The phylogenetic and global distribution of extracellular bioplastic degrading genes

V. R. Viljakainen,[1] L. A. Hug[1]†

[1]University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1

† Address correspondence to Laura A. Hug, laura.hug@uwaterloo.ca.

Abstract

Polyhydroxyalkanoates (PHAs) are a family of microbially-made polyesters that have been commercialized as biodegradable plastics. PHA production rates are predicted to increase rapidly as global concerns around environmental plastic contamination and limited fossil fuel resources have increased the importance of bio-based plastic alternatives. PHAs are meant to quickly degrade in the environment, but this degradation is reliant on microbially-secreted PHA depolymerases, whose taxonomic and environmental distribution have not been well-defined. As a result, the impact of increased PHA production and disposal on global environments is unknown. Here we used 3,842 metagenomes to analyze the distribution of PHA depolymerase genes in microbial communities from diverse aquatic, terrestrial and waste management systems. Our results indicate that extracellular PHA depolymerases are globally widespread but unevenly distributed, with certain environments showing little to no evidence for this activity. In tandem, we screened 5,290 metagenome-assembled genomes to describe the phylogenetic distribution of this trait, which is substantially broader compared to current cultured representatives. We identified members of the Proteobacteria and Bacteroidetes as key lineages with PHA biodegradation potential and predict this activity in members of the Actinobacteria, the Candidate phylum Rokubacteria, Firmicutes, Planctomycetes and Spirochaetes.

Importance: Environmental concerns alongside legislation banning single-use petroleum-based plastics are expected to promote the production of bio-based plastics, including PHAs. PHAs represent a novel and emerging waste stream. If PHA disposal follows the precedent set by conventional plastics, a significant portion will be littered into the environment, or improperly discarded into landfills instead of composting facilities. Traditionally, the identification of

bioplastic degrading enzymes and organisms has relied on culture-dependent assays. As a result, the PHA degradation capabilities of the "unculturable" fraction of microorganisms remain largely unexplored. Here, we leverage large amounts of environmental sequence data to assess which environments harbor PHA-degrading organisms and to determine the taxonomic affiliations of bioplastic degraders. Our analyses inform our understanding of the biodegradation potential in the environment, with implications for the impact of bioplastic pollution. We identify enzymes and organisms that may be suitable for future bioremediation, chemical processing or biotechnological applications.

Introduction

Widespread environmental impacts associated with liberal plastic use have led to a demand for sustainable biodegradable alternatives, including polyhydroxyalkanoates (PHAs) (1). Conventional plastics rely on fossil resources and are extremely recalcitrant to degradation, causing them to accumulate in the environment (2). Annual plastic production now exceeds 380 million tons (3) and accounts for approximately 6% of global fossil feedstocks (4). Plastics often serve only a short service lifespan before being discarded (1). It is estimated that, as of 2015, humans had generated 6,300 million metric tons of plastic waste (3). According to a recent global assessment, only 21% of plastic waste has been incinerated or recycled; the remaining 79% has been discarded in landfills or littered into the environment (3). Plastics in the environment can harm organisms through ingestion or entanglement causing ecological damage (5). The effects of plastic pollution in the environment is a complex problem and the extent of these effects remains largely unknown (6).

New legislation aimed at mitigating the environmental impacts of plastics alongside growing public concern has generated a growing need for new biodegradable alternatives (7–9). PHAs have garnered substantial academic and industrial interest, and commercially available PHAs have been developed, for reviews see (10–12). PHAs are considered environmentally favorable to conventional plastics because they can be made from renewable carbon resources and have been shown to degrade in many different environments (13–20). 2.11 million tons of bioplastics were produced in 2018 (8). PHAs currently represent 1.4% of the global bioplastic market with production rates expected to more than double in the next five years (8).

PHAs are a family of polyesters that are synthesized naturally by many types of bacteria and archaea as an intracellular energy and carbon storage compound (13, 21). There are over 150 different types of PHA monomers which can be combined in different ways to yield polymers with a variety of chemical and structural properties (22). PHAs are broadly classified as either short-chain length (PHAscl) or medium chain length (PHAmcl) depending on the number of carbon atoms found in the monomer. PHAscl have 3-5 carbon atoms per monomer and PHAmcl have 6-15 carbon atoms per monomer. PHAs occur in two different biophysical states, which impacts which enzymes can break them down (13). Intracellular PHAs are amorphous and are found in protein-associated granules sometimes referred to as carbonosomes (23); extracellular PHAs become denatured upon extraction and take on a semi-crystalline state (13).

PHA biodegradation in the environment is mediated by microorganisms that produce PHA depolymerases (24). PHA depolymerases are a diverse family of intracellular and extracellular carboxylesterases and are members of the alpha/beta-hydrolase fold family (25). To

date, approximately 35 PHA depolymerases have been isolated and biochemically characterized (Supp Data 1 and references within), from organisms within the phyla Ascomycetes, Firmicutes, Proteobacteria (most) and Actinobacteria. PHA depolymerases are classified based on subcellular location (extracellular or intracellular), substrate specificity (mcl [EC 3.1.1.76] or scl [EC 3.1.1.75]), and features of the catalytic domain (13, 25). Intracellular PHA (inPHA) depolymerases mobilize amorphous endogenous PHA stores whereas extracellular PHA (ePHA) depolymerases degrade exogenous semi-crystalline PHAs (24). ePHA depolymerases allow microorganisms to scavenge PHAs that have been released into the environment from ruptured cells (13) and are the main avenue for biodegradation of commercial PHAs. Here, we focused our analyses on ePHA depolymerases because we were interested in enzymes that are secreted into the environment and have the potential to degrade commercial PHAs.

All known ePHA depolymerases contain a catalytic triad composed of a serine (embedded in a lipase box motif), histidine and aspartic acid (13, 25). Extracellular short chain length (ePHAscl) depolymerases typically consist of a signal peptide, a catalytic domain, a linker domain and a C-terminal substrate binding domain (25). There are two recognized ePHAscl depolymerase subtypes (denoted type 1 [T1] and type 2 [T2]) which differ in the arrangement of their catalytic domains (13). In ePHAsclT1 depolymerases the oxyanion hole is N-terminal to the lipase box, whereas in ePHAsclT2 depolymerases the oxyanion hole is C-terminal to the lipase box (25). Characterized extracellular medium-chain length (ePHAmcl) depolymerases do not exhibit a C-terminal substrate binding domain; instead the N-terminal region is presumed to mediate substrate binding (13). ePHAmcl depolymerases share no significant homology with ePHAscl depolymerases except for the catalytic triad (25), which is non-specific as it is shared

by all carboxylesterases (27). However, ePHAsclT1 depolymerases do share significant homology with some biochemically characterized intracellular PHAscl (inPHAscl) depolymerases (28) and ePHAsclT2 depolymerases share significant similarity with a periplasmic PHA depolymerase specific for amorphous forms of PHA (29).

Rates of plastic degradation are influenced by polymer properties (e.g., available surface area, monomeric composition, presence of additives), the extant microbial population, and abiotic factors (e.g., temperature, UV, pH) (30). PHAs are considered biodegradable, however the rate of this process depends on the polymer properties and the environmental context (1). Thus, the term "biodegradable" is ambiguous unless the conditions and rate of degradation are reported (31). Standardized tests exist to quantify degradation under specified conditions, such as industrial composts (see ASTM D5338-15) (32), but these tests cannot be applied to non-engineered environmental systems. Field studies on biodegradability have shown conflicting results, and there is a growing concern that biodegradable plastics are often not as biodegradable as claimed (33).

Here, we have focused specifically on determining which environments harbour microbes with the genetic potential to mediate PHA biodegradation. PHA biodegradation rates are strongly influenced by the abundance and diversity of microorganisms encoding PHA depolymerases (24). With PHA production levels expected to increase, a strong understanding of the distribution of PHA-degrading enzymes and organisms is crucial for understanding the impact of PHAs on the environment (24). Early culture-based studies have isolated PHAscl degrading microorganisms from different environments including soil, compost, sewage and aquatic

139 environments, but indicated that PHAmcl degraders appear to be relatively rare (13, 24).
140 Reflecting these findings, relatively few ePHAmcl depolymerases have been characterized
141 compared to ePHAscl depolymerases (25, 33).

142 Most PHA biodegradation research has relied heavily on culture-based approaches or
143 black-box burial degradation trials (30). Culture-based research has developed the foundational
144 understanding of the enzymes and organisms mediating PHA biodegradation reviewed by (13).
145 However, a global summary of the environmental distribution of PHA depolymerases has not
146 been reported. Many microbial lineages are currently unamenable to culturing and are only
147 recognized from culture-independent analyses (35). With the advent of high-throughput
148 sequencing and metagenomics, there has been a proliferation of publicly available DNA
149 sequence data from microbial communities inhabiting diverse global environments (36, 37),
150 enabling broad surveys of trait distributions, see (38, 39). Here, we employed a culture-
151 independent approach using publicly available environmental sequence data to characterize the
152 phylogenetic and environmental distribution of PHA-degrading enzymes. We screened 5,290
153 metagenomes assembled genomes (MAGs) and 3,842 metagenomes sequenced from globally
154 distributed natural environments and waste-management systems. Biochemically characterized
155 ePHA depolymerases were used as biomarkers to assess the potential for PHA degradation
156 activity in different environments and across the bacterial tree of life.

