UC Riverside

UC Riverside Previously Published Works

Title

Novel putative polyethylene terephthalate (PET) plastic degrading enzymes from the environmental metagenome

Permalink

https://escholarship.org/uc/item/7v15z2qp

Journal

Proteins Structure Function and Bioinformatics, 90(2)

ISSN

0887-3585

Authors

Karunatillaka, Isuru
Jaroszewski, Lukasz
Godzik, Adam

Publication Date

2022-02-01

DOI

10.1002/prot.26245

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Proteins. Author manuscript; available in PMC 2023 February 01.

Published in final edited form as:

Proteins. 2022 February ; 90(2): 504–511. doi:10.1002/prot.26245.

Novel putative polyethylene terephthalate (PET) plastic degrading enzymes from the environmental metagenome

Isuru Karunatillaka[1], **Lukasz Jaroszewski**[2], **Adam Godzik**[2,*]

[1]Undergraduate Research Project, College of Natural and Agricultural Sciences, University of California Riverside, 900 University Ave., Riverside, CA, 92521, USA

[2]Biosciences Division, University of California Riverside School of Medicine, 900 University Ave., Riverside, CA, 92521, USA

Abstract

Several plastic degrading enzymes have been described in the literature, most notably PETases that are capable of hydrolyzing polyethylene terephthalate (PET) plastic. One of them, the PETase from *Ideonella sakaiensis*, a bacterium isolated from environmental samples within a PET bottle recycling site, was a subject of extensive studies. To test how widespread PETase functionality is in other bacterial communities, we used a cascade of BLAST searches in the JGI metagenomic datasets and showed that close homologs of *I. sakaiensis* PETase can also be found in other metagenomic environmental samples from both human-affected and relatively pristine sites. To confirm their classification as putative PETases, we verified that the newly identified proteins have the PETase sequence signatures common to known PETases and that phylogenetic analyses group them with the experimentally characterized PETases. Additionally, docking analysis was performed in order to further confirm the functional assignment of the putative environmental PETases.

Keywords

polyethylene terephthalate (PET) degradation; PETase; plastic degradation; plastic recycling; environmental microbiome

Introduction

Polyethylene terephthalate (PET) plastic is one of the most widely used plastic materials for consumer products and packaging[1]. The widespread use of PET has resulted in its harmful accumulation in the environment, endangering wildlife and human health alike. With the recycling rate of PET bottles and jars being only 29.1% in 2017, there is a large commercial interest in more efficient means of biological recycling in order to increase PET recycling rates[2].

PETases are the esterase class of enzymes that catalyze the hydrolysis of PET into mono-2-hydroxyethyl terephthalate (MHET), with minor amounts of terephthalic acid (TPA)

*Corresponding author: adam.godzik@medsch.ucr.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and bis(2-hydroxyethyl) TPA (BHET). These enzymes can further hydrolyze BHET into MHET[3]. Some enzymes with PET hydrolysis activity can hydrolyze MHET into TPA and ethylene glycol, but a separate esterase, MHETase, can be used instead for this hydrolysis [5]. The end products can be utilized by the bacterial organism for growth. Importantly, they can also be used to create new plastics or numerous other polyester consumer products, therefore PETases could serve as a major solution to plastic pollution in both terrestrial and aquatic habitats [4]. PETases are a subset of a broader protein family of cutinases, enzymes that hydrolyze cutin, which is one of two waxy polymers that are the main components of the cuticle that covers all aerial surfaces of plants, and they belong to an even larger family of alpha-beta hydrolases [5]. Despite their close homology, PETases can be distinguished from classical cutinases by the presence of specific sequence motifs and by phylogeny, which will be elaborated upon further in this paper.

While enzymes able to hydrolyze PET polymers have been known for a while [6, 7], one of the most efficient PETases discovered thus far comes from the bacteria *Ideonella sakaiensis* strain 201-F6, which was cloned from environmental samples taken from a PET bottle recycling site in Sakai, Osaka Prefecture in Japan [8]. After the sequencing of its genome, the enzyme responsible for the PETase activity was identified as a product of the gene ISF6_4831 and was shown to be capable of degrading PET film of roughly 15% crystallinity (commercial soft drink bottles typically have PET crystallinity of 15.7% at pH 7.2 [9]). The *I. sakaiensis* PETase has a PET film degradation rate of 0.13 mg cm^{-2} day^{-1} at 30°C [8]. Its structure was later determined by multiple research groups using X-ray crystallography, with structures available from the Protein Data Bank (PDB). In the analyses performed here, we will use the *I. sakaiensis* PETase structure with the PDB code 6EQD as an example, since all the structures are very similar.

Aside from the *I. sakaiensis* PETase, at least 69 PETase-like enzymes from a variety of terrestrial and aquatic soil bacteria are currently known, however the *I. sakaiensis* enzyme remains the most efficient [3].

