The abundance of mRNA transcripts of bacteroidetal polyethylene terephthalate (PET) esterase genes may indicate a role in marine plastic degradation

Hongli Zhang
University of Hamburg https://orcid.org/0000-0002-4871-2790
Robert Dierkes
University of Hamburg
Pablo Pérez-García
https://orcid.org/0000-0003-2248-3544
Sebastian Weigert
University of Bayreuth https://orcid.org/0000-0002-0469-1545
Stefanie Sternagel
University of British Columbia https://orcid.org/0000-0002-4518-1875
Steven Hallam
University of British Columbia https://orcid.org/0000-0002-4889-6876
Thomas Schott
Leibniz Institute of Baltic Sea Research
Klaus Juergens
Leibniz Institute for Baltic Sea Research Warnemünde (IOW)
Christel Vollstedt
University of Hamburg
Cynthia Chibani
Kiel University
Dominik Danso
University of Hamburg
Patrick Buchholz
University of Stuttgart
Jürgen Pleiss
University of Stuttgart https://orcid.org/0000-0003-1045-8202
Alexandre Almeida
EMBL-EBI https://orcid.org/0000-0001-8803-0893
Birte Hocker
Universität Bayreuth https://orcid.org/0000-0002-8250-9462

Ruth Schmitz
Christian-Albrechts-University Kiel https://orcid.org/0000-0002-6788-0829
Jennifer Chow
University of Hamburg https://orcid.org/0000-0002-7499-5325
Wolfgang Streit (wolfgang.streit@uni-hamburg.de)
University of Hamburg https://orcid.org/0000-0001-7617-7396

Article

Keywords: HMM, hydrolases, metagenome, metagenomic screening, PET degradation, Polyethylene terephthalate (PET), Bacteroidetes, Flavobacteriaceae

Posted Date: August 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-567691/v2

Version of Record: A version of this preprint was published at Frontiers in Microbiology on January 5th, 2022. See the published version at https://doi.org/10.3389/fmicb.2021.803896.

The abundance of mRNA transcripts of bacteroidetal polyethylene terephthalate (PET) esterase genes may indicate a role in marine plastic degradation

Hongli Zhang[1], Robert Dierkes[1], Pablo Perez-Garcia[1,2], Sebastian Weigert[3], Stefanie Sternagel[4], Steven J. Hallam[4,5], Thomas Schott[6], Klaus Jürgens[6], Christel Vollstedt[1], Cynthia Maria Chibani[2], Dominik Danso[1], Patrick C.F. Buchholz[7], Juergen Pleiss[7], Alexandre Almeida[8,9], Birte Höcker[3], Ruth A. Schmitz[2], Jennifer Chow[1] and Wolfgang R. Streit[1*]

[1]Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany

[2]Molecular Microbiology, Institute for General Microbiology, Kiel University, Kiel, Germany

[3]Department of Biochemistry, University of Bayreuth, Bayreuth, Germany

[4]Department of Microbiology & Immunology, University of British Columbia, Canada

[5]ECOSCOPE Training Program, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.

[6] Department of Biological Oceanography, Leibniz Institute of Baltic Sea Research, Rostock-Warnemünde, Germany

[7] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

[8] European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK

[9] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

Keywords: HMM, hydrolases, metagenome, metagenomic screening, PET degradation, Polyethylene terephthalate (PET), Bacteroidetes, Flavobacteriaceae

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

Polyethylene terephthalate (PET) is an important synthetic polymer accumulating in nature and recent studies have identified microorganisms capable of degrading PET. While the majority of known PET hydrolases originate from the Actinobacteria and Proteobacteria, here we describe the first functional PET-active enzymes from the Bacteroidetes phylum. Using a PETase-specific Hidden-Markov-Model (HMM)-based search algorithm we identified two promiscuous and cold-active esterases derived from *Aequorivita* sp. (PET27) and *Chryseobacterium jeonii* (PET30) acting on PET foil and powder. Notably, one of the enzymes (PET30) was able to hydrolyze PET at temperatures between 4° - 30°C with a similar turnover rate compared to the well-known *Ideonella sakaiensis* enzyme (IsPETase).

PET27 and PET30 homologues were detected in metagenomes encompassing a wide range of different global climate zones. Additional transcript abundance mapping of marine samples imply that these promiscuous enzymes and source organisms may play a role in the long-term degradation of microplastic particles and fibers.

IMPORTANCE

Polyethylene terephthalate (PET) has accumulated in our environment throughout the last six decades, presumably with no major degradation paths identified. While several PET hydrolases have been described, their abundance and activity in nature remains unconstrained. In this study, homology-based mining of metagenomic datasets demonstrated for the first time that functional PET hydrolases are affiliated with the Bacteroidetes phylum. We prove that these enzymes have significant capability to hydrolyze PET at temperatures ranging from 4 °C - 30°. Moreover, they are distributed globally and, in several instances, expressed at high levels. While these enzymes have relatively low turnover rates, they may significantly contribute to PET removal because of their global distribution and expression. This work provides a greater understanding of the phylogenetic diversity, biochemical and structural traits of PET hydrolases and sheds new light on their potential impact on plastics degradation in the environment.

INTRODUCTION

PET is one of the most common plastics used in many consumer products. The worldwide PET resin production amounted to 27.8 million tons in 2015 [1,2]. However, only a small fraction of PET is recycled,

and it is estimated that 58 % ends up in the landfills and in the ocean [3,4]. Our knowledge of microbial degradation of most plastics is rather limited. Degradation is, however, initiated by UV light and or mechanical grinding through waves and other movements generating microplastics (< 5mm) [5]. Thereby, it can be assumed that the microparticles allow better microbial attachment [6,7]. In the case of PET, recent research has demonstrated that some bacteria are able to degrade the polymer. Although it is unclear the larger crystalline fibers are degraded by bacteria, it is well known that cutinases (EC 3.1.1.74), lipases (EC 3.1.1.3) and carboxylesterases (EC 3.1.1.1) can act on amorphous and low crystalline PET. These enzymes break the ester bond in the polymer to either produce bis-hydroxyethyl terephthalate (BHET), mono-hydroxyethyl terephthalate (MHET) or terephthalic acid (TPA) and ethylene glycol (EG). MHET can subsequently be cleaved with a specific MHETase and the TPA monomers degraded via cleavage of the aromatic ring structure using known aryl pathways [8,9].

To date, only a limited number of bacterial and fungal species have been identified that are capable of breaking down PET to oligomers or even monomers. Most bacterial isolates with verified enzymatic PET degradation are affiliated with the Gram-positive phylum Actinobacteria [10]. The best characterized examples belong to the genera *Thermobifida* or *Thermomonospora* [10-15]. Further, the leaf compost-derived cutinase LCC is closely related to the Actinobacterial enzymes and is currently one of the best described and most active PETases [16,17].

More recently, the complete degradation of amorphous PET materials was described for the Gram-negative Betaproteobacterium *Ideonella sakaiensis* 201-F6, which is capable of using PET as a major energy and carbon source [18]. *I. sakaiensis'* genome also encodes a tannase that appears to be unique, and which is designated MHETase as it is capable to degrade MHET. Besides these, a number of other PETases affiliated with the Proteobacterial phylum have been identified [19-22].

In a previous study, we identified potential PET esterases affiliated with the Bacteroidetes phylum using HMM profile database searches [9,19]. These enzymes were mainly recovered from marine environments and annotated solely on the basis of homology. However, we had not verified the enzymatic function, the environmental distributions and expression of the predicted enzymes within that framework, and because of their global occurrence, we now sought to determine, if the predicted enzymes are indeed acting on the PET polymer. Bacteroidetes representatives can be found in nearly all ecological niches including soils, oceans and fresh water and are part of the microbiome of many animals, especially as

inhabitants of the intestinal tract [23-26]. The Bacteroidetes phylum, however, is highly heterogenous and contains at least four classes of bacteria (*e.g.* Bacteroidia, Flavobacteria, Sphingobacteria, and Cytophagia) with each class having several thousand described species. The phylum contains non-spore forming and rod shaped aerobic but often anerobic microorganisms with an enormous metabolic diversity [24]. The global distribution of Bacteroidetes representatives is probably due to their ability to decompose a very wide variety of bio-based polymers such as cellulose, chitin or algal cell walls. In particular, the decomposition of polysaccharides (cellulose and hemicellulose) by Bacteroidetes inhabiting the intestinal tract of humans and animals has been a well-studied property of the gut microbiome [27].

Here, we provide the first experimental evidence that different Bacteroidetes representatives have evolved promiscuous esterases that degrade the PET polymer. We show that at least the two Bacteroidetes genera, Aequorivita and Chryseobacterium (recently renamed to Kaistella), harbor PET-active enzymes and that these enzymes are widespread, expressed and likely distributed through horizontal gene transfer. Given their abundance and diversity we speculate that the described PETases have considerable potential to impact long-term removal of PET from the marine environment.

.

RESULTS

Profile Hidden Markov Model (HMM) searches identify potential bacteroidetal PETases

Previously, we identified PET-active genes and enzyme candidates affiliated with phylum Bacteroidetes [19]. In this study, we initiated work to enrich the diversity of these genes encoding PET-active enzymes and to validate their catalytic function using experimental approaches. To achieve these goals, we performed global database searches using publicly available data from single bacterial genomes and different metagenomes available through NCBI GenBank. In addition, we searched several private datasets harboring human- and environmental-affiliated Bacteroidetes sequences (TABLE 1). Searches were done from January until March 2019. This global search initially resulted in the identification of nine novel potential PETases affiliated with Bacteroidetes including candidates sourced from either Seaweed [28], an Antarctic moss [28], river sediment [29], an aquaculture (unpublished data from our lab) and the human gut microbiome [30,31] (TABLE 1). Most of these candidates were affiliated with the Flavobacteriaceae genus *Aequorivita* sp. (PET27-29, PET31 and PET53). PET29 and PET31 were highly similar (< 98% identity on amino acid level)

but differed in the length of their sequence by 10 amino acids (aa). PET30, annotated as a potential lipase was derived from the published genome sequence of *Chryseobacterium jeonii* NCTC 13459. The predicted PETases PET57-59 were derived from bacteria affiliated with the genus *Porphyromonas* sp. (Porphyromonadaceae), while the predicted enzyme PET38 was derived from the species *Fluviicola taffensis* (Cryomorphaceae).