157 **Methods**

158 *Identification of ePHA depolymerases*

159 A reference set of biochemically validated ePHA depolymerases were gathered from the
160 literature (Supp Data 1) and used as queries in a BLAST search (threshold $e < 1x10^{-10}$) against

3,842 metagenomes (Supp Data 2) and 5,290 MAGs (Supp Data 3) that were publicly available on the Integrated Microbial Genomes (IMG) system (https://img.jgi.doe.gov) (37). Associated IMG metadata was used to classify each dataset by environment. Datasets screened were publicly available, but permissions were requested for PIs associated with the data. Many metagenomes came from datasets associated with publications, and we acknowledge the work of the scientific community in generating such a rich database (40–55). A series of curation steps were taken to increase PHA depolymerase prediction accuracy and to reduce or eliminate non-functional homologs. Hits identified in metagenomes and MAGs were curated separately. The results of the curation process applied to MAG hits (smaller, more tractable dataset) were used to develop a computationally feasible curation protocol for metagenome hits.

ePHA depolymerase hit curation:

Unique ePHA depolymerase blast hits (threshold < $1x10^{-10}$) were gathered and aligned with the corresponding reference sequences (7 ePHAmcl, 14 ePHAsclT1, and 7 ePHAsclT2) (Supp Data 4). All alignments were performed using MUSCLE v. 3.8.3 (56) as implemented in Geneious v. 10.2.6 (https://www.geneious.com) with default parameters. Hits were first sorted and divided into groups of similarly sized sequences before alignments were performed. This served to reduce the computational load of performing a single alignment with many dissimilar sequences and facilitated identification of areas of shared conservation with reference enzymes. An alignment-based curation was performed to identify hits that shared key catalytic residues with reference sequences (serine embedded in a lipase box motif [GxSxG], aspartic acid, and histidine, as well as the oxyanionic histidine for the ePHAscl depolymerases). To prevent removal of true PHA depolymerases containing insertions or deletions, higher scoring blast hits

184 (threshold e< $1x10^{-30}$) were retained for further analyses in addition to those retained from the
185 alignment-based curation. Genes that were highly gapped in the alignment or which were
186 truncated within the catalytic domain were removed. Next, placement within a maximum
187 likelihood (ML) tree was used to distinguish PHA depolymerases from other types of closely
188 related enzymes. ML trees were constructed using RAxML-HPC2 (57) with automatic
189 bootstopping on the CIPRES webserver (58; Supp. Data 5). ModelTest-NG (59) was
190 implemented for model selection assuming a uniform rate of evolution. PROTGAMMA WAG,
191 VT+F, and WAG+F models were used for ePHAmcl, ePHAsclT1 and ePHAsclT2 hits
192 respectively.
193
194 ML-based curation was performed using tree topology and branch length to assess
195 relatedness to known PHA depolymerases. Representative MAG-encoded sequences for each
196 major clade occurring in different regions of the ML tree were used as queries in a blast search
197 against (1) the uniprot / swissprot database (https://www.uniprot.org/) (60); (2) NCBI's refseq
198 database (https://www.ncbi.nlm.nih.gov/refseq/) (61); and (3) the pdb database
199 (www.wwpdb.org) (62). Top blast hits were retrieved and used to predict clade annotations.
200 Many MAG-encoded genes could not be confidently assigned to a PHA depolymerase subtype
201 based on tree placement alone, due to overlapping phylogenetic signals between (1) ePHAsclT1
202 depolymerases and inPHAscl depolymerases and (2) ePHAsclT2 depolymerases and a
203 periplasmic PHA depolymerase.
204
205 *Ribosomal protein tree*

206 A concatenated ribosomal protein tree was constructed to phylogenetically place MAGs with
207 predicted bioplastic degrading genes. 16 ribosomal proteins (rpL2, L3, L4, L5, L6, L14, L15,
208 L16, L18, L22, L24, S3, S8, S10, S17 and S19) were gathered from 143 MAGs encoding
209 predicted PHA depolymerases. Ribosomal proteins were independently aligned to a previously
210 described bacterial reference set (35). Alignments were masked to remove columns comprised of
211 95%+ gaps and trimmed to remove ambiguously aligned positions at the N and C termini. The
212 16 ribosomal protein alignments were concatenated to form a final alignment composed of 2,491
213 positions and 1,941 taxa. In tandem, MAG quality was estimated using checkM v1.0.13 (63).
214 Thirteen poor quality MAGs (>10% redundant and/or <50% complete) were excluded from the
215 final alignment. Fifteen MAGs were also excluded from the final alignment because their
216 ribosomal protein complement resulted in <50% of the expected number of ungapped residues in
217 the final alignment. A total of 115 MAGs encoding predicted PHA depolymerases were included
218 in the final alignment. Phylogenetic placement was inferred using FastTree (64) as implemented
219 in Geneious v.10.2.6. (https://www.geneious.com) and the tree was visualized using iTOL
220 (https://itol.embl.de/) (65).
221
222 *Environmental ePHA depolymerase hit curation*
223 Hits with an e-value $<1x10^{-25}$ for PHAmcl depolymerases (1,196) and $<1x10^{-30}$ for
224 ePHAscl (T1: 54,986, T2: 14,042) depolymerases were gathered and aligned to reference
225 sequences as for the MAG-derived genes. E-value thresholds were set at least five orders of
226 magnitude lower than the lowest quality hit that had ultimately been retained during the MAG hit
227 curation, to ensure a broad search. ePHAscl depolymerase hits missing either the lipase box
228 motif (GxSxG) and the oxyanion hole (HGC) motif were removed before alignments were

229 performed. Alignments and phylogenetic tree placement-based curation were performed as for
230 the MAG-derived sequences, with the following modifications to accommodate large dataset
231 sizes. High quality blast hits, defined as $<1x10^{-40}$ for ePHAmcl and $1<10^{-50}$ for ePHAscl datasets,
232 were retained for inclusion in the tree curation steps. CD-HIT (66) was used to cluster
233 ePHAsclT1 hits to 90% identity (word size 5) to reduce computational load of alignments and
234 tree-based curation. After tree-based curation, near-full-length sequences occurring within the
235 same clustered set for retained predicted ePHAsclT1 depolymerases were identified using a
236 custom script (Supp Data 6) and included in the final ePHAsclT1 sequence set.
237
238 Final sets of curated predicted ePHA depolymerases were independently aligned to the
239 relevant reference enzymes using MAFFT v.7 (67) (ePHAsclT1) and MUSCLE v. 3.8.3 (56)
240 (ePHAsclT2 and ePHAmcl). MAFFT was used for the final predicted ePHAclT1 alignment due
241 to large dataset size. Final protein trees were independently constructed for predicted ePHAmcl,
242 ePHAsclT1 and ePHAsclT2 depolymerases using FastTree (68) due to dataset size. Final trees
243 were visualized and annotated using ITOL (https://itol.embl.de/) (65).
244
245 **Results**
246 **Environmental distribution of PHA depolymerases**
247 To characterize the environmental capacity for PHA biodegradation, 3,842 publicly
248 available assembled and annotated metagenomes from diverse aquatic (n=1,982), terrestrial
249 (n=1,596), wastewater (n=230) and solid waste management systems (n=34) were screened for
250 ePHA depolymerases (Supp Data 2). Metagenomes sequenced from all seven continents were
251 included in our dataset but were unequally represented, with certain regions having more

sampled locations and sample types than others (Supp Fig 1). Biochemically validated ePHA depolymerase genes were used as biomarkers for biodegradation potential (Supp Data 1). Homology-based searches using biomarker enzymes followed by rigorous curation of predicted enzymes resulted in a global distribution for PHA depolymerases (Fig 1a). Metagenomes from North America were the most heavily represented in our dataset and nearly every environmental subtype showed some evidence for PHA degradation activity within this region (Fig 1b). A total of 13,869 PHA depolymerases were predicted in 1,295 metagenomes, sampled from aquatic (417/1,982 = 21.0%), solid waste (23/34 =67.6 %), terrestrial (764/1,596 =47.8%) and wastewater (91/230 = 39.5%) environments (Fig 2).

ePHAmcl depolymerization potential was sparsely represented across environments. Only 222 ePHAmcl depolymerases were predicted out of the 13,869 ePHA identified. ePHAmcl depolymerases were identified from 153 metagenomes from aquatic (41/1,982 = 2.1%), terrestrial (94/1,596 = 5.9%), and wastewater (18/230 = 7.8%) environments (Fig 2). Predicted ePHAmcl depolymerases were placed in a ML tree anchored with biochemically characterized ePHAmcl enzymes (Supp Fig 2a). ePHAmcl depolymerases recently isolated from *Streptomyces* species share little sequence homology with other ePHAmcl biomarker enzymes and form a novel subgroup (69). Based on the ML tree, 14 predicted ePHAmcl depolymerases placed with the three ePHAmcl depolymerases isolated from *Streptomyces* (69, 70) while the other 208 predicted enzymes formed a second clade with ePHAmcl biomarker enzymes isolated from species of *Bdellovibrio* (34) and *Pseudomonas* (71, 72). Based on tree placement, only a handful of predicted enzymes were closely related to biomarker enzymes; the majority of predicted

enzymes formed more distantly related clades. If functional, these enzymes significantly expand the known sequence diversity of ePHAmcl depolymerases.

