# Title: Rapid Evolution of Plastic-degrading Enzymes Prevalent in the Global Ocean

**Authors:** Intikhab Alam[1*], Nojood Aalismail[2], Cecilia Martin[2], Allan Kamau[1], Francisco J. Guzmán-Vega[5], Tahira Jamil[1], Afaque A. Momin[5], Silvia G. Acinas[3], Josep M. Gasol[3], Stefan T. Arold[5,6], Takashi Gojobori[1], Susana Agusti[4] and Carlos M. Duarte[1,2*].

**Affiliations:**

[1]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

[2]Red Sea Research Centre (RSRC) and Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

[3]Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar-CSIC, Pg. Marítim de la Barceloneta 37-49, 08003, Barcelona, Spain

[4]Red Sea Research Centre (RSRC) and Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

[5]King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Biological and Environmental Science and Engineering, Thuwal, 23955-6900, Saudi Arabia

[6]Centre de Biochimie Structurale, CNRS, INSERM, Université de Montpellier, 34090 Montpellier, France

.

*Correspondence to: intikhab.alam@kaust.edu.sa, carlos.duarte@kaust.edu.sa

**Short title: Plastic-degrading Enzymes Prevalent in the Global Ocean**

**Abstract:**

Estimates of marine plastic stocks, a major threat to marine life (1), are far lower than expected from exponentially-increasing litter inputs, suggesting important loss factors (2, 3). These may involve microbial degradation, as the plastic-degrading polyethylene terephthalate enzyme (PETase) has been reported in marine microbial communities (4). An assessment of 416 metagenomes of planktonic communities across the global ocean identifies 68 oceanic PETase variants (oPETase) that evolved from ancestral enzymes degrading polycyclic aromatic hydrocarbons. Twenty oPETases show predicted efficiencies comparable to those of laboratory-optimized PETases, suggesting strong selective pressures directing the evolution of these enzymes. We found oPETases in 90.1% of samples across all oceans and depths, particularly abundant at 1,000 m depth, with a strong dominance of *Pseudomonadales* containing putative highly-efficient oPETase variants in the dark ocean. Enzymatic degradation may be removing plastic from the marine environment while providing a carbon source for bathypelagic microbial communities.

Exponential growth in plastic production along with poor waste management practices have led to over 150 Tg plastic waste delivered to the ocean since 1950 (1), where it harms marine life, from zooplankton to whales (5). Synthetic plastic polymers are derived from oil hydrocarbons and designed to be durable in the environment (6, 7), being largely resistant to microbial degradation. However, a newly-evolved plastic-degrading enzyme, polyethylene terephthalate hydrolase (PETase), was recently discovered in a Japanese waste processing plant (8). This PETase was inferred to have evolved since 1970, when the polyethylene terephthalate polymer was introduced at industrial scale (9). The PETase catalyzes the hydrolysis of polyethylene terephthalate (PET) plastic to monomeric mono-2-hydroxyethyl terephthalate (MHET), which is ultimately broken down into non-hazardous monomers, terephthalate and ethylene glycol, by MHETases. The efficiency of currently known PETases seems, however, to be low, and it has been suggested that in the environment it only achieves a breakdown from micro- to nanoplastics, rather than full degradation (10).

The discovery, and subsequent biotechnological improvement (11) of this spontaneously-evolved plastic-degrading enzyme offers hope that plastic waste can be degraded in waste facilities, thereby preventing it from entering the ocean (12). However, this still leaves the fate of the > 150 Tg plastic waste already delivered to the ocean unresolved. Recent estimates show that 99% of the plastic that entered the oceans cannot be accounted for, since inventories of plastic floating on the surface of the ocean account to ~ 1% of the expected load (2), pointing at the operation of processes responsible for major losses of plastic at sea (3). One such process may involve the microbial degradation of plastics. Microplastics, the numerically dominant marine debris, have developed a characteristic microbial community, referred to as the marine plastisphere (10). If the PETase enzyme evolved in waste treatment plants from ancestral *Ideonella sakaiensis* (*I. sakaiensis*) genes (8), and provided the current prevalence of polyethylene terephthalate as debris in the marine environment, it is possible that other PETases may have evolved independently in plastisphere ocean microbes as well. The huge population of prokaryotes in the global ocean, estimated to contribute $10^{29}$ cells distributed among $10^{10}$ different OTUs (13), provides enough

population size, genetic diversity and rate of growth for mutations(14) to independently evolve PETase activity, which would enable microorganisms to access a novel, widely available organic substrate. Indeed, a recent assessment of PETase abundance in marine and terrestrial environments, confirmed the presence of the PETase gene in 31 of the 108 environmental marine samples examined, all from surface waters (4). However, a global assessment of the presence, abundance and diversity of PETases in the whole ocean, extending to the deep ocean, the largest microbial habitat in the biosphere (15), is yet to be conducted.

## Results

Here we report the prevalence of PETases in the global ocean microbiome and provide a phylogeny suggesting Oceanospirillales to be the parental clade leading to 68 identified variants of marine PETase. We do so on the basis of a thorough exploration of PETase through gene catalogs from two global-ocean expeditions, *Tara* Oceans (16) and Malaspina (17), delivering 416 metagenomes collected from 204 individual ocean locations across the Indian, Pacific and Atlantic Oceans from surface to the deep sea (Fig. 1).