Amino acid sequence and structural analyses identify unique traits of bacteroidetal PETases.

To further resolve candidate enzymes properties, all predicted PETases were subject to more detailed bioinformatic inspection. The predicted molecular weights ranged from 36 to 48 kDa with an average of 330 aa. Remarkably, the predicted PETases PET27- PET38 showed a type IX secretion system (T9SS)/PorC type sorting domain-containing part at the C-terminus. This C-terminal domain had been described earlier by a profile HMM from the TIGRFAM database (TIGR04183). T9SS sorting domains are involved in protein transport across the bacterial outer membrane and have so far been described as a bacteroidetal-specific secretion system [32-34]. The predicted domain encompassed 62-64 aa in the cases of PET27-38. PET57-58 carried truncated sorting domains ranging from 42-55 aa in length. This observation implies that these enzymes are most likely exoenzymes (TABLE 2, and FIGUREs 1&S1).

In line with these observations, each candidate contained N-terminal signal domains for protein transport to the periplasm as predicted with SignalP 5.0 [35], further supporting the notion that these are secreted proteins (TABLE 2). Further analyses of the amino acid sequences identified a G-x-S-x-G motif which is typical for α/β serine hydrolases [36] (FIGUREs 1&S1). The catalytic triad consists of the residues Asp-His-Ser and a potential substrate binding site was identified containing the aa Phe-Met-(Trp/Tyr/Ala). The latter differed from the known IsPETase, the LCC and PET2 binding sites in which a Tyr was reported in the first position and position 3 was occupied by a Trp (TABLE 2). PET57 is the only exception with a Trp-Met-Tyr binding site.

For a more detailed structural inspection, we modeled the structures of all predicted PETases using the IsPETase (PDB code 6QGC) as backbone. These modeling experiments suggested that catalytic parts of the predicted PETases have minor differences in their 3D structures (FIGURE 1 and FIGURE S2). However, the C-terminal part affiliated with the T9SS domain differed largely. It is not present in the

IsPETase and was in some cases 100 aa in length (TABLE 2) and consisted of up to seven predicted β-sheets and, occasionally, a few α-helices.

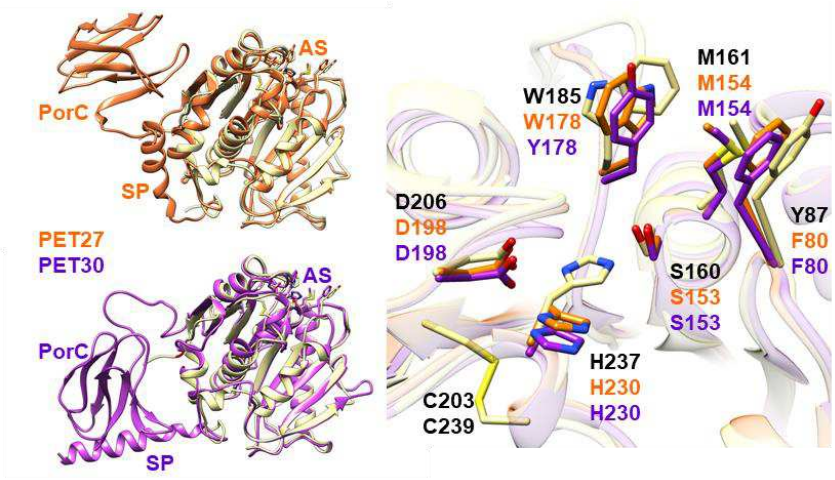

FIGURE 1: The model structures of two PETases affiliated with the Bacteroidetes phylum resemble the crystal structure of the IsPETase - with unique features. a Next to an N-terminal Signal Peptide (SP; 1x α-helix) and the core PETase domain containing the Active Site (AS), PET27 and PET30 (orange and purple, respectively) present a C-terminal PorC domain (7x β-strands) for protein secretion via the Bacteroidetes-specific Type IX Secretion System (T9SS) that is not present in the IsPETase (light yellow). **b** All three enzymes present the typical residues of Ser-hydrolases at the catalytically active positions (Ser, His and Asp), but PET27 and PET30 differ in some of the amino acids associated with PET binding. The residues of IsPETase are indicated in black. They also lack a disulfide bridge in the proximity of a catalytic loop. 3D structures were modeled using the Robetta server [37] using the IsPETase crystal structure (6QGC) as a backbone. Figure S1 provides the position of these residues in details on the amino acids level.

Recombinant PET27 and PET30 hydrolyze PET foil.

To expand on our bioinformatic analysis, we cloned and expressed the predicted PETases in *Escherichia coli* for functional testing. The nine candidate genes were synthesized and cloned into the expression vector pET21a(+) (Biomatik, Wilmington, DA, USA) and expressed in *E. coli* BL21 and T7-SHuffle (TABLE S1). The resulting clones were verified by PCR for carrying the respective insert using the primers indicated in TABLE S2. Finally, all obtained clones were sequenced using commercial services to verify the correctness of the inserts (data not shown). Initial tests using recombinant purified proteins and tributyrin (TBT)-containing agar plates indicated that the genes PET27-30 coded for active esterases. The remaining enzymes PET38, PET53, PET57 and PET58 were inactive and were either produced as insoluble proteins and/or only at very low amounts (TABLE 1). Because of these obvious difficulties affiliated with their expression, these four predicted enzymes were not further characterized.

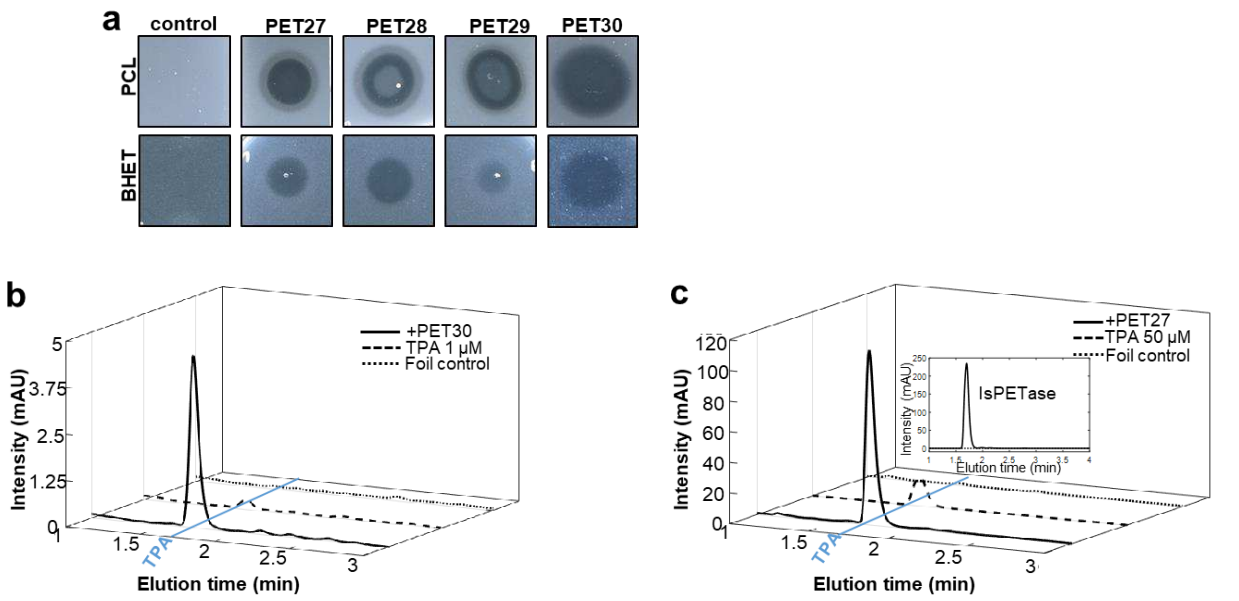

FIGURE 2: Hydrolytic activities of PET27 and PET30 on BHET, PCL and PET foil. a Activities of PET27-30 on PCL and BHET using agar plates. 10 µl containing 10-100 µg of purified enzyme were applied to agar plates containing 500 mg/l of either PCL or 5 mM of BHET. Clearing zones were observed after 12 hours. Control indicates plates without enzymes but treated with 10 µl buffer; **b** UHPLC profiles of PET30 after incubation on PET foil for 120 hours. 1 mg/ml of recombinant and purified enzymes were applied to amorphous foil and incubated over 120 hours at 30 °C. Graphs shown are representative graphs and were repeated at least three times; **c** UHPLC profiles of PET27 after incubation on PET foil for 120 hours. 1 mg/ml of recombinant and purified enzymes were applied to amorphous foil and incubated over 120 hours at 30 °C. The IsPETase was included for reasons of benchmarking in B at 1 mg/ assay (upper right corner).

Additional tests with PET27 and PET30 indicated that these enzymes hydrolyzed the esters *para*-nitrophenol- (*p*NP) hexanoate (C6) and, *p*NP-decanoate (C10, TABLE 1). All four enzymes were able to hydrolyze bis-hydroxyethyl terephthalate (BHET) and polycaprolactone (PCL) (TABLE 1, FIGURE 2a). PCL was used as a model substrate as hydrolysis of this compound indicates possible activities on the more complex PET. The recombinant enzymes produced clear halos on agar plates containing PCL or BHET after overnight incubation (FIGURE 2a, TABLE 1). Further UHPLC analyses confirmed the above findings for the enzymes PET27 and PET30 and with respect to activities on amorphous PET foil as substrate in a 200 µl reaction volume. In these tests, 1 mg ml^{-1} PET27 released 871.8 ± 200.4 µM of TPA in 120 h at 30°C (FIGURE 2c, TABLE 3). Surprisingly, under the same conditions, PET30 released only 15.9 ± 9.5 µM TPA (FIGURE 2b, TABLE 3). When we benchmarked these data with self-produced recombinant IsPETase, 1 mg ml^{-1} of IsPETase released under the same conditions 4055.7 ± 516.9 µM of TPA. Thus, the IsPETase is 4.7-fold more active compared to PET27 and approximately 253-fold more active compared to PET30.

While these data clearly demonstrate the capability of both enzymes to act on amorphous PET, the observed differences may be related to a single amino acid substitution in the predicted substrate binding pocket of PET27 and PET30 (TABLE 2). Notably, the IsPETase carries a Tyr-Met-Trp motif in the known and experimentally verified PET binding site. PET27, however, has the Tyr replaced with a Phe in its predicted binding site and PET 30 has in addition the Trp in position 3 replaced with a Tyr (TABLE 2).

