

Plastic degradation by enzymes from uncultured deep sea microorganisms

Daniel J. Acosta                                           <img alt="Wiley

representative of postconsumer PET products. The enzyme was found to have an optimal reaction temperature of 60°C and a melting temperature of 84°C, which are suited for PET depolymerization. Previous studies have used bioinformatic strategies such as the employment of hidden Markov models (HMMs) to screen metagenomic databases [14,15]. By using the sequences of known PETases, it is possible to construct a profile of the amino acid sequence of a PETase and use that to predict new putative PETases from metagenomic data. We recently obtained massive high-throughput metagenomic sequence data from five sediment cores spanning various thermal gradients harvested from four sites of the basin. Here we investigate the presence of PETases and other PET active enzymes from hydrothermally impacted sediments in GB and experimentally validate novel enzymes belonging to the globally distributed archaea, Bathyarchaeia, and other uncultured bacteria.

Materials and methods

Metagenomic sequencing, assembly, and binning

Sequencing through *de novo* binning has been previously described in detail for these samples [16]. Briefly, GB sediment push cores were collected by the human-operated vessel Alvin and research vessel Atlantis in November 2018 (AT42-05). Onshore, we extracted DNA for each core subsection across four sites. Library preparation and short-read sequencing was conducted by North Carolina State Genomics Sciences Laboratory (NovaSeq S4 [Illumina]) and University of Delaware (NextSeq [Illumina]). After initial quality control, reads were interleaved and assembled with BBTools Reformat v38.18 [17] and MEGAHIT v1.2.9 [18], respectively. We used assembled scaffolds >2.5 kbp for bioprospecting plastic-degrading enzymes and binning via MaxBin v2.2.7 [19], CONCOCT [20], and MetaBAT [21] via MetaWRAP v1.3.2 [22]. DAS Tools v1.1.2-2 provided a non-redundant set of bins for further analysis [23].

The archaeal PETase (GuaPA) encoding MAG was obtained from a southern mat mound (Latitude: 27°00.37' N; Longitude: 111°24.57' W and 2007.2 m below sea level [mbsl]), 15–20 cm deep with a thermal gradient ~20–38°C. The uncultured bacterial MAGs were extracted from Aceto Balsamico microbial mat (Latitude: 27°00.47' N; Longitude: 111°24.44' W and 2008.87 mbsl) and Cathedral Hill (Latitude: 27°00.67' N; Longitude: 111°24.26' W and 2012 mbsl), respectively. The Aceto Balsamico site (Marker 14) was previously named for high acetate and methane concentrations, supporting a diverse community [24,25]. The genome encoding the first BHETase (B1) was from an Aceto Balsamico site, 20–30 cm deep and had temperature readings between ~74.1 and 115°C. Whereas, the genome encoding the second BHETase (B4) was from a Cathedral Hill site 5–13 cm deep core with temperature of 89°C measured at a 25 cm core depth.

Computational bioprospecting

Using Prodigal v2.6.3, we predicted all the open reading frames of the GB 2.5 kb scaffolds (GE3) described previously [16]. These 7 782 778 million proteins became the database used to search for potential enzymes. To develop the search criteria, we used the PlasticDB [26], a database of known plastic-degrading enzymes, as the input enzymes. Using MMseqs2 [27], we clustered proteins listed under PET in PlasticDB by 30% sequence similarity, generating three clusters of PET-degrading enzymes. The largest cluster contained IsPETase, Leaf Compost Cutinase (LCC), and 55 other enzymes, whereas the other two clusters were sparse, containing three and two sequences, respectively. We computed

a multiple sequence alignment (MSA) of proteins in the largest cluster using MUSCLE v3.8.1551 [28], then generated HMM profiles from the resulting MSA using the HMMER v3.3 software suite hmmbuild function. Using these HMM models, we searched the previously generated GB protein database for hits $\leq 1E-5$, identifying 362 unique candidate enzymes. These hits were aggregated and filtered using custom python scripts available at github.com/marcottelab/GuaPA. Then, the candidate enzymes were clustered by 40% sequence similarity using CD-HIT v4.8.1 [29,30]. From the 32 resulting clusters, candidates with the lowest E-values from the initial HMM search were selected, and their protein structures were predicted using ESMFold [31], a large language model-based sequence-to-structure predictor that has previously been used for the efficient prediction of millions of metagenomic protein structures. Supplementary Fig. S1 depicts a phylogenetic tree showing the distribution of the candidates. Foldseek was then used to compare the structures of these candidates to the canonical enzymes IsPETase and LCC. The best 22 structural matches were then selected to move forward.

Prior to synthetic expression, the deep learning model TMBed [32], the software packages SignalP 6.0 [33], and Unbiased Organism-agnostic Signal Peptide Network (USPNet) [34] were applied to the 22 chosen sequences to predict signal peptides. Manual comparison of the output from these three methods determined the exact sequences to remove from these candidates. The predicted signal peptides extracted from the N-termini of the 22 candidate proteins are provided in Supplementary Fig. S2 and the removed sequence on GuaPA is highlighted in Supplementary Fig. S3.

Media and microorganisms

LB Broth Miller and LB Agar Miller were purchased from Fisher BioReagents. Terrific Broth was purchased from Sigma-Aldrich. Escherichia coli strains BL21 (DE3) and DH5 α chemically competent cells were purchased from New England BioLabs.

Cloning and protein expression

The 22 putative PETases (Supplementary Table S1) were ordered as gene fragments from TWIST Biosciences (South San Francisco, CA, USA) codon optimized for *E. coli* and flanked by Bsal cut sites. The gene fragments were cloned into pET28 α which was amplified with primers to incorporate Bsal cut sites which inserted the putative PETase genes in-frame with the NcoI Start codon and 6x-C-terminal His tag present in the original vector. DH5 α chemically competent cells were transformed with the assemblies, and transformants were selected on LB agar + 50 μ g/ml of kanamycin. For protein expression, BL21 (DE3) chemically competent cells were freshly transformed with sequence verified plasmid. Single colonies were used to inoculate overnight cultures of LB + kanamycin, which were used to inoculate flask cultures of Terrific Broth + kanamycin at a 1:100 ratio of medium to culture. Flask cultures maintaining a medium-to-flask volume ratio of 1:5 were grown at 37°C and 220 RPM until an OD₆₀₀ 0.4–0.6 was reached (in ~2–4 h). Flasks were placed in a 4°C cold room for 15 min, then induced with IPTG to a final concentration of 1 mM and allowed to shake at 220 RPM and 18°C for 18 h or overnight.

Protein purification

Cells, which had been induced and allowed to express protein for 18 h or overnight, were harvested via centrifugation at 6000 \times g at 4°C for 25 min. The cell pellets were resuspended in 30 ml of resuspension/lysis buffer (50 mM phosphate buffer pH 7.5,

300 mM NaCl, 20 mM imidazole, 0.1% Igepal CO-630 (Sigma-Aldrich), and 5 mM MgSO₄. Resuspended cells were lysed via sonication using 40% amplitude and 4 min of sonication time with 1 s on 4 s off. The cell lysate was clarified by ultracentrifugation at 4°C for 30 min at 35 000 × g. The clarified cell lysate was applied to a disposable 10 ml gravity column with a 1–2 ml resin bed (column volume (cv)) of Ni-NTA resin (Thermo) which had been equilibrated with 20 cv of equilibration buffer (50 mM phosphate buffer, pH 7.5, 300 mM NaCl, and 20 mM imidazole). The sample column was washed with 20 cv equilibration buffer and 5 cv wash buffer (50 mM phosphate buffer, pH 7.5, 300 mM NaCl, and 50 mM imidazole). Protein was eluted using 5 ml of elution buffer (50 mM phosphate buffer, pH 7.5, 300 mM NaCl, and 250 mM imidazole). The eluted protein was concentrated with an Amicon Ultra-15 Centrifugal Filter MWCO 10 000 spin filter and desalting by adding 15 ml of PBS to the concentrated sample and concentrating again. The concentrated, desalting protein was stored at 4°C for up to 2 weeks or mixed 1:1 with glycerol and stored at –20°C.