In contrast to ePHAmcl depolymerases, ePHAscl depolymerases were more prevalent across environments. A total of 10,245 ePHAsclT1 depolymerases were predicted from 1,136 metagenomes, from aquatic (338/1,982 =17.1%), solid waste (13/34 = 38.2 %), terrestrial (713 /1,596 = 44.7%), and wastewater (72/230 = 31.3%) environments (Fig 2). For ePHAsclT2 depolymerases, a total of 3,402 enzymes were predicted from 908 metagenomes from aquatic (264/1,982 = 13.3%), solid waste (22/34 = 64.7%), terrestrial (540/1,596 = 33.8%), and wastewater (82/230 = 35.6%) environments (Fig 2). Predicted ePHAscl T1 and T2 depolymerases were placed in protein trees anchored with biomarker enzymes (Supp Fig 2b and 2c). For both predicted ePHAscl T1 and T2 depolymerases, considerable sequence diversity was seen relative to biomarker enzymes, with large clades of putative PHA depolymerases containing no biomarker enzymes. As for ePHAmcl enzymes, this indicates the known diversity of ePHAscl enzymes may be a marked underrepresentation of the true protein family.

Aquatic Environments

Plastic contamination is ubiquitous in aquatic environments (6, 73). Plastic debris, either macroplastics or microplastics (typically <0.5mm) can harm aquatic organisms through entanglement and ingestion (74). Microplastics are of particular concern because they can be ingested by organisms and serve as a vector for chemical contaminants (74). As bioplastics become more popular, it is expected that they will become a larger component of aquatic contamination. PHAs have been shown to degrade in aquatic environments, however rates can

vary widely depending on polymer properties, temperature, and microbial population, and can be quite slow (20, 75, 76). For example, in freshwater ponds, samples of PHAscl showed 7% mass loss after six months (20). In marine water, PHAscl bottles showed only minor disintegration after a year (76). Relatively little research has characterized the degradation of PHAmcl in aquatic environments, and PHAmcl degraders are typically scarce relative to PHAscl degrading organisms (69, 77).

Here, we screened 1,982 aquatic metagenomes for evidence of microbial PHA biodegradation potential. Our analyses included 946 freshwater, 452 marine, 87 non-marine saline and alkaline, 114 sediment and 383 thermal spring metagenomes. Less than half of the aquatic metagenomes we screened contained any type of predicted ePHA depolymerases (Fig 2). The highest proportion of aquatic metagenomes with predicted PHA depolymerization potential were from sediments (55/114 = 48.2%) followed by non-marine saline and alkaline (31/87= 35.6%) and marine environments (140/452=31.0%). Only 19.2% of freshwater metagenomes (182/946) and 2.3% of thermal spring metagenomes (9/383) contained predicted ePHA depolymerases.

Very few aquatic metagenomes showed evidence for ePHAmcl depolymerization potential. Only 1.9% (18/946) freshwater, 3.5% (16/452) marine, 2.3% (2/87) non-marine saline and alkaline, 3.5% (4/114) sediment, and 0.3% (1/383) thermal spring metagenomes had predicted ePHAmcl depolymerases. In contrast to ePHAmcl, all aquatic environmental subtypes had predicted ePHAscl depolymerization potential to some degree (ePHAsclT1 and ePHAsclT2 depolymerases). 43.9% (50/114) of sediment, 24.1% (21/87) of non-marine saline and alkaline,

20.4% (92/452) of marine, and 17.5% (166/946) of freshwater metagenomes contained predicted ePHAsclT1 depolymerases. Very few thermal spring metagenomes (9/383 = 2.34%) had these predicted depolymerases. For ePHAsclT2 depolymerases, the highest proportion of metagenomes encoding this activity was sediments (45/114 = 39.5%), followed by 21.7% (98/452) marine, 20.7% (18/87) non-marine saline and alkaline environments, and only 0.5% (2/383) thermal spring metagenomes. Thermal spring metagenomes had the lowest PHA degradation potential across all aquatic environmental subtypes.

In aquatic environments, plastics provide a surface for microbial colonization and biofilm formation (78). Few studies have characterized how microbial assemblages change and respond to bioplastic waste, including PHAs. Pinnell & Turner (2017) found that PHA-associated biofilms in benthic coastal marine environments were distinct from non-biodegradable polyethylene terephthalate controls and were dominated by sulfate-reducing microorganisms (79). These results indicate that PHA waste may alter microbial communities and have unintentional impacts on biogeochemical activities in this environment through the stimulation of sulfate reducers. Our analyses indicate that the majority of aquatic environments are not equipped with an abundance of PHA degrading organisms. This may suggest that bioplastics (particularly PHAmcl products) may persist longer than anticipated in these environments, or their presence may slowly enrich for bioplastic-degrading organisms, altering the local microbial community structure.

Terrestrial Environments

Soil microbial communities have been extremely well-sampled globally relative to other environments. From our screen, 51.5% (740/1,437) of soil metagenomes show ePHA depolymerization potential (Fig 2). Other terrestrial environments with evidence for this activity were deep subsurface environments (21/143= 14.7%) and rock-dwelling communities (3/3 = 100.0%). Loam (0/3), peat (0/5) and volcanic (0/5) microbial communities showed no evidence for ePHA depolymerization potential, although this may be an artifact of small sample sizes for these environments.

Rock-dwelling metagenomes were not predicted to contain any ePHAmcl depolymerases, whereas 0.7% (1/143) of deep subsurface and 6.5% (93/1,437) of soil metagenomes did encode one or more predicted ePHAmcl depolymerases. 13.3% (19/143) deep subsurface, 100% (3/3) rock-dwelling, and 48.1% (691/1,437) of soil metagenomes screened positive for ePHAsclT1 depolymerases. ePHAsclT2 depolymerases were predicted in 5.6% (8/143) of deep subsurface, 66.7% (2/3) of rock-dwelling environments, and 36.9% (530/1437) of soil metagenomes.

Recently, culture-independent approaches have revealed the diversity and heterogeneity of soil microbial communities. Soil can have distinct microbial communities across micrometers or millimeters in space (80). PHAscl biodegradation in soil has been well-studied and many PHA degrading organisms, mostly polyhydroxybutyrate (PHB) degraders (the most common type of PHA monomer), have been isolated from these environments (14, 17, 30, 81). Again, less work has characterized PHAmcl degradation in soils relative to PHAscl products. However, in tropical soils, PHAmcl has been showed to degrade much more slowly compared to PHB films (77).

Solid Waste Environments

Over half of metagenomes screened from solid waste environment subtypes contained predicted ePHAscl depolymerases, with 66.7% (20/30) of compost metagenomes and 75.0% (3/4) of landfill metagenomes showing ePHA depolymerization potential. When parsed by ePHAscl depolymerase subtype, 33.3% (10/30) of compost and 75.0% (3/4) of landfill samples had predicted ePHAsclT1 depolymerases and 63.3% (19/30) of compost and 75.0% (3/4) of landfill metagenomes had predicted ePHAsclT2 depolymerases. Notably, our analyses did not predict any ePHAmcl depolymerases from solid waste environments, which may have implications for bioplastic waste-management.

Composting is considered the ideal way to handle certain types of biodegradable plastic waste because it allows for energy and material recovery (31). Furthermore, since plastics are broken down, they are removed as a potential physical threat to organisms in the environment. Composting expands the end-of-life options available for bioplastics and is particularly beneficial when mechanical recycling is not an option (31). However, composting requires existing infrastructure and collection systems and access to these varies based on locale (82). Commercial PHA products are certified as compostable under various standards, including the European standard (EN 13432) which stipulates that under industrial composting conditions 90% of the plastic must be converted to CO_2 after six months (31). It was concerning that a third of composting samples showed no predicted PHA depolymerases. This may reflect a lack of bioplastics in the current waste streams, where low abundance degraders are below current detection limits, or may indicate a future requirement for targeted amendment of bioplastic degrading organisms to composting facilities to ensure degradation occurs at mandated rates.

A large fraction of plastic waste is disposed of in landfills. Despite landfills being a likely site of disposal, very little characterization of the biodegradation of biodegradable plastics in landfills has taken place (82). Very few plastics are recycled due to low consumer compliance, lack of infrastructure, and contamination of recycling streams (3). Bioplastics will likely follow this precedent, although because they can be composted, a larger proportion may be diverted away from landfills. Landfills are not an ideal solution for bioplastic waste; plastics in landfills represent a lost opportunity as their stored energy cannot be recovered and they take up valuable space (1, 83). Landfills are complex heterogenous environments and landfill microbial communities are not well-characterized (84). The landfill metagenomes included in our analyses originated from a Canadian site and showed the potential for PHAscl degradation, but not PHAmcl degradation. It is thus possible that PHAmcl polymers will behave similarly to conventional plastics and remain relatively inert over a long timeframe. This poses little threat to the environment provided landfills are well-maintained and do not disperse wastes into the surrounding environments, but this cannot be assumed for 100% of landfilled bioplastics. Our identification of predicted ePHAscl depolymerases in landfill metagenomes coupled with data from earlier studies indicates that ePHAscl plastics biodegrade in landfills if amenable abiotic conditions (temperature, pH, moisture) are maintained. One concern is that anoxic degradation of PHAs generates greenhouse gases in the form of CO_2 and CH_4 (24). However, most modern landfills have infrastructure to capture methane as biogas, generating energy and reducing greenhouse gas emissions, which would mitigate harmful by-products of landfill PHA degradation (85).