Despite its relatively high efficiency, the *I. sakaiensis* PETase is still too slow for industrial scale PET plastic recycling. Since an industrialized biological recycling of PET plastic would have a much lower cost than current PET plastic recycling methods, eliminating roadblocks to scalability is paramount. Eliminations of such roadblocks can occur by either finding ways to improve the efficiency of the *I. sakaiensis* PETase or finding new, more efficient, PETases.

PETases have conserved Ser-His-Asp catalytic triads indicating catalytic mechanisms shared with other esterases [3, 10]. Additionally, two binding subsites and a stabilizing disulfide bridge within the binding site were identified as being specific to PETases. Known PETases were classified into three types (type I, IIa, and IIb) based on amino acid sequence and structural similarity. Type IIb PETases include the *I. sakaiensis* PETase. The other enzymes of type IIb are predicted to have similar PET-degrading activities as the *I. sakaiensis* PETase [3].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In this publication, we assume that the conservation of the residues that constitute the catalytic triad, the two subsites, and the disulfide bridge common to all known PETases can be used in combination with a cascade of sequence-based homology searches to identify candidates for novel PETases and, relying on this assumption, identify candidate PETases in the Joint Genome Institute (JGI) terrestrial metagenome microbiome samples [11]. We show that PETases could be widely distributed in different environments and suggest that systematic searches for new PETases in the environmental metagenomes may yield many novel enzymes with PETase activity.

Results

Putative PETases found in environmental samples

Using the search strategy described in detail in the Methods section, we identified 27 putative PETases (further denoted PETen1–27) from 24 environmental metagenomics samples in the JGI Integrated Microbial Genomes & Microbiomes (IMG/M) resource [11]. All of the candidate PETases (see the full list included in supplementary table S1) share identical or similar residues, with few exceptions, in the positions constituting the catalytic site, binding subsite I and subsite II, in all currently known PETases/PETase-like enzymes. Since these putative PETases come from metagenomic sequencing projects, some of the sequences were incomplete and did not include regions where some of the aforementioned functionally important residues should be found. Here we focused on two candidate PETases (PETen1 and PETen2) whose sequences are sufficiently complete to include all of their functionally important motifs (see Table 1).

Our data shows that most of the residues/ positions in the catalytic triad, subsite I, subsite II, and additional disulfide bond of *I. sakaiensis* PETase are conserved in the newly found environmental metagenomic putative PETases. Below we analyze in depth their conservation in PETen1 and 2. In PETen1, the only substitution is from polar T88 to nonpolar L88, while PETen2, has Y87F, A89E, and W159H substitutions. It has been shown by Joo et. al. that W159H variant decreases PET-degrading activity [3], but site-directed mutagenesis of T88L, Y87F, nor A89E variants have not been reported. The other binding site residues of importance that are critical for defining efficient PETases are present in all the metagenomic putative PETases whose sequences include these positions, including the catalytic triad and an additional disulfide bridge (C203 and C239). The additional disulfide bond is indicative of a structure conformation that keeps H237 engaged in the catalytic triad, stabilizing subsite II, and improving thermal stability. This disulfide bridge, which so far has been unique to *I. sakaiensis* PETase and other type II PETases, is found within both PETen1 and 2. Joo et. al. emphasizes the importance of an extended loop constituting positions S242, G243, N244, S245, N246, Q247 for holding together subsite II and the catalytic triad. This extended loop, which is also unique to *I. sakaiensis* PETase and other type II PETases, is conserved within PETen1 and 2 (an exception being Q247 which is mutated to alanine in PETen1 and to threonine in PETen2). Additionally, PETen1 is unique in that it has a natural R280A mutation, which has been shown to improve the catalytic rate in *I. sakaiensis* PETase mutants over the wildtype by 22.4% in 18 hours and 32.4% in 36 hours [1 12]. The fact that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the majority of the binding site residues in PETen1 and 2 match those indicative of type IIb PETases, strongly suggests that their function is the same as in the *I. sakaiensis* PETase.

In regard to the other environmental putative PETases, PETen 1, 2, 17, 18, and 20 contain all of the motif residues of each binding site feature and seem to be full length sequences (see supplementary table S1). In PETen 17, all the functionally important residues are fully conserved with *I. sakaiensis* PETase. In PETen 3, 8, 10, 12, 14, 15, 16, 19, 21, 23, and 24 most of the functional motif residue positions are conserved.