Enzyme assays

Enzymes concentrations were determined using Bradford assays. Enzyme activity was assayed using 10 mM BHET or a 6 mm hole punch of PET film, either beancake film or Goodfellow low crystallinity PET film as specified. Reactions were performed in a total of 600 µl using 100 mM potassium phosphate buffer pH 8 in protein low-bind tubes unless otherwise specified. BHET reactions were performed at 30°C for 24 h. BHET reactions were also performed after pre-incubating the tube with no BHET at 70°C for an hour and adding BHET once the tube had cooled. For PET assays, tubes were incubated for 48 h at either 30°C or 60°C unless otherwise specified. All reactions were performed in triplicate and with substrate and buffer only as a negative control. To stop the reactions, the entire reactions or aliquots of the reactions were mixed 1:1 with DMSO [35] (Supplementary Fig. S4).

High-performance liquid chromatography quantification of polyethylene terephthalate and bishydroxyethylene terephthalate hydrolysis

To quantify hydrolysis of PET and BHET, samples from enzyme assays were filtered using 0.2 µM syringe or 96-well plate centrifuge filters. These samples were then measured using a Vanquish HPLC system (Thermo Fisher Scientific) with a 260 nM UV detector. The mobile phase buffers consisted of 0.1% formic acid in water or 0.1% formic acid in acetonitrile over the course of 30 min with a fixed flow rate of 0.8 ml/min. The mobile phase used the following conditions: a solvent gradient of 1%–5% organic (vol/vol) for 5 min, a solvent gradient of 5%–100% organic (vol/vol) for 8 min, 100% organic solvent for 10 min, a solvent gradient of 100%–5% organic (vol/vol) for 2 min, and finally a solvent mix of 5% organic (vol/vol) for 5 min. Two-standard curves of BHET, MHET, and TPA were prepared for high and low ranges of detection in the linear ranges of 200 µM to 1.2 mM and 1.2 mM to 12 mM.

For BHET hydrolysis assays, the quantity of MHET and TPA was measured by subtracting the calculated mean quantity of TPA and MHET in the enzyme free negative control reactions from the calculated concentration of MHET and TPA in the samples.

Polyethylene terephthalate catabolism pathway structural homology search

We obtained protein sequences for monomer catabolism from UniProt [36] and searched the MAGs containing B1, B4, and GuaPA to determine pathway completeness (E-value ≤ 1e-5). In parallel, we downloaded the predicted protein structures of the known

enzymes necessary for the metabolic process from the AlphaFold Database [37] and used ESMFold [31] to predict the structures of the proteins present in the MAGs. We then used Foldseek with an E-value cutoff of 1e-5 to search the MAGs for any structural similarity.

Polyethylene terephthalate-hydrolyzing enzymes phylogenomic analyses

Database identifiers for 105 known PETase enzymes were downloaded from the PAZy database [15] and used to download the corresponding amino acid sequence from GenBank, UniProt, or MGnify. The sequences were then combined with candidate PETase enzymes from the GB samples. A MSA was computed via MAFFT v7.453 [38,39] and trimmed with ClipKIT v1.3.0 [40], and the phylogenetic tree was computed with IQ-TREE v1.6.12 [41]. The position of the 22 GB candidate enzymes in a more global phylogenetic tree was further determined by searching for homologs in the NCBI non-redundant database [42] (ca. September, 2024), retaining the top 50 hits for each sequence above an E-value of 1E-5. Duplicate sequences were removed, the sequences realigned with MAFFT auto v7.526 [38], and 50% of the gaps masked with Geneious Prime 2023.2.1 [43]. This overall dataset was then used to infer a phylogeny using IQ-TREE v2.0.7 [44]. The best-fit model WAG + R10 was chosen according to Bayesian information criterion. Phylogenies were visualized with iTOL v6 [45].

Phylogenetic analyses of polyethylene terephthalate-hydrolyzing enzyme candidate-encoding genomes

GTDB-tk v2.4.0 [46] initially classified the genomes encoding B1, B4, and GuaPA as *Poribacteria*, *Thermotogota*, and *Thermoproteota*, respectively. To verify the predicted taxonomic assignment, we selected publicly available bacterial genomes from (NCBI) belonging to Deinococcota, Synergistetes and Thermotogae (DST) phyla (*Deinococcota*, *Synergistota*, *Lindowbacteria*, *Spirochaetota*, *Elusimicrobiota*, *Desantibacteria*, *Mcinerneyibacteriota*, *Fusobacteriota*, *Poribacteria*) and basal Gracilicutes phyla (*Caldisericota*, *Lithacetigenota*, *Coprothermobacterota*, *Bipolaricaulota*, and *Thermotogota*). Archaeal references were chosen from *Thermoproteota* classes (*Bathyarchaeia*, *Korarchaeia*, *Methanomethyllicia*), orders (*Brockarchaeales*, *Caldarchaeales*, *Conexivisphaerales*, *Gearchaeales*, *Geothermarchaeales*, *Marsarchaeales*, *Nitrososphaerales*, *Panguiarchaeales*, *Sulfolobales*, *Thermofilales*, *Thermoproteales*), and an outgroup genus *Methanobacterium*. We extracted 37 cross-domain markers via the PhyloSift [47] pipeline to infer the phylogenetic placement of the archaeal and bacterial PETase candidate-encoding genomes within the reference set. The resulting concatenated alignments were realigned with MAFFT v7.490 [38,39] (auto) and masked (50% gaps) with Geneious Prime 2023.2.1 [43]. Phylogenetic trees were inferred with IQ-TREE v2.0.7 [44] for the domains separately based on 1000 ultrafast bootstrap inferences and optimized with hill-climbing nearest neighbor interchange. The archaeal and bacterial trees were rooted with *Methanobacterium* and between *Terrabacteria* and *Gracilicutes*, respectively. The resulting phylogenies were visualized with iTOL v6 [45].

Results

Identification of polyethylene terephthalate-hydrolyzing enzymes from thermophilic ocean floor microorganisms

Recent studies have identified an increasing number of novel PETase enzymes from a variety of non-marine sources [15]. Given

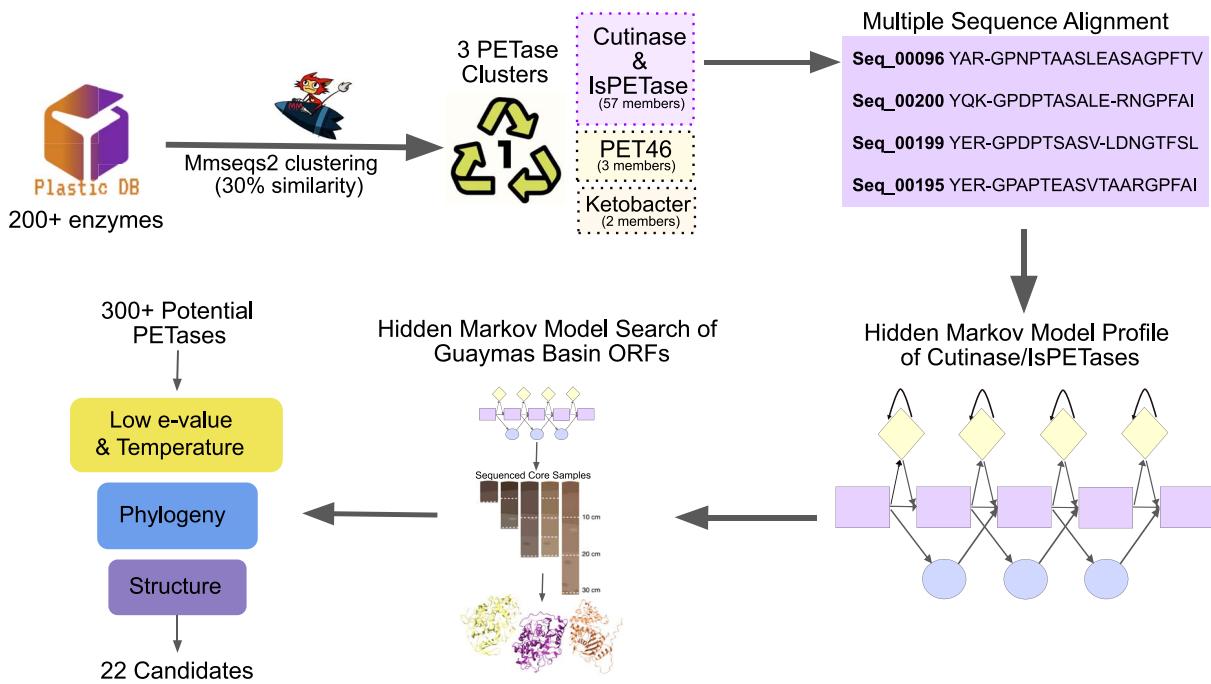


Figure 1. Computational pipeline picturing the process taken for aggregating known plastic degrading sequences, constructing a HMM profile, searching GB metagenomic data and subsequently downselecting potential enzyme candidates.