Wastewater Environments

Within wastewater environments, the majority of anaerobic digestors (51/54 = 94.4%) and nutrient removal communities (27/35=77.1%) were predicted to conduct ePHA degradation, whereas only a small fraction of activated sludge communities showed evidence for this activity (13/141=9.2%).

ePHAmcl depolymerases were predicted in 7.8% (11/141) of activated sludge, 18.5% (10/54) of anaerobic digestors, and 57.1% (20/35) of nutrient removal systems. A high proportion of anaerobic digestors (38/54=70.4%) and nutrient removal systems (21/35 = 60.0%) had predicted ePHAsclT1 depolymerases whereas only 9.2% (13/141) of activated sludge environments had predicted enzymes. ePHAsclT2 depolymerases were also well-represented in anaerobic digestors (47/54 = 87.0%) and nutrient removal systems (27/35=77.1%), but only occurred in 8/141, or 5.7% of activated sludge metagenomes.

Interestingly, wastewater can be used as a carbon source for PHA production , and appears to be a promising avenue for biopolymer production (86). Given the potential presence of PHA degraders in the majority of wastewater systems, there is the possibility of a circular system, where wastewater acts as both feedstock and waste processing components of the bioplastic lifecycle.

Phylogenetic distribution of predicted PHA depolymerases

We screened 5,290 MAGs available on IMG for ePHA depolymerases, comprising 4,927 bacterial and 363 archaeal genomes. Screened MAGs were reconstructed from metagenomes originating from engineered (443), environmental (4,659), host-associated (186), and unclassified (2) systems (Supp Data 3). We identified 143 MAGs (2.7%) encoding a total of 231

predicted extracellular PHA depolymerases. CheckM was used to assess MAG quality, with 130 MAGs >50% percent complete and <10% redundant, within which 122 MAGs were >70% complete and <10% redundant (Supp Data 7). All MAGs encoding predicted ePHA depolymerases were from the bacterial domain. A number of distantly related genes were identified in archaeal MAGs but these did not pass our curation steps. These archaeal sequences may represent true PHA depolymerases that are distantly related from reference enzymes and thus may be interesting targets for future characterization (Supp Data 8).

A concatenated ribosomal protein tree was created to phylogenetically place the 115 predicted ePHA degrading MAGs that were >50% complete and <10% redundant and had an appropriate complement of ribosomal proteins (see methods) (Fig 3). From this, 6 pairs of curated MAGs had identical concatenated ribosomal protein sequences. These MAG pairs often originated from related samples (e.g., time series samples of the same environment), indicating they may represent the same bacterial populations (Supp Data 7). From the curated set of 115 MAGs, we identified putative PHA depolymerases from members of the Proteobacteria (92), Bacteroidetes (14), Actinobacteria (3), Firmicutes (2), Candidate phylum Rokubacteria (2), Spirochaetes (1), and Planctomycetes (1). Notably, we did not detect PHA biodegradation potential in any MAGs associated with the Candidate Phyla Radiation/Patescibacteria. Prior to our screen, bacterial PHA depolymerases were reported from the Proteobacteria, the Actinobacteria and the Firmicutes, as well as one fungal phylum, the Ascomycetes. Our screen extends the phylogenetic distribution of this activity to four additional bacterial phyla, doubling its taxonomic diversity. The distribution of fungal PHA depolymerases was not investigated in

our screen, but is an important avenue for future characterization to holistically characterize PHA depolymerization potential.

The vast majority of the predicted PHA degrading organisms were Proteobacteria, with 13 Alphaproteobacterial, 51 Betaproteobacterial, 8 Deltaproteobacterial, 18 Gammaproteobacterial and two Oligoflexia MAGs identified. Alphaproteobacterial MAGs with ePHA depolymerases were from marine, freshwater, and soil environments, with members of the families Rhizobiales (7), Sphingomonadales (4), and Rhodobacterales (1), and two unclassified Alphaproteobacteria. Many Betaproteobacterial PHA degraders have been identified through culture-based studies, particularly from soils (13). As a result, many of the biomarker enzymes used in our screen were from this class (Supp Data 1 and references within). The Betaproteobacterial MAGs with predicted PHA degradative capacity were predominantly from the Burkholderiales (31), but also from the Rhodocyclales (2) and unclassified lineages (13). The Burkholderiales *Paucimonas lemoignei* is a model organism for PHA degradation and encodes at least five PHA depolymerases (87). It is not clear whether PHA depolymerases from this class were so abundant in our screen due to a detection bias based on the closely related biomarker enzymes in our reference set or because this family is genuinely enriched for this function compared to other lineages. Eight Deltaproteobacteria with PHA depolymerization potential were identified, including six Myxococcales, a Desulfobacterales and an unclassified bacterium. A total of 18 Gammaproteobacteria including species from the Alteromonadales (6), Chromatiales (1), Methylococcales (1), Oceanospirillales (3), Pseudomonadales (1), Xanthomonadales (2), and four unclassified organisms were predicted to encode ePHA depolymerases.

Beyond the Proteobacteria, 13 MAGs from the Bacteroidetes encoded predicted ePHAsclT1 depolymerase(s). All encoded a single predicted ePHAsclT1 except for two unclassified bacteria from soil and groundwater which each encoded two predicted ePHAsclT1 enzymes that shared high-level of sequence identities. Bacteroidetes with predicted ePHA degradation capacities included an unclassified Cryomorphaceae (Flavobacteria) and an unclassified Saprospiraceae (Saprospiria) from a wastewater nutrient removal system, and nine unclassified organisms from groundwater. The three Actinobacteria MAGs were predicted to encode ePHAsclT1 depolymerases. Two were unclassified members of the Pseudonocardiales from soil. One of these MAGs encoded a single ePHAsclT1 depolymerase whereas the other encoded two. The third Actinobacteria MAG was from an organism from the Actinomycetales. Two Firmicutes from the Clostridiales sequenced from an anerobic bioreactor were predicted to have PHAmcl depolymerization potential. One unclassified Spirochaetes bacterium from groundwater was predicted to encode three closely related ePHAsclT1 depolymerases. Two MAGs from the Candidate phylum Rokubacteria from groundwater and soil had predicted ePHAsclT1 depolymerases. One nutrient removal wastewater system-derived Planctomycetes MAG from the Phycisphaerales encoded a predicted ePHAsclT1. For a detailed overview of the PHA depolymerase complement for each MAG, please see Supp Data 7.

MAGs encoding multiple predicted PHA depolymerases

Of the 115 high-quality MAGs with PHA biodegradation potential, 67 were predicted to encode one PHA depolymerase, 24 were predicted to encode two, 23 were predicted to encode three, and one, a Burkholderiales, was predicted to encode nine PHA depolymerases (Supp Data 7).

The Burkholderiales MAG encoding nine putative PHA depolymerases originated from soil and was 98% complete and 5% redundant. Predicted depolymerases included two closely related ePHAsclT1, four ePHAsclT2 and three closely related ePHAmcl depolymerases.

MAGs from the Proteobacteria (22) and Spirochaetes (1) encoded three predicted ePHA depolymerases. One Alpha-, 18 Beta-, one Delta- and two Gammaproteobacteria all encoded three predicted PHA depolymerases. An Alphaproteobacterial MAG (100% complete, 1% redundant) that placed near the Bradyrhizobiaceae of the Rhizobiales encoded three closely related ePHAsclT1 enzymes. From the Betaproteobacterial MAGs, eight closely related *Candidatus* Accumulibacter organisms encoded two ePHAsclT2 enzymes and one ePHAsclT1 enzyme (all >84% complete, <5% redundant). PHA cycling is a core phenotype to the *Candidatus* Accumulibacter lineage, though most focus has been on PHA synthesis to date (49). Nine Burkholderiales encoded three ePHAscl enzymes and one Rhodocyclales (98% complete and 0% redundant) encoded one predicted ePHAsclT1 and two predicted ePHAsclT2 enzymes. For the Gammaproteobacteria, two near-identical *Halomonas* MAGs of the Oceanospirillales (both 99% complete, 1% redundant) encoded two ePHAsclT1 and one ePHAsclT2. The Spirochaetes MAG (90% complete, 0% redundant) had three predicted ePHAsclT1depolymerases that clustered closely within the ML tree.