Verification of the putative PETases by modeling and docking calculations

Homology modeling and docking were performed to further examine the validity of classifying both PETen1 and 2 as putative PETases. To accomplish this, tertiary structures of both proteins were predicted using Robetta [13] and docking experiments were performed using AutoDock Vina [14]. As expected, the predicted structures of both PETen1 and 2 have significant structural similarity to *I. sakaiensis* PETase. The overall predicted structure of PETen1 compared to *I. sakaiensis* PETase has an RMSD of 0.94Å and the functionally important binding site residues have an RMSD of 0.75Å. The overall predicted structure of PETen2 compared to *I. sakaiensis* PETase has an RMSD of 0.80Å and the functionally important binding site residues have an RMSD of only 0.40Å. All RMSD calculations made using UCSF Chimera [20].

It should be noted that *Fusarium solani* f. sp. *pisi* has been shown to conduct PET surface modifications at optimal conditions, but the enzyme is unable to degrade the inner block of PET, thus we will consider it a “classical cutinase” in comparison to PET degrading enzymes [15]. As shown by the docking results (see Table 2), PETen1, PETen2, and *I. sakaiensis* PETase had similar and consistently lower (more favorable) binding energies for PET compared to the classical cutinase *Fusarium solani* f. sp. *pisi* [16]. Additionally, the classical cutinase had a lower binding energy for cutin than either PETen1, PETen2 or *I. sakaiensis* PETase. Surprisingly, the *Fusarium solani* f. sp. *pisi* cutinase did not show as strong of a preference for the cutin molecule as might be expected. This could be due to the oligo-polyester differing from the actual cutin substrate that the enzyme naturally degrades, especially considering the vast range of different existing cutin polymers. It should also be noted that the orientations of the docked 2-HE$(MHET)_4$ in relation to PETen1 and 2 were very similar to its docked orientation in *I. sakaiensis* PETase (Figure 1) due to their unobstructed binding site clefts. On the other hand, the *Fusarium solani* f. sp. *pisi* binding site cleft is not continuous - in a very similar fashion to *Tf*Cut2 in Figure 5 of Joo et. al.[1] - therefore the shortened PET molecule docked in an unrealistically convoluted manner. Due to the binding preference and the more realistic docking conformations that the two metagenomic proteins and *I. sakaiensis* PETase have with PET, these data support the classification of PETen1 and PETen2 as putative PETases rather than classical cutinases.

Weblogo Comparison

In order to further compare the metagenomic putative PETases and experimentally verified PETases, we prepared weblogos[17] representing key functional residues (Figure 2).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The Weblogos show that for subsite I, Q119, M161 and W185 are conserved in the metagenomic putative PETases as well as known PETases, while Y87 is commonly replaced with phenylalanine in both groups. For subsite II positions, polar T88 is frequently replaced by nonpolar L88 in the metagenomic putative PETases and in known PETases. Nonpolar A89 is not as conserved as the other binding site residues and is replaced by acidic glutamate in a small percentage of metagenomic sequences. The W159 residue (adjacent to the serine nucleophile) is commonly replaced with histidine in both the metagenomic sequences and previously known PETases. Polar S238 and N241 are strongly conserved in the metagenomic putative PETases. The catalytic triad (S160, D206, and H237) is conserved within all the metagenomic putative PETases.

Comparison of the Weblogos shows that, overall, *I. sakaiensis* PETase binding and active site residues are well conserved in the new metagenomic proteins. S238 found in *I. sakaiensis* PETase is much more dominant in the metagenomic proteins than the previously determined PETases. S238, along with W159, play a crucial role in higher PET-degrading ability by making the binding cleft more shallow and the catalytic site more accessible compared to the binding cleft in low activity PET hydrolases [3]. Also, T88 and A89 in I. *sakaiensis* have higher conservation in the metagenomic proteins than in the previously determined PETases.

The critical additional disulfide bridge residues (C203 and C239) are found in all of the metagenomic putative PETases with sufficiently complete sequences confirming likelihood of PET degradation ability.

Additionally, PETens 1, 2, 8b, 14, 15a, and 22 of the metagenomic sequences contain an extended loop (N244, S245, N246) which is important for maintaining a continuous binding cleft [1]. This extended loop is exclusive to only *I. sakaiensis* PETase and other type II PETases. The residue similarity between the metagenomic sequences and the experimentally verified PETases at the functionally important positions strongly suggests that the metagenomic sequences collected in this study are indeed putative PETases.

The sequences of the putative PETases were checked by reciprocal BLAST [18] in the NCBI NR datatase and for all of them the *I. sakaiensis* PETase or one of the other annotated type II PETases were the top BLAST hits.