the global spread of plastic waste, we hypothesized that PETases might be found in deep sea microbial sources. We therefore plumbed a genomic catalog from the GB, a hydrothermal vent system in the Gulf of California [16,48] known to provide a wealth of information regarding new phyla and metabolic capabilities [48]. In particular, PETase sequences were mined from previously described samples from the GE3 [16] expedition of November 2018. During this expedition researchers collected sediment cores from ~2000 mbsl using a manned submersible, Alvin. The core samples were frozen at -80°C until DNA was extracted onshore [16] and sequenced using a combination of the NovaSeq S4 and NextSeq (Illumina) platforms, depending on DNA quantity. Protein-coding sequences from this metagenomic dataset were identified and annotated with Prodigal v2.6.3 [49], ultimately generating a set of 7 782 778 unique protein sequences (see Materials and methods).

A HMM profile was built from known plastic-degrading enzymes and used to search the compiled protein sequences (Fig. 1). Some 300 unique putative enzymes were identified and filtered to a top 22 candidates based on their thermophilic potential, as determined by the temperature of the sampling site where the DNAs were acquired, phylogenetic novelty, and structural similarity to known PETases (Supplementary Fig. S1). The majority of candidates chosen had less than 30% sequence identity with more well-known enzymes such as *Ideonella sakaiensis* PETase (IsPETase) and LCC, and at most 65% identity to known PETases listed in the PAZy database [15], with an overall average identity of 27.6% (Supplementary Fig. S5).

Screening of candidate enzymes on bishydroxyethylene terephthalate and amorphous polyethylene terephthalate

The 22 candidate enzymes were codon optimized, a C-terminal hexa-histidine tag added, and genes were expressed following cloning into pET28a and transformation into *E. coli* BL21 DE3. The putative PETases were purified using a nickel chelate column and

protein was obtained from all 22 identified sequences, in amounts ranging from 0.4 to 20 mg/100 ml culture. The proteins were estimated to be >90% pure via polyacrylamide gel electrophoresis (PAGE) (Supplementary Fig. S6).

Candidate enzyme activities were determined by applying enzyme samples to amorphous, beancake film PET (BC-PET), a relatively easy-to-degrade PET film, as described previously [4], and examining plastic breakdown products using high-performance liquid chromatography (HPLC). In an initial screen each of the 22 candidate PETases at 30°C no evidence of PET hydrolysis was observed (Fig. 2A). Knowing that some thermophilic enzymes might prove inactive at mesophilic temperature ranges, the entire PETase candidate list was rescreened on BC-PET at 60°C. Enzyme candidate 7, from *Bathyarchaeia* (Supplementary Fig. S7; GuaPA), was found to hydrolyze BC-PET, releasing a combined 4.4 mM of the depolymerization products TPA, MHET, and BHET after 48 h (Fig. 2A). GuaPA is thus the only archaeal enzyme that has been shown to depolymerize PET film.

To ascertain whether any of the candidate enzymes could also degrade BHET or MHET, all 22 enzymes were further assayed using BHET as a model substrate, initially at 30°C because BHET undergoes significant autohydrolysis at higher temperatures [50]. GuaPA was not able to hydrolyze BHET; however, Guaymas enzymes 1 and 4 (B1 and B4) were found to hydrolyze BHET to yield combined MHET and TPA products (2.7 and 4.0 mM, respectively, Fig. 2B). The production of TPA indicates that both B1 and B4 hydrolyze BHET to MHET and can further hydrolyze MHET to yield TPA. A pre-incubation of the enzymes at 70°C, followed by incubation with BHET at 30°C revealed that both B1 and B4 remained active following the thermal pulse and were still able to hydrolyze BHET (Fig. 2B). Given that the three active enzymes identified from the putative list of PETases appeared to be thermophilic, or at least thermotolerant, we decided to investigate their respective melting temperatures using differential scanning fluorimetry (DSF, the Biotium GloMelt system). The melting temperatures of GuaPA, B1, and B4 were 79, 87, and 80°C, respectively (Supplementary Fig. S8).

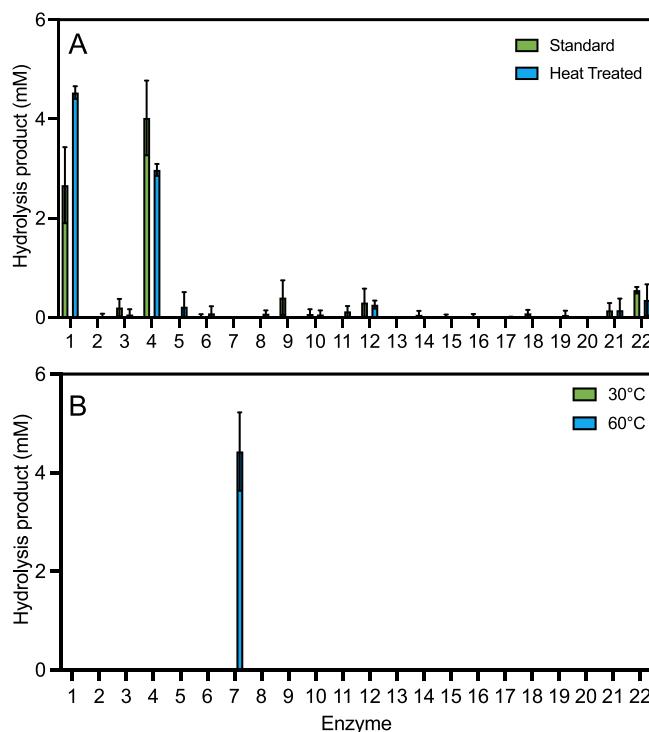


Figure 2. Hydrolysis products of 22 putative PETases using beancake film amorphous PET (A) and BHET (B) as substrates. All reactions were run in 600 μ l 100 mM KPB pH 8 with an enzyme concentration of 200 nM using either a 6 mm hole punch of PET film or 10 mM BHET. PET reactions were incubated for 48 h, while BHET reactions were incubated for 24 h.

Optimizing polyethylene terephthalate hydrolysis by GuaPA

We sought to optimize GuaPA's performance. Reaction parameters such as salt concentration, pH, enzyme concentration, and temperature were initially optimized; each was varied independently, using BC-PET as a substrate. The best reaction conditions were 200 mM KPB, pH 8, at 60°C (Supplementary Fig. S9A). The higher salt concentration preferred by GuaPA, 200 mM, compared to other PETases may be related to the enzyme's relatively low isoelectric point 4.9 or its oceanic origin. GuaPA contained what appeared to be a 24 amino acid signal peptide (Supplementary Fig. S3). When expressed with and without this leader the truncated version appeared to perform slightly better, but the differences were not statistically significant (Supplementary Fig. S9B).

We observed that GuaPA was most active at lower enzyme concentrations (100 nM). This contrasts with other PET-degrading enzymes, such as LCC and its engineered variants, which are typically used at micromolar concentrations. There is no loss of activity upon dilution of GuaPA from higher concentration (Supplementary Fig. S10), suggesting that the phenomenon is due to reversible interactions between the enzyme and the plastic or between enzyme molecules.