Distribution of PHA depolymerase subtypes

ePHAmcl depolymerases: We predicted 10 ePHAmcl depolymerase genes from 8 MAGs (Fig 4a). A soil Burkholderiales bacterium from the Betaproteobacteria encoded 3 predicted enzymes. The remaining putative ePHAmcl depolymerases were found as singletons within MAGs, including a Pseudomonadales (Gammaproteobacteria) from soil, two Myxococcales

525 (Deltaprotebacteria) from a wastewater bioreactor and a marine environment, two *Sandaracinus*
526 spp. (Deltaprotebacteria) from a marine environment, and a *Clostridium* (Firmicutes) from an
527 anaerobic bioreactor (Fig. 4a). To our knowledge, ePHAmcl depolymerases have not previously
528 been identified in the Firmicutes. Several *Clostridia* species are human pathogens and obligate
529 anaerobes (88) which may have hindered recognition of PHA degrading activity in cultivation-
530 based studies, which have largely been conducted aerobically. Based on ML analyses, the
531 putative *Clostridium* ePHAmcl depolymerases were more closely related to the *Streptomyces*
532 ePHAmcl depolymerase subgroup, whereas the other predicted PHAmcl enzymes were more
533 closely related to those isolated from *Pseudomonas* and *Bdellovibrio* (Fig 4a).
534
535 ***ePHAsclT1 depolymerases:*** We predicted 156 ePHAsclT1 depolymerases in 120 unique MAGs
536 from the Actinobacteria (7), Bacteroidetes (17), Candidate phylum Rokubacteria (2),
537 Planctomycetes (1), Proteobacteria (125), Spirochaetes (3) and Tenericutes (1) (Fig 4b).
538 ePHAsclT1 depolymerases are the most well-characterized type of PHA depolymerase (13, 24).
539 Biochemically validated enzymes in our reference set originated from species of *Pseudomonas*
540 (Gammaproteobacteria), *Ralstonia* (Betaprotebacteria)*, Paucimonas* (Betaprotebacteria)*,*
541 *Alcaligenes* (Betaprotebacteria), and *Bacillus* (Firmicutes) (Supp Data 1 and references within).
542 Our screen extends the distribution of ePHAsclT1 enzymes to a further six phyla. An ePHAsclT1
543 depolymerase was also identified from a Tenericute MAG, however due to MAG quality (49%
544 complete, 0% redundant), it was excluded from the final concatenated ribosomal protein tree.
545
546 ***ePHAsclT2 depolymerases****:* We predicted 65 ePHAsclT2 depolymerases in 48 MAGs from
547 members of the Alpha, Beta and Gammaproteobacteria (Fig 4c). Most ePHAsclT2 biochemically

validated enzymes originated from members of the Burkholderiales family (Betaproteobacteria). Additional reference enzymes came from *Streptomyces* (Actinobacteria) and a *Penicillium* (Ascomycota), neither of which had related sequences in our environmentally-derived ePHAsclT2 depolymerases (Supp Data 1 and references within).

Summary of environmental distribution data

Screening of metagenomic sequence data from diverse aquatic, terrestrial and waste-management environments revealed that, while PHA-degrading bacteria are globally distributed, some environments show very little evidence for this activity. In particular, less than a third of the thermal springs, freshwater and marine environments we screened showed evidence for degradative capacity. This patchy distribution of PHA depolymerases may have impacts for pollution mitigation. If biodegradable plastics do not degrade in the environment at an appreciable rate, they have the potential to cause ecological damage analogous to conventional plastics. Alternatively, the presence of bioplastic waste may select for low-abundance community members with PHA degradation capacity and significantly alter microbial communities, which may have downstream consequences on global biogeochemical cycling.

Notably, PHAmcl depolymerases were poorly represented in the global dataset, indicating degradation of bioplastics with medium chain monomers may not be robust in most environments. There are very few reports of isolated PHAmcl degraders (69). PHAmcls are not easily emulsified into microbiological media, which may have hindered their detection in clear zone assays (13). This is an unfortunate gap because the more flexible and less brittle properties of PHAmcl make them more suitable than PHAscl for certain applications (89). Our screen

confirms the sparse environmental distribution of mcl PHA degrading organisms, with less than 5% of environmental samples showing evidence for PHAmcl degradation capacity. This suggests that commercial PHAs may be substantially more recalcitrant to degradation than current forecasts.

Summary of phylogenetic distribution data

We predicted PHA depolymerization activity for organisms from seven different phylum-level lineages, where existing reference enzymes derive from only three bacterial phyla. The majority of PHA depolymerases were identified from Proteobacteria, and specifically from the Betaproteobacteria. The majority of MAGs encoded one or two ePHA depolymerases, with a similarly low proportion of PHAmcl enzymes as were predicted from the environmental metagenomes. Our ePHAmcl reference set originated from only two Proteobacterial classes and the Actinobacteria. However, our analyses expanded the predicted distribution of PHAmcl depolymerases to the Firmicutes as well as an additional class of Proteobacteria. One organism, a Burkholderiales, encoded a complement of nine ePHA depolymerases, suggesting it may be an active, efficient, and/or highly versatile degrader.

Conclusions

This work presents the first summary of the global environmental and phylogenetic distribution of ePHA depolymerases using cultivation-independent approaches. Our analyses present a carefully curated set of predicted PHA depolymerases which provide a conservative estimate of the existing sequence diversity and environmental distribution of these protein families. Our screen predicts that many environments are not well-suited for PHA degradation,

which is a concern as this class of bioplastics increases in production. To fully determine bioplastic fate in global environments, a combination of metagenomic predictions from public datasets, burial degradation trials with functional profiling, and biochemical validation of predicted enzymes (particularly in enzymes distantly related to biomarker enzymes) is needed. This comprehensive examination would help establish the true diversity of functional PHA depolymerases and allow for more accurate activity predictions in the future.

Our analyses provide insights into the bioplastic degradation potential of microbial communities found around the world in unique and relevant environments. Understanding the limitations for bioplastic degradation in global environments is critical to ensure that the proper waste-management solutions are developed and enforced. Biodegradable plastics that meet compost certifications degrade under specified conditions that may not be present in a natural environment, and therefore rates of biodegradation may be highly variable. It is important that consumers do not assume that biodegradable plastics can be littered into the environment without consequence. Polymeric properties such as monomeric composition, surface area, and crystallinity strongly influence PHA biodegradability, but the underlying microbial community and the enzymes they secrete is critical for mediating bioplastic depolymerization (1, 13). The key bioplastic degraders occurring in the environment may aid in the development of future bioremediation options for these plastics and may include enzymes and organisms suitable for use as waste-management tools in the processing, degradation or chemical recycling of these plastics.

Acknowledgements

We would like to thank Angus Hilts for help with technical infrastructure. We would also like to extend our gratitude to the scientific community sequencing environmental metagenomes for their part in developing global data repositories. LAH is supported by a Tier II Canada Research Chair, and this work was funded by an NSERC Discovery grant (2016-03686) to LAH. A significant proportion of the datasets used were derived from grant support to Jill Banfield under DOE Office of Science BER grant DOE-SC10010566 and National Science Foundation grant CZP EAR-1331940; and to James Tiedje under DOE Office of Science BER DE-FC02-07ER64494.

Figures

Figure 1: 13,869 PHA depolymerases were predicted in 1,295 globally distributed metagenomes from aquatic (417), terrestrial (764), solid waste (23) and wastewater (91) environments. **A)** World map showing geocoordinates of samples with metagenomes containing one or more predicted PHA depolymerase, colored by environmental subtype. **B)** Detailed view of North America, which had many samples predicted to contain PHA depolymerases.

Figure 2: Environmental distribution of 13,869 PHA depolymerases (222 ePHAmcl, 10,245 ePHAsclT1 and 3,402 ePHAsclT2 depolymerases) identified by screening 3,842 metagenomes. Graph depicts the percentage of metagenomes that contained one or more predicted PHA depolymerase.

Figure 3: The predicted phylogenetic distribution of PHA degrading bacteria. Two concatenated ribosomal protein trees inferred using FastTree show the placement of 115 high-quality MAGs

encoding predicted PHA depolymerases within the bacterial domain. **A)** All predicted PHA degrading organisms place within select phyla in a shared region (purple) of bacterial diversity, excluding the Candidate Phylum Radiation/Patescibacteria. **B)** Detailed phylogeny of the purple subset of bacteria from (A) containing all MAGs encoding predicted ePHA depolymerases. Phyla containing members with predicted PHA depolymerases are indicated by the coloured inner ring, and the number of predicted scl and mcl PHA depolymerases in each MAG are indicated by the size of the circles in the outer rings.

Figure 4: Environmental origin and phylogenetic relationships of the 231 PHA depolymerases predicted from 143 MAGs. Outer rings are colored by environmental origin of predicted PHA depolymerases. Inner rings are colored by phylum of the MAG encoding the predicted PHA depolymerase. ML trees showing **A)** 10 (8 unique) ePHAmcl depolymerases **B)** 156 (133 unique) ePHAscl T1 depolymerases and **C)** 65 (56 unique) ePHAscl T2 depolymerases.

Supplementary Figures

Supplementary Figure 1: World map showing geocoordinates of sampling locations for the 3,842 metagenomes screened in this analysis, colored by environmental subtype. Several samples share geocoordinates and represent samples taken at different depths, samples taken within close proximity to each other, or at different time points.

Supplementary Figure 2: Environmental origin and phylogenetic relationships of the 13,869 predicted PHA depolymerases identified in 1,295 metagenomes. Outer colored ring indicates the

environmental origin of the metagenome encoding each predicted PHA depolymerase. Biomarker enzymes are indicated with black arrows. Protein trees were inferred with FastTree and show the sequence diversity of A) 222 predicted ePHAmcl depolymerases B) 10,245 predicted ePHAsclT1 depolymerases and C) 3,402 predicted ePHAsclT2 depolymerases, each with biochemically confirmed biomarker proteins included as references.