Phylogenetic Analysis

In order to further verify that the selected metagenomic sequences represent putative PETases, we analyzed their phylogenetic relations to previously identified PETases, fungal cutinases, and to a group of bacterial homologs of *I. sakaiensis* PETase which do not have many of the functionally important residues that are indicative of type I, IIa, or IIb PETases (for details, see Methods section). All taxa highlighted by a bold font in Figure 3 were identified by Joo et. al. [1]. The representative group of fungal cutinases, except the cutinase from *Humicola insolens*, were taken from Uniprot by randomly selecting proteins with the name “cutinase” within the UniProtKB database (their sequences can be found in supplementary table S4), none of them have documented PET degrading ability [19]. The fungal cutinase from *Humicola insolens* was experimentally shown to degrade

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

low-crystallinity PET films [20]. We will refer to the *Humicola insolens* sequence associated with PDB code 4OYY [21]. The current literature does not clearly identify specific bacterial cutinases with no PETase activity. In order to select representative bacterial cutinases with very low likelihood of PET degrading activity, three bacterial homologs of *I. sakaiensis* PETase were taken from PSI-BLAST searches of the NR database based on whether at least four of the eleven functionally important positions (excluding catalytic triad) contain residues different from any type of PETase. We will assume that bacterial homologs which contain such residue deviations would not be effective at degrading PET since they do not fit within any PETase type defined by Joo et. al.

As can be seen in the phylogenetic tree, fungal cutinases form a separate branch from all the bacterial proteins, with a bootstrap value of 100%. With the exception of PETen4 (Deep Shales in Ohio, USA: 3300009576), all of the metagenomic sequences phylogenetically group with the type IIb PETase-like enzymes and can be considered *I. sakaiensis* PETase orthologs. Metagenomic PETen4 may be considered a type I PETase-like enzyme based on sequence motifs of type I PETases described by Joo et. al. None of the PETens share the same branch with the type IIa PETase-like enzymes. Our tree shows that the metagenomic sequences of closest homology to *I. sakaiensis* PETase are PETen14 (Indian Creek, Utah, USA: 3300017965) and PETen15b (Lamont, Oklahoma, USA (b): 3300032080). Furthermore, the phylogenetic tree reinforces our classification of PETen1 and PETen2 as putative type IIb PETases. All of the metagenomic enzymes are promising candidates for PET hydrolysis based on phylogenetic distribution within PETase-like types. PETen1 and PETen2 are especially promising because they are full length.

The seemingly independent appearance of PET hydrolysis activity in fungal and bacterial cutinases could suggest functional convergent evolution, but further research is necessary. Although we do not provide evidence of convergent evolution between fungal and bacterial PET degrading enzymes, our tree suggests convergent evolution could be at work within the bacterial branch. The bacterial homologs, *Actinoplanes toevensis* lipase and *Amycolatopsis cihanbeyliensis* serine aminopeptidase, have different residues than any PETase type at four of the eleven functionally important positions (*A. toevensis*: A89N, W159Y, W185Y, N241T; *A. cihanbeyliensis*: T88I, A89E, Q119F, C239V). All position numbering is in respect to *I. sakaiensis* PETase PDB code 6EQD. Since these two sequences differ at four of the critical positions, they may be considered unlikely to degrade PET. It is interesting that these two bacterial homologs are the taxa most closely related to experimentally verified type I PETases, despite not having many critical residues. PETen4 on the other hand does show the residues characteristic of type I PETases despite being less phylogenetically related. This apparent jump in residue conservation can be explained by convergent evolution. The possibility of convergent evolution of PET degrading enzymes may be an important consideration in further studies and in identifying more metagenomic PETase-like enzymes.

Discussion

In this work we have identified 27 putative PETases in metagenomic datasets, from which we have used two (PETen1 and PETen2) to further argue for their classification

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

as PETases through sequence and docking analysis. Additionally, we have shown how the new metagenomic sequences are highly similar to known PETases by comparative Weblogos and phylogenetic analysis.

Surprisingly, in contrast to the *I. sakaiensis* PETase, which was discovered in a plastic recycling center, nearly all of the proposed metagenomic putative PETases were found in areas that are relatively pristine and less affected by plastic pollution. This indicates that the selection that led to the emergence of PETase activity may have been in response to a natural polymer, possibly a cutin variant, with similar molecular structure to PET polymers. As our data suggests, bacterial PETases may have branched independently from classical cutinases much earlier than previously thought and, most likely, without the specific impact of human-made PET polymers. With no notable commonalities between the samples' environments, this raises the question of what specific environmental substrate led to PETases branching off from classical cutinases?

It is noteworthy that our searches only included the terrestrial JGI microbiome metagenomes (7,375 samples). We did not perform searches in aquatic, airborne, engineered, host-associated, and undetermined origin metagenomic samples which constitute a total of 28,112 samples.

Overall, we have demonstrated that there are a significant number of previously uncharacterized proteins in the metagenomes that can be reliably predicted to have the capability of PET degradation. Before this publication, 69 possible PETase-like enzymes were known, all from known organisms; our work suggests that the number of PETases could be expanded by systematic analysis of metagenomic samples. The future steps would include crystallization of the full-length putative type IIb PETases, including the two analyzed in this paper, in order to more reliably assess their PET-degrading capabilities.

