterial phyla and most bacterial isolates with the potential of PET degradation are members of the gram-positive phylum of Actinobacteria (7). Thereby, the best characterized examples originate from the genera Thermobifida or Thermomonospora (7-12) (TABLE 1). The enzymes involved in the degradation (e.g. PET hydrolase and tannase) are typical serine hydrolases e.g. cutinases (EC 3.1.1.74), lipases (EC 3.1.1.3) and carboxylesterases (EC 3.1.1.1). These enzymes possess a typical α/β-hydrolase fold and the catalytic triad is composed of a serine, a histidine and an aspartate residue (13,14).

More recently, polyethylene (PE) degrading bacteria were reported in insect guts. In this recent study, Enterobacteria and Bacillus strains had been isolated and were capable to degrade polyesters (15, 16).

Furthermore, a complete degradation of amorphous PET materials was described for the Gram-negative bacterium Ideonella sakaiensis 201-F6, that is able to use PET as a major...
Danso et al., 2018; PET hydrolases from metagenomes

energy and carbon source (17). In addition to the hydrolase, the *I. sakaiensis* genome codes for a second enzyme that appears to be unique and which is designated as a tannase capable of degrading mono(2-hydroxyethyl) terephthalic acid. Thereby, the secreted PET hydrolase produces the intermediate mono(2-hydroxyethyl) terephthalic acid (MHET). MHET is presumably internalized by the cell and hydrolyzed by the MHETase. The resulting monomers are then degraded in a downstream process and used for the bacterial metabolism. *I. sakaiensis* is affiliated with the phylum of Betaproteobacteria and belongs to the Burkholderiales.

In this work, our intention was the mining of metagenomes for the detection of novel genes involved in PET degradation and to establish an overview on their taxonomic distribution within the different bacterial phyla. Therefore, we have developed a Hidden Markov Model (HMM) to search existing genome and metagenome databases for the presence of potential PET hydrolases (FIGURE 1 and FIGURE 3B). Using this approach, we identified >500 potential PET hydrolases in the UniProtKB database. In addition, 349 sequence homologs were obtained from several public metagenome datasets deposited on the IMG server and four of the identified candidate genes were functionally verified. Altogether these results imply that PET hydrolase genes are globally distributed in marine and terrestrial metagenomes. Further, we provide evidence that in the marine environments, the PET hydrolases originate mainly from the phylum of Bacteroidetes and in the terrestrial metagenomes from Actinobacteria.

RESULTS

Construction of a Hidden Markov Model (HMM) for PET hydrolases

Only few well characterized PET hydrolases are currently known. The most prominent examples are PET hydrolases from *T. fusca* and *I. sakaiensis* (see references in TABLE 1, FIGURE 2). Within this manuscript, we set out to increase the diversity of this intriguing group of
Danso et al., 2018; PET hydrolases from metagenomes

To identify potential novel PET hydrolases, a T-Coffee amino acid sequence alignment of nine already known examples was constructed. The enzyme sequences used for the model have all verified activity on PET-based substrates (TABLE 1). Of the proteins used for the model, seven sequences originated from the phylum of Actinobacteria (e.g. sequences #1-7, TABLE 1), one from the phylum of Proteobacteria (Betaproteobacteria) (#8, TABLE 1) and one sequence (WP_54022242.1) was a protein with metagenomic origin and not yet assigned to any phylum (#9, TABLE 1). A comparison with the well described PET active cutinase TICut2 from Thermobifida fusca allowed the identification of the location of the catalytic triad and other residues which are commonly involved in binding of the substrates (FIGURE 3A). Next to the serine of the catalytic triad, a methionine residue was found in every sequence, which is of importance for forming an oxyanion hole together with an aromatic residue. This aromatic residue is also part of an aromatic clamp together with similar amino acids like tryptophan, tyrosine, histidine and phenylalanine (13, 18). Terminal cysteine residues are present in all examples and may be important for the thermostability of these enzymes (19, 20). The alignment was used for the construction of a “Hidden Markov Model”. For visualization, a HMM-Logo was created via the “skyline.org” online tool (FIGURE 3B). A subsequent visual analysis and conservation prediction using the “JS divergence scoring method” revealed at least eight conserved regions (FIGURE 3B).