Supplementary Data Files

Supplementary Data 1: Supp_Data_1_PHA_depolymerases_references.xls

Supplementary Data 2: Supp_Data_2_PHA_depolymerase_reference_set_191118.fasta

Supplementary Data 3: Supp_Data_3_3842_IMG_metadata.xls

Supplementary Data 4: Supp_Data_4_5290_IMG_MAGs_metadata.xls

Supplementary Data 5: Supp_Data_5_final_trees_in_newick_format.txt

Supplementary Data 6: Supp_Data_6_cd_hit_uncluster_using_lengths.txt

Supplementary Data 7: Supp_Data_7_PHA_depolymerase_MAGs_CheckM_env.xls

Supplementary Data 8: Supp_Data_8_unique_archaeal_IMG_MAG_hits_e10.xls

References

1. Shah AA, Hasan F, Hameed A, Ahmed S. 2008. Biological degradation of plastics: A comprehensive review. Biotechnology Advances 26:246–265.
2. Barnes D, Galgani F, Thompson RC, Barlaz M. 2009. Accumulation and fragmentation of plastic debris in global environments. Philosophical Transactions of the Royal Society B: Biological Sciences 364:1985–1998.
3. Geyer R, Jambeck JR, Law KL. 2017. Production, use, and fate of all plastics ever made. Science Advances 3:25–29.
4. Ellen MacArthur Foundation. 2016. The New Plastics Economy: Rethinking the future of plastics. World Economic Forum, Cologny, Switzerland.
5. Li WC, Tse HF, Fok L. 2016. Plastic waste in the marine environment: A review of sources, occurrence and effects. Science of the Total Environment. Elsevier B.V.
6. Law KL. 2017. Plastics in the marine environment. Annual Review of Marine Science 1:63–67.

7. Excell C, Salcedo-La Viña C, Worker J, Moses E. 2018. Legal Limits on Single-Use Plastics and Microplastics: A Global Review of National Laws and Regulation. United Nations Environment Programme, Nairobi, Kenya.
8. European Bioplastics. 2018. Bioplastics market data 2018.European Bioplastics, Berlin, Germany.
9. Aeschelmann F, Carus M. 2015. Biobased building blocks and polymers in the world: Capacities, production, and applications-status quo and trends towards 2020. Industrial Biotechnology 11:154–159.
10. Koller M, Maršálek L, de Sousa Dias MM, Braunegg G. 2017. Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner. New Biotechnology 37:24–38.
11. Chen GQ, Jiang XR. 2017. Engineering bacteria for enhanced polyhydroxyalkanoates (PHA) biosynthesis. Synthetic and Systems Biotechnology 2:192–197.
12. Park SJ, Kim TW, Kim MK, Lee SY, Lim SC. 2012. Advanced bacterial polyhydroxyalkanoates: Towards a versatile and sustainable platform for unnatural tailor-made polyesters. Biotechnology Advances 30:1196–1206.
13. Jendrossek D, Handrick R. 2002. Microbial Degradation of Polyhydroxyalkanoates. Annual Review of Microbiology 56:403–432.
14. Boyandin AN, Prudnikova S v., Karpov VA, Ivonin VN, Dỗ NL, Nguyễn TH, Lê TH, Filichev NL, Levin AL, Filipenko ML, Volova TG, Gitelson II. 2013. Microbial degradation of polyhydroxyalkanoates in tropical soils. International Biodeterioration and Biodegradation 83:77–84.
15. Nishida H, Tokiwa Y. 1993. Distribution of Poly(ß-hydroxybutyrate) and Poly(€-caprolactone) Aerobic Degrading Microorganisms in Different Environments. Journal of Environmental Polymer Degradation 1:227–233.
16. Dilkes-Hoffman LS, Lant PA, Laycock B, Pratt S. 2019. The rate of biodegradation of PHA bioplastics in the marine environment: A meta-study. Marine Pollution Bulletin 142:15–24.
17. Altaee N, El-Hiti GA, Fahdil A, Sudesh K, Yousif E. 2016. Biodegradation of different formulations of polyhydroxybutyrate films in soil. SpringerPlus 5:1–12.
18. Volova TG, Boyandin AN, Vasil'ev AD, Karpov VA, Kozhevnikov I v., Prudnikova S v., Rudnev VP, Xuản BB, Dũng VV, Gitel'zon II. 2011. Biodegradation of polyhydroxyalkanoates (PHAs) in the South China Sea and identification of PHA-degrading bacteria. Microbiology 80:252–260.
19. Voinova O, Gladyshev M, Volova TG. 2008. Comparative study of PHA degradation in natural reservoirs having various types of ecosystems. Macromolecular Symposia 269:34–37.
20. Mergaert J, Wouters A, Anderson C, Swings J. 1995. *In situ* biodegradation of poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in natural waters. Canadian Journal of Microbiology 41:154–159.
21. Fernandez-Castillo R, Rodriguez-valera F, Gonzalez-Ramos J, Ruiz-Berraquero F. 1986. Accumulation of Poly(B-Hydroxybutyrate) by Halobacteria 51:214–216.
22. Kadouri D, Jurkevitch E, Okon Y. 2003. Poly β-hydroxybutyrate depolymerase (*PhaZ*) in *Azospirillum brasilense* and characterization of a *phaZ* mutant. Archives of Microbiology 180:309–318.

23. Bresan S, Sznajder A, Hauf W, Forchhammer K, Pfeiffer D, Jendrossek D. 2016. Polyhydroxyalkanoate (PHA) granules have no phospholipids. Scientific Reports 6:1–13.
24. Tokiwa Y, Calabia BP. 2004. Degradation of microbial polyesters. Biotechnology Letters 26:1181–1189.
25. Knoll M, Hamm TM, Wagner F, Martinez V, Pleiss J. 2009. The PHA Depolymerase Engineering Database: A systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. BMC Bioinformatics 10:1–8.
26. Sood S, Sharma A, Sharma N, Kanwar SS. 2018. Carboxylesterases: Sources, Characterization and Broader Applications. Insights in Enzyme Research 01.
27. Abe T, Kobayashi T, Saito T. 2005. Properties of a novel intracellular poly(3-hydroxybutyrate) depolymerase with high specific activity (*PhaZd*) in *Wautersia eutropha* H16. Journal of Bacteriology 187:6982–6990.
28. Handrick R, Reinhardt S, Kimmig P, Jendrossek D. 2004. The "intracellular" poly(3-hydroxybutyrate) (PHB) depolymerase of *Rhodospirillum rubrum* is a periplasm-located protein with specificity for native PHB and with structural similarity to extracellular PHB depolymerases. Journal of Bacteriology 186:7243–7253.
29. Tokiwa Y, Calabia BP, Ugwu CU, Aiba S. 2009. Biodegradability of plastics. International Journal of Molecular Sciences 10:3722–3742.
30. European Bioplastics. 2019. Frequently Asked Questions on Bioplastics. European Bioplastics, Berlin, Germany.
31. ASTM International. 2015. ASTM D5338-15: Standard Test Method for Determining Aerobic Biodegradation of Plastic Materials Under Controlled Composting Conditions, Conditions, Incorporating Thermophilic Temperatures. ASTM Standards, Pennsylvania, United States.
32. Haider TP, Volker J, Kramm J, Landfester K, Wurm FR. 2019. Plastics of the Future? The Impact of Biodegradable Polymers on the Environment and on Society. Angewandte Chemie - International Edition 58:50–62.
33. Martínez V, de la Peña F, García-Hidalgo J, de la Mata I, García JL, Prieto MA. 2012. Identification and biochemical evidence of a medium-chain-length polyhydroxyalkanoate depolymerase in the *Bdellovibrio bacteriovorus* predatory hydrolytic Arsenal. Applied and Environmental Microbiology 78:6017–6026.
34. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nature Microbiology 1:1–6.
35. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:1–8.
36. Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG/M: The integrated metagenome data management and comparative analysis system. Nucleic Acids Research 40:123–129.
37. Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo A v., Somenahally AC, Elias DA. 2015. Global prevalence and distribution of genes and microorganisms involved in mercury methylation. Science Advances 1:1–13.