Materials and Methods

Data Collection

Metagenomic sequences were obtained by conducting BLAST searches on 7,375 terrestrial metagenomic samples in the JGI Integrated Microbial Genomes and Microbiomes database [11] with *I. sakaiensis* PETase used as the query. Only proteins with ≥60% sequence identity to *I. sakaiensis* PETase were considered for further analysis. Sequences of these 27 proteins were then checked by reciprocal BLAST (each PETen used as a query) in the NCBI NR database in order to make sure their top BLAST hit is the *I. sakaiensis* PETase or another type IIb PETase. The sequences were manually analyzed for the presence of the various sequence motifs (see Figure 2). PETen2 has an extra N-terminal region which appears to be an artifact of the metagenomic sequencing, as no known PETase or cutinase contains a homologous region [22]. This extra N-terminal region which extends to Gly178 was consequently removed from the sequence used in our analysis.

Structure Prediction and Docking

The structures for the two selected putative PETases (PETen1 and PETen2) were predicted by using Robetta [13]. FATCAT was then used in order to verify that the structures are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

similar to *I. sakaiensis* PETase as expected [23]. The docking binding energies of the two metagenomic proteins, *I. sakaiensis* PETase, and the cutinase from *Fusarium solani f. sp. pisi* were calculated using Autodock Vina [14] implemented as a part of UCSF Chimera [24].

Weblogo

The weblogos were prepared using Berkeley Weblogo 2.8.2 [17] after aligning the sequences using MUSCLE [25]. The sequences were trimmed to only include the residues of interest in the weblogos.

Phylogenetic Tree

In order to construct the phylogenetic tree, the web server IQ-TREE [26] was used to generate the maximum likelihood tree with columns containing >50% of gaps removed from the MUSCLE multiple alignment [25]. To assess for convergent evolution between the bacterial PET degrading enzymes, three extra bacterial homologs were found by performing a PSI-BLAST search with two iterations of the NR database using *I. sakaiensis* PETase as the query [17]. The collected sequences were then aligned to *I. sakaiensis* PETase using MUSCLE and trimmed down to the functionally important residues [24]. Three sequences with at least four of the eleven residues (excluding catalytic triad) being different from any residues found in either type I, IIa, or IIb were extracted. The bootstrap values were calculated using 1000 samplings and are shown as ratios. The tree was visualized using iTol Tree of Life v5.0 [27].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

We acknowledge help of all the members of the Godzik lab

Funding.

This research was funded in part by National Institute of Allergy and Infectious Diseases contract HHSN272201700060C (CSGID) and National Institute of General Medical Sciences Award GM118187.

References

1. Andrady AL and Neal MA, Applications and societal benefits of plastics. Philos Trans R Soc Lond B Biol Sci, 2009. 364(1526): p. 1977–84. [PubMed: 19528050]
2. EPA, E.P.A. Plastics: Material-Specific Data. 2020; Available from: www.epa.gov/factsand-figures-about-materials-waste-and-recycling/plastics-material-specific-data.
3. Joo S, et al. , Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. Nat Commun, 2018. 9(1): p. 382. [PubMed: 29374183]
4. Glaser J,A,, Biological Degradation of Polymers in the Environment, in Plastics in the Environment, Gomiero A, Editor. 2019, IntechOpen.
5. Holmquist M, Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr Protein Pept Sci, 2000. 1(2): p. 209–35. [PubMed: 12369917]
6. Tokiwa Y and Suzuki T, Hydrolysis of polyesters by lipases. Nature, 1977. 270(5632): p. 76–8. [PubMed: 927523]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Gumargalieva KZ, et al. , Effect of infections on the degradation of polyethylene terephthalate implants. Biomaterials, 1982. 3(3): p. 177–80. [PubMed: 6214288]
8. Yoshida S, et al. , A bacterium that degrades and assimilates poly(ethylene terephthalate). Science, 2016. 351(6278): p. 1196–9. [PubMed: 26965627]
9. Austin HP, et al. , Characterization and engineering of a plastic-degrading aromatic polyesterase. Proc Natl Acad Sci U S A, 2018. 115(19): p. E4350–E4357. [PubMed: 29666242]
10. Wallace AC, Laskowski RA, and Thornton JM, Derivation of 3D coordinate templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. Protein Sci, 1996. 5(6): p. 1001–13. [PubMed: 8762132]
11. Chen IA, et al. , IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res, 2019. 47(D1): p. D666–D677. [PubMed: 30289528]
12. Taniguchi I, et al. , Biodegradation of PET: Current Status and Application Aspects. ACS Catalysis, 2019. 9(5): p. 4089–4105.
13. Kim DE, Chivian D, and Baker D, Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res, 2004. 32(Web Server issue): p. W526–31. [PubMed: 15215442]
14. Trott O and Olson AJ, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 2010. 31(2): p. 455–61. [PubMed: 19499576]
15. Nimchua T, Punnapayak H, and Zimmermann W, Comparison of the hydrolysis of polyethylene terephthalate fibers by a hydrolase from Fusarium oxysporum LCH I and Fusarium solani f. sp. pisi. Biotechnol J, 2007. 2(3): p. 361–4. [PubMed: 17136729]
16. Longhi S, et al. , Atomic resolution (1.0 A) crystal structure of Fusarium solani cutinase: stereochemical analysis. J Mol Biol, 1997. 268(4): p. 779–99. [PubMed: 9175860]
17. Crooks GE, et al. , WebLogo: a sequence logo generator. Genome Res, 2004. 14(6): p. 1188–90. [PubMed: 15173120]
18. Altschul SF, et al. , Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 1997. 25(17): p. 3389–402. [PubMed: 9254694]
19. UniProt C, UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res, 2019. 47(D1): p. D506–D515. [PubMed: 30395287]
20. Ronkvist ÅM, et al. , Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate). Macromolecules, 2009. 42(14): p. 5128–5138.
21. Kold D, et al. , Thermodynamic and structural investigation of the specific SDS binding of Humicola insolens cutinase. Protein Sci, 2014. 23(8): p. 1023–35. [PubMed: 24832484]
22. Jaroszewski L, et al. , FFAS03: a server for profile--profile sequence alignments. Nucleic Acids Res, 2005. 33(Web Server issue): p. W284–8. [PubMed: 15980471]
23. Ye Y and Godzik A, Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics, 2003. 19 Suppl 2: p. ii246–55. [PubMed: 14534198]
24. Pettersen EF, et al. , UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem, 2004. 25(13): p. 1605–12. [PubMed: 15264254]
25. Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res, 2004. 32(5): p. 1792–7. [PubMed: 15034147]
26. Trifinopoulos J, et al. , W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res, 2016. 44(W1): p. W232–5. [PubMed: 27084950]
27. Letunic I and Bork P, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res, 2016. 44(W1): p. W242–5. [PubMed: 27095192]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