Biochemical characterization and activity on esterase substrates

Because of the relatively low turnover rates observed for PET27 and PET30 on PET foil, we speculated that PET is not the preferred substrate of both enzymes and especially not for PET30. Therefore, and to further characterize and identify preferred substrates of the bacteroidetal enzymes, we characterized both enzymes in more detail. A substrate spectrum was recorded with *p*NP-esters, which had an acyl chain length of 4 to 18 C-atoms. Remarkably, PET30 revealed a relatively narrow spectrum of substrates it could degrade. The highest activities were observed with *p*NP- hexanoate (C6, FIGURE 3). Significantly lower activities were measured with short (C4) and long (C8-12) acyl chain lengths. Only very low activities were observed with *p*NP-esters with a chain length of C14-18.

The kinetic parameters for PET30 were determined with *p*NP-C6 at 30 °C and pH 8 according to Michaelis-Menten. Thereby, we observed a v_{max} of 2.3 nmol min^{-1}, a k_{cat} of 8.9 s^{-1}, a K_m of 0.3 mM and a k_{cat}/K_m value of 26,136.11 M^{-1} s^{-1}. Altogether, these findings imply that PET30 prefers short chain fatty acid-substrates rather than larger PET polymers.

Using 1 mM *p*NP- hexanoate as substrate, the recombinant enzyme PET30 revealed a remarkably broad temperature spectrum. The highest activity was observed at 30°C, but 80% of activity was observed at 20 °C and between 40 and 50 °C. Surprisingly, at 10 °C, the enzyme still showed a relative activity of 65 %. It remained active at a temperature of 4°C showing a relative activity of 42% on *p*NP- hexanoate (FIGURE 3).

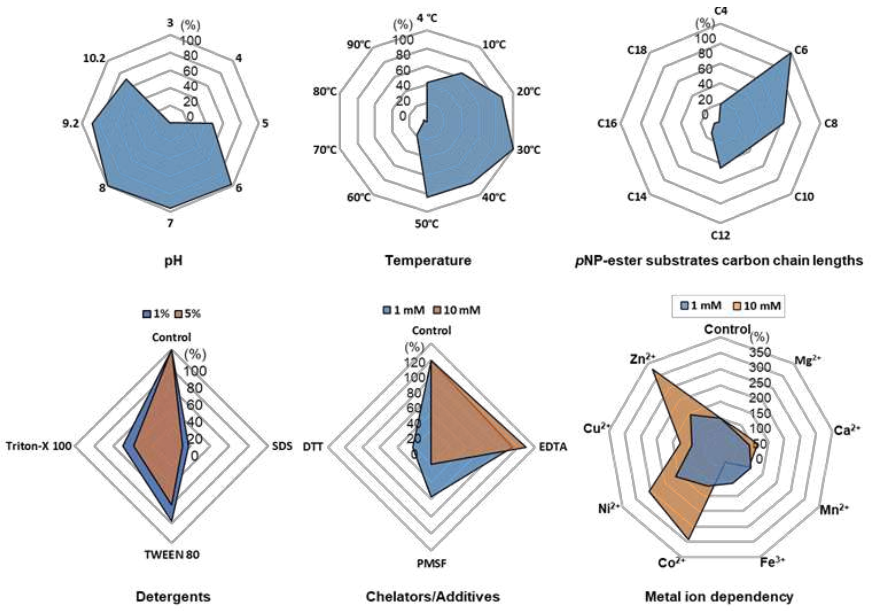

FIGURE 3: Biochemical characterization of PET30 using various *p*NP-substrates. Data represent mean values of at least three independent samples. Data obtained with a *p*NP-assay are shown in net diagrams. Substrate preferences, temperature optimum and pH optimum were tested. All tests besides substrate preferences were carried out with *p*NP-hexanoate.

To assess thermostability, the enzyme was incubated at 50 and 60 °C for 3 hours, after which the enzyme retained only 23 and 5 % of its original activity, respectively (supplementary FIGURE S3b).

PET30 was most active at pH 8.0 when tested in 0.1 M PBS and with 1 mM *p*NP- hexanoate as substrate at its temperature optimum of 30 °C. However, it also retained relatively high activities at pH 6 and 10 °C, when more than 50% residual activity was observed. To further characterize the effects of metal ions, different ions (Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mg^{2+}, Mn^{2+}, Ni^{2+} and Zn^{2+}) were added to the assays at 1 and 10 mM final concentrations. Activity was measured with *p*NP- hexanoate and compared with a metal-free control. The activity of PET30 decreased in the presence of most of these ions. However, addition of Co^{2+}, Zn^{2+} and Ni^{2+} resulted in an increase of the activity. Up to a threefold increase in activity was recorded in the presence of these metal ions (FIGURE 3).

Further, EDTA, DTT and PMSF were applied in final concentrations of 1 and 10 mM (FIGURE 3). The presence of DTT and PMSF (1 and 10 mM) inactivated PET30 almost completely. Whereas EDTA at 1 and 10 mM had no large impact on the enzyme's activity. Finally, we tested the sensitivity of PET30 towards detergents. A concentration of 1 and 5 % of the detergents Triton X-100, Tween 80 and SDS strongly affected the enzyme activities (FIGURE 3).

Finally, we asked if the C-terminal sorting domain is of importance for its catalytic activities. To answer this question, we constructed a deletion mutant designated PET30_Δ300-366 that lacked the sorting sequence. Biochemical tests implied that it was not affected in its activities using *p*NP-hexanoate or PET foil (TABLEs 2 & 3). The enzyme released similar amounts of TPA as it was observed for the wildtype enzyme (TABLE 3).

In summary, these data imply that PET30 is a promiscuous mesophilic esterase with highest activities on carboxylic esters with C6-acyl chains, which get increased in the presence Co^{2+}, Zn^{2+} and Ni^{2+}. While PET30 is a mesophilic enzyme with a temperature optimum at 30 °C, it revealed a high residual activity of 42% at 4 °C (FIGURE 3).

Additionally, a biochemical analysis was performed for PET27 (FIGURE S3a). Recombinant and purified PET27 was active on *p*NP-ester substrates with a chain length ranging from C4-C18. It preferred *p*NP-octanoate (C8) (TABLE 1; Supplementary FIGURE S3a). Using the same conditions as outlined above, PET27 revealed a v_{max} of 4.9 nmol min^{-1}, a k_{cat} of 19.08 s^{-1}, a K_m of 1.37 mM and a k_{cat}/K_m value of 13,859.27 M^{-1} s^{-1}. The enzyme was active at a broad pH spectrum ranging from 4-10. Similarly, the enzyme was active over a wide range of different temperatures. It appeared to be relatively active at lower temperatures but was also active at a temperature of 90°C. It retained 73 % of its activity at 10° compared to the optimum temperature at 40°C. Interestingly, at 90° C it still retained 45% of its activity at 40°C. The activity of PET27 increased in the presence of metal ions. Addition of Ca^{2+} resulted in a 1.5-fold increase of the activity (FIGURE S3).

The observation here that both enzymes were still active at lower temperatures raised the question whether they would also turn over PET foil at these low temperatures. Therefore, TPA release on foil was assayed at 4 °C over a time of 30 days in a 200 µl reaction volume. Under these conditions, 1 mg ml^{-1} of PET30 released an average of 6.1 µM of TPA. Interestingly, IsPETase released under the same conditions a similar amount of TPA (5.9 µM TPA). Notably, under these conditions PET27 released no detectable amounts of TPA within 30 days.

Bacteroidetal PET esterases forming two phylogenetic subclusters are globally occurring enzymes

Using the amino acid sequences of published and functionally verified PETases and employing the RAxML-NG autoMRE algorithm via TreeSAPP[38] we observed that they formed multiple phylogenetic

clusters roughly corresponding to Actinobacteria, Proteobacteria, Firmicutes and Ascomycota (FIGURE 4). Thereby our phylogenetic analysis using most of the currently known PETases suggests that the putative and now confirmed PETases from the phylum of the Bacteroidetes form a fifth and distinct cluster. All putative and confirmed enzyme candidates are grouped within this bacteroidetal but polyphyletic cluster. Interestingly, the PET-active enzyme from *Bacillus subtilis* was most closely related to the Bacteroidetes enzymes possibly indicating a horizontal gene transfer (FIGURE 4). Furthermore, the predicted but functionally not verified enzymes from the genus *Porphyromonas* (PET57-PET59) formed a subcluster within the Bacteroidetes cluster. Similarly, the two enzymes PET27 and PET30, shown to be active on PET foil, were part of a subcluster that consisted of predicted and functional enzymes affiliated with the genus *Aequorivita* and *Chryseobacterium*. Interestingly, this subcluster harbored only sequences of aquatic and environmental origin, while the cluster formed by the enzymes PET57-PET59 contained gut-affiliated sequences. Since we were only able to verify activities for the environmental cluster, it is likely that related enzymes can degrade PET-based microplastics in the environment.

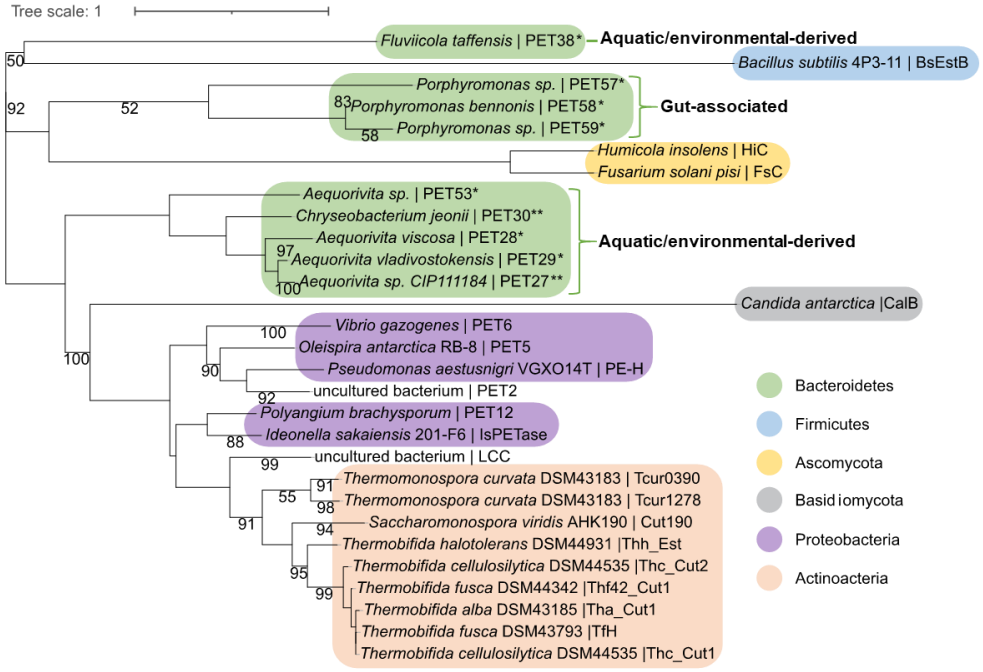

FIGURE 4: Phylogenetic tree of previously validated PETases including PETases affiliated with the Bacteroidetes phylum . The tree was constructed using the RAxML-NG autoMRE algorithm [39] with the treesapp create command implemented in TreeSAPP [38] with maximum bootstraps set at 1000 (see methods section for more details). GenBank entries of the putative and verified PETases affiliated with the Bacteroidetes phylum are listed in TABLE 1. GenBank entries and identifiers of all other PETases included in the tree are indicated in (TABLE S1).