An initial HMMER online tool database search with the model against the UniProtKB-Database revealed a total of 10,854 significant query matches, with a highest Bit score value of 441.6. Of these, a subset was chosen that showed a Bit score value of >180. In addition, a BLAST search was performed using the newly discovered potential PET hydrolases from the HMM search as initial query sequences against the non-redundant and the metagenomic datasets available at the NCBI database in May 2017. This resulted in the detection of 504 potential PET hydrolase candidate genes. From the obtained homologous sequences, 13 potential PET hydrolase homologs (FIGURE 2) were manually chosen due to their sequence
Danso et al., 2018; PET hydrolases from metagenomes similarity to known PET hydrolases (PET1-PET13). They were used for initial verification and further in silico and/or biochemical characterization. These novel predicted PET hydrolases are summarized in TABLE S1 together with their UniProt entries and pfam domain similarities. Thereby, it was of interest to select mainly non-actinobacterial proteins, in order to diversify the HMM. The 13 sequences were added to the alignment and used for a modified and refined HMM. The 13 initially identified putative PET hydrolases harbor the above mentioned and in TABLE 2 listed residues and motifs.

Classification of PET-hydrolases and taxonomic assignments

A NCBI conserved domain search in the CDD databases showed that the nine active PET hydrolases previously known (TABLE 1) as well as the thirteen novel (TABLE S1) possible homologs harbor domains belonging to the superfamily of $\alpha/\beta$-hydrolases_5 (pfam12695). Of these, only PET9 and PET12 showed specific hits for the superfamily of acetyl xylan esterases (AXE1, pfam05448). Additional to the specific hits, several unspecific domain hits were obtained (TABLE S1).

A further alignment and subsequent tree calculation with the above identified 504 potential PET hydrolases allowed the assignment of all enzymes in 17 subclasses (FIGURE 4). Interestingly, two sequences (A0A1N6SMU6 and A0A168EN35) did form individual subclusters and could not be assigned to other clusters. The majority of the subclasses were mainly affiliated with Actinobacteria and only one subclass (XII) with Proteobacteria. For the subclass XVII, no clear assignment was possible. The Thermobifida PET hydrolase sequences are clustered within subcluster XV together with the PET hydrolase from Saccharomonospora. The “Leaf Compost Cutinase” (LCC) sequence was found in subcluster XI and the Thermomonospora curvata sequence is located in group XII. The PET hydrolase from Ideonella sakaiensis is located in subcluster VI.
Experimental verification of the HMM and characterization of selected novel PET hydrolases

Since the bioinformatic approach only delivered potential PET hydrolases enzymes, we initiated work to verify a small number of the identified candidate genes with respect to their function. Therefore, we chose the enzymes PET 2, 5, 6 and 12 (TABLE S1). The respective genes were either synthesized or amplified from genomic DNA using vectors and primers as outlined in TABLE 3 and TABLE 4 and cloned into the expression vectors. Initial tests indicated that all genes coded for active enzymes. On agar plates containing PET nanoparticles or polycaprolactone (PCL) (21, 22), all active clones produced halos after overnight incubation and were compared to the PET hydrolase from Thermobifida fusca as a positive control (FIGURE S1). PCL was used as a model substrate as hydrolysis of this compound indicates possible activities on the more complex PET. From these active enzymes, we choose two enzymes for a more detailed biochemical characterization. These were the two enzymes PET 2 and PET 6. PET 2 was derived from a marine metagenomics data set (23) and PET 6 was derived from Vibrio gazogenes DSM-21264 (24). After successful expression and purification of the two enzymes in sufficient amounts, the obtained enzymes were further characterized using para-Nitrophenyl esters (pNP-esters) (FIGURE 5, FIGURE S2 and FIGURE S3). Both enzymes showed best activity against short chain pNP-esters (C2-C4), but were able to convert long chain substrates as well (> C10). Their temperature optimum was 55 °C and 70 °C for PET6 and PET2 respectively. Remarkably, PET 2 retained 80 % of its relative activity at 90 °C after incubation for > 5 hours. Both enzymes preferred alkaline pH values of 8-9. Rubidium had a strong effect by increasing the activity by 50 % at a concentration of 1 mmol/L in case of PET2. The same, but significantly smaller effect was observed in case of PET6. Both enzymes showed reduced hydrolytic activity in the presence of 5 % SDS, 10 mM PMSF and 30 % acetonitrile. Additional HPLC analyses confirmed the above findings for PET2. In tests using 14 mg...
Danso et al., 2018; PET hydrolases from metagenomes

amorphous PET foil as substrate, 100 μg PET 2 was able to release 900 μM of terephthalic acid after 24 hours of incubation (FIGURE S4).