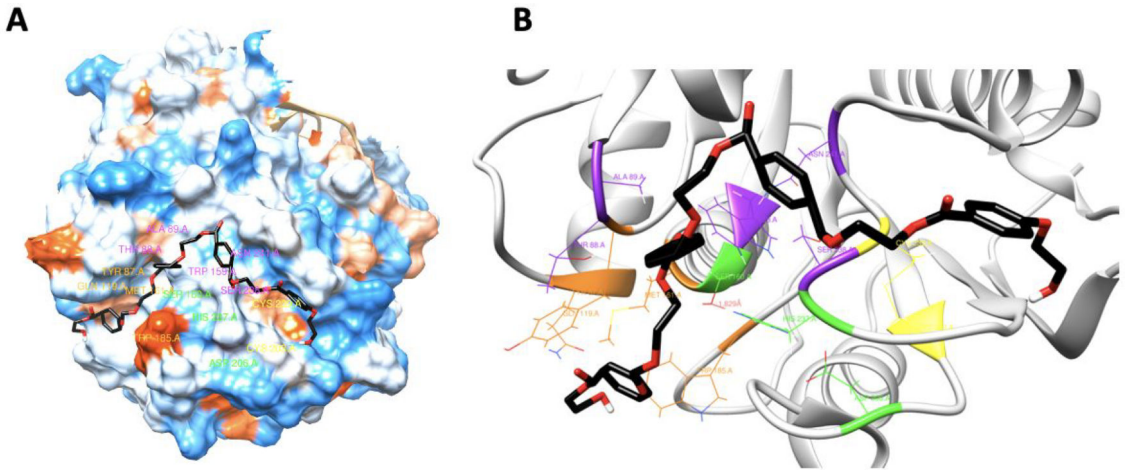

Figure 1.

Binding mode of PET molecule A) Shortened PET molecule 2-HE(MHET)4 docked within the binding site cleft of I. sakaiensis PETase. Surface is colored according to hydrophobicity. B) detailed view of the binding site of I. sakaiensis PETase with docked 2-HE(MHET)4 shortened PET molecule. The catalytic triad residues are colored green, subsite I residues are colored orange, subsite II residues are colored purple, and the disulfide bridge cysteines are colored yellow. This color scheme (consistent with the Joo et. al. 2) is used throughout the paper to indicate these groups of binding site residues.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