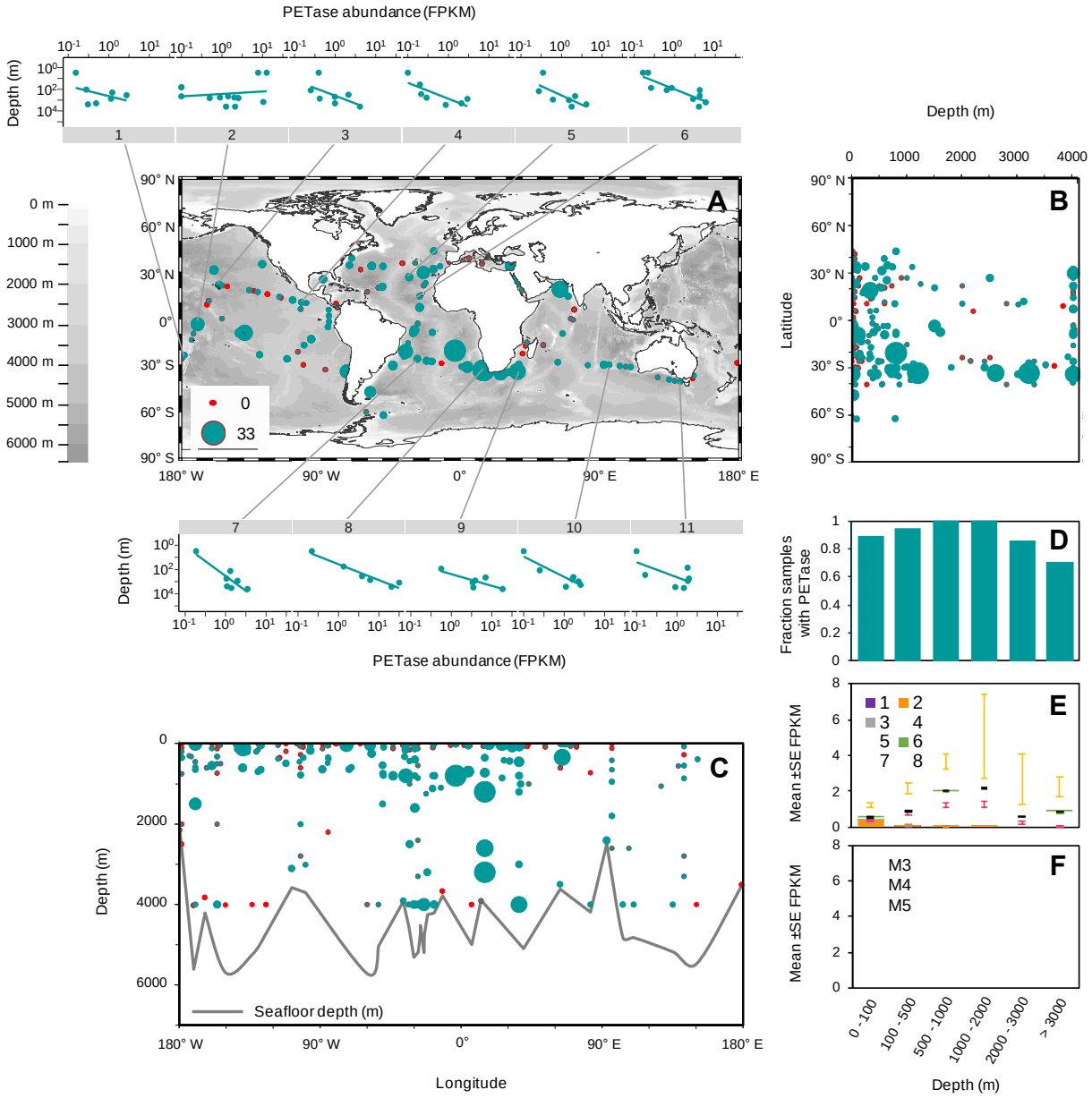



**Fig. 1** (**A**) Global distribution of PETase (in FPKM) in marine metagenomes. The scale also applies to panels B and C. The 11 small plots show the distribution of PETase genes across depth (m) in 11 profiles collected during the Malaspina Expedition. **(B)** Distribution of PETase along depth and latitude and **(C)** along depth and longitude. The grey line in panel C shows the seafloor depth (m) as it compares to the deepest sampling point. **(D)** Fraction of samples with PETase and mean ±SE PETase abundance (in FPKM) per **(E)** taxonomic group and **(F)** Abundance of oPETases assigned with Motif groups M3-5 (set of motifs designed to capture PETase characteristic catalytic triad and other features) at 6 depth layers (0 - 100 m, 100 - 500 m, 500 - 1000 m, 1000 - 2000 m, 2000 - 3000 m and > 3000 m). Taxonomic groups, identified as numbers from 1 to 8 in panel E, correspond to the following bacterial orders: 1. Candidatus Poribacteria_UO; 2. Flavobacteriales; 3. Gammaproteobacteria_UO; 4. Gemmatimonadetes_UO; 5. Oceanospirillales; 6. Phycisphaerales; 7. Planctomycetales; 8. Pseudomonadales.

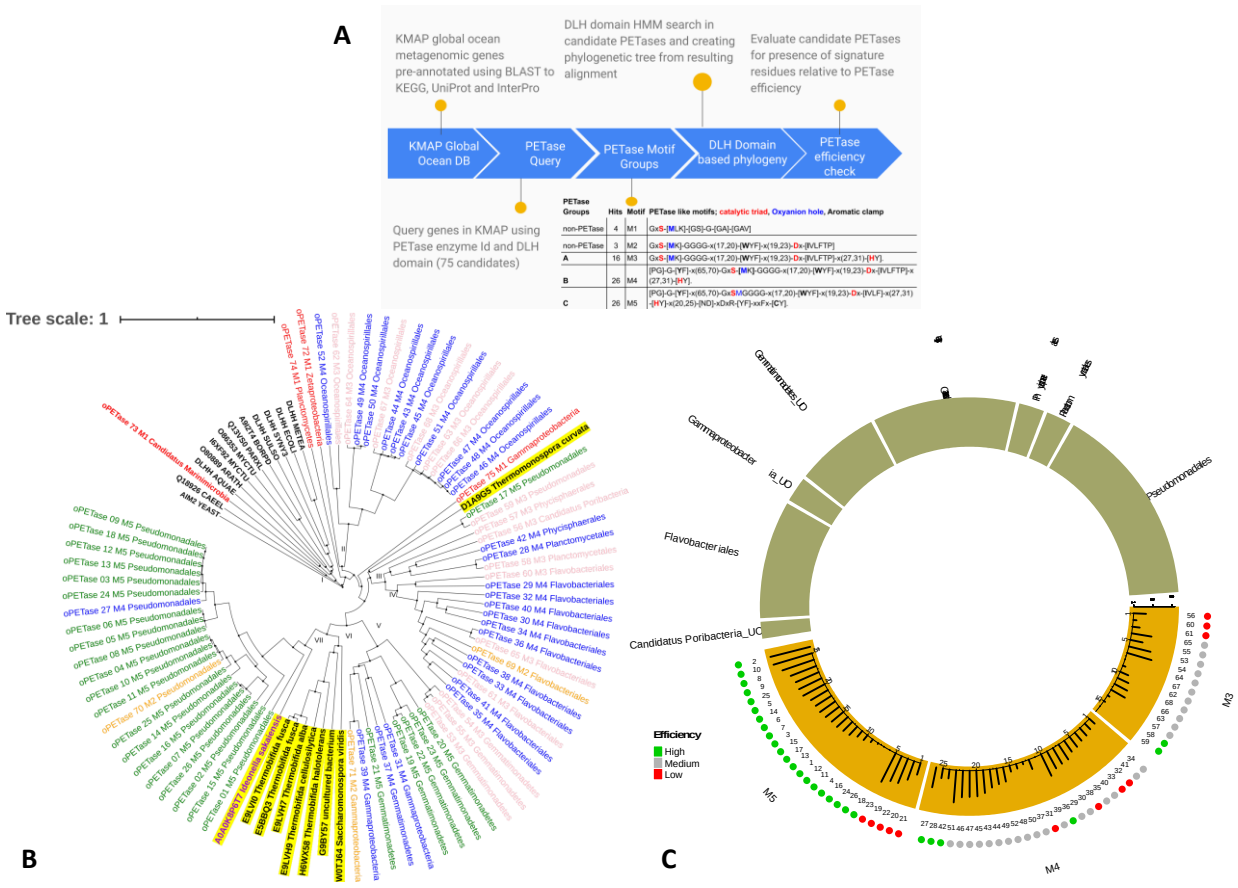
Using the KAUST Metagenomic Analysis Platform (KMAP) we queried our pre-annotated *Tara* Oceans and Malaspina gene catalogs for the presence of *Ideonella sakaiensis* PETase (EC 3.1.1.101)-like enzymes containing the Dienelactone hydrolase (DLH) domain. DLH is a generic α/β-hydrolase domain present in many hydrolases, but in the case of PETase it contains a signature catalytic triad (https://pfam.xfam.org/protein/PETH_IDESA, Fig 2a) (4). The DLH domain is present in more than 17,000 sequences in the Pfam database (https://pfam.xfam.org/family/PF01738, accessed Feb 2, 2020). However, most of these sequences are related to genes coding for enzymes involved in the degradation of hydrocarbons. Only one DLH form, harboring specific substitutions in the catalytic triad (motif residues S, D and H) has been experimentally confirmed to degrade PET (4), Fig 2a).