Altogether, the above presented data indicate that the developed search algorithm is useful for the identification of novel and functionally active PET hydrolases from single genomes and metagenomes.

Global distribution of PET hydrolases and their significance in marine and terrestrial environments

After the successful construction of a reliable HMM and the identification as well as partial characterization of new PET hydrolases, we asked, if and to what extent these enzymes could be identified on a global level. To evaluate the environmental distribution of sequences encoding PET hydrolases, data of 108 marine and 25 terrestrial metagenomes were taken into account and downloaded from the IMG-Database (25) (TABLE S2). Criteria for the selection of marine data were sample depth (max. 2m), assembly status, global distribution of sample locations, size and availability of the dataset. The same criteria except the sample depth were chosen for terrestrial metagenomes. The size of assembled metagenome data in case of marine metagenomes ranged from 10.85 Mb up to 7.99 Gb. In the case of terrestrial metagenomes, the number of assembled bases ranged from 58 Mb to 9.2 Gb. The modified HMM was used to find PET hydrolase homologs in the sequence data of those metagenomes on a global scale. The searches identified possible PET hydrolase homologs in 31 marine and 11 terrestrial metagenomes. A total of 349 hits was observed for these 42 samples. The number of hits per sample was normalized, calculated as hits per Mb, and visualized on a global map representing the geographical location as well as the frequency of PET hydrolase homologs (FIGURE 6).

Within the marine and terrestrial metagenomes, PET hydrolase frequencies ranged from 0.004 to 0.92 hits/Mb and 0.0001 to 1.513 hits/Mb, respectively.
The combined genome sizes of terrestrial metagenomes are nearly 2.5-fold higher than those of the marine metagenomes and harbor 157 PET hydrolase homologs in average. In contrast, the marine metagenomes harbor 42 PET hydrolases in average. The terrestrial metagenome with the highest abundance of potential PET hydrolases contains 135 sequence hits and was derived from the sediment core of a heavy oil reservoir in Canada (IMG Genome ID: 3300001197). In case of the marine metagenomes, 31 hits were the maximum and found within metagenome data of a sample from Delaware coast in the USA (FIGURE 6).

We further observed that within the terrestrial habitats, the Actinobacteria were the main hosts for the terrestrial-derived enzymes. However, in the marine samples, most predicted PET hydrolases originated from the phylum of the Bacteroidetes (FIGURE 7). Bacteroidetes in the marine samples were affiliated with 43 % of all hits. The phyla of Proteobacteria was the second most abundant in both datasets with 23 % in marine and 20 % in terrestrial data (FIGURE 7).
Within this work, we have developed a search algorithm, allowing the in-silico identification of PET hydrolase gene candidates from genomes and metagenomes. Altogether, we were able to identify 504 novel possible enzyme candidates in the UniProtKB and non-redundant (RefSeq) and the metagenomic database available at the NCBI database. In addition, we identified 349 candidate genes and enzymes from marine and terrestrial metagenomes available at the IMG platform. This is by far the largest collection of PET hydrolase enzyme candidates currently available. A first classification of the PET hydrolases derived from UniProtKB/GenBank due to their protein sequence similarities and occurrence of conserved homologs enabled the formation of 17 enzyme clusters within this work (FIGURE 4). An additional search in the global metagenomes revealed that the PET hydrolases occur in both marine and terrestrial habitats. However, the frequencies (hits per Mb) are comparably low. The lowest hit rate was observed for a metagenome from a Kansas prairie soil sample and the highest hit rate was observed for a metagenome derived from a heavy oil reservoir sediment core. Since we included over 100 metagenomes in this analysis, the data give a reliable picture of the overall occurrence of these enzymes. But the data do not allow estimations on the expression of these genes in the native environment. However, the overall low gene frequencies might suggest that bacterial evolution did not yet allow the spreading of this trait. This also implies that the overall degrading potential of the oceans is rather low compared to other habitats and enzymes involved in the breakdown of natural polymers like starch or cellulose (26,27).