Improving polyethylene terephthalate degradation with enzyme cocktails

Previous research has shown that the inclusion of BHETases can improve PET hydrolysis by removing BHET and MHET intermediates that can occupy the PETase active site but are not efficiently hydrolyzed [51]. To assess this possibility with our newly discovered enzymes, we combined equimolar amounts of GuaPA and the

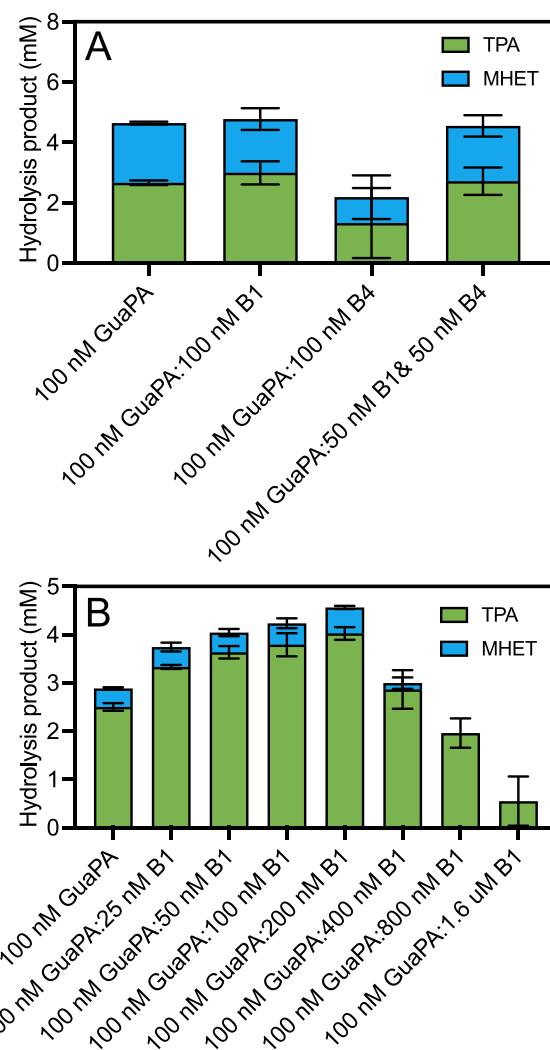


Figure 3. Depolymerization of PET films using two enzyme reactions either. (A) Depolymerization of BC film by GuaPA or combinations of GuaPA and BHETases. (B) Optimization of the ratio of GuaPA and B1 for increasing PET depolymerization of goodfellow low crystallinity PET.

BHETases B1, B4, or a combination of B1 and B4 (Fig. 3A). B1 or a combination of B1 and B4 showed no difference in hydrolysis, B4 seemed to significantly hinder PET hydrolysis.

To discern if B1 may improve PET hydrolysis with a more crystalline, and thus difficult to depolymerize substrate, B1 assays were carried out with 8% crystalline PET from Goodfellow. Although this is still low crystallinity PET, it is more difficult to degrade than the BC-PET which is 1%–2% crystallinity. On Goodfellow PET, GuaPA depolymerized an average of 2.9 mM of product, which consisted of 86% TPA and 14% MHET over 48 h (Fig. 3B). When combined at a ratio of one molecule of GuaPA to two molecules of B1, maintaining the concentration of GuaPA at 100 nM, hydrolysis is increased by 1.7-fold to a total hydrolysis of 4.6 mM combined TPA and MHET, with 88% of the release product again being TPA. As the proportion of B1 further increases, overall yields begin to fall with TPA eventually becoming 100% of the released product.

With both Goodfellow film and BC PET GuaPA releases depolymerization products not previously seen during the enzymatic depolymerization of PET films, whether by FAST PETase, LCC, or other PETases [4]. LC-MS analysis determined that the products

are small oligomers of PET (Supplementary Fig. S11), suggesting that GuaPA may function primarily as an endo-PETase. This is consistent with GuaPA's inability to hydrolyze BHET (Fig. 2B). This pattern of substrate preference has not been demonstrated among known PETases. These products disappear with the inclusion of B1, suggesting that B1 may be able to degrade PET oligomers larger than BHET.

Predicted structural properties of GuaPA

When an ESMFold [31] model was used to predict the structure of GuaPA it did not fall neatly into the Type I, Type II classification system previously used for categorizing PETases [52], and instead contained distinct structural features that differentiated it from other previously described PETases (Fig. 4). GuaPA is most different from other PETases in its extended loop region: this loop, which is 3–6 amino acids in Type I and Type II enzymes, is 12 amino acids in GuaPA. Docking a PET trimer in the active site of GuaPA using Chai-1 [53] shows that this loop expansion increases the number of hydrophobic amino acids proximal to the PET chain, potentially enlarging the hydrophobic cleft region. This expansion has been shown to be beneficial in other PETases by creating additional hydrophobic interactions that draw the PET chain into the active site [3,52,54]. The surface of GuaPA is very negatively charged (Fig. 5A and C), a property which was previously deemed to be a hindrance to PET hydrolysis [55]. The hypothesis that a negatively charged enzyme or active site might not be able to approach carboxylates on the PET surface was chemically suspect, and the existence of GuaPA suggests that PETases with large may have a variety of ways in which they can achieve plastics breakdown. Emphasizing such new possibilities, the demonstrated thermostability of GuaPA was attained even in the absence of a disulfide bond, a feature that is unique compared to other thermostable PETases [8,56]. We speculate that in the absence of disulfide bonds GuaPA may be stabilized by a network of π - π interactions. Using the Contacts of Aromatic Residues in Proteins analysis pipeline [57], GuaPA was found to have 33 predicted aromatic-aromatic residue interactions (Supplementary Fig. S12), whereas LCC had only 24 and IsPETase only 13 (Supplementary Fig. S12 and Supplementary Tables S2–S4). This collection of unique structural features makes GuaPA an especially appealing candidate for future enzyme engineering, potentially by incorporating the beneficial attributes of canonical PETases into its structure.

Predicted metabolism of polyethylene terephthalate derived monomers by enzyme containing metagenomes

The different enzymatic features of GuaPA led us to speculate that the organisms from which it was derived might also have unique PET degradation pathways. Given the ability of B1 and B4 to hydrolyze BHET, we wondered whether the metagenomic samples might also contain pathways capable of metabolizing liberated TPA and EG. The metabolic pathways to fully metabolize PET have been described previously [58], but searching for similar sequences via BLAST did not return homologous catabolic pathways. Nonetheless, many enzymes that might participate in homologous pathways were found using Foldseek structural similarity searches (Zenodo Repository) Bathyarchaeia MAG encoding GuaPA contains all of the genes necessary to import PET and degrade it to TPA, and from there to protocatechuate (Fig. 6), as well as pathways for the further integration of TPA into central metabolism via either the 2,3-meta-cleavage or the

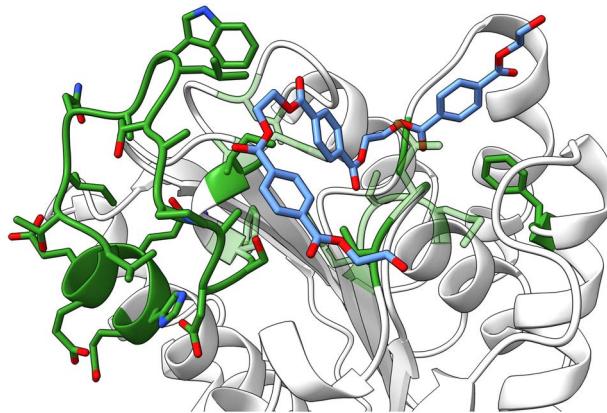


Figure 4. Residue differences in the structure of GuaPA in the context of a PET trimer compared to previously described PETases. Residues which are part of the catalytic triad, Subsite I, Subsite II, the extended loop, and the additional disulfide bond are shown as stick models. Those residues that do not fit into either of the traditional Type I or Type II PETases are opaque whereas those that do are transparent.

4,5-meta-cleavage pathway, similar to other metabolically diverse prokaryotes like *Comamonas testosteroni* [58].