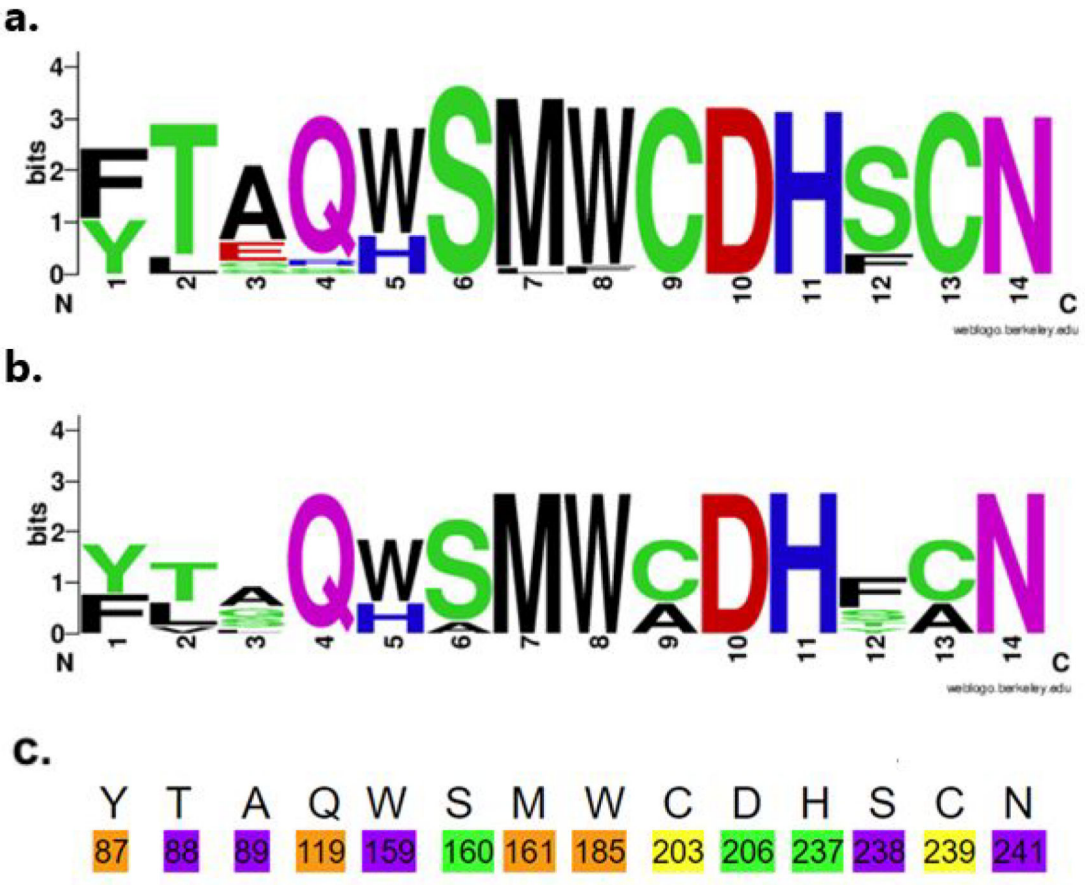

Figure 2.

The Weblogos of binding site positions of a) metagenomic putative PETases, b) previously determined PETases, c) The mapping of Weblogo positions to the sequence numbering of *I. sakaiensis* PETase where orange indicates subsite I, violet indicates subsite II, and green indicates catalytic triad. The colors of residues in a) and b) should not be associated with an indication of subsite I, subsite II, or the catalytic triad. See Methods section for the description of how the Weblogos were prepared. For full sequences, refer to table S3.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