We verified the usefulness of the developed algorithm by cloning four novel PET hydrolase genes and expressing them heterologously in *E. coli*. All enzyme clones were active and supported the notion that the developed search algorithm is useful. The enzyme properties of the newly characterized PET hydrolases fit well into the overall picture of known PET
hydrolyzing enzymes with an optimum pH at slightly alkaline values and the preference of substrates with a short chain length (19) (28). Overall these novel enzymes revealed comparable activities to those already previously characterized (see Table 1 and References herein).

To our surprise, both enzymes characterized in more detail (PET2 and PET6) show traits of thermostability (FIGURE 5) (23). PET 2 was stable up to 90 °C, with measurable residual activity of more than 50 %. This is more stable than the LCC derived from a compost metagenome (19).

All the newly identified PET hydrolases originated mainly from three bacterial phyla: *Proteobacteria*, *Actinobacteria*, and *Bacteroidetes*. Within this framework, it is notably, that the *Bacteroidetes* have so far not been associated with PET degradation, but *Bacteroidetes* have been described as very potent degraders of other polymers and they harbour a multitude of hydrolases and binding modules (29-31).

The restriction to few bacterial phyla could indicate that the metabolic capability has only rather recently been evolved and it is thus limited to very few phylogenetic groups. The observation here, that in the marine habitat, the phylum of the *Bacteroidetes* is the main host of PET hydrolases is new and intriguing out of several reasons. First, using classical searches and biochemical characterization the *Actinobacteria* and *Proteobacteria* were considered as the main hosts for these enzymes (TABLE 1). Second, the searches in UNIProtKB and other databases implemented in the NCBI webpage underlined the presence of PET hydrolases in the phylum of the *Actinobacteria* and *Proteobacteria*. Only when we extended our search for metagenomes of mainly non-cultivated bacterial phyla, we identified the *Bacteroidetes* as main host for these enzymes in the marine environment.

The recent findings on PET hydrolases described in this publication will significantly extend the knowledge of these enzymes and provide promising candidates for biotechnological applications.
In summary, the over 800 enzyme candidates identified in this work will build the basis for a global repository and database of this urgently needed enzyme class.
MATERIAL AND METHODS

Bacterial strains, plasmids and primers

Bacterial strains, plasmids and primers used in this study are listed in TABLE 3 and TABLE 4. If not mentioned otherwise, *Escherichia coli* clones were grown in LB medium (1 % tryptone/peptone, 0.5 % yeast extract, 1 % NaCl) supplemented with appropriate antibiotics (25 μg/ml kanamycin, or 100 μg/ml ampicillin) at 37 °C for 18 hours.

Databases used in this study and bioinformatic analysis

Nucleotide and amino acid sequences of putative PET hydrolases were acquired from databases integrated into the NCBI (https://www.ncbi.nlm.nih.gov/), UniProt (http://www.uniprot.org/) and IMG (JGI, http://jgi.doe.gov/) webpages (32), (25, 33). Sequences were compared to others deposited in the NCBI databases using BLAST alignment tools (34). Amino acid sequence HMM search was carried out using the HMMER (http://hmmer.org) webpage or a local version of the software (v3.1b2) with downloaded datasets. Structural information on the enzymes was retrieved from the RCSB-PDB (35) database.

Sequence data were processed using BioEdit and Clone Manager Suite9 (Sci-Ed, Denver, USA). Neighbor-joining phylogenetic trees based on amino acid sequence alignments were constructed using MEGA6 (36). Nine known and activity-confirmed bacterial PET hydrolase sequences were obtained from NCBI, aligned with T-coffee (37) and manually revised. Afterwards, the alignment was used to construct a profile HMM with the hmmbuild function of the HMMER package (http://hmmer.org). After the identification of PET hydrolase homologs, the obtained sequences were included into the above-mentioned alignment and a the
HMM was refined. HMM logo was visualized using the skyline online tool (38). Metagenomic data was downloaded from IMG using a Globus endpoint and was further analyzed using hmmsearch from the HMMER package. Phylogenetic assignment was done via a local diamond-blast (39) against the non-redundant protein database (34) and subsequent analysis with MEGAN6 (40). The map representing the frequency and geographical distribution of PET hydrolases in metagenomes (FIGURE 6) was constructed using QGis Desktop 2.18.5 (http://www.qgis.org).