Discussion

The newly identified thermophilic PETases from deep sea sediments may prove to be excellent candidates for engineering and translation [59], given their already very high thermostolerance, atypical isoelectric point, apparent preference for interior ester bonds, and ability to completely degrade PET to TPA and MHET. For example, GuaPA is significantly more thermostable than IsPETase, and only slightly less thermostable than LCC, one of the natural PETases with the highest reported melting temperatures [4,8,60].

The discovery of this enzyme in deep sea sediments may reveal a bounty of new opportunities for further biomining these sediments, given that computational analyses reveal unique enzyme features (Figs 4 and 5) and potentially unique pathways for the catabolism of PET degradation products (Fig. 6). In the validated PAZy database, there is only one other archaeal PETase (PET46) from an uncultivated GB Bathyarchaeia [61]. That said, GuaPA is the only archaeal enzyme that has been shown to depolymerize PET film and has a unique position in the phylogenetic tree of PETases. Additionally, B1 is from an isolate of the phylum *Poribacteria*, which are known commensals of marine sponges [62], and B4 is from the genus *Thermotogota*, which are known to be a part of general macromolecule degrading communities in the GB [63] (Supplementary Fig. S13). Like archaea, these bacteria are less well represented marine microorganisms in plastics degradation enzyme databases.

The identification of PETases from the deep ocean should at some level come as no surprise. Plastic pollution is permeant on the planet, and the deep ocean has been shown to be a major sink, harboring up to 11 million metric tons of plastic [64], a mixture of oceanic plastic from fishing equipment and terrestrial plastic deposited in the deep ocean by currents, with PET, nylon, and polyethylene [65,66] making up large percentages. Plastic degrading enzymes have been catching up with this rich new source of carbon [67]. Because the evolution of plastic degradation is still in its infancy, however, it is likely that there will be many parallel enzymes and pathways invented to supplement organismal growth.

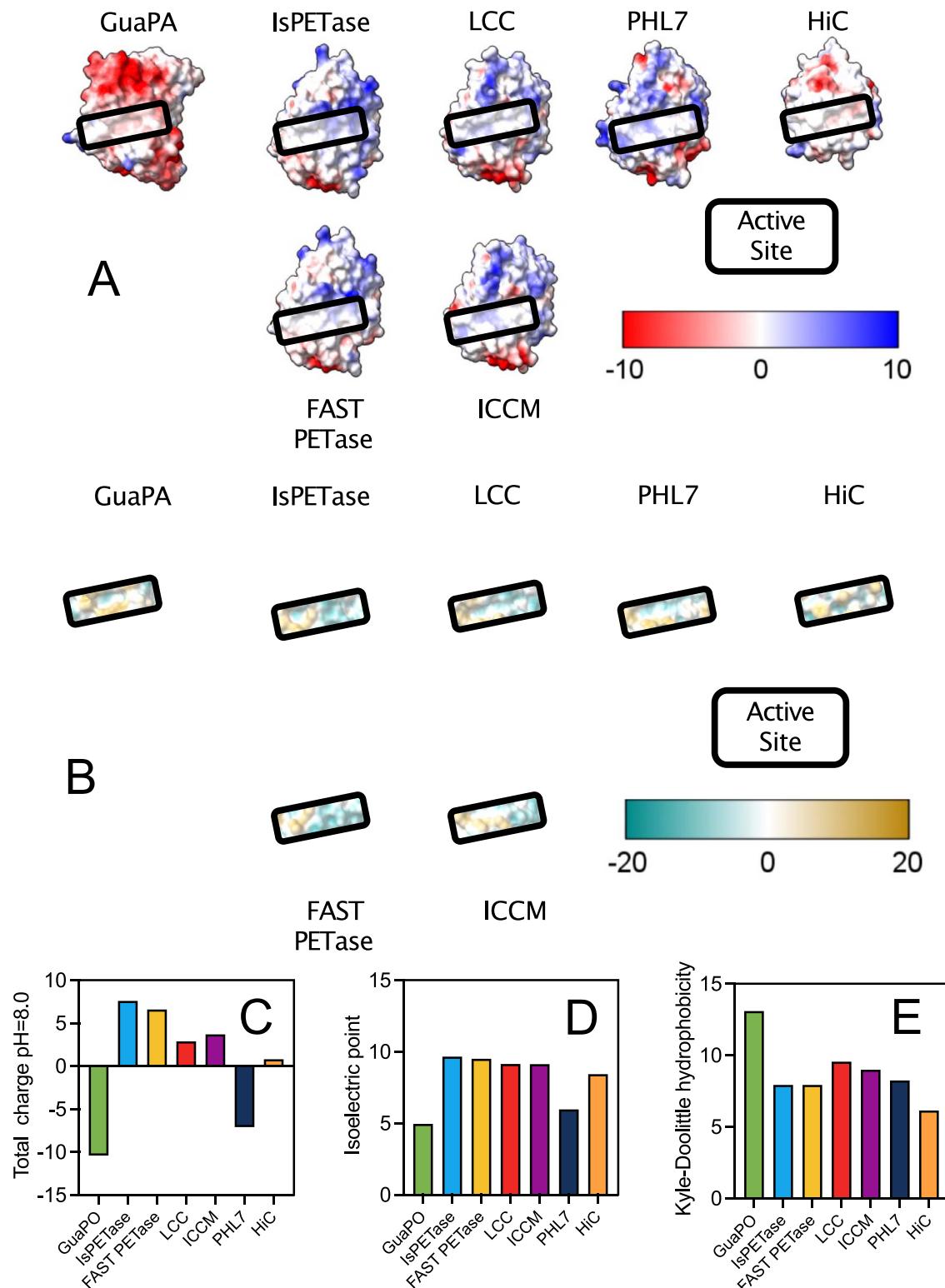


Figure 5. Surface properties of GuaPA compared to IsPETase, FAST PETase, LCC, ICCM, PHL7, and HiC. (A) Surface charge map of the catalytic face of 7 PETases demonstrating the highly negatively charged face of GuaPA in comparison to other PETases. (B) Hydrophobicity map of 7 PETases demonstrating the slightly larger hydrophobic pocket surrounding the active site of GuaPA. The net charge (C), isoelectric point (D), and hydrophobicity (E) of 7 PETases show GuaPA's unique combination of low charge, low isoelectric point, and higher hydrophobicity.

GuaPA is likely just the tip of the iceberg in terms of identifying PETases and other plastic-degrading enzymes. Phylogenetic analysis of enzymes related to GuaPA identified 93 further archaeal homologs from diverse sources, suggesting the archaeal diversity

of PETases remains largely untapped (Fig. 7). Broadly, the origins of GuaPA are far less clear than for other PETases, which are largely derived from cutinases (Fig. 7A and B). GuaPA is found to cluster with alpha/beta hydrolases, a serine peptidase, and