## Diversity and efficiency of oceanic PETase

Our analysis yielded 75 candidate PETases in the *Tara* Oceans and Malaspina gene catalogues. These gene catalogs were annotated based on full sequence comparisons to public databases. Based on manual inspection of alignments from these 75 oceanic candidate oPETases and 8 known homologs (4) of *I. sakaiensis* PETase we defined five overlapping generic to specific PHI-BLAST searchable motifs, M1-5 (Fig 2a, 2d). Motifs M1 and M2 detect PETase-like sequences but miss the catalytic triad, the main component of known PETases (4). Motif M3 include the catalytic triad, conserved central motif GGGG as well as the oxyanion hole residues and the aromatic clamp. Two exceptions are oPETase 72 and 75 where catalytic triad appears to be available but the conserved motif GGGG is modified to GG-[AG]-[GA]. M4 is similar to M3 but contains the [PG]-G-[YF] motif that defines an additional aromatic clamp and oxyanion hole upstream of the catalytic triad. M5 is the largest motif. It includes M3 and M4 and additionally contains a conserved DxDxR(Y)xxF(L)C sequence, preceding a region providing thermostability to the enzyme (Fig 2a, 2d). Because of the requirement for the catalytic triad of PETases degrade PET, putative PETases containing only motifs M1 and M2 were discarded. Only those with motifs including the catalytic triad (M3, M4 or M5) were considered as functional PETases. This selection yielded 68 variants of putative functional PETases present in the *Tara* Ocean and Malaspina gene catalogues (Suppl. Table 1). Alignments of these oPETases from motif groups M3-5 and shown in Figure 2-suppliment 1.

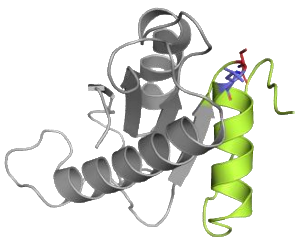
Scoring of candidate oPETases (see Methods and Suppl. Table S2-3) consistently designated most of the oPETases containing Motif 5 as having a predicted efficiency score close to or slightly above that of the reference *I. sakaiensis* PETase (4), which also contains M5 (Suppl. Table S2). Most of the key residues that are essential for PETase activity are largely conserved across the 68 functional variants. Most of the oPETase variants predicted to be efficient contained M5 (Fig. 2c). Moreover, we found that 20 of the newly identified oPETases, many of which had the M5 motif, already displayed residue substitutions that were identified as enhancing PETase activity in engineered laboratory variants of *I. sakaiensis* PETases (18) (Suppl. Table S3).

To support and inform the scoring process, high-confidence structural homology models were alos produced for all putative PETases, using SWISS-MODEL (19), based on up to 50% sequence identity with the crystal structure of the IsPETase in a complex with HEMT (PDB ID 5xh3 (20)). The structural analysis of potential oPETases can be described using one of the high scoring oPETase example such as oPETase_02. This selected putative PETase contains the catalytic triad (S160, D206, H237) and the disulfide bond linking residues C203 and C239, demonstrated to be essential for PETase function (20, 21)(Fig. 2d). In addition, it harbors two modifications that were shown to increase the catalytic activity of IsPETase (22), namely R90A and L117F. The R90A modification is thought to reduce the steric hindrance around the active site and increase the hydrophobic area to facilitate the accumulation of substrate near the substrate-binding cleft. The L117F modification stabilizes Y87 (F87 in this case), which enhances the interaction between the enzyme and the substrate, as shown by a lower Michaelis constant, *Km* of the L117F mutant as compared to the wild-type PETase. The S214H variant found in some oPETases has been seen to reduce PETase activity, possibly because the larger His restricts the movement of the adjacent W185 (20). Also, other variants that have not been experimentally tested were found in these oPETases (see Supp. Table S1), and their effect on the catalytic activity of the protein remains uncertain. The variants Y87F, T88V and S238Y/F increase the hydrophobicity around the active site, which is expected to be beneficial for PETase activity, as were other substitutions increasing the hydrophobicity of this region, as they may facilitate the interaction of the enzyme with the substrate. The R280Q substitution in most of the higher-scoring variants is conservative and not expected to influence catalysis, and other important positions such as W159, W185 and M161 remain unchanged in these variants. Overall, the compound effects of these modifications make these variants good PETase candidates with a chance of having a PET degrading activity close to or potentially better than IsPETase. It is however important to note that our predicted efficiency scores of these oPETases to rank potentially better or worse cases are adapted based on the presence of favorable or inhibiting mutations only, and that an experimental validation is needed to confirm real efficiency of these oPETases.

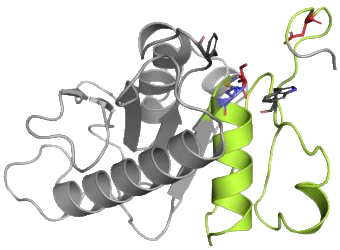



170

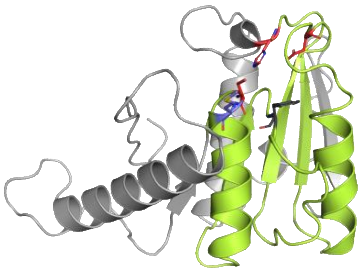
oPETase 73 – M1



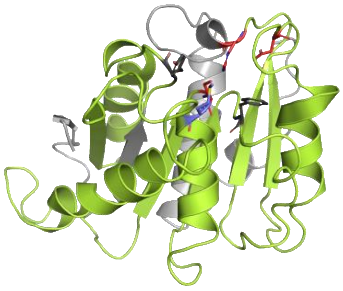
oPETase 71 – M2



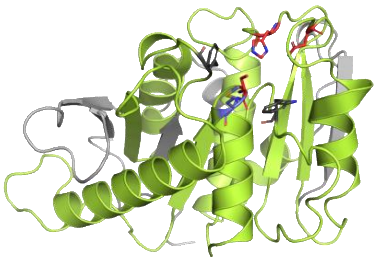
oPETase 64 – M3



oPETase 52 – M4



oPETase 02 – M5



D

171

**Fig. 2**. Overview, Phylogenetics and efficiency analysis of candidate Ocean PETases (oPETases). **(A)** An overview of our workflow in exploring oPETases. **(B)** Phylogenetic analysis of DLH-containing PETases. Here sequence names are colored according to associated Motifs M1-5 (M1 in red, M2 in orange, M3 in pink, M4 in blue and M5 in green). DLH seed sequences are in black, similarly known PETases are in black but highlighted in yellow, the original *I. sakaiensis* PETase sequence is in purple and highlighted in yellow. Roman numbers indicate the 7 clades in which PETases are grouped **(C)** Analysis of candidate oPETases is shown for motif groups M3, M4 and M5 (lower half of the circle) presenting potential for efficiency (outer lower circle) as green, grey and red dots depicting high, medium and low efficiency categories, respectively. For potential efficiency to degrade PET, based on the evaluation of only critical residues, see Suppl. Table S2 and efficiency scoring scheme. Total gene abundance for each of these oPETases in ocean is shown in respective bar graphs with orange background. Associated taxonomic information (upper half of the circle) is linked to oPETase motif groups through colored ribbons (M3 is green, M4 is blue and M5 is shown in pink). (D) An illustration of protein structures representative of each motif category, with the corresponding motif M1 to M5 colored in lime green. The models were obtained by SWISS-MODEL based on the crystal structure of IsPETase in complex with HEMT (PDB ID 5xh3) as a template. Key residues are shown as stick models, color-coded as follows: red, the catalytic triad; blue, the Met residue involved in the oxyanion hole; dark grey: residues forming the aromatic clamp, where [YF]87 is part of both the aromatic clamp and the oxyanion hole.