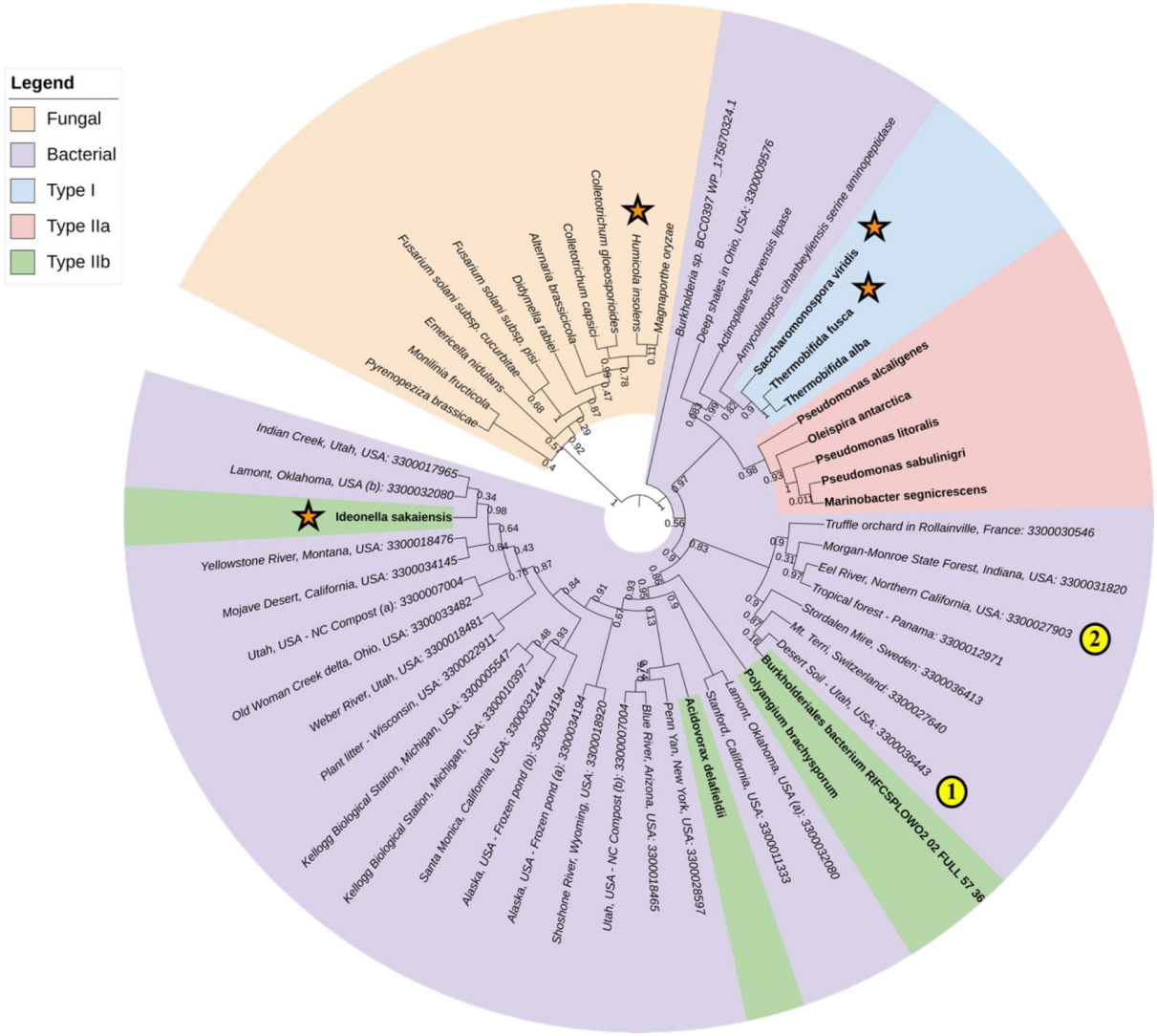

Figure 3.

Phylogenetic tree of proteins discussed in this manuscript. Figure legend displays respective groupings of each taxon. Light brown indicates fungal enzymes, purple indicates bacterial enzymes. Within bacterial enzymes: blue indicates type I PETase-like enzymes, red indicates type IIa PETase-like enzymes, and green indicates type IIb PETase-like enzymes (this classification is only shown for enzymes identified by Joo et. al. as belonging to a specific type). Orange stars indicate enzymes with significant experimentally verified PET degrading ability. The metagenomic sequences are labeled according to their corresponding location and IMG genome ID. PETen1 and 2 are labeled by “1” and “2” respectively.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 1.

The conservation of functionally important PETase residues in the two most complete putative metagenomic PETase sequences. The top row shows residues and position numbers from *I. sakaiensis* PETase. *I. sakaiensis* PETase residues which are not conserved in metagenomic putative PETases are colored red. The full sequences of all proposed novel PETases can be found in the supplementary table S2.

Genome / Sample name (link), IMG Genome ID IMG Gene ID (Internal ID)	catalytic triad			subsite I				subsite II					S-S bridge	
	S 160	D 206	H 237	Y 87	Q 119	M 161	W 185	T 88	A 89	W 159	S 238	N 241	C 203	C 239
Desert soil microbial communities from Utah, United States - 20190125_19, 3300036443 Ga0401344_00007_285137_286432 (PETen1)	S	D	H	Y	Q	M	W	L	A	W	S	N	C	C
Vadose zone soil microbial communities from the Eel River Critical Zone Observatory, Northern California, USA - Rivendell_Oct2014_Saprolite_2_DNA_Bulk_2 (SPAdes), 3300027903 Ga0209488_100034395 (PETen2)	S	D	H	F	Q	M	W	T	E	H	S	N	C	C

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Docking Results

The PET molecule used was the shortened tetramer PET molecule 2-HE(MHET)$_4$. The cutin molecule used was linear $C_{24}H_{42}O_8$, an oligo-polyester which mimics some cutin molecules.

	PETen1	**PETen2**	***I. sakaiensis* PETase**	**Classical Cutinase (*Fusarium solani pisi*)**
Binding Energy (kcal mol^{-1}) - PET	−7.2	−7.3	−7.7	−5.4
Binding Energy (kcal mol^{-1}) - Cutin	−4.0	−3.6	−4.2	−5.8