Cloning and heterologous expression of PET2, PET5, PET6, PET12 in Escherichia coli T7-Shuffle

Cloning of PET hydrolase genes into expression vectors pET21a(+) and pET28a(+) was accomplished after amplification of genomic DNA using specific primer pairs with underlined homolog regions to the vector or restriction sites. The sequence of PET2 was obtained from NCBI (ACC95208) and synthesized after codon usage optimization for E. coli (MWG Eurofins, Germany). Obtained DNAs were cloned into expression vectors and the constructs were transformed into E. coli T7-Shuffle cells. The cultures were grown aerobically in auto-induction medium (ZYM-5052) (41) containing 100 μg/ml ampicillin or 25 μg/ml kanamycin, for pET21a(+) and pET28a(+) respectively, at 37 °C until they reached an OD<sub>600</sub> of 1.0. The proteins were expressed afterwards at 17 °C for 16-20 h harboring a C- or N-terminal histidine tag. The cells were harvested and lysed with pressure using a French-press. Afterwards, the proteins were purified with nickel-ion affinity chromatography using Ni-NTA agarose (Qiagen, Hilden, Germany) and analyzed by SDS-PAGE. The elution buffer was exchanged against 0.1 mM potassium phosphate buffer pH 8.0 in a 10 kDa Amicon Tube (GE Health Care, Solingen, Germany).
Biochemical characterization of PET2 and PET6

For activity tests, both enzymes were assayed using purified recombinant protein. Unless otherwise indicated, the enzymes were added to a substrate solution containing 190 μL of either 0.2 M sodium phosphate buffer or 0.1 M Tris-HCl with a defined pH between 7 and 8 and 0.5 mM pNP-substrate dissolved in isopropanol, incubation time ranged from 15 min. to 30 min. As substrates, we tested pNP-esters with chain length of C2, C4, C6, C8, C10, C12, C14, C16 and C18. After incubation at defined temperature, the color change from colorless to yellow was measured at 405 nm in a plate reader (Biotek, Winooski, USA). All samples were measured in triplicate. For determination of the optimal temperature, samples were incubated between 17 °C and 90 °C for 15 min. The impact of pH conditions on the activity of each enzyme was measured in citrate phosphate- (pH 3.0, 4.0 and 5.0), potassium phosphate (pH 6.0, 7.0 and 8.0) and carbonate bicarbonate buffer (pH 9.2 and 10.2). The influence of possible cofactors, solvents, detergents and inhibitors was assayed at different concentration levels. After 1 h incubation in the presence of the following substances, the residual activity was determined after 15 min. incubation at optimal temperature with pNP-octanoate and optimal pH. The possible cofactors Ca²⁺, Co²⁺, Cu²⁺, Fe³⁺, Mg²⁺, Mn²⁺, Rb²⁺ and Zn²⁺ with a final concentration of 1 and 10 mM were used. To determine the solvent stability, DMSO, isopropanol, methanol, DMF, acetone, acetonitrile and ethanol with a final concentration of 10 % and 30 % (v/v) was added to the reaction. Detergent stability was assayed with SDS, Triton X-100 and Tween 80 at 1 % and 5 % (w/v, v/v) concentration. The inhibitory effect of EDTA, DTT and PMSF was tested at 1 and 10 mM concentration. Substrate analysis using the HPLC LaChrom Elite® system from Hitachi (Tokyo, Japan) with a Lichrospher® 100 RP-18e column (VWR International GmbH, Darmstadt, Germany), consisting of particles with 5 μm diameter, were done as previously published (19). 14 mg Low-crystallinity PET film (Goodfellow GmbH, Bad Nauheim, Germany) was used as substrate. For enzymatic hydrolysis, up to 50 μg of protein was incubated at 60 °C with
continuous shaking at 500 rpm. As a mobile phase, acetonitrile (A) and water with 0.1 % TFA (B) were used in a isocratic method with 20 % acetonitrile (A). The reaction buffer was 0.1 M Tris-HCl, pH 7.5 with an injection volume of 99 μl. Detection was performed at 241 nm.
ACKNOWLEDGEMENTS

This work was in part supported by the BMBF within the programs FUPOL and MetaCat and the EU project INMARE at the University of Hamburg.
REFERENCES


Danso et al., 2018; PET hydrolases from metagenomes


FIGURE LEGENDS

FIGURE 1. Workflow used in this study to identify and partially characterize novel PET hydrolases from data bases and global metagenomes.