## Phylogeny and evolution of efficient oceanic PETase

We established a phylogenetic tree of all PETase sequences, including all Motifs 1 to 5, to investigate the evolutionary pathway from the ancestral DLH domain to fully functional PETase domains including the signature catalytic triad. The phylogenetic tree included the 75 candidate PETases identified in the *Tara* Oceans and Malaspina gene catalogues, the original *I. sakaiensis* PETase sequence (8), the eight land-derived homologues reported from the uniprot database (4), and 12 non-PETase DLH domain core sequences from the Pfam database as an outgroup (Fig. 2b). The resulting phylogenetic tree identified the 12 non-PETase DLH domain core sequences together with two out of three of those containing Motif 1 as a parental group from which PETases evolved, hereafter defined as clade I (Fig. 2b). All other clusters defining six additional taxonomic PETase clades of fully-functional PETases containing M3 to M5 (Fig. 2b), as well as three sequences containing non-functional PETase M2 motifs, interspaced among M3 to M5-containing sequences (Fig. 2b). Clade II is dominated by Oceanospirialles, Clade III contains a mixture of taxa, Clade IV contains Flavobacteriales, Clade V contains Gemmatimonadetes and Gammaproteobacteria, Clade VI contains the eight previously defined homologues of *I sakaiensis* PETase from land microbes (4), with an additional such homologue (thermomonospora curvata, D1A9G5) clustering with clade III (Fig. 2b). Cluster VII contains the original *I. sakaiensis* PETase sequence along with Pseudomonadales sequences, all except two, containing the Motif 5 (Fig. 1b).

In summary, the order of clusters in the phylogenetic tree represents a putative evolutionary progression from DLH-containing ancestors to the eight bacterial orders containing putatively functional oPETases (Fig. 2b). Specifically, the phylogenetic tree obtained suggests that PETases evolved from M1 and M2 DLH-containing sequences of genes involved in the degradation of naturally-occurring hydrocarbons present in the ocean prior to synthetic PET addition to the ocean. The PETases appear to have evolved in parallel in six major independent clades (clades II to VII), all of which include functional PETase motifs (M3, M4 and M5). The phylogenetic tree shows an enrichment of the clades in Motif 5 from clade V, where it first appears, to the clade VII, where all (except one *Pseudomonadales* sequence) include M5. Our analysis points at

*Pseudomonadales* containing efficient variants of PETase, as the fundamental outcome of the processes driving the evolution of efficient PETase in the ocean, here oPETase variants 2 and 10 were found to be most abundant (Fig 2c).

## Distribution and abundance of PETases in the global ocean

PETases were prevalent in marine metagenomes, present in 90.1% of the samples assessed, across all oceans and depths sampled (Fig. 1). PETases were present from the surface to the deep sea (Fig. 1b, c), with both prevalence and abundance reaching a maximum at the base of the permanent thermocline (1,000 m depth, Fig. 1d, e), thereby showing a tendency to increase toward the ocean interior (mixed-effect model, slope for depth = 2.979, P = 0.0133, 11 plots in Fig. 1a). This analysis is based on 11 sampling points, where samples were collected at different depths leading to 11 profiles, 10 of which show that PETase abundance increases with depths. PETase abundance in the upper ocean (< 200 m) was highest around South America, South Africa, India, French Polynesia and the Red Sea (Fig. S2a), and PETase abundance in the deep-sea was greatest in the Atlantic Ocean, followed by the Indian Ocean (Fig. S2). Some of these areas, including Southern Africa (23) and Indian subcontinent (24), have been reported to be areas of high abundance of microplastics.

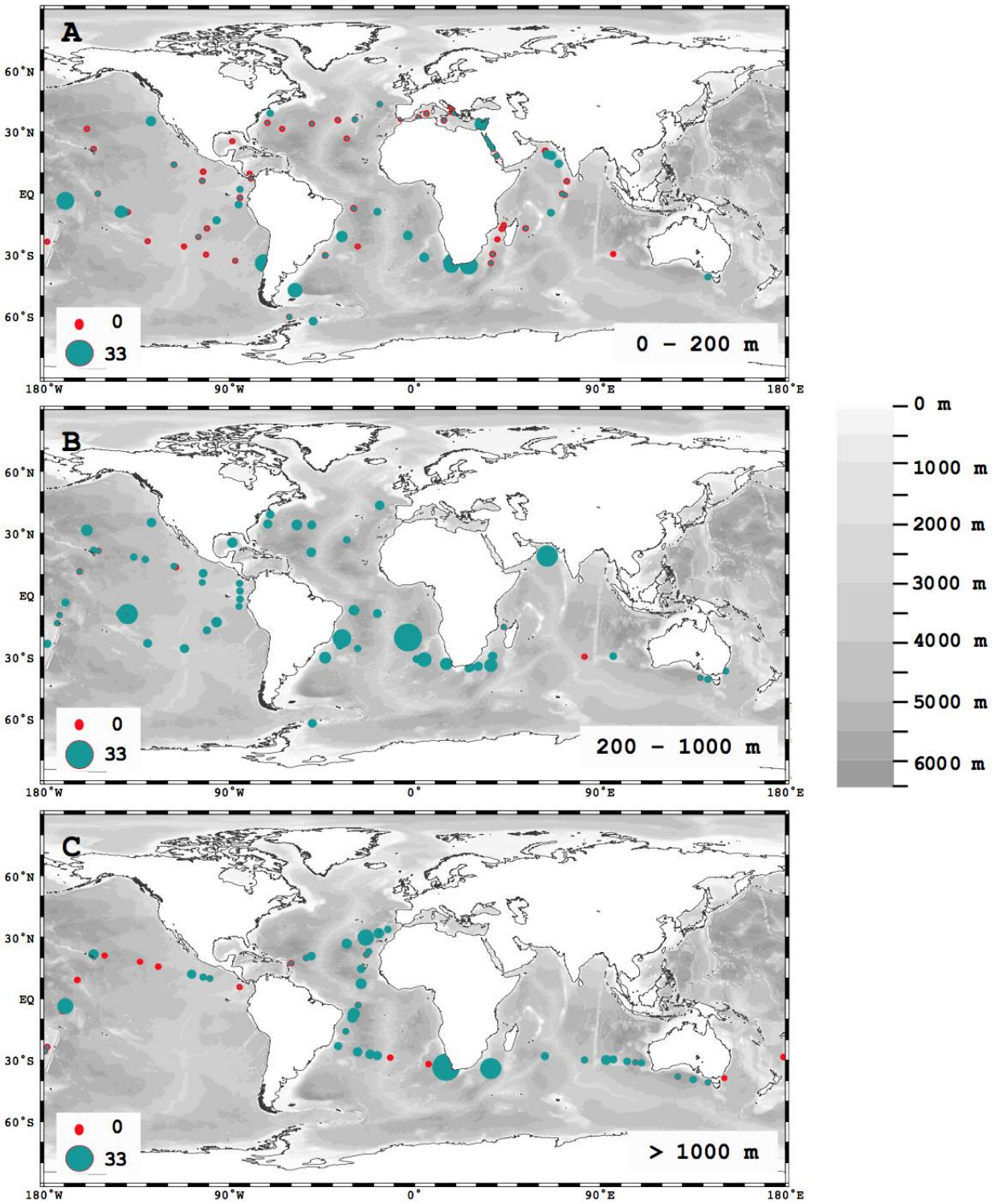



**Fig. 3**. Global distribution of PETase (in FPKM) in marine metagenomes of **(A)** surface (0 - 100 m), **(B)** mesopelagic (200 - 1,000) and **(C)** deep waters (> 1000m).