The diversity of Bacteroidetal PETases raised the question to what extent these enzymes could impact plastic degradation in the environment. To address this question in part, we analyzed the global distribution of PET27 and PET30 and their homologues. For this purpose, we used the protein sequences of PET27 and PET30 and analyzed their occurrence and frequency in global databases available in IMG/M ER [29,40]. Using both enzymes for a BLASTp-based search (cutoffs 50% identity; 80% coverage), we were initially able to identify very few (<10) possible homologs in the global databases analyzed and affiliated with the genera *Aequorivita* and *Chryseobacterium* (FIGURE 5a). Interestingly, when we extended our search to the Flavobacterium-Cytophaga-Bacteroidetes (FCB), we were able to identify 98 possible homologs in our global searches including single cell amplified genomes (SAGs) from the Baltic Sea (TABLE S4). 47 hits were affiliated with the genus of the *Marinimicrobia*, indicating a potential role for these ubiquitous and abundant marine microorganisms in PET degradation while others were more closely associated with Bacteroidetes [41]. As expected, the majority of these homologs were associated with marine and aquatic samples (FIGURE 5b).

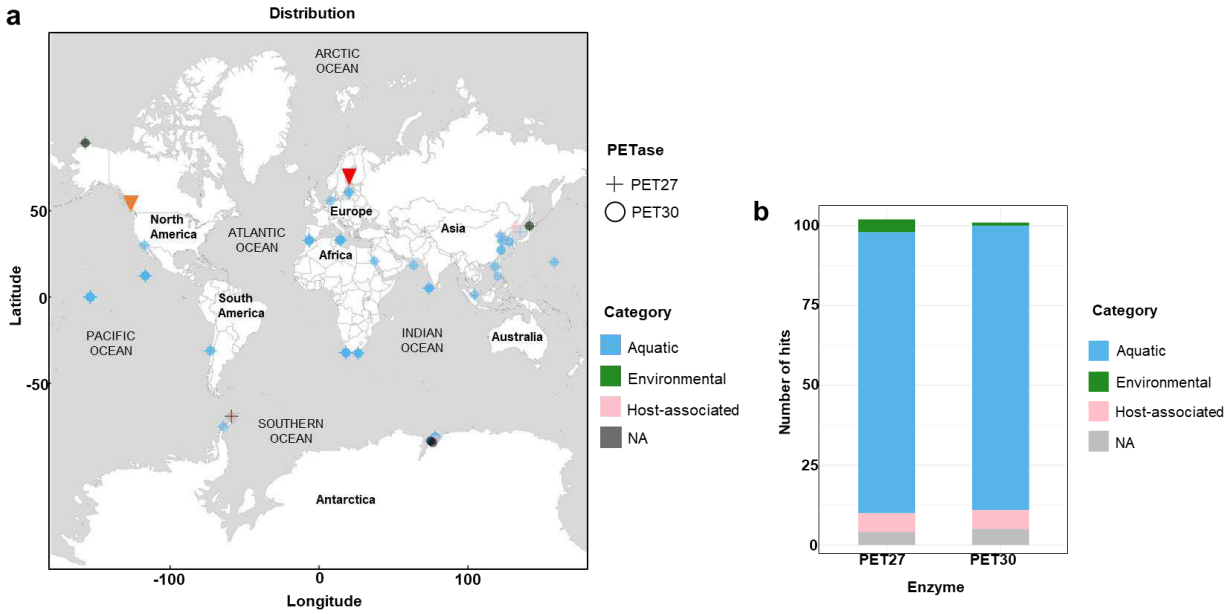

FIGURE 5: Global distribution of PET27 and PET30 homologues. a PET27 and PET30 homologues containing metagenomes were visualized on a world map containing circles for the different metagenomes. The cut off in the similarity searches was set to 50%. Data depicted include only hits to bacteria affiliated with the Flavobacteria-Cytophaga-Bacteroidetes (FCB) group. The metagenomes searched and included in this figure are listed in TABLE S2 in the supplementary material. Red and orange triangles indicate the sample sites used for metagenomic and transcriptomic analysis in the Baltic Sea and Saanich Inlet (NESAP) in FIGURE 7, respectively; **b** Number of hits observed in the same global metagenomes. Color code indicates the type of habitat: air, aquatic, terrestrial host-associated and/or engineered.

Notably, Marinimicrobia has also been named Marinisomatota in GTDB. Interestingly, many of the Saanich Inlet metagenome-assembled genomes (MAG) sequences (see FIGURE S4) at the root of the Aequorivita cluster were attributed to Marinisomatota (Marinimicrobia), suggesting that Marinisomatotal and Bacteroidetal enzymes may be interdispersed phylogenetically. Also, many of the sequences from the environmental metagenomes that originally mapped within the PET27/30 subcluster actually mapped to the same cluster, also suggesting they are very closely related, and it is difficult to resolve Bacteroidetal from Marinimicrobial sequences/enzymes.

To further evaluate the phylogenetic placement of identified PETase homologues, metagenome assembled genomes (MAGs) derived from Saanich Inlet water column contigs were mapped to the tree (FIGURE S5), and it was confirmed that hits mapping to the Bacteroidetes subclusters were primarily affiliated with Bacteroidetes MAGs (FIGURE S4, TABLE S5). Moreover, many of the sequences assigned to the *B. subtilis* PETase and a few that were mapped near the fungal sequences were also affiliated with Bacteroidetes, Actinobacterial and Proteobacterial MAGs. Additional sequences mapping to these subclusters were affiliated with Verrucomicrobiota, Marinisomatota (also known as candidate phylum Marinimicrobia), Planctomycetota, Gemmitamonidota, and Chloroflexota MAGS (FIGURE S5, TABLE S5), providing further evidence of widespread horizontal gene transfer of PETases within the marine environment.

Bacteroidetal PETases are possibly transcribed in nature

Based on these observations, we asked if expression of PET27 and PET30 homologous could be detected in metatranscriptomic data sets. Because one of our database hits was associated with a Bacteroidetes SAG from the Baltic Sea, we started searching in the Baltic Sea by mapping a selection of the recently established Baltic Sea Reference Metagenome (BARM) metagenomic sequence assemblies, or contigs, from discrete depths over the redoxcline from October 2014 [42,43]. While the majority of hits were not affiliated with the PET27-30 subcluster, one hit was recovered with high sequence similarity to the Bacteroidetes SAG (FIGURE S5a). Based on this observation we expanded our HMM search to the entire BARM data set [43] (TABLE S3). Thereby, we identified at least 9 hits of potential bacteroidetal PETase genes with a bit score of above 200 and an e-value smaller than 1.5E-63, which also showed expression in several water depths in a separate Baltic Sea metatranscriptome data set from July 2015. For one gene (gene id:

k99_34994859_1, TABLE S4) which showed high expression values at several stations and water depths, an example of the expression profile is shown in FIGURE 6a. At the time of sampling, the area was still affected by a major Baltic inflow event from end of 2014 which transported water with higher salinity and oxygen concentration into the deep anoxic basins [44]. This resulted in elevated oxygen concentrations in the bottom water of the Gotland basin (FIGURE 6a). Remarkably, PETase expression peaked in the water depths where oxygen concentration increased, both above and below the anoxic, sulfidic zone, reaching up to 3-14 x 10^5 transcripts L^{-1} (FIGURE 6a).