FIGURE 2. Neighbor joining tree of manually chosen, potential PET hydrolase sequences found in this work. Sequences were obtained from a HMM search in the UniProtKB database and named PET1-13. The tree was calculated using Mega6. Besides the 13 newly found PET hydrolase (TABLE S1) sequences, 9 already known PET hydrolases (TABLE 1) were added to the tree in order to visualize the phylogenetic distribution and similarity of the PET hydrolase sequence homologs.

FIGURE 3. Amino acid sequence alignment of described PET hydrolases. (A) An alignment of PET hydrolase sequences listed in TABLE 1 revealed the position of binding relevant residues and conserved regions. (B) Hidden Markov Model (HMM) of PET hydrolase amino acid motifs. The amino acid alignment from (A) was used to calculate a HMM profile. The HMM was consequently visualized as a logo with information content above background (skylign.org). Eight sequence motifs are shown in total. Motif 2-8 include amino acids crucial for thermostability, substrate binding and/or catalytic activity.

FIGURE 4. Classification and phylogenetic tree of 504 novel and potential PET hydrolases obtained by HMM searches. Sequences were obtained from the the UniProtKB database. A total of 504 sequences identified with the constructed HMM and a bitscore > 180 were visualized, whereby the sequences of PET1-13 (TABLE S1) as well as 9 already described PET hydrolases (TABLE 1) represent a subset of the new found potential enzymes.
FIGURE 5. Biochemical characterization of PET2 and PET6 with different pNP-substrates. Data obtained with a pNP-assay are shown in net diagrams for PET2 and PET6. Substrate preferences, temperature optimum and pH optimum were tested. All tests besides substrate preferences were carried out with pNP-octanoate.

FIGURE 6. Global distribution of PET hydrolases in available metagenomes. Potential PET hydrolase containing metagenomes were visualized on a world map containing circles for marine and triangles for terrestrial metagenomes. Blue and red color shading indicates the frequency of PET hydrolase genes in hits/Mb for marine and terrestrial metagenomes, respectively. Red and green boxes magnify regions with overlapping spots (sample sites).

FIGURE 7. Phylogenetic affiliation of 349 predicted PET hydrolases from 31 marine and 11 terrestrial metagenomes. Colored and stacked bars represent the number of hits per phylum. Data were normalized per Mb of assembled DNAs for the analyzed samples.
### TABLE 1: Currently known and partially characterized PET hydrolases

Names and protein database (PDB) entry numbers of currently known PET hydrolases used in this work as references. Sequence data of those examples were used for the initial construction of the HMM.

<table>
<thead>
<tr>
<th>Number</th>
<th>PDB Entry</th>
<th>Gene names</th>
<th>Organism</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>W0TJ64</td>
<td>Cut190</td>
<td><em>Saccharomonospora viridis</em></td>
<td>(42)</td>
</tr>
<tr>
<td>2</td>
<td>E9LV10</td>
<td>cut1</td>
<td><em>Thermobifida fusca (Thermomonospora fusca)</em></td>
<td>(43)</td>
</tr>
<tr>
<td>3</td>
<td>E9LBQ3</td>
<td>cut-2</td>
<td><em>Thermobifida fusca (Thermomonospora fusca)</em></td>
<td>(11)</td>
</tr>
<tr>
<td>4</td>
<td>D1A8G5</td>
<td>Tcur_1278</td>
<td><em>Thermomonospora curvata</em></td>
<td>(21)</td>
</tr>
<tr>
<td>5</td>
<td>E9LVH7</td>
<td>cut1</td>
<td><em>Thermobifida alba</em></td>
<td>(9)</td>
</tr>
<tr>
<td>6</td>
<td>H9WX98</td>
<td>N/A</td>
<td><em>Thermobifida halotolerans</em></td>
<td>(12)</td>
</tr>
<tr>
<td>7</td>
<td>E9LH9</td>
<td>cut2</td>
<td><em>Thermobifida cellulosa</em></td>
<td>(7)</td>
</tr>
<tr>
<td>8</td>
<td>A0A0K8P6T7</td>
<td>ISF6_4831</td>
<td><em>Ideonella sakaiensis</em></td>
<td>(17)</td>
</tr>
<tr>
<td>9</td>
<td>G9BY57</td>
<td>LCC</td>
<td>uncultured bacterium</td>
<td>(19)</td>
</tr>
</tbody>
</table>
TABLE 2: Detailed list of determined searching criteria for the identification of PET hydrolase candidate genes in data bases.