38. Holert J, Cardenas E, Bergstrand LH, Zaikova E, Hahn AS, Hallam SJ, Mohn WW. 2018. Metagenomes reveal global distribution of bacterial steroid catabolism in natural, engineered, and host environments. mBio 9:1–18.
39. Alteio L., Schulz F, Seshadri R, Varghese N, Rodriguez-Reillo W, Ryan E, Goudeau D, Eichorst SA, Malmstrom RR, Bowers RM, Katz LA, Blanchard JL, Woyke T. 2020. Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. mSystems 5:1–18.
40. Schulz F, Alteio L, Goudeau D, Ryan EM, Yu FB, Malmstrom RR, Blanchard J, Woyke T. 2018. Hidden diversity of soil giant viruses. Nature Communications 9:1–9.
41. Sharrar A, Crits-Christoph A, Meheust R, Diamond S, Starr E, Banfield JF. 2019. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. bioRxiv 818815.
42. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, Lane KR, Thomas BC, Pan C, Northen TR, Banfield JF. 2019. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nature Microbiology 4:1356–1367.
43. Bouma-Gregson K, Olm MR, Probst AJ, Anantharaman K, Power ME, Banfield JF. 2019. Impacts of microbial assemblage and environmental conditions on the distribution of anatoxin-a producing cyanobacteria within a river network. ISME Journal 13:1618–1634.
44. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, Sunagawa S, Hingamp P, Ogata H, Lima-Mendez G, Roux S, González JM, Arrieta JM, Alam IS, Kamau A, Bowler C, Raes J, Pesant S, Bork P, Agustí S, Gojobori T, Bajic V, Vaqué D, Sullivan MB, Pedrós-Alió C, Massana R, Duarte CM, Gasol JM. 2019. Metabolic Architecture of the Deep Ocean Microbiome. bioRxiv 635680.
45. He S, Malfatti SA, McFarland JW, Anderson FE, Pati A, Huntemann M, Tremblay J, de Rio TG, Waldrop MP, Windham-Myers L, Tringe SG. 2015. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. mBio 6:1–15.
46. Butterfield CN, Li Z, Andeer PF, Spaulding S, Thomas BC, Singh A, Hettich RL, Suttle KB, Probst AJ, Tringe SG, Northen T, Pan C, Banfield JF. 2016. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ 2016:1–28.
47. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature Communications 7:1–11.
48. Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. 2016. Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus* Accumulibacter phosphatis. ISME Journal 10:810–822.
49. Yao Q, Li Z, Song Y, Wright SJ, Guo X, Tringe SG, Tfaily MM, Paša-Tolić L, Hazen TC, Turner BL, Mayes MA, Pan C. 2018. Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. Nature Ecology and Evolution 2:499–509.

50. Flores-Núñez VM, Fonseca-García C, Desgarennes D, Eloe-Fadrosh E, Woyke T, Partida-Martínez LP. 2020. Functional Signatures of the Epiphytic Prokaryotic Microbiome of Agaves and Cacti. Frontiers in Microbiology 10:1–13.
51. Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, Wilkins MJ, Williams KH, Singh A, Banfield JF. 2016. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environmental Microbiology 18:159–173.
52. Lavy A, McGrath DG, Matheus Carnevali PB, Wan J, Dong W, Tokunaga TK, Thomas BC, Williams KH, Hubbard SS, Banfield JF. 2019. Microbial communities across a hillslope-riparian transect shaped by proximity to the stream, groundwater table, and weathered bedrock. Ecology and Evolution 9:6869–6900.
53. Hartman WH, Ye R, Horwath WR, Tringe SG. 2017. A genomic perspective on stoichiometric regulation of soil carbon cycling. ISME Journal 11:2652–2665.
54. Kimbrel JA, Ballor N, Wu YW, David MM, Hazen TC, Simmons BA, Singer SW, Jansson JK. 2018. Microbial community structure and functional potential along a hypersaline gradient. Frontiers in Microbiology 9:1–15.
55. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.
56. Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
57. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees2010 Gateway Computing Environments Workshop, GCE 2010.
58. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37:291–294.
59. Bateman A. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research 47:506–515.
60. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44:733–745.
61. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo L di, Christie C, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranovic V, Guzenko D, Hudson BP, Liang Y, Lowe R, Peisach E, Periskova I, Randle C, Rose A, Sekharan M, Shao C, Tao YP, Valasatava Y, Voigt M, Westbrook J, Young J, Zardecki C, Zhuravleva M, Kurisu G, Nakamura H, Kengaku Y, Cho H, Sato J, Kim JY, Ikegawa Y, Nakagawa A, Yamashita R, Kudou T, Bekker GJ, Suzuki H, Iwata T, Yokochi M, Kobayashi N, Fujiwara T, Velankar S, Kleywegt GJ, Anyango S, Armstrong DR, Berrisford JM, Conroy MJ, Dana JM, Deshpande M, Gane P, Gáborová R, Gupta D, Gutmanas A, Koča J, Mak

L, Mir S, Mukhopadhyay A, Nadzirin N, Nair S, Patwardhan A, Paysan-Lafosse T, Pravda L, Salih O, Sehnal D, Varadi M, Văreková R, Markley JL, Hoch JC, Romero PR, Baskaran K, Maziuk D, Ulrich EL, Wedell JR, Yao H, Livny M, Ioannidis YE. 2019. Protein Data Bank: The single global archive for 3D macromolecular structure data. Nucleic Acids Research 47:520–528.

62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25:1043–1055.

63. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5:1–10.

64. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research 44:1–4.

65. Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

66. Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic acids research 41:22–28.

67. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5.

68. Martínez V, de Santos PG, García-Hidalgo J, Hormigo D, Prieto MA, Arroyo M, de la Mata I. 2015. Novel extracellular medium-chain-length polyhydroxyalkanoate depolymerase from *Streptomyces exfoliatus* K10 DSMZ 41693: a promising biocatalyst for the efficient degradation of natural and functionalized mcl-PHAs. Applied Microbiology and Biotechnology 99:9605–9615.

69. Santos M, Gangoiti J, Keul H, Möller M, Serra JL, Llama MJ. 2013. Polyester hydrolytic and synthetic activity catalyzed by the medium-chain-length poly(3-hydroxyalkanoate) depolymerase from *Streptomyces venezuelae* SO1. Applied Microbiology and Biotechnology 97:211–222.

70. Kim DY, Hyun CK, Sun YK, Young HR. 2005. Molecular characterization of extracellular medium-chain-length poly(3-hydroxyalkanoate) depolymerase genes from *Pseudomonas alcaligenes* strains. Journal of Microbiology 43:285–294.

71. Schirmer A, Jendrossek D. 1994. Molecular characterization of the extracellular poly(3-hydroxyoctanoic acid) [P(3HO)] depolymerase gene of *Pseudomonas fluorescens* GK13 and of its gene product. Journal of Bacteriology 176:7065–7073.

72. Lebreton LCM, van der Zwet J, Damsteeg JW, Slat B, Andrady A, Reisser J. 2017. River plastic emissions to the world's oceans. Nature Communications 8:1–10.

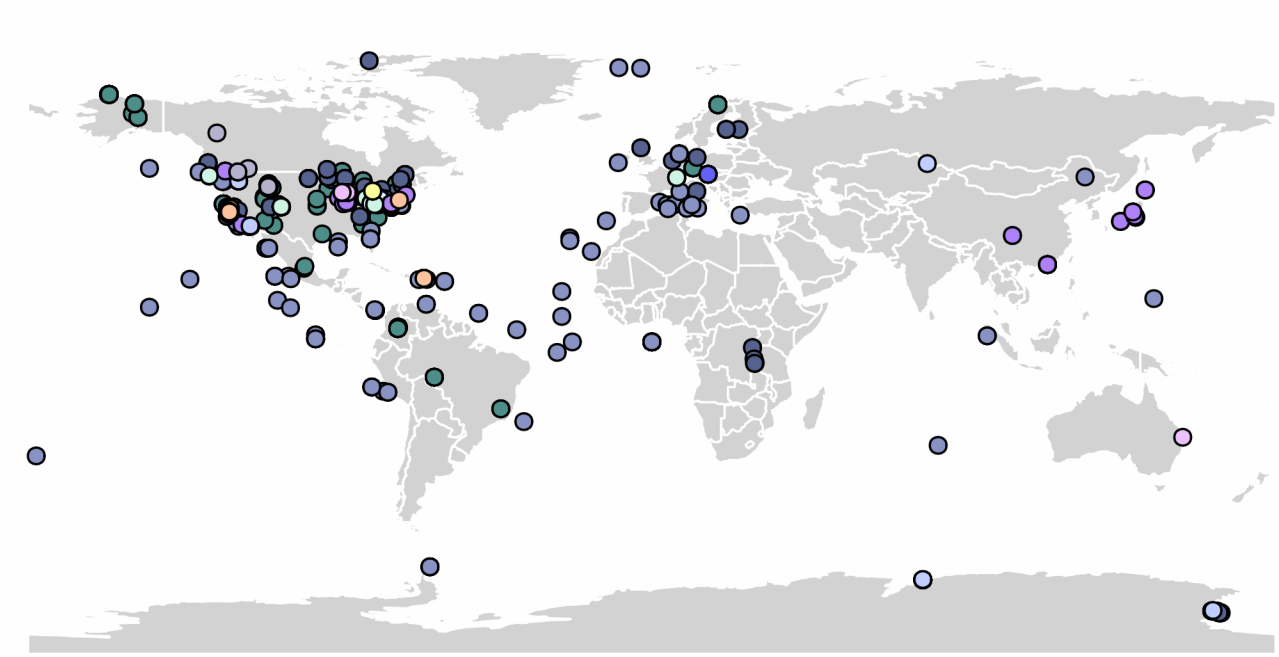
73. Wright SL, Thompson RC, Galloway TS. 2013. The physical impacts of microplastics on marine organisms: a review. Environmental Pollution 178:483–492.