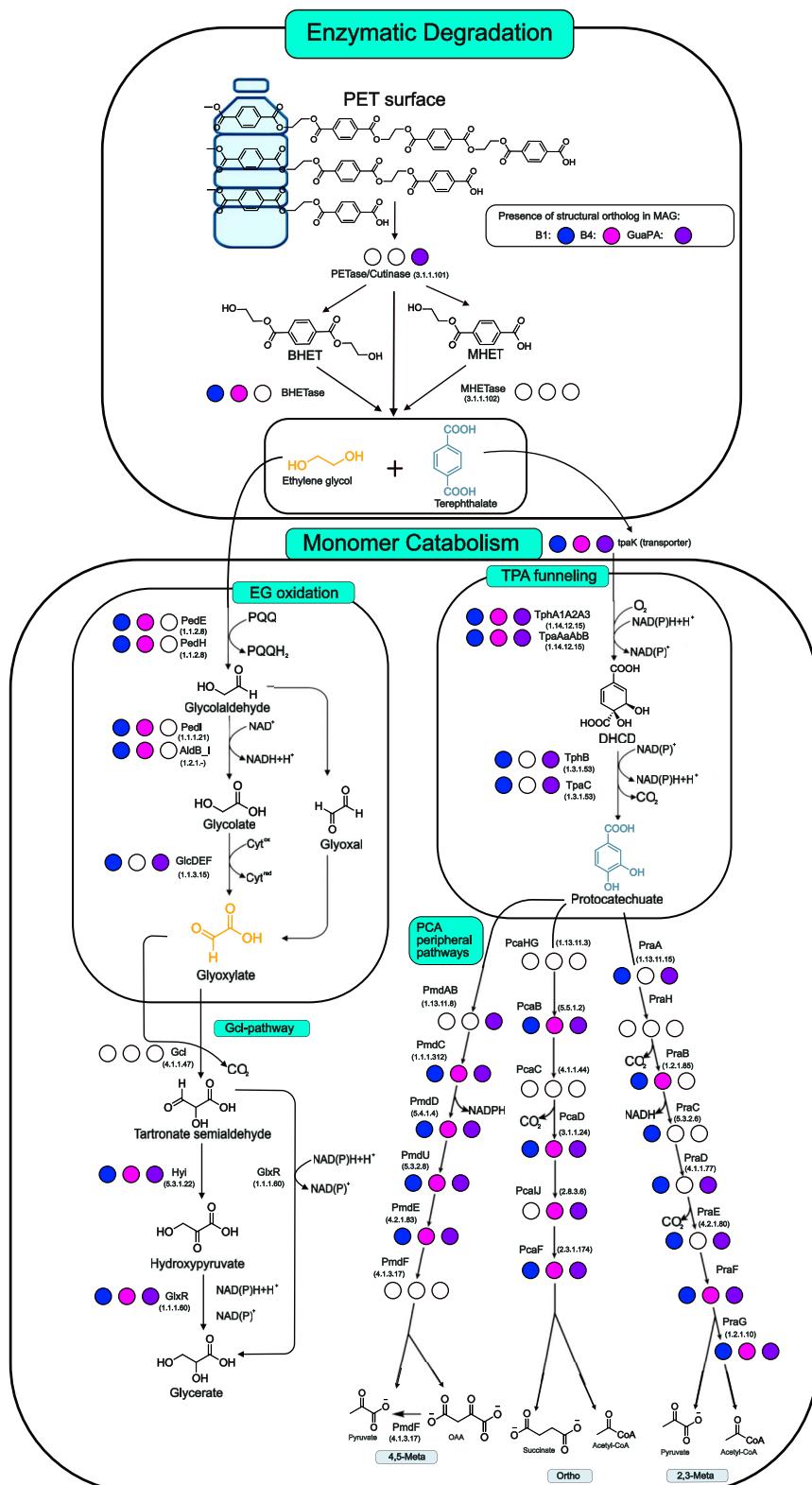


Figure 6. PET degradation and monomer catabolism pathway. Filled in circles represent whether a structural homolog for the given enzyme is present in the organism where B1, B4, or GuaPA resides. EC numbers are listed below each of the named enzymes in parentheses.

dienelactone hydrolases (Fig. 7E). Dienelactone hydrolases have previously been shown to have BHETase-like activity [68]. The identification of new PETases that are not derived from cutinases ultimately tracks with the relative lack of cutin in oceanic

environments [69], as opposed to the abundance of cutin in terrestrial plants [70].

The integration of sequence-based discovery methods, structural predictions, and phylogenetic data has unearthed new

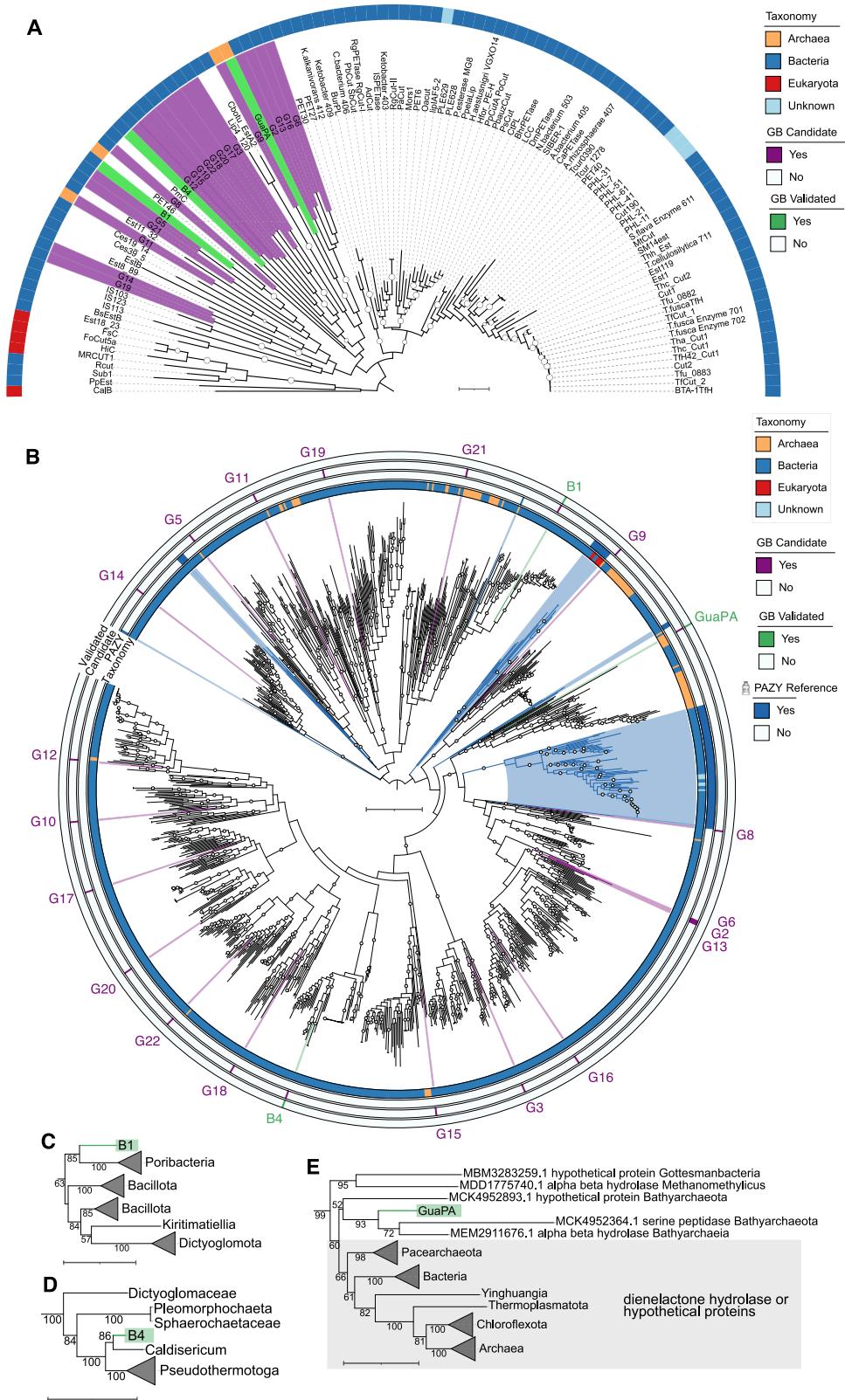


Figure 7. Phylogenetic distribution of candidate and validated Guaymas Basin (GB) PETases. (A) Unrooted phylogenetic tree of five published wild-type PET-hydrolyzing enzymes from the PAZy database along with the 22 GB candidates. The colored blocks indicate the taxonomic group from which the enzymes originate, according to PAZy. We highlighted the unvalidated and the active GB enzymes on PET or BHET. Tree scale is 0.1. (B) To increase phylogenetic support, we added unvalidated non-reference hits from the NCBI non-redundant sequence database for each of the 22 GB candidates, including the three validated enzymes. This expanded analysis includes 94 PAZy reference sequences and 998 blast sequences. We inferred the maximum likelihood phylogeny with IQ-TREE v2.0.7, using the best-fit model WAG + R10 chosen according to Bayesian information criterion. The concentric rings highlight taxonomy, PAZy references, GB candidates, and GB validated enzymes. Ultrafast bootstrap support values ≥ 90 are shown and midpoint rooted. Insets show sequences clustering with the validated GB enzymes: (C) B1, (D) B4, and (E) GuaPA. Ultrafast bootstrap support values ≥ 50 are shown for the insets. For (B-E), the tree scale is equivalent to 1.