The distribution of PETase contained within *Proteobacteria* was relatively uniform across depth layers (Fig. 1e), however the distribution of PETase contained within *Gemmatimonadetes* increases with depth up to 2,000 m and declines slightly in deeper waters. PETases contained

within *Bacteroidetes* were restricted to water depths shallower than 1,000 m (Fig. 1e). PETase-containing *Proteobacteria*, the prokaryote class with the highest abundance and prevalence of PETases, was, in turn, dominated by *Pseudomonadales,* containing efficient M5 motifs, and *Oceanospirillales* orders, with an unidentified order within *Gemmatimonadetes* also contributing considerably to PETase abundance in the mesopelagic ocean (Fig. 1e, Table S1). There was no clear difference in PETase abundance and prevalence between particle attached (> 3 μm) and free-living communities (<3 μm, Wilcoxon Rank Sum Test, W = 2892.5, p-value = 0.46). PETase variants containing the efficient M5 module were dominant across all depth layers (Fig. 1f), with PETase variants containing the M3 variant being the least abundant in oceanic microbial communities (Fig. 1f).

The taxonomy of oPETases differs from that of PETase containing bacteria in other environments. A previous search of 8 homologues of *I. sakaiensis* PETases, all containing the M5 motif, in the public NCBI non-redundant protein database retrieved ~500 PETase-like sequences (4), largely corresponding to microbial communities sampled from terrestrial habitats and related to *Actinobacteria*. An updated search in NCBI (accession date, December 2019), resulted in twice the number of sequences (1138 sequences), also mostly corresponding to soil *Actinobacteria* (1040 sequences). However, ~86 hits appear to be related with proteobacteria (Suppl. Table 4). Our analysis of the phylogeny and abundance of oPETases compared with the results from NCBI clearly show that the PETase-containing ocean microbial community differs, and that PETases likely evolved independently on land and in the ocean.

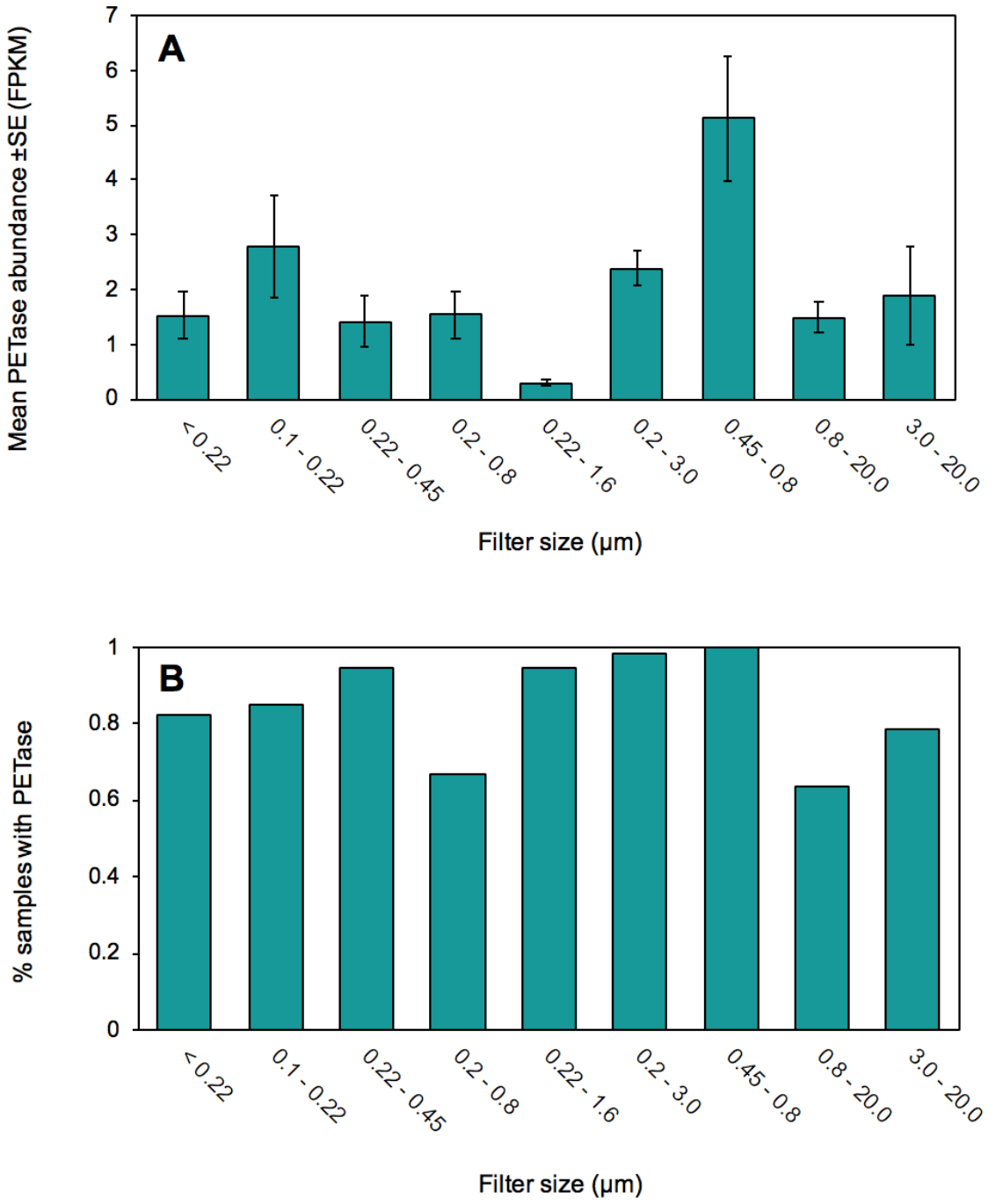



Fig. 4. **(A)** Mean PETase abundance ±SE (FPKM) and **(B)** prevalence of samples where PEtase occurred in each of 9 filter sizes (µm).

Our sampling of metagenomes does not allow to directly link the PETase containing microbes to plastic particles. However, our samples are likely to all contain plastic particles. Indeed, as the global metagenomes assessed were retrieved by filtering large volumes of seawater onto filters of different sizes (Fig. 4), any plastic particle, whether marine micro- (< 5 mm) or nanoplastics (< 0.1 µm) (25), present in the water would also be retained. Especially, filters would have retained synthetic fibers (about 0.2 to several mm long and less than 50 µm in diameter), which

are the most abundant debris in the ocean, present in the water at concentrations of about 100 to 300 fibers per liter (26). A review of the microbial communities found in the oceanic plastisphere shows that *Pseudomonadales* and *Flavobacteriales*, which our analysis showed that may contain the efficient M5 and M4/M3 PETase, respectively (Fig. 2c), have been detected to form biofilms on marine microplastics sampled in the North Atlantic (27) and North Sea (28). In addition, *Gammaproteobacteria*, which our analysis showed contain the M4-containing PETases (Fig. 2c), have been also reported growing in microbial films on marine microplastics from the Baltic Sea (29), the Caribbean Sea (30), at the coastline of Singapore (31), and through the great Pacific garbage patch (32). However, *Oceanospirillales*, which our analysis showed contain M3 and M4 variants of the PETase, have been identified only in marine plastic-associated bacterial communities in the North Atlantic (27).