The letter x indicates an non-conserved position within the sequence pattern. Brackets indicate a less conserved position within the sequence pattern. Numbering of amino acids is according to the HMM (FIGURE 3A).

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GxSMGGGG</td>
<td>Serine of catalytic triade and methionine for oxyanionhole formation</td>
</tr>
<tr>
<td>2</td>
<td>F,Y62</td>
<td>Amino acids for oxyanionhole formation and aromatic clamp</td>
</tr>
<tr>
<td>3</td>
<td>W,Y157</td>
<td>Amino acids for oxyanionhole formation and aromatic clamp</td>
</tr>
<tr>
<td>4</td>
<td>I,V180</td>
<td>Optional aromatic amino acid for aromatic clamp formation</td>
</tr>
<tr>
<td>5</td>
<td>F,W211</td>
<td>C-terminal cysteine residues for thermostability supporting disulfide bond formation</td>
</tr>
<tr>
<td>6</td>
<td>C255 C262</td>
<td>C-terminal cysteine residues for thermostability supporting disulfide bond formation</td>
</tr>
<tr>
<td>7</td>
<td>DxDxR(Y)xxF(L)C</td>
<td>Conserved sequence prior to first thermostability giving cysteine</td>
</tr>
</tbody>
</table>
TABLE 3: Bacterial strains, plasmids used in this work

<table>
<thead>
<tr>
<th>Strains</th>
<th>properties</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli DH5α</td>
<td>supE44 ΔlacU169 (Φ80 lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1</td>
<td>(44)</td>
</tr>
<tr>
<td>E. coli BL21 (DE3)</td>
<td>F, ompT, hsdS B (r~ m~) gal, dcm, FDE3</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>E. coli T7Shuffle Express</td>
<td>flhA2 lacZ::T7 gene1 [lon] ompT  ahpC gal λatt::pNEB3-r1-cDsbC</td>
<td>NEB (Frankfurt am Main, Germany)</td>
</tr>
<tr>
<td>Deinococcus maricopensis</td>
<td>type strain, DSM-2121</td>
<td>DSMZ (Braunschweig, Germany)</td>
</tr>
<tr>
<td>Vibrio gazogenes</td>
<td>type strain, DSM-21264</td>
<td>DSMZ (Braunschweig, Germany)</td>
</tr>
<tr>
<td>Polyangium brachysporum</td>
<td>type strain, DSM-7029</td>
<td>DSMZ (Braunschweig, Germany)</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pET21α(+)</td>
<td>Expression vector, lacI, Amp4, T7-lac- promoter, C-terminal His6-tag coding sequence</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>pET28α(+)</td>
<td>Expression vector, lacI, Amp8, T7-lac- promoter, C-terminal His6-tag and N-terminal coding sequence</td>
<td>Novagen/Merck (Darmstadt, Germany)</td>
</tr>
<tr>
<td>pEX-A2</td>
<td>Cloning vector, Amp8, PspnlacZ, pUC ori</td>
<td>Eurofins MWG Operon (Ebersberg, Germany)</td>
</tr>
</tbody>
</table>
### TABLE 4: Primers used in this work