plastic-degrading enzymes that likely would have been missed by a sequence-only approach. Not only can new PETases be discovered, but also structural homology searches have enabled the identification of new and potentially unique downstream enzymes for PET degradation (Fig. 6). As but one key example, B1 and B4 would not have been found if phylogenetic relatedness had been the sole criteria for selection. The genomic context of the newly discovered PETases may further reveal previously unknown plastic degradation pathways. For instance, the GuaPA genome appears capable of incorporating TPA into central metabolism via a multitude of pathways (Fig. 6), all of which may be repurposed for the next generation of PET biotransformation, with the same being true of the EG incorporation pathways present in the MAGs of B1 and B4 (Fig. 6). Further interrogation of deep sea sediments may thus well continue to turn up entirely new solutions to PET degradation that differ significantly from their terrestrial counterparts, and may yield even more tractable engineering solutions for PET valorization; e.g. the successful combination of GuaPA and B1, enzymes from a *Bathyarchaeia* and *Poribacteria*, respectively, may foreshadow functional syntrophic relationships between GB organisms that could be exploited by creating synthetic pathways for plastic degradation. As another example of how nature can lead biotechnology, combining the metabolic capacities of the GuaPA for TPA and B1 for EG could potentially improve not only the depolymerization of PET (Fig. 3), but also the downstream biotransformation of PET into glycerate and pyruvate, ultimately remediating the plastic while generating useful chemical feedstocks for additional transformations.

Acknowledgments

This research was funded by grants from the National Institute of General Medical Sciences (R35GM122480 to E.M.M.), the Army Research Laboratory (Cooperative Agreement Number W911NF-23-2-0089 to A.D.E.), the Simons Foundation (Investigator award LI-SIAME-00002001 to B.J.B.), and the Welch Foundation (F-1515 to E.M.M.; F-1654 to A.D.E.). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the US Government. The authors acknowledge additional support generously provided by Tito's Handmade Vodka and H-E-B's Our Texas Our Future initiative. The authors acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing high-performance computing resources that have contributed to the research results reported within this paper.

Supplementary material

Supplementary material is available at *The ISME Journal* online.

Conflicts of interest

The authors declare that they have no competing interests.

Funding

None declared.

Data availability

All sequence data and GB sample information are available at NCBI under BioProject ID PRJNA1112871. All raw data underlying phylogenomic (alignments and resulting phylogenetic trees) and

structural analyses (pdb structures) have been deposited into Zenodo ([10.5281/zenodo.14262448](https://doi.org/10.5281/zenodo.14262448)) along with the MAGs containing B1, B4, and GuaPA. Custom code and python scripts are available at github.com/marcottelab/GuaPA.

References

1. Plastics—the Facts 2022 • Plastics Europe. *Plastics Europe*. <https://plasticseurope.org/knowledge-hub/plastics-the-facts-2022-2>.
2. Yoshida S, Hiraga K, Takehana T. et al. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* 2016;351: 1196–9. <https://doi.org/10.1126/science.aad6359>
3. Knott BC, Erickson E, Allen MD. et al. Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proc Natl Acad Sci USA* 2020;117:25476–85. <https://doi.org/10.1073/pnas.2006753117>
4. Lu H, Diaz DJ, Czarnecki NJ. et al. Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* 2022;604:662–7. <https://doi.org/10.1038/s41586-022-04599-z>
5. Acosta DJ, Alper HS. Advances in enzymatic and organismal technologies for the recycling and upcycling of petroleum-derived plastic waste. *Curr Opin Biotechnol* 2023;84:103021. <https://doi.org/10.1016/j.copbio.2023.103021>
6. Werner AZ, Clare R, Mand TD. et al. Tandem chemical deconstruction and biological upcycling of poly(ethylene terephthalate) to β -ketoadipic acid by *Pseudomonas putida* KT2440. *Metab Eng* 2021;67:250–61. <https://doi.org/10.1016/j.ymben.2021.07.005>
7. Enzymatic recycling. Carbios. <https://www.carbios.com/en/enzymatic-recycling>.
8. Tournier V, Topham CM, Gilles A. et al. An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* 2020;580:216–9. <https://doi.org/10.1038/s41586-020-2149-4>
9. Cui Y, Chen Y, Liu X. et al. Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. *ACS Catal* 2021;11:1340–50. <https://doi.org/10.1021/acscatal.0c05126>
10. Eiamthong B, Meesawat P, Wongsatit T. et al. Discovery and genetic code expansion of a polyethylene terephthalate (PET) hydrolase from the human saliva metagenome for the degradation and bio-functionalization of PET. *Angew Chem Int Ed* 2022;61:e202203061. <https://doi.org/10.1002/anie.202203061>
11. Sonnendecker C, Oeser J, Richter PK. et al. Low carbon footprint recycling of post-consumer PET plastic with a metagenomic polyester hydrolase. *ChemSusChem* 2022;15:e202101062. <https://doi.org/10.1002/CSSC.202101062>
12. Teske A. The Guaymas Basin—a hot spot for hydrothermal generation and anaerobic microbial degradation of hydrocarbons. *Int Biodeterior Biodegradation* 2024;186:105700. <https://doi.org/10.1016/j.ibiod.2023.105700>
13. Perez-Garcia P, Chow J, Costanzi E. et al. *Commun Chem* 2023;6: 1–13. <https://doi.org/10.1038/s42004-023-00998-z>
14. Erickson E, Gado JE, Avilán L. et al. Sourcing thermotolerant poly(ethylene terephthalate) hydrolase scaffolds from natural diversity. *Nat Commun* 2022;13:7850. <https://doi.org/10.1038/s41467-022-35237-x>
15. Buchholz PCF, Feuerriegel G, Zhang H. et al. Plastics degradation by hydrolytic enzymes: the plastics-active enzymes database—PAZy. *Proteins* 2022;90:1443–56. <https://doi.org/10.1002/prot.26325>
16. Appler KE, Lingford JP, Gong X. et al. Oxygen metabolism in descendants of the archaeal-eukaryotic ancestor. *bioRxiv* 2024. <https://doi.org/10.1101/2024.07.04.601786>