In summary, the results obtained confirm that PETases are prevalent and abundant in the global ocean microbiome, where they must have evolved recently, following the mass production of polyethylene terephthalate (9). PETases are widespread from the surface to the deep ocean, reaching a maximum abundance and prevalence at about 1,000 m, where *Pseudomonas* has evolved an efficient PETase. PETases were found in high abundance in some areas known to receive high loads of plastic waste, such as waters off the Indian subcontinent, Brazil, and Southern Africa. The phylogeny constructed also shows a putative pathway for the evolution of oPETases from ancestral genes involved in hydrocarbon degradation. The increased efficiency of the PETase along the phylogenetic sequences, along with the high abundance of bacteria containing efficient PETases, points at an ongoing evolutionary process driven by selection pressures providing advantages to bacteria able to use an increasingly available resource, plastic polymers, in the deep ocean, mainly where all other naturally-occurring organic substrates are extremely diluted (33). The widespread prevalence, 90.1%, of PETases across the ocean together with the acute carbon limitation in bathypelagic waters suggests that the ocean microbiome is rapidly evolving to degrade plastic waste, providing a hopeful, nature-based solutions for plastic already in the marine environment.

## Materials and Methods

### Data collection and DMAP based annotation

Metagenome data was retrieved from two global expeditions, the TARA Oceans expedition, which sampled the upper layer of the ocean from sail-boat *Tara* (16) and the Malaspina Expedition, which sampled open ocean waters extending down to 4,000 m depth from spanish R/V *Hesperides* (17). The metagenomic data-sets included 243 metagenome samples from TARA and 173 metagenomes from the MALASPINA expedition, ranging from surface down to 4,000 m depth and including samples from all oceans (Fig. 1). These samples were used to investigate the prevalence of PETase genes. Samples were filtered through different filter sizes (<0.22, 0.1 - 0.22, 0.22 - 0.45, 0.2 - 0.8, 0.22 - 1.6, 0.2 - 3, 0.45 - 0.8, 0.8 - 3.0, 3.0 - 20.0 3 μm) to explore the abundance and prevalence of PETase genes in each fraction and especially in particle attached (> 3 μm) and free-living communities (< 3 μm). Non-redundant Gene catalogs from MP and MD samples were obtained by assembly using MegaHit (34) and gene prediction for each of the samples, followed by pooling of sample specific genes through CD-HIT based clustering (95% nucleotide identity and 80% overlap of the shorter sequence). TARA oceans

317 microbial gene catalog was already available with 40 million genes. MP and MD gene catalogs
318 produced a non-redundant set of 32 and 11 million genes, respectively. These gene catalogs were
319 annotated using DMAP where taxonomic affiliation of every gene is investigated using Least
320 Common Ancestor estimation through BLAST against UniProtKB database. Functional
321 annotation is obtained through BLAST matches to KEGG sequences with defined KEGG
322 Orthologs (KO) and identification of domains using InterPro (35). These Taxonomic and
323 Functional annotations including cross references, such as Enzyme Classification (E.C), Gene
324 Ontology (GO) and BLAST statistics (Evalue, Percent Identity and Percent Coverage) are saved
325 in to Gene Information Table (36) format, ready to be indexed for browsing and comparison
326 through DMAP compare module. To estimate gene abundance counts as Fragments per Kilobase
327 normalized to one Million (FPKM), reads were mapped on to their respective gene catalogs from
328 individual samples in strictly paired mode using bbmap.

329

330 **Analysis of plastic degrading-related genes**

331 **PETase abundance in global ocean metagenomes**

332 We queried TARA, MP and MD gene catalogs to obtain list of non-redundant genes considering
333 enzyme label 3.1.1.101, KEGG Ortholog (KO) K21104 restricted to hits with Pfam signature
334 domain 'Dienelactone hydrolase family' with PFAM ID PF01738. Table S1 shows the count of
335 non-redundant genes found as a result of this search. This same query was used in DMAP to
336 obtain gene abundance estimates from individual samples.

337

338 **Phylogeny analysis**

339

340 A protein sequence set of all non-redundant PETase related 75 genes, obtained through querying
341 DMAP, was combined with *I. sakaiensis* PETase gene, its 8 homologs (from (4)) and as an
342 outgroup 12 seed sequences from DLH domain were included. An alignment of sequences was
343 performed using hmmsearch with DLH domain as a query. Resulting alignment was converted
344 into fasta format and a phylogenetic tree was built using mafft phylogeny option at
345 https://mafft.cbrc.jp/alignment/server/phylogeny.html . The alignment was colored based on
346 different motif groups, *Ideonella sakaiensis* PETase with homologs and seed sequences from
347 DLH domain.

348

349 **Efficiency scoring and homology modelling**

350

351 A scoring system was developed to rank the candidate oPETases, 8 known PET hydrolases, and
352 12 seed sequences containing the DLH domain. This scoring was made by comparison to the
353 PETase sequence from *I. sakaiensis* (IsPETase), taking into account some of the key residues
354 that are known to contribute to the activity and stability of IsPETase, as well as variants and
355 mutations that have been experimentally tested and yielded either an increase or decrease in its
356 ability to degrade PET (Supplementary Methods and Suppl. Table S2-3). The scoring was then
357 performed with the following considerations: Given the difficulties to predict how the presence

of additional varying residues in the sequence with respect to IsPETase might affect the performance of the experimentally tested mutations as reported in the literature, the scoring was applied as a count or sum of the number of positive mutations (i.e. enhancing PETase activity) minus the number of negative mutations (i.e. inhibiting PETase activity). Additionally, a higher weight was given to the residues composing the catalytic triad and the PETase-specific disulfide bond, and a higher penalty to their absence. A labelling of PETases based on resulting score binned into low, medium and high efficiency, depicted by red, grey and green colors, respectively, is shown in Fig. 1c. After several rounds of scoring varying different parameters, such as the number of mutations/key amino acids considered and the weights given to each residue, the ranking of the analyzed sequences was proven to be robust.

**Acknowledgments:** We are thankful to KAUST Supercomputing Laboratory (KSL) for providing computational resources need to carry out this work, **Funding:** The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR), under award number URF/1/1976-21-01, and the Malaspina Circumnavigation Expedition funded by the Spanish Ministry of Economy and Competitiveness (CSD2008-00077),; **Author contributions:** IA and CMD conceived the study, SGA. JMG, TG, CMD and SA contributed data, IA, AK, CMD, NA, CM and TJ analyzed the metagenome data, FJGV, AAM and STA conducted the structural modelling and efficiency assessment, IA, CMD, NA and CM wrote the first draft of the paper and all coauthors contributed to improve the paper and approved the submission; **Competing interests:** Authors declare no competing interests; and **Data and materials availability:** All data are available (see sequence data information in M&M).