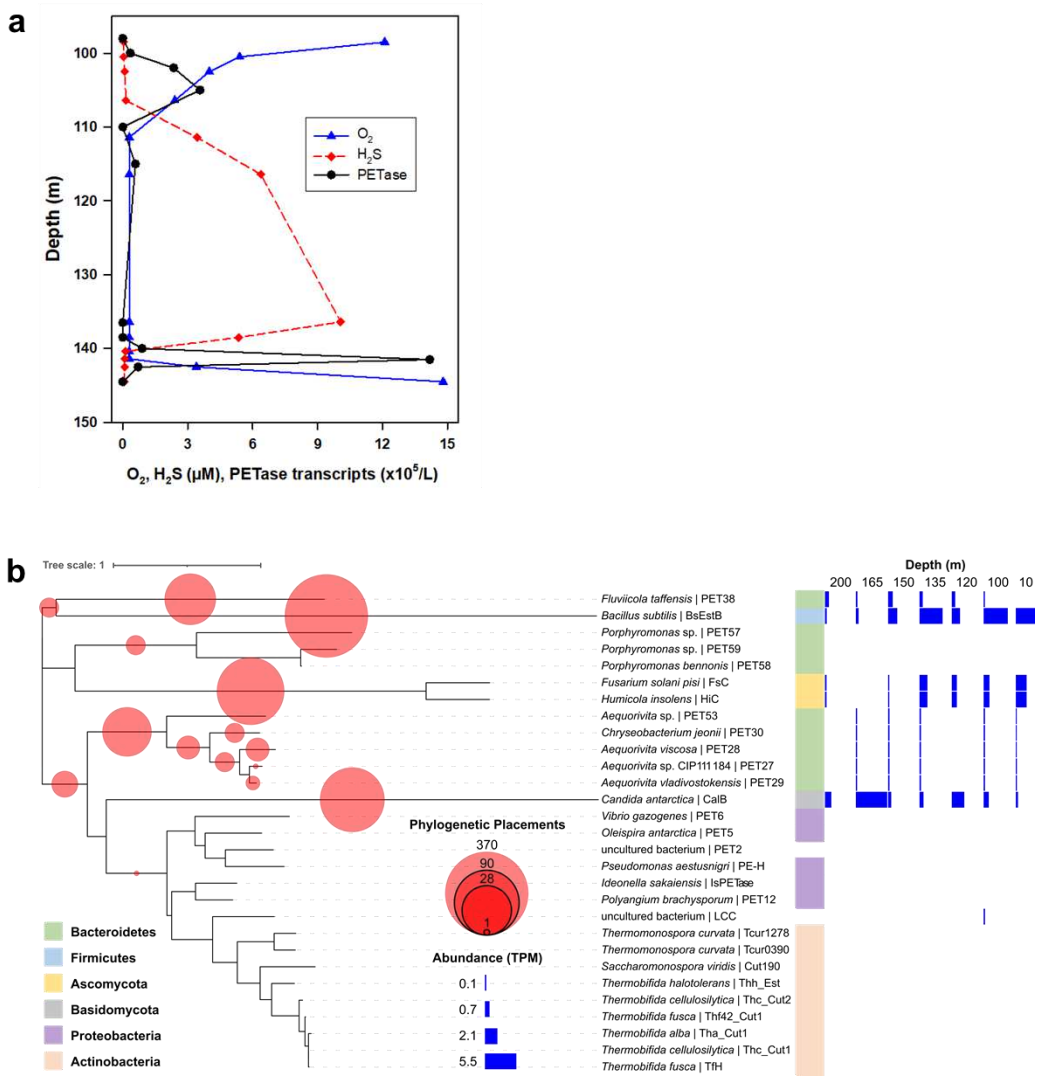

FIGURE 6: Abundance and expression of PET27 and PET30 homologues along defined redox gradients in marine samples. **a** Water column profile (showing the 50 m above the sea floor) of oxygen, sulfide and transcripts of PETase homologue k99_34994859_1 at station Trkl04 (57°6.01'N, 19°36.64'E) on 22 July 2015 in the central Gotland basin (Baltic Sea) and using samples collected during the cruise EMB107 of the German R/V "Elisabeth Mann Borgese"; **b** Sequences from metagenomic sequence assemblies and metatranscriptomic read abundances at seven depths spanning the redoxcline of Saanich Inlet during Cruise 72 (SI072) on 12 August 2012 at station SI03 (latitude: 48.59166667°, longitude: -123.505°) mapped to the phylogenetic tree of 29 known

and predicted PETase sequences shown in FIGURE 4. Light red circles represent the phylogenetic position and relative number of sequence hits from the concatenated set of SI072 metagenomic contigs; blue bars indicate the relative abundance (TPM) of metatranscriptomic reads matching metagenomic sequence hits from contigs at the cognate depth. Tree construction, sequence assignment, and abundance calculations performed using TreeSAPP and visualized with iTOL.

We went on to expand our gene-centric exploration using both the BARM data set and previously published metagenomic and metatranscriptomic data sets from Saanich Inlet, a seasonally anoxic fjord off the East Coast of Vancouver Island, British Columbia in the Northeast Subarctic Pacific (NESAP) ocean [45,46] using TreeSAPP [38] (FIGURE 6b). Mirroring the methods for the BARM redoxcline, metagenomic assemblies from discrete depths spanning the redoxcline of Saanich Inlet from August 2012 were also mapped to PETase tree (FIGURE 4), and then both metagenomic (FIGURE S5b) and metatranscriptomic read abundance (FIGURE 6b) was calculated for at each corresponding depth. Again, numerous PETase homologues were identified with the majority of sequences mapping to the known PETase from *Bacillus subtilis* and several fungal enzymes (FIGURE S5b, FIGURE 6b). Notably however, ten hits mapped to Bacteroidetes subclusters, with two placed within the PET27-PET30 subcluster (FIGURE S5b, FIGURE 6b). Transcripts corresponding to PET27 and PET30 homologues were detected at five of seven depths peaking at 150 m, though at a relatively low frequency (FIGURE S5b, FIGURE 6b). Similar to what was found in the Baltic Sea *e.g.* lack of expression of PET27 and PET30 homologues at 200 m corresponded to increased sulfide concentration at this depth [46,47]. In contrast, transcripts affiliated with the *B. subtilis* and fungal PETases, especially that for *Candida antarctica*, were identified at 200 m pointing to potential activity under anaerobic conditions (FIGURE 6b).

DISCUSSION

Today's global plastic pollution problem makes the study of plastic degrading microorganisms and enzymes integral to the development of biotechnological solutions. Currently, there are only a handful of known bacterial phyla encoding active PET esterases (FIGURE 4, and references TABLE S3).

Here, we have identified and partially characterized two novel functional PETases affiliated with *Chryseobacterium* and *Aequorivita* genera within the Bacteroidetes phylum. Bacteria affiliated with the genus *Chryseobacterium* are globally occurring aerobic organisms colonizing a wide range of different habitats including plants, soil, fish, human gut and sea water. Within the genus *Chryseobacterium,* over one hundred species have been described of which few are pathogens, but many are beneficial and host-

associated [48,49]. Only a few species have been identified within the genus *Aequorivita*, mainly belonging to marine or fresh-water organisms that are mostly psychrotolerant and aerobic [50]. Notably, Bacteroidetes have been described as very potent degraders of polymers and they harbor a multitude of hydrolases and binding modules [27,51,52].

The two enzymes, which we have characterized, are both typical esterases (*i.e.* serine hydrolases) belonging to the EC 3.1. Both appear to be secreted enzymes as they carry an N-terminal secretion signal and one secretion PorC-like motif. The N-terminal secretion signal is linked to the transport into the periplasm [53] and the PorC-like domain is part of the type IX secretion system (T9SS) [33,34]. The T9SS is composed of several outer membrane, periplasmic and inner membrane proteins, whereby it is affiliated with the secretion of pathogenicity factors, hydrolases but also with gliding motility in the Bacteroidetes phylum [33,34].

PET27 and PET30 were active on PET foil but differed strongly in their overall activities. We speculate that these different activities are the result of the exchange of one amino acid residue in the substrate binding sites (TABLE 2). Notably, PET27 reveals a Phe-Met-Trp motif and PET30 a Phe-Met-Tyr. The most active enzymes such as LCC and IsPETase both carry a Tyr-Met-Trp consensus binding motif. Thus, these marked changes may in part explain the different overall activities. The non-active enzymes PET38, PET53, PET57 PET58 and PET59 revealed either a Phe-Met-Ala, a Phe-Met-Trp or Trp-Met-Tyr substrate binding motif possibly explain their lack of activities.

Benchmarking activities of polymer active enzymes with literature values is not trivial since most studies use different types of foils with different degrees of crystallinity and distinct assay conditions. To partially overcome this challenge, we produced our own recombinant wildtype enzymes of the IsPETase and compared its activities with PET27 and PET30. As expected, IsPETase was 4.7-fold more active at 30°C than PET27 and up to 253-fold more active than PET30. With respect to the overall activity of the IsPETase, however, our data are in line with published data for this enzyme [54]. The observation here that the activities of the PET27 and PET30 enzymes are relatively low compared to the IsPETase and certainly with respect to the published values of the even more active LCC may imply that PET27 and PET30 are no PET esterases *in sensu strictu.* However, our data imply that both are short-chain fatty acid acting esterases revealing some promiscuity in their substrate profile (TABLE 2, and FIGURE 3). Notably, esterases are well known to be promiscuous enzymes that can convert many different substrates. The substrates not fitting

well into the active site or the binding site are turned over at lower rates compared to the substrates that fit best [55-57].

Intriguingly, the observations that both enzymes were catalytically active on PET foil implies perhaps a wider role in the degradation of PET and especially PET nanoparticles. Because of the significant activities even at 4 °C, these enzymes may in fact play a heretofore unknown role in PET microparticle degradation, also in cold environments. This hypothesis is supported by our observation that homologs of both enzymes can be found on a global level covering a wide range of climate zones (FIGURE 5) and the observation of transcripts in a cold climate marine environment such as the Baltic Sea and Saanich Inlet (FIGURE 6). The transcripts of these genes are likely derived from diverse phylogenetic groups that have not been previously linked to PET degradation. This observation indicates that water column PET degradation is possible and likely occurring throughout the global ocean on spatiotemporal scales that remain to be determined. In addition to these data, we provided first evidence that that few of the PET27 and PET30 homologs were actively expressed in nature. This observation may indicate that PET degradation in the marine environment is possible and occurring. It, however, does not yet allow us to estimate the global turnover rate.

In summary, our biochemical results significantly extend the knowledge of PETase enzymes and provides promising candidates for biotechnological applications at low temperatures. Furthermore, the data presented here will help to advance our knowledge on the ecological role of the Bacteroidetes in the decomposition of marine PET litter and enable the development of an expanded phylogenetic framework for identifying the diversity of putative PETases in diverse marine microbial groups throughout the global ocean.

MATERIALS AND METHODS

Bacterial strains, plasmids and primers

Bacterial strains, plasmids and primers used in this study are listed in TABLE S1 and TABLE S2. 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 and confirmed PETases 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/) servers [58-60]. Human gut sequences were retrieved from the Unified Human Gastrointestinal Protein (UHGP) catalog (PMID:32690973). Sequences were compared to others deposited in the NCBI databases using BLAST alignment tools [61]. Amino acid sequence HMM search was carried out using the HMMER (http://hmmer.org) webpage or a local version of the software (v3.1b2) [62] with downloaded datasets. Structural information on the enzymes was retrieved from the RCSB-PDB [63] database.