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' → 3')</th>
<th>Length (bp)</th>
<th>Tm (°C)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 promotor</td>
<td>TAATACGACTCACTATAGGG</td>
<td>20</td>
<td>53.2</td>
<td>Eurofins MWG (Ebersberg, Germany)</td>
</tr>
<tr>
<td>T7 terminator</td>
<td>CTAGTTATTGCTCAGCGT</td>
<td>19</td>
<td>54.5</td>
<td>Eurofins MWG (Ebersberg, Germany)</td>
</tr>
<tr>
<td>PET5_for</td>
<td>CGCCGCCATATGAATAAATCTATTCTAAAAAANCTCTC</td>
<td>38</td>
<td>68</td>
<td>this work</td>
</tr>
<tr>
<td>PET5_rev</td>
<td>CGATTCCGCGGCGGCGCGTAATTACATGTGACG</td>
<td>34</td>
<td>77</td>
<td>this work</td>
</tr>
<tr>
<td>PET6_for</td>
<td>CGTAGTCAATATGCACCCTGTCTGGGACTG</td>
<td>29</td>
<td>69</td>
<td>this work</td>
</tr>
<tr>
<td>PET6_rev</td>
<td>CAGCCGCGGCTTAATAGTAACTACAGTTGTCTC</td>
<td>34</td>
<td>73</td>
<td>this work</td>
</tr>
<tr>
<td>PET12_for</td>
<td>CGCCATATGCAGAACCACCTACACAGCGAGGCC</td>
<td>35</td>
<td>80</td>
<td>this work</td>
</tr>
<tr>
<td>PET12_rev</td>
<td>CTTCGCCGCCCTCAGATCGGCGAGCGCGCTCGGCTGCGGTACTCC</td>
<td>39</td>
<td>84</td>
<td>this work</td>
</tr>
</tbody>
</table>
FIGURE 1. Workflow used in this study to identify and partially characterize novel PET hydrolases from data bases and global metagenomes.
FIGURE 2. Neighbor joining tree of manually chosen, potential PET hydrolase sequences found in this work. Sequences were obtained from a HMM search in the UniProtKB database and named PET1-13. The tree was calculated using Mega6. Besides the 13 new found PET hydrolase (TABLE S1) sequences, 9 already known PET hydrolases (TABLE 2) were added to the tree in order to visualize the phylogenetic distribution and similarity of the PET hydrolase sequence homologs.
FIGURE 3. Amino acid sequence alignment of described PET hydrolases. (A) An alignment of PET hydrolase sequences listed in TABLE 2 revealed the position of binding relevant residues and conserved regions. (B) Hidden Markov Model (HMM) of PET hydrolase amino acid motifs. The amino acid alignment from (A) was used to calculate a HMM profile. The HMM was consequently visualized as a logo with information content above background (skyline.org). Eight sequence motifs are shown in total. Motif 2-8 include amino acids crucial for thermostability, substrate binding and/or catalytic activity. The information content (bits) is represented on the ordinate. The three values provided under each amino acid give information on the occupancy, insert probability and insert length, respectively.
FIGURE 4. Classification and phylogenetic tree of 504 novel and potential PET hydrolases obtained by HMM searches. Sequences were obtained from the UniProtKB database. A total of 504 sequences identified with the constructed HMM and a bitscore > 180 were visualized, whereby the sequences of PET1-13 (TABLE S1) as well as 9 already described PET hydrolases (TABLE 2) represent a subset of the new found potential enzymes.
FIGURE 5. Biochemical characterization of PET2 and PET6 with different pNP-substrates. Data obtained with a pNP-assay are shown in net diagrams for PET2 and PET6. Substrate preferences, temperature optimum and pH optimum were tested. All tests besides substrate preferences were carried out with pNP-octanoate.
FIGURE 6. Global distribution of PET hydrolases in available metagenomes. Potential PET hydrolase containing metagenomes were visualized on a world map containing circles for marine and triangles for terrestrial metagenomes. Blue and red color shading indicates the frequency of PET hydrolase genes in hits/Mb for marine and terrestrial metagenomes, respectively. Red and green boxes magnify regions with overlapping spots (sample sites).
FIGURE 7. Phylogenetic affiliation of 349 predicted PET hydrolases from 31 marine and 11 terrestrial metagenomes. Colored and stacked bars represent the number of hits per phylum. Data were normalized per Mb of assembled DNAs for the analyzed samples.