17. BBMap. SourceForge. <https://sourceforge.net/projects/bbmap/>. (5 November 2024, date last accessed).
18. Li D, Liu CM, Luo R. et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; **31**:1674–6. <https://doi.org/10.1093/bioinformatics/btv033>
19. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**:605–7. <https://doi.org/10.1093/bioinformatics/btv638>
20. Alneberg J, Bjarnason BS, de Bruijn I. et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–6. <https://doi.org/10.1038/nmeth.3103>
21. Kang DD, Li F, Kirton E. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; **7**:e7359. <https://doi.org/10.7717/peerj.7359>
22. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**:158. <https://doi.org/10.1186/s40168-018-0541-1>
23. Sieber CMK, Probst AJ, Sharrar A. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 2018; **3**:836–43. <https://doi.org/10.1038/s41564-018-0171-1>
24. Teske A, de Beer D, McKay L. et al. The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front Microbiol* 2016; **7**. <https://doi.org/10.3389/fmicb.2016.00075>
25. Zhuang G-C, Montgomery A, Samarkin VA. et al. Generation and utilization of volatile fatty acids and alcohols in hydrothermally altered sediments in the Guaymas Basin, Gulf of California. *Geophys Res Lett* 2019; **46**:2637–46. <https://doi.org/10.1029/2018GL081284>
26. Gambarini V, Pantos O, Kingsbury JM. et al. PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database* 2022; **2022**:baac008. <https://doi.org/10.1093/database/baac008>
27. Steingger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017; **35**:1026–8. <https://doi.org/10.1038/nbt.3988>
28. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**:1792–7. <https://doi.org/10.1093/nar/gkh340>
29. Fu L, Niu B, Zhu Z. et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012; **28**:3150–2. <https://doi.org/10.1093/bioinformatics/bts565>
30. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; **22**:1658–9. <https://doi.org/10.1093/bioinformatics/btl158>
31. Lin Z, Akin H, Rao R. et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv* 2022; 2022.07.20.500902.
32. Bernhofer M, Rost B. TMbed: transmembrane proteins predicted through language model embeddings. *BMC Bioinform* 2022; **23**:326. <https://doi.org/10.1186/s12859-022-04873-x>
33. Teufel F, Almagro Armenteros JJ, Johansen AR. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022; **40**:1023–5. <https://doi.org/10.1038/s41587-021-01156-3>
34. Shen J, Yu Q, Chen S. et al. Unbiased organism-agnostic and highly sensitive signal peptide predictor with deep protein language model. *Nat Comput Sci* 2024; **4**:29–42. <https://doi.org/10.1038/s43588-023-00576-2>
35. Wang H, Bergeson AR, Lu H. et al. Evaluating the effectiveness of a lysine mutation and its portability across different poly(ethylene terephthalate)-hydrolyzing enzymes. *Biochem Eng J* 2025; **214**:109573. <https://doi.org/10.1016/j.bej.2024.109573>
36. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* 2023; **51**:D523–31. <https://doi.org/10.1093/nar/gkac1052>
37. Varadi M, Bertoni D, Magana P. et al. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* 2023; **52**:D368–75. <https://doi.org/10.1093/nar/gkad1011>
38. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**:772–80. <https://doi.org/10.1093/molbev/mst010>
39. Katoh K, Misawa K, Kuma K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**:3059–66. <https://doi.org/10.1093/nar/gkf436>
40. Steenwyk JL, Buida TJ, Li Y. et al. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* 2020; **18**:e3001007. <https://doi.org/10.1371/journal.pbio.3001007>
41. Nguyen L-T, Schmidt HA, von Haeseler A. et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32**:268–74. <https://doi.org/10.1093/molbev/msu300>
42. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 2001; **17**:282–3. <https://doi.org/10.1093/bioinformatics/17.3.282>
43. Geneious | Bioinformatics Software for Sequence Data Analysis. *Geneious*. <https://www.geneious.com>.
44. Minh BQ, Schmidt HA, Chernomor O. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020; **37**:1530–4. <https://doi.org/10.1093/molbev/msaa015>
45. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024; **52**:W78–82. <https://doi.org/10.1093/nar/gkae268>
46. Chaumeil P-A, Mussig AJ, Hugenholtz P. et al. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020; **36**:1925–7. <https://doi.org/10.1093/bioinformatics/btz848>
47. Darling AE, Jospin G, Lowe E. et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014; **2**:e243. <https://doi.org/10.7717/peerj.243>
48. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun* 2018; **9**:4999. <https://doi.org/10.1038/s41467-018-07418-0>
49. Hyatt D, Chen GL, LoCascio PF. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 2010; **11**:119. <https://doi.org/10.1186/1471-2105-11-119>
50. Schubert S, Schaller K, Bååth JA. et al. Reaction pathways for the enzymatic degradation of poly(ethylene terephthalate): what characterizes an efficient PET-hydrolase? *Chembiochem* 2023; **24**:e202200516. <https://doi.org/10.1002/cbic.202200516>
51. Li A, Sheng Y, Cui H. et al. Discovery and mechanism-guided engineering of BHET hydrolases for improved PET recycling and

- upcycling. *Nat Commun* 2023;14:4169. <https://doi.org/10.1038/s41467-023-39929-w>
52. Joo S, Cho IJ, Seo H. et al. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat Commun* 2018;9:382. <https://doi.org/10.1038/s41467-018-02881-1>
 53. Chai Discovery team, Boitreaud J, Dent J. et al. Chai-1: decoding the molecular interactions of life2024. *bioRxiv* 2024; 2024.10.10.615955.
 54. von Haugwitz G, Han X, Pfaff L. et al. Structural insights into (terephthalate-ester hydrolysis by a carboxylesterase and its role in promoting PET depolymerization. *ACS Catal* 2022;12: 15259–70. <https://doi.org/10.1021/acscatal.2c03772>
 55. Brinch-Pedersen W, Keller MB, Dorau R. et al. Discovery and surface charge engineering of fungal cutinases for enhanced activity on poly(ethylene terephthalate). *ACS Sustain Chem Eng* 2024;12:7329–37. <https://doi.org/10.1021/acssuschemeng.4c00060>
 56. Joho Y, Royan S, Caputo AT. et al. Enhancing PET degrading enzymes: a combinatory approach. *Chem Bio Chem* 2024; 25:e202400084. <https://doi.org/10.1002/cbic.202400084>
 57. Lobanov MY, Pereyaslavets LB, Likhachev IV. et al. Is there an advantageous arrangement of aromatic residues in proteins? Statistical analysis of aromatic interactions in globular proteins. *CSBJ* 2021;19:5960–8. <https://doi.org/10.1016/j.csbj.2021.10.036>
 58. Wilkes RA, Waldbauer J, Carroll A. et al. Complex regulation in a *Comamonas* platform for diverse aromatic carbon metabolism. *Nat Chem Biol* 2023;19:651–62. <https://doi.org/10.1038/s41589-022-01237-7>
 59. Rao TE, Imchen M, Kumavath R. Chapter Seven—Marine enzymes: production and applications for human health. In: Kim S.-K., Toldrá F. (eds.), *Advances in Food and Nutrition Research*, Academic Press, volume 80, 2017, 149–163. <https://doi.org/10.1016/bs.afnr.2016.11.006>
 60. Bell EL, Smithson R, Kilbride S. et al. Directed evolution of an efficient and thermostable PET depolymerase. *Nat Catal* 2022;5: 673–81. <https://doi.org/10.1038/s41929-022-00821-3>
 61. An archaeal lid-containing feruloyl esterase degrades polyethylene terephthalate. *Commun Chem*. <https://www.nature.com/articles/s42004-023-00998-z>
 62. Fieseler L, Horn M, Wagner M. et al. Discovery of the novel candidate phylum “Poribacteria” in marine sponges. *Appl Environ Microbiol* 2004;70:3724–32. <https://doi.org/10.1128/AEM.70.6.3724-3732.2004>
 63. Pérez Castro S, Borton MA, Regan K. et al. Degradation of biological macromolecules supports uncultured microbial populations in Guaymas Basin hydrothermal sediments. *ISME J* 2021;15:3480–97. <https://doi.org/10.1038/s41396-021-01026-5>
 64. Zhu X, Rochman CM, Hardesty BD. et al. Plastics in the deep sea—a global estimate of the ocean floor reservoir. *Deep-Sea Res I: Oceanog Res Pap* 2024;206:104266. <https://doi.org/10.1016/j.dsr.2024.104266>
 65. Choy CA, Robison BH, Gagne TO. et al. The vertical distribution and biological transport of marine microplastics across the epipelagic and mesopelagic water column. *Sci Rep* 2019;9:7843. <https://doi.org/10.1038/s41598-019-44117-2>
 66. Zhang D, Liu X, Huang W. et al. Microplastic pollution in deep-sea sediments and organisms of the Western Pacific Ocean. *Environ Pollut* 2020;259:113948. <https://doi.org/10.1016/j.envpol.2020.113948>
 67. Zrimec J, Kokina M, Jonasson S. et al. Plastic-degrading potential across the global microbiome correlates with recent pollution trends. *MBio* 2021;12:02155–21. <https://doi.org/10.1128/MBIO.02155-21>
 68. Almeida DV, Ciancaglini I, Sandano ALH. et al. Unveiling the crystal structure of thermostable dienelactone hydrolase exhibiting activity on terephthalate esters. *Enzym Microb Technol* 2024;180:110498. <https://doi.org/10.1016/j.enzmictec.2024.110498>
 69. Feng X, Gustafsson Ö, Holmes RM. et al. Multi-molecular tracers of terrestrial carbon transfer across the pan-arctic: comparison of hydrolyzable components with plant wax lipids and lignin phenols. *BG* 2015;12:4841–60. <https://doi.org/10.5194/bg-12-4841-2015>
 70. Philippe G, Sørensen I, Jiao C. et al. Cutin and suberin: assembly and origins of specialized lipidic cell wall scaffolds. *Curr Opin Plant Biol* 2020;55:11–20. <https://doi.org/10.1016/j.pbi.2020.01.008>