Sequence data were processed and analyzed using ChromasPro 2.1.8 (Technelysium, Brisbaine Australia) or SnapGene (GSL Biotech LLC, San Diego CA, USA). Amino acid alignment was constructed using structural alignments with T-Coffee [64] and was further visualized with Bioedit [65]. The model structures of bacteroidetal PETases were modeled with the Robetta server [37] using the IsPETase crystal structure (6QGC) as a backbone. A phylogenetic tree was constructed using the RAxML-NG autoMRE algorithm [39] with the treesapp create command implemented in TreeSAPP [38] with maximum bootstraps set at 1000. RAxML-NG has recently to return the best scoring tree for highest number of datasets when compared against other fast maximum likelihood (ML) methods [39], allowing a large number of maximum bootstraps to be used to produce as conservative a tree as possible. Sequences were assigned NCBI lineages according to source organisms listed in TABLE 1 and TABLE S1, and colors were assigned to the tree at the phylum level using the treesapp colour command. This reference tree was also built for and used with TreeSAPP-based to analyses of marine metagenomic and metatranscriptomic data sets to gain an understanding of Bacteriodetal PETase potential role in the environment. The UHPLC profiles were plotted and edited using MATLAB version R2020b. Scanning IMG/M was done on 19/November/2020 for PET30 and on 14/January/2021 for PET27. Geo locations were used as provided whenever available. In case no Geo location was available, whenever possible, information about isolation source/location/city/country were used to look up Geo coordinates on GeoHack (https://geohack.toolforge.org). The map representing the frequency and geographical distribution of PET hydrolases in metagenomes (FIGURE 5) was constructed using QGis Desktop 2.18.5 (http://www.qgis.org).

436 BARM Redoxcline metagenomic sequence assemblies from samples collected on October 26, 2014
437 at TF0271 sampling station were downloaded from EMBL-EBI under the study accession number
438 PRJEB34883 [42]. The corresponding metagenomic reads were obtained from EMBL-EBI under the study
439 accession number PRJEB22997 [43]. Four of eight samples from depths of 100 m, 120 m, 140 m, and 200 m
440 (samples P2236_102, P2236_104, P2236_106, and P2236_108) were chosen from the BARM Redoxcline
441 dataset as they most closely mirrored the Saanich Inlet data also used for environmental analysis described
442 below. For the transcriptome analyses of the Baltic Sea, water samples from the central Baltic Sea were
443 taken during cruise EMB107 of the German R/V "Elisabeth Mann Borgese" in July 2015 with an automatic
444 flow injection sampler (AFIS) attached to a conductivity-temperature-depth (CTD)-rosette. The *in-situ*
445 fixation by the AFIS system has been shown to provide relatively unbiased gene expression data from
446 oxygen-deficient water layers [66]. RNA extraction and further processing were described previously [67]. For
447 downstream absolute quantification of RNA transcripts, standards were added before the start of RNA
448 extraction [68]. Bioinformatic analyses followed previously published protocols [67] except that
449 metatranscriptome reads were mapped onto the BARM metagenome [43] with very sensitive settings. The
450 recovery rate of the internal standard molecules in the sequencing data reads was used to calculate the
451 absolute number of transcripts that were in the respective water samples [68,69]. The above described HMM
452 search was used to identify potential bacteroidetal PETase genes in the BARM metagenome and compared
453 to those which showed expression in the Baltic Sea metatranscriptome data set.
454 Water samples were collected from seven discrete depths spanning the redoxcline of Saanich Inlet
455 from sampling station S3 (SI03; latitude: 48.59166667°, longitude: -123.505°) during Cruise 72 (SI072) on
456 August 12, 2012 as part of an ongoing oceanographic time series program. DNA and RNA was extracted
457 from the 0.22 µm Sterivex filters with the biomass from each depth and Illumina metagenomic and
458 metatranscriptomic shotgun libraries were generated and paired end sequenced at the Joint Genome
459 Institute as described previously [70].
460 The resulting forward and reverse reads were processed, quality controlled and filtered with
461 Trimmomatic (v.0.35) [71]. Trimmomatic evoked using the trimmomatic command applied to each of the seven
462 depths with -phred33 flag and the following parameters: ILLUMINACLIP:…/Trimmomatic-
463 0.35/adapters/TruSeq3-PE.fa:2:3:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. The
464 resulting trimmed FASTQ files were used in contig assembly and to calculate read abundance with

TreeSAPP. MEGAHIT (v.1.1.3) [72] was used to assemble the filtered metagenomic reads into contigs. MEGAHIT used via the megahit command applied with the following parameters: -m 0.5 –k-min 27 –k-step 10 –min-contig-len 500.

MetaWRAP (v1.2.4) [73] was used to generate metagenome-assembled genomes (MAGs) from the assemblies and filtered reads. MetaWRAP leverages multiple binning software (we chose to use MetaBAT2 (v2.12.1) [74] and MaxBin (v2.2.7) [75]) to create a non-redundant set of MAGs that are better quality than those from any single software. The quality of the resulting bins – assessed by their completeness, contamination, and strain-heterogeneity – was calculated with MetaWRAP's implementation of CheckM (v1.0.12) [76]. The metawrap binning command was used to invoke MetaWRAP on the assembled contigs with the following parameters: -m 64 -l 1500 --metabat2 --maxbin2. The resulting MAGs were then passed through GTDB-Tk(v1.4.0) classify workflow with the reference data version r95 [77] via the gtdbtk classify_wf command. After updating the headers for resulting 219 bins with the sample IDs, the medium plus quality (completeness >50%, contamination <10%) bins were concatenated together into one FASTA file to use with TreeSAPP [38].

For metagenomic and metatranscriptomic analyses of the BARM and Saanich redoxcline data sets, TreeSAPP (v11.0) [38] was used to create a PETase reference package (FIGURE 4) from 29 PETase protein sequences (TABLE 1, TABLE S1) in FASTA format. The reference package includes a profile HMM, a multiple sequence alignment, a phylogenetic tree (described above) and a taxonomy table. For validation, the original 29 sequences were assigned back to the tree to visualize reference package construction in iTOL [78] with treesapp assign using the –trim_align option. To phylogenetically place PETase homologs recovered from the entire SI072 metagenome and the selected BARM redoxcline metagenome, the concatenated FASTA of contig assemblies from all depths for each data set was mapped to the reference tree using TreeSAPP (v11.0) [38] using the treesapp assign command with the –trim_align flag. Metagenomic contigs were then assigned separately by depth alongside their corresponding metagenomic and/or transcriptomic reads in FASTQ format using treesapp assign with the –abundance option to calculate corresponding transcript abundance (transcripts per million, TPM) for metagenomic hits at each depth. SI072 medium plus quality MAG sequences were placed in the reference tree using the same methodology as for the contigs described above to better understand and interpret phylogenetic placements from Saanich Inlet. All trees were then visualized in iTOL [78]. It is important to note TreeSAPP abundance calculations as

visualized in iTOL maps read abundance at the tips of the reference tree. If a sequence maps closer to the root, that abundance gets split evenly among the children in that cluster, and abundances shown represent a combination of reads that mapped to all nodes feeding into that particular leaf. Multiple sequences can also be placed at the same location in the tree, and reads mapping to all sequences will contribute to the abundance calculation.

Heterologous expression of putative PET esterase genes in Escherichia coli BL21 (DE3)

The putative PETases were extracted from metagenomic datasets (Table 1), therefore the gene sequences were optimized for expression in *E. coli* and synthesized into pET21a(+) vector at Biomatik (Wilmington, USA). The obtained constructs were sequenced at Eurofins (Ebersberg, Germeny) and checked for correctness by comparing to the original sequences. Chemical competent *E. coli* T7-Shuffle or *E. coli* BL21(DE3) were used for heterologous expression of possible PETases. IsPETase gene in pMAL-p4x was provided by Sebastian Weigert (University of Bayreuth, Germany) and purified by maltose-binding Tag. The cultures were grown aerobically in auto-induction medium (ZYM-5052) [79] containing 100 µg/ml ampicillin for pET21a(+) at 37 °C until they reached an OD_{600} of 1.0. The proteins were expressed afterwards at 22 °C for 16-20 h harboring an 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 PET27 and PET30

For activity tests, both enzymes were assayed using purified recombinant protein. Unless otherwise indicated, a total amount of 0.1 µg to 1 µg of the enzymes were added to a substrate solution containing 190 µl of either 0.2 M sodium phosphate buffer or 0.1 M potassium phosphate with a defined pH between 7 and 8 and 10 µl of 0.1 mM *p*NP-substrate dissolved in isopropanol. After incubating the samples for 10 min, the assay was stopped by adding 200 mM of Na_2CO_3. Afterwards, the samples were centrifuged at 4 °C, 13,000 rpm for 3 min. As substrates, we tested *p*NP-esters with chain lengths of C4, C6, C8, C10, C12, C14, C16 and C18. After incubation at defined temperatures, 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. To determine the optimal temperature, samples were incubated between 10 °C and 90 °C for 10 min. The influence 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 impact of cofactors, solvents, detergents, and inhibitors was assayed at different concentration levels. The possible cofactors Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mg^{2+}, Mn^{2+}, Rb^{2+} and Zn^{2+} with a final concentration of 1 and 10 mM were used. 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. After 1 h incubation in the presence of these substances, the residual activity was determined after 10 min incubation at the optimal temperature with *p*NP-hexanoate and at the optimal pH.

For the verification of enzymatic PET hydrolysis, a 7 mg platelet of low-crystallinity PET film (Goodfellow GmbH, Bad Nauheim, Germany) was folded in half and used as substrate together with 200 µg of enzyme in 200 µl of 100 mM potassium phosphate buffer at pH 8.0. Incubation was carried out under continuous shaking at 400 rpm in 1.5 ml microcentrifuge tubes at 30°C if not stated otherwise.

Analysis of breakdown products was performed with an UltiMate™ 3000 UHPLC system from Thermo Scientific (Waltham, MA, USA) using a Triart C18 column (YMC Europe GmbH, Dinslaken, Germany) with a dimension of 100 × 2.0 mm containing particles with 1.9 µm diameter. Isocratic elution was performed using a mobile phase consisting of 20:80 (v/v) acetonitrile and water (acidified with 0.1% vol trifluoroacetic acid) at a flowrate of 0.4 ml min^{-1}. UHPLC samples were prepared by mixing 50 µl of incubation supernatant with 200 µl acetonitrile (acidified with 1% vol trifluoroacetic acid), followed by centrifugation at 10,000 × g for 3 minutes and transferring 200 µl of the supernatant into 600 µl water. 15 µl of sample were injected per measurement and detection was performed at 254 nm with a VWD-3400 detector from Thermo Scientific (Waltham, MA, USA). Quantification of peak areas was performed using data analysis software supplied with the Compass HyStar software package from Bruker (Billerica, MA, USA).