74. Volova TG, Prudnikova S v., Vinogradova ON, Syrvacheva DA, Shishatskaya EI. 2017. Microbial Degradation of Polyhydroxyalkanoates with Different Chemical Compositions and Their Biodegradability. Microbial Ecology 73:353–367.

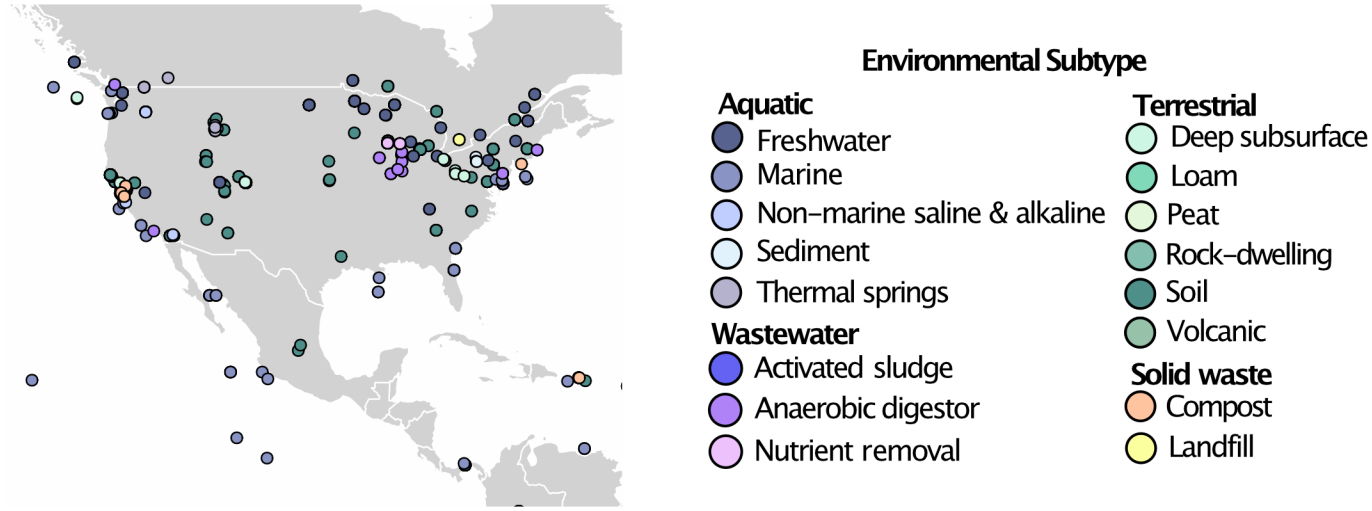
75. California Department of Resources Recycling and Recovery. 2012. Report Topic: PLA and PHA biodegradation in the marine environment. CalRecycle: California Department of Resources Recycling and Recovery, California, United States.

76. Lim SP, Gan SN, Tan IKP. 2005. Degradation of medium-chain-length polyhydroxyalkanoates in tropical forest and mangrove soils. Applied Biochemistry and Biotechnology 126:23–33.
77. Zettler ER, Mincer TJ, Amaral-Zettler LA. 2013. Life in the “plastisphere”: Microbial communities on plastic marine debris. Environmental Science and Technology 47:7137–7146.
78. Pinnell LJ, Turner JW. 2019. Shotgun metagenomics reveals the benthic microbial community response to plastic and bioplastic in a coastal marine environment. Frontiers in Microbiology 10:1–15.
79. Fierer N. 2017. Embracing the unknown: Disentangling the complexities of the soil microbiome. Nature Reviews Microbiology 15:579–590.
80. Mergaert J, Swings J. 1996. Biodiversity of microorganisms that degrade bacterial and synthetic polyesters. Journal of Industrial Microbiology and Biotechnology 17:463–469.
81. Emadian SM, Onay TT, Demirel B. 2017. Biodegradation of bioplastics in natural environments. Waste Management 59:526–536.
82. Quecholac-Piña X, Hernández-Berriel M del C, Mañón-Salas M del C, Espinosa-Valdemar RM, Vázquez-Morillas A. 2020. Degradation of Plastics under Anaerobic Conditions: A Short Review. Polymers 12:1–18.
83. Stamps BW, Lyles CN, Suflita JM, Masoner JR, Cozzarelli IM, Kolpin DW, Stevenson BS. 2016. Municipal solid waste landfills harbor distinct microbiomes. Frontiers in Microbiology 7:1–16.
84. Spokas K, Bogner J, Chanton JP, Morcet M, Aran C, Graff C, Golvan YM le, Hebe I. 2006. Methane mass balance at three landfill sites: What is the efficiency of capture by gas collection systems? Waste Management 26:516–525.
85. Pittmann T, Steinmetz H. 2017. Polyhydroxyalkanoate production on waste water treatment plants: Process scheme, operating conditions and potential analysis for German and European municipal waste water treatment plants. Bioengineering 54:1–24.
86. Handrick R, Reinhardt S, Focarete ML, Scandola M, Adamus G, Kowalczuk M, Jendrossek D. 2001. A New Type of Thermoalkalophilic Hydrolase of *Paucimonas lemoignei* with High Specificity for Amorphous Polyesters of Short Chain-length Hydroxyalkanoic Acids. Journal of Biological Chemistry 276:36215–36224.
87. Hatheway CL. 1990. Toxigenic Clostridia. Clinical Microbiology Reviews 3:66–98.
88. Pérez-Rivero C, López-Gómez JP, Roy I. 2019. A sustainable approach for the downstream processing of bacterial polyhydroxyalkanoates: State-of-the-art and latest developments. Biochemical Engineering Journal 150:1–14.

A)

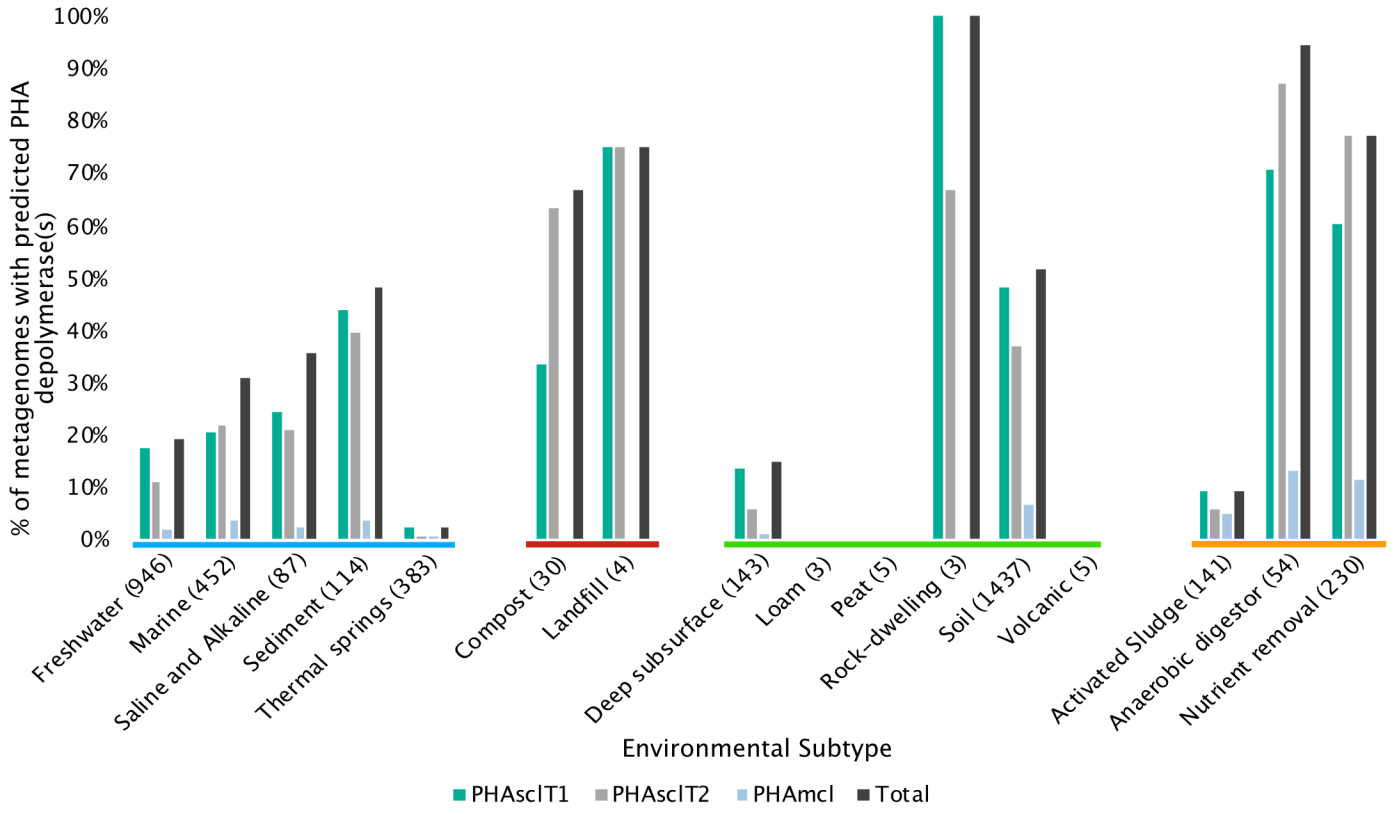