382

383 References

384

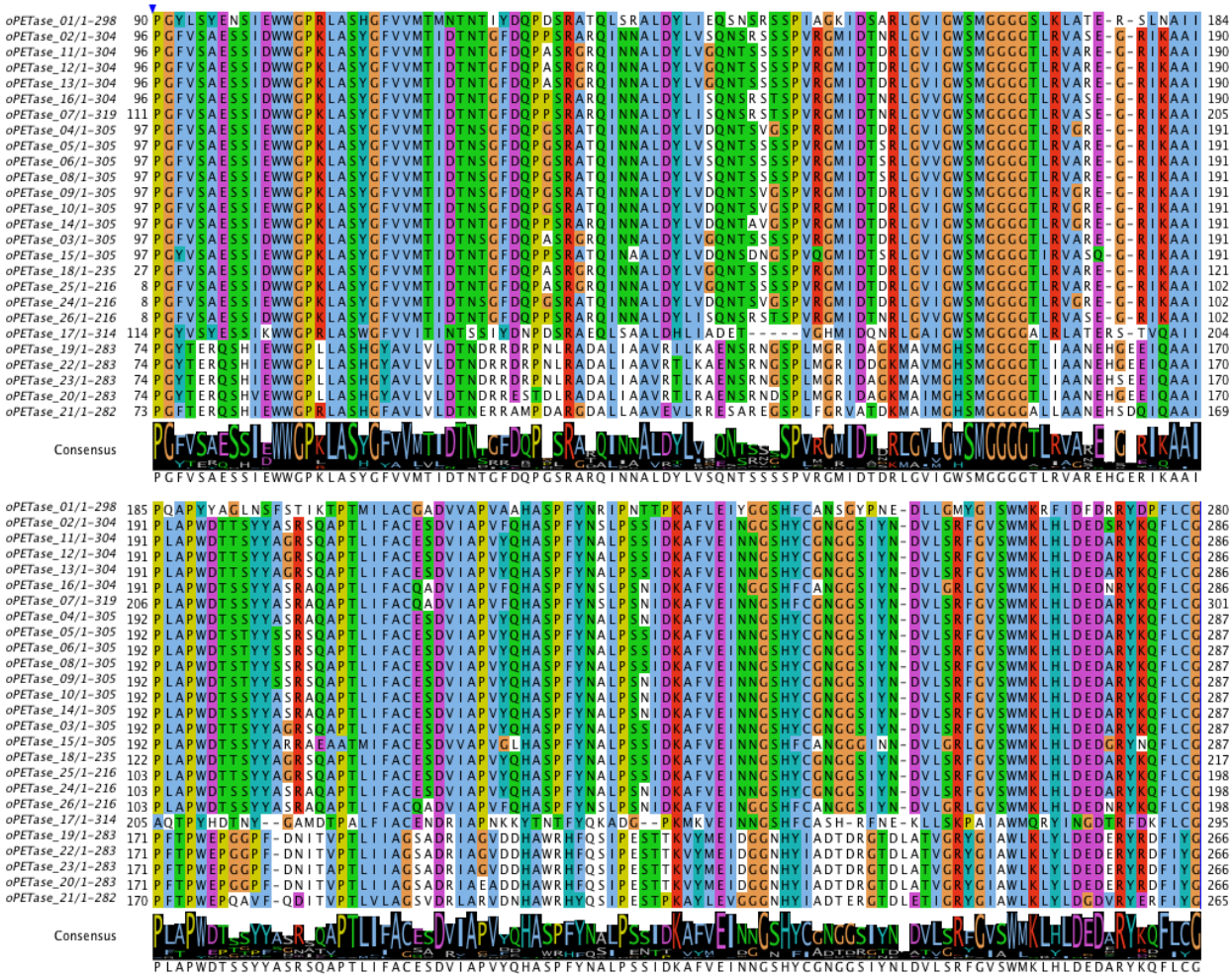
385 1. F. Gallo *et al.*, Marine litter plastics and microplastics and their toxic chemicals
386 components: the need for urgent preventive measures. *Environ Sci Eur* **30**, 13 (2018).
387 2. E. van Sebille *et al.*, A global inventory of small floating plastic debris. *Environmental*
388 *Research Letters* **10** (2015).
389 3. M. W. Ryberg, M. Z. Hauschild, F. Wang, S. Averous-Monnery, A. Laurent, Global
390 environmental losses of plastics across their value chains. *Resources, Conservation and*
391 *Recycling* **151** (2019).
392 4. D. Danso *et al.*, New Insights into the Function and Global Distribution of Polyethylene
393 Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial
394 Metagenomes. *Appl Environ Microbiol* **84** (2018).
395 5. S. C. Gall, R. C. Thompson, The impact of debris on marine life. *Mar Pollut Bull* **92**,
396 170-179 (2015).
397 6. D. K. Barnes, F. Galgani, R. C. Thompson, M. Barlaz, Accumulation and fragmentation
398 of plastic debris in global environments. *Philos Trans R Soc Lond B Biol Sci* **364**, 1985-
399 1998 (2009).
400 7. B. Gewert, M. M. Plassmann, M. MacLeod, Pathways for degradation of plastic
401 polymers floating in the marine environment. *Environ Sci Process Impacts* **17**, 1513-
402 1521 (2015).
403 8. S. Yoshida *et al.*, A bacterium that degrades and assimilates poly(ethylene terephthalate).
404 *Science* **351**, 1196-1199 (2016).
405 9. H. P. Austin *et al.*, Characterization and engineering of a plastic-degrading aromatic
406 polyesterase. *Proc Natl Acad Sci U S A* **115**, E4350-E4357 (2018).
407 10. L. A. Amaral-Zettler, E. R. Zettler, T. J. Mincer, Ecology of the plastisphere. *Nat Rev*
408 *Microbiol* **18**, 139-151 (2020).
409 11. M. Furukawa, N. Kawakami, A. Tomizawa, K. Miyamoto, Efficient Degradation of
410 Poly(ethylene terephthalate) with Thermobifida fusca Cutinase Exhibiting Improved
411 Catalytic Activity Generated using Mutagenesis and Additive-based Approaches. *Sci Rep*
412 **9**, 16038 (2019).
413 12. M. U. Sheth *et al.*, Bioengineering a Future Free of Marine Plastic Waste. *Frontiers in*
414 *Marine Science* **6** (2019).
415 13. K. J. Locey, J. T. Lennon, Scaling laws predict global microbial diversity. *Proc Natl*
416 *Acad Sci U S A* **113**, 5970-5975 (2016).
417 14. D. L. Kirchman, Growth Rates of Microbes in the Oceans. *Ann Rev Mar Sci* **8**, 285-309
418 (2016).
419 15. B. N. Orcutt, J. B. Sylvan, N. J. Knab, K. J. Edwards, Microbial ecology of the dark
420 ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* **75**, 361-422 (2011).
421 16. E. Karsenti *et al.*, A holistic approach to marine eco-systems biology. *PLoS Biol* **9**,
422 e1001177 (2011).
423 17. C. M. Duarte, Seafaring in the 21st century: the Malaspina 2010 circumnavigation
424 expedition. *Limnology and Oceanography Bulletin* **24**, 11-14 (2015).
425 18. H. Seo *et al.*, Production of extracellular PETase from Ideonella sakaiensis using sec-
426 dependent signal peptides in E. coli. *Biochemical and Biophysical Research*
427 *Communications* **508**, 250-255 (2019).

19. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303 (2018).
20. X. Han *et al.*, Structural insight into catalytic mechanism of PET hydrolase. *Nat Commun* **8**, 2106 (2017).
21. S. Joo *et al.*, Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat Commun* **9**, 382 (2018).
22. Y. Ma *et al.*, Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering. *Engineering* **4**, 888-893 (2018).
23. C. Verster, K. Minnaar, H. Bouwman, Marine and freshwater microplastic research in South Africa. *Integrated Environmental Assessment and Management* **13**, 533-535 (2017).
24. F. Shahul Hamid *et al.*, Worldwide distribution and abundance of microplastic: How dire is the situation? *Waste Management & Research* **36**, 873-897 (2018).
25. A. A. de Souza Machado, W. Kloas, C. Zarfl, S. Hempel, M. C. Rillig, Microplastics as an emerging threat to terrestrial ecosystems. *Global Change Biology* **24**, 1405-1416 (2018).
26. M. A. Browne *et al.*, Accumulation of Microplastic on Shorelines Woldwide: Sources and Sinks. *Environmental Science & Technology* **45**, 9175-9179 (2011).
27. E. R. Zettler, T. J. Mincer, L. A. Amaral-Zettler, Life in the "Plastisphere": Microbial Communities on Plastic Marine Debris. *Environmental Science & Technology* **47**, 7137-7146 (2013).
28. I. V. Kirstein, A. Wichels, G. Krohne, G. Gerdts, Mature biofilm communities on synthetic polymers in seawater - Specific or general? *Marine Environmental Research* **142**, 147-154 (2018).
29. K. Kesy, S. Oberbeckmann, B. Kreikemeyer, M. Labrenz, Spatial Environmental Heterogeneity Determines Young Biofilm Assemblages on Microplastics in Baltic Sea Mesocosms. *Frontiers in Microbiology* **10** (2019).
30. K. L. Dudek, B. N. Cruz, B. Polidoro, S. Neuer, Microbial colonization of microplastics in the Caribbean Sea. *Limnology and Oceanography Letters* **5**, 5-17 (2020).
31. E. Curren, S. C. Y. Leong, Profiles of bacterial assemblages from microplastics of tropical coastal environments. *Science of The Total Environment* **655**, 313-320 (2019).
32. J. A. Bryant *et al.*, Diversity and Activity of Communities Inhabiting Plastic Debris in the North Pacific Gyre. *mSystems* **1** (2016).
33. A. C. Leri, L. M. Mayer, K. R. Thornton, B. Ravel, Bromination of marine particulate organic matter through oxidative mechanisms. *Geochimica et Cosmochimica Acta* **142**, 53-63 (2014).
34. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).
35. S. Hunter *et al.*, InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**, D211-D215 (2009).
36. M. Yiğitoğlu, M. Arslan, Selective removal of Cr(VI) ions from aqueous solutions including Cr(VI), Cu(II) and Cd(II) ions by 4-vinly pyridine/2-hydroxyethylmethacrylate monomer mixture grafted poly(ethylene terephthalate) fiber. *Journal of Hazardous Materials* **166**, 435-444 (2009).

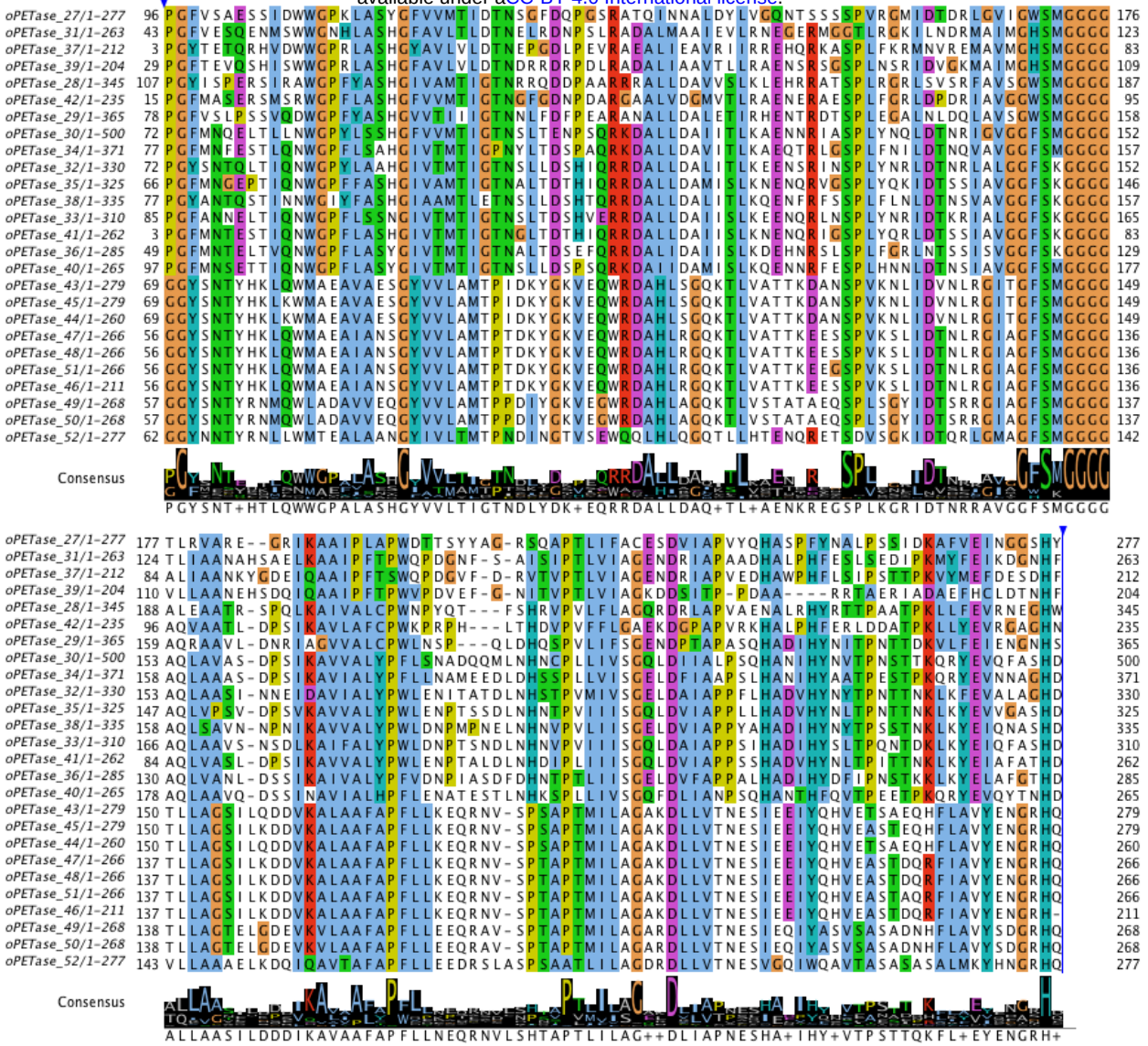
473
474

Figure 2 supplement 1.

**A.**




**B.**




**C.**

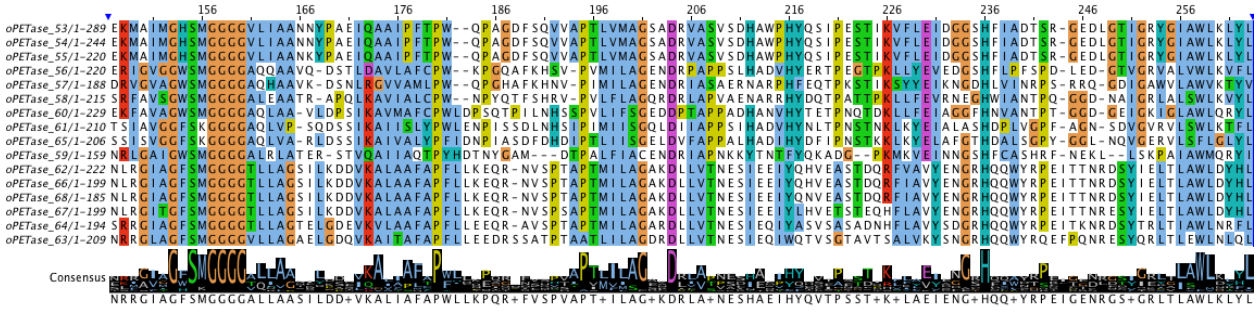



**Fig 2 Supplementary 1.** Alignments of ocean PETase groups M5 (A), M4(B) and M3 (C) showing conservation and variation in key PETase residues particularly catalytic tirade, aromatic clamp and oxyanion hole.