Data availability

DNA sequences of the identified and in part functionally verified Bacteroidetes affiliated esterases (TABLE 1) are all available at GenBank or MGnify: WP_111881932 WP_073216622 WP_052671284 WP_039353427 WP_083800582.1/ GCA_000194605.1, k99_709705_13, GUT_GENOME137663_00143, GUT_GENOME065712_01381 GUT_GENOME243617_00165. BARM Redoxcline metagenomic sequence assemblies are available from EMBL-EBI under the study accession number PRJEB34883 [42] and metagenomic reads under the study accession number PRJEB22997 [43]. Saanich Inlet metagenomic and metatranscriptomic data sets are available via the JGI IMG/M portal (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) under the study name "Marine microbial communities from expanding oxygen minimum zones in the northeastern subarctic Pacific Ocean" [45,46].

ACKNOWLEDGEMENTS

This work was in part supported by the BMBF within the programs MarBiotech (031B0562A) MetagenLig (031B0571B), MethanoPEP (031B0851B) LipoBiocat (031B0837B) and PlastiSea (031B867B) MetagenLig (031B0571A) at the Universities of Hamburg, Kiel and Stuttgart. Additional support was derived from the University of Bayreuth funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 391977956 – SFB 1357, the US Department of Energy (DOE) Joint Genome Institute, an Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, the Natural Sciences and Engineering Research Council (NSERC) of Canada, the G. Unger Vetlesen and Ambrose Monell Foundations, the Canada Foundation for Innovation (CFI) and Compute Canada through grants awarded to S.J.H.. A.A. is funded by EMBL core funds.

Author contribution statement

W.R.S., J.C. and P.P.G. designed the study, coordinated manuscript writing and bioassays. H.Z., R.D., C.V. contributed to planning, writing and data collection. P.P.G., J.C. and H.Z. were involved in enzyme structural work, bioinformatic and initial phylogenetic analyses. R.A.S and C.C were involved in global data base searches. P.C.F.B. and J.P. were involved in NCBI and IMG data mining. A.A. was involved in gut microbiome data mining. B.H. and S.W. contributed to structural predictions and biochemical analyses. S.H, T.S., S.S. and K.J. coordinated and contributed transcriptome data and mining in environmental data sets

and S.S. and S.H. were involved in phylogenetic analyzes. All authors contributed to manuscript writing and editing.

Competing interests

The authors declare to have no financial or non-financial conflict of interest.

TABLES

TABLE 1: Key traits of predicted bacteroidetal PET esterases. TBT, tributyrin; BHET, bis-hydroxyethyl terephthalate; PCL, polycaprolactonate; pNP-C6/C10, para-nitrophenyl esters with chain length C6 or C10; aa, amino acids; MW, molecular weight. N.C. not determined. +, active; -, not active. PET57-59 were extracted from the gut genomes available at: https://www.ebi.ac.uk/metagenomics/genomes/MGYG-HGUT-01059 (PET57); https://www.ebi.ac.uk/metagenomics/genomes/MGYG-HGUT-01060 (PET58) and https://www.ebi.ac.uk/metagenomics/genomes/MGYG-HGUT-00764 (PET59).

Predicted PETase	GenBank entry / MGY identifier	Phylogenetic Affiliation	aa / MW	Derived from	Expression level / solubility	Active on						
						TBT	*p*NP-C6	*p*NP-C10	PCL	BHET	PET-foil	PET particles
PET27	WP_111881932	*Aequorivita sp. CIP111184*	364 / 37.8	Antarctic source (28)	High / majority in inclusion bodies	+	+	+	+	+	+	+
PET28	WP_073216622	*Aequorivita viscosa*	365 / 38.3	Seaweed (28)	High / majority in inclusion bodies	+	+	+	+	+	-	-
PET29	WP_052671284	*Aequorivita vladivostokensis*	365 / 39.3	Troitsa bay, Sea of Japan (28)	High / majority in inclusion bodies	+	+	+	+	+	-	-
PET30	WP_039353427	*Chryseobacterium (Kaistella) jeonii*	366 / 37.4	Antarctic moss (28)	High / majority soluble	+	+	+	+	+	+	+
PET38	WP_083800582.1/ GCA_000194605.1	*Fluviicola taffensis*	447 / 40.4	River, UK (29)	Low	-	-	-	-	-	-	-
PET53	k99_709705_13	*Aequorivita* sp.	294 / 37.8	Marine aquaculture fish tank metagenome/unpublished data University of Hamburg	Low	-	-	-	-	N.D.	N.D.	N.D.
PET57	GUT_GENOME137663_00143	*Porphyromonas* sp.	323 / 36.3	Human gut (30, 31)	High / majority soluble	-	-	-	+	N.D.	N.D.	N.D.
PET58	GUT_GENOME065712_01381	*Porphyromonas bennonis*	338 / 37.6	Human gut (30, 31)	High / majority in inclusion bodies	-	-	-	-	N.D.	N.D.	N.D.
PET59	GUT_GENOME243617_00165	*Porphyromonas* sp.	345/ 38.4	Human gut (30, 31)	High / majority soluble	-	-	-	-	N.D.	N.D.	N.D.

TABLE 2: Conserved motifs and structural features identified in the predicted bacteroidetal PET esterases. The *Ideonella sakaiensis* PETase (IsPETase, PDB: 6EQE; [18,41]), the LCC (4EB0; [17]) and PET2 [19]) were included for benchmarking purposes. SP: Signal Peptide; α, α -helix; ß, ß-sheet; N/A, not identified; *, verified and predicted disulfide bonds; PorC, Por secretion system C-terminal sorting domain.

	N-terminus						C-terminus			
Predicted PETase	**Alignment 1st aa**	**Length [N]**	**SP cleavage site**	**Catalytic triad**	**Substrate binding site**	**Disulfide bonds***	**Alignment last aa**	**Length [N]**	**Secondary structure**	**Conserved domain**
IsPETase	T39	39	27-28	Asp-His-Ser	Tyr-Met-Trp	2x	A273	17	N/A	N/A
LCC	D53	53	21-22	Asp-His-Ser	Tyr-Met-Trp	1x	C275	18	N/A	N/A
PET2	S60	60	27-28	Asp-His-Ser	Tyr-Met-Trp	2x	C289	19	N/A	N/A
PET27	T29	29	23-24	Asp-His-Ser	Phe-Met-Trp	N/A	T266	98	7xβ	PorC
PET28	T29	29	23-24	Asp-His-Ser	Phe-Met-Trp	N/A	T266	99	6xβ	PorC
PET29	T29	29	23-24	Asp-His-Ser	Phe-Met-Trp	N/A	T266	99	α, 4xβ, α, 2xβ	PorC
PET30	T29	29	23-24	Asp-His-Ser	Phe-Met-Tyr	N/A	T267	99	7xβ	PorC
PET38	S7	7	19-20	Asp-His-Ser	Phe-Met-Ala	1x	T292	167	loops, 5xβ, α, 2xβ	PorC
PET53	T29	29	22-23	Asp-His-Ser	Phe-Met-Trp	1x	L269	85	4xβ	N/A
PET57	E27	27	25-26	Asp-His-Ser	Trp-Met-Tyr	N/A	K290	33	α + loops	N/A
PET58	E26	26	24-25	Asp-His-Ser	Phe-Met-Tyr	N/A	V294	44	loops + α	N/A
PET59	D41	41	24-25	Asp-His-Ser	Phe-Met-Tyr	N/A	F295	50	α + semi-α	N/A

TABLE 3: TPA (µM) released by different PET active enzymes. The different recombinant and purified enzymes were incubated at a concentration of 1mg x ml-1 for a time period of 120 h at 30°C. For the tests a circular piece of PET foil (5mm, and as specified in material and methods) was employed and folded once in the middle. Incubations were carried out in a reaction volume of 200 µl. Data are mean values with standard deviations of a minimum of 3 and up to 6 measurements per sample.

Enzyme	**Released TPA [µM]**
PET30	15.9 ± 9.48
PET27	871.8 ± 200.4
IsPETase	4,055.7 ± 516.9
PET30_Δ300-366	23.3 ± 9.2

REFERENCES

1 Plasticseurope. PlasticsEurope, Plastics—The Facts 2018 : An Analysis of European Plastics Production, Demand and Waste Data (PlasticsEurope, 2018. (2018).

2 plasticsinsight. *Global PET resin production capacity*, <https://www.plasticsinsight.com/global-pet-resin-production-capacity> (2016).

3 Jambeck, J. R. *et al.* Plastic waste inputs from land into the ocean. *Science* **347**, 768-771, doi:10.1126/science.1260352 (2015).

4 Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Science Advances* **3**, e1700782, doi:10.1126/sciadv.1700782 (2017).

5 Sang, T., Wallis, C. J., Hill, G. & Britovsek, G. J. P. Polyethylene terephthalate degradation under natural and accelerated weathering conditions. *European Polymer Journal* **136**, 109873, doi:https://doi.org/10.1016/j.eurpolymj.2020.109873 (2020).

6 Amaral-Zettler, L. A., Zettler, E. R. & Mincer, T. J. Ecology of the plastisphere. *Nature Reviews Microbiology* **18**, 139-151, doi:10.1038/s41579-019-0308-0 (2020).

7 Kirstein, I. V., Wichels, A., Gullans, E., Krohne, G. & Gerdts, G. The Plastisphere – Uncovering tightly attached plastic "specific" microorganisms. *PLOS ONE* **14**, e0215859, doi:10.1371/journal.pone.0215859 (2019).

8 Wei, R. & Zimmermann, W. Microbial enzymes for the recycling of recalcitrant petroleum-based plastics: how far are we? *Microbial Biotechnology*, n/a-n/a, doi:10.1111/1751-7915.12710 (2017).

9 Danso, D., Chow, J. & Streit, W. R. Plastics: Microbial Degradation, Environmental and Biotechnological Perspectives. *Appl Environ Microbiol*, doi:10.1128/AEM.01095-19 (2019).

10 Acero, E. H. *et al.* Enzymatic surface hydrolysis of PET: Effect of structural diversity on kinetic properties of cutinases from *Thermobifida*. *Macromolecules* **44**, 4632-4640, doi:10.1021/ma200949p (2011).

11 Kleeberg, I., Hetz, C., Kroppenstedt, R. M., Muller, R. J. & Deckwer, W. D. Biodegradation of aliphatic-aromatic copolyesters by *Thermomonospora fusca* and other thermophilic compost isolates. *Applied and Environmental Microbiology* **64**, 1731-1735 (1998).

12 Hu, X., Thumarat, U., Zhang, X., Tang, M. & Kawai, F. Diversity of polyester-degrading bacteria in compost and molecular analysis of a thermoactive esterase from *Thermobifida alba* AHK119. *Appl Microbiol Biotechnol* **87**, 771-779, doi:10.1007/s00253-010-2555-x (2010).

13 Wei, R., Oeser, T. & Zimmermann, W. Synthetic Polyester-Hydrolyzing Enzymes From Thermophilic Actinomycetes. *Advances in Applied Microbiology, Vol 89* **89**, 267-305, doi:10.1016/B978-0-12-800259-9.00007-X (2014).

14 Ribitsch, D. *et al.* A New Esterase from Thermobifida halotolerans Hydrolyses Polyethylene Terephthalate (PET) and Polylactic Acid (PLA). *Polymers* **4**, 617-629, doi:10.3390/polym4010617 (2012).

15 Chen, S. *et al.* Identification and characterization of bacterial cutinase. *J Biol Chem* **283**, 25854-25862, doi:10.1074/jbc.M800848200 (2008).

16 Sulaiman, S. *et al.* Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Appl Environ Microbiol* **78**, 1556-1562, doi:10.1128/AEM.06725-11 (2012).

17 Sulaiman, S., You, D. J., Kanaya, E., Koga, Y. & Kanaya, S. Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. *Biochemistry* **53**, 1858-1869, doi:10.1021/bi401561p (2014).

18 Yoshida, S. *et al.* A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**, 1196-1199, doi:10.1126/science.aad6359 (2016).

19 Danso, D. *et al.* New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. *Appl Environ Microbiol* **84**, doi:10.1128/AEM.02773-17 (2018).

20 Bollinger, A. *et al.* A Novel Polyester Hydrolase From the Marine Bacterium Pseudomonas aestusnigri - Structural and Functional Insights. *Front Microbiol* **11**, 114, doi:10.3389/fmicb.2020.00114 (2020).

21 Haernvall, K. *et al.* Polyol Structure Influences Enzymatic Hydrolysis of Bio-Based 2,5-Furandicarboxylic Acid (FDCA) Polyesters. *Biotechnol J* **12**, doi:10.1002/biot.201600741 (2017).

22 Ronkvist, Å. M., Xie, W., Lu, W. & Gross, R. A. Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate). *Macromolecules* **42**, 5128-5138, doi:10.1021/ma9005318 (2009).

23 Hahnke, R. L. *et al.* Genome-Based Taxonomic Classification of Bacteroidetes. *Front Microbiol* **7**, 2003, doi:10.3389/fmicb.2016.02003 (2016).

24 Krieg, N., Ludwig, W., Euzéby, J. & Whitman, W. in *Bergey's Manual of Systematics of Archaea and Bacteria* (eds M.E. Trujillo *et al.*) 1-2 (2015).

25 Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, 593-621, doi:10.1128/cmr.00008-07 (2007).

26 Munoz, R., Rosselló-Móra, R. & Amann, R. Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov. *Syst Appl Microbiol* **39**, 281-296, doi:10.1016/j.syapm.2016.04.004 (2016).

27 Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut *Bacteroidetes*: the food connection. *Frontiers in Microbiology* **2**, 16, doi:10.3389/fmicb.2011.00093 (2011).

28 Li, N. *et al.* The Type IX Secretion System Is Required for Virulence of the Fish Pathogen Flavobacterium columnare. *Applied and Environmental Microbiology* **83**, e01769-01717, doi:10.1128/aem.01769-17 (2017).

29 Woyke, T. *et al.* Complete genome sequence of the gliding freshwater bacterium Fluviicola taffensis type strain (RW262). *Stand Genomic Sci* **5**, 21-29, doi:10.4056/sigs.2124912 (2011).

30 Mitchell, A. L. *et al.* MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research* **48**, D570-D578, doi:10.1093/nar/gkz1035 (2019).

31 Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology* **39**, 105-114, doi:10.1038/s41587-020-0603-3 (2021).

32 Shoji, M. *et al.* Por secretion system-dependent secretion and glycosylation of Porphyromonas gingivalis hemin-binding protein 35. *PloS one* **6**, e21372-e21372, doi:10.1371/journal.pone.0021372 (2011).

33 de Diego, I. *et al.* The outer-membrane export signal of Porphyromonas gingivalis type IX secretion system (T9SS) is a conserved C-terminal β-sandwich domain. *Scientific Reports* **6**, 23123, doi:10.1038/srep23123 (2016).

34 Sato, K. *et al.* A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences* **107**, 276-281 (2010).

35 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).

36 Ollis, D. L. *et al.* The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211 (1992).

37 Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531, doi:10.1093/nar/gkh468 (2004).

38 Morgan-Lang, C. *et al.* TreeSAPP: the Tree-based Sensitive and Accurate Phylogenetic Profiler. *Bioinformatics* **36**, 4706-4713 (2020).

39 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455 (2019).

40 Mukherjee, S. *et al.* Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Research* **49**, D723-D733, doi:10.1093/nar/gkaa983 (2020).

41 Austin, H. P. *et al.* Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc Natl Acad Sci U S A* **115**, E4350-E4357, doi:10.1073/pnas.1718804115 (2018).

42 Alneberg, J. *et al.* Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Communications biology* **3**, 1-10 (2020).

43 Alneberg, J. *et al.* BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. *Scientific data* **5**, 1-10 (2018).

44 Mohrholz, V., Naumann, M., Nausch, G., Krüger, S. & Gräwe, U. Fresh oxygen for the Baltic Sea—An exceptional saline inflow after a decade of stagnation. *Journal of Marine Systems* **148**, 152-166 (2015).

45 Hawley, A. K. *et al.* A compendium of multi-omic sequence information from the Saanich Inlet water column. *Scientific data* **4**, 1-11 (2017).

46 Hallam, S. J., Torres-Beltrán, M. & Hawley, A. K. Monitoring microbial responses to ocean deoxygenation in a model oxygen minimum zone. *Scientific data* **4**, 1-3 (2017).

47 Torres-Beltrán, M. *et al.* A compendium of geochemical information from the Saanich Inlet water column. *Scientific data* **4**, 1-11 (2017).

48 Bernardet, J. F. *et al.* Polyphasic study of Chryseobacterium strains isolated from diseased aquatic animals. *Syst Appl Microbiol* **28**, 640-660, doi:10.1016/j.syapm.2005.03.016 (2005).

49 Loch, T. P. & Faisal, M. Emerging flavobacterial infections in fish: A review. *J Adv Res* **6**, 283-300, doi:10.1016/j.jare.2014.10.009 (2015).

50 Bowman, J. P. & Nichols, D. S. Aequorivita gen. nov., a member of the family Flavobacteriaceae isolated from terrestrial and marine Antarctic habitats. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1533-1541, doi:https://doi.org/10.1099/00207713-52-5-1533 (2002).

51 Dodd, D., Mackie, R. I. & Cann, I. K. O. Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. *Molecular Microbiology* **79**, 292-304, doi:10.1111/j.1365-2958.2010.07473.x (2011).

52 Foley, M. H., Cockburn, D. W. & Koropatkin, N. M. The Sus operon: a model system for starch uptake by the human gut *Bacteroidetes*. *Cellular and Molecular Life Sciences* **73**, 2603-2617, doi:10.1007/s00018-016-2242-x (2016).

53 Desvaux, M., Hébraud, M., Talon, R. & Henderson, I. R. Outer membrane translocation: numerical protein secretion nomenclature in question in mycobacteria. *Trends in microbiology* **17**, 338-340 (2009).

54 Son, H. F. *et al.* Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. *ACS Catalysis* **9**, 3519-3526, doi:10.1021/acscatal.9b00568 (2019).

55 Martinez-Martinez, M. *et al.* Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chem Biol* **13**, 225-234, doi:10.1021/acschembio.7b00996 (2018).

56 Hult, K. & Berglund, P. Enzyme promiscuity: mechanism and applications. *Trends in Biotechnology* **25**, 231-238, doi:10.1016/j.tibtech.2007.03.002 (2007).

57 Leveson-Gower, R. B., Mayer, C. & Roelfes, G. The importance of catalytic promiscuity for enzyme design and evolution. *Nature Reviews Chemistry* **3**, 687-705, doi:10.1038/s41570-019-0143-x (2019).

58 Coordinators, N. R. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **45**, D12-D17, doi:10.1093/nar/gkw1071 (2017).

59 The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

60 Markowitz, V. M. *et al.* IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* **40**, D115-122, doi:10.1093/nar/gkr1044 (2012).

61 Agarwala, R. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **44**, D7-D19, doi:10.1093/nar/gkv1290 (2016).

62 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* **41**, e121-e121, doi:10.1093/nar/gkt263 (2013).

63 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242, doi:DOI 10.1093/nar/28.1.235 (2000).

64 Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**, 205-217, doi:10.1006/jmbi.2000.4042 (2000).

65 Hall, T. A. in *Nucleic acids symposium series.* 95-98 ([London]: Information Retrieval Ltd., c1979-c2000.).

66 Feike, J. *et al.* Measuring unbiased metatranscriptomics in suboxic waters of the central Baltic Sea using a new in situ fixation system. *The ISME journal* **6**, 461-470 (2012).

67 Schulz-Vogt, H. N. *et al.* Effect of large magnetotactic bacteria with polyphosphate inclusions on the phosphate profile of the suboxic zone in the Black Sea. *The ISME journal* **13**, 1198-1208 (2019).

68 Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods in enzymology* **531**, 237-250 (2013).

69 Beier, S. *et al.* A metatranscriptomics ‐ based assessment of small ‐ scale mixing of sulfidic and oxic waters on redoxcline prokaryotic communities. *Environmental microbiology* **21**, 584-602 (2019).

70 Hawley, A. K. *et al.* Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nature Communications* **8**, 1507, doi:10.1038/s41467-017-01376-9 (2017).

71 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

72 Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).

73 Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 1-13 (2018).

74 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

75 Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).

76 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25**, 1043-1055 (2015).

77 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. (Oxford University Press, 2020).

78 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* **47**, W256-W259 (2019).

79 Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234, doi:10.1016/j.pep.2005.01.016 (2005).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTABLES.pdf
- SupplemntaryFIGURESVersionC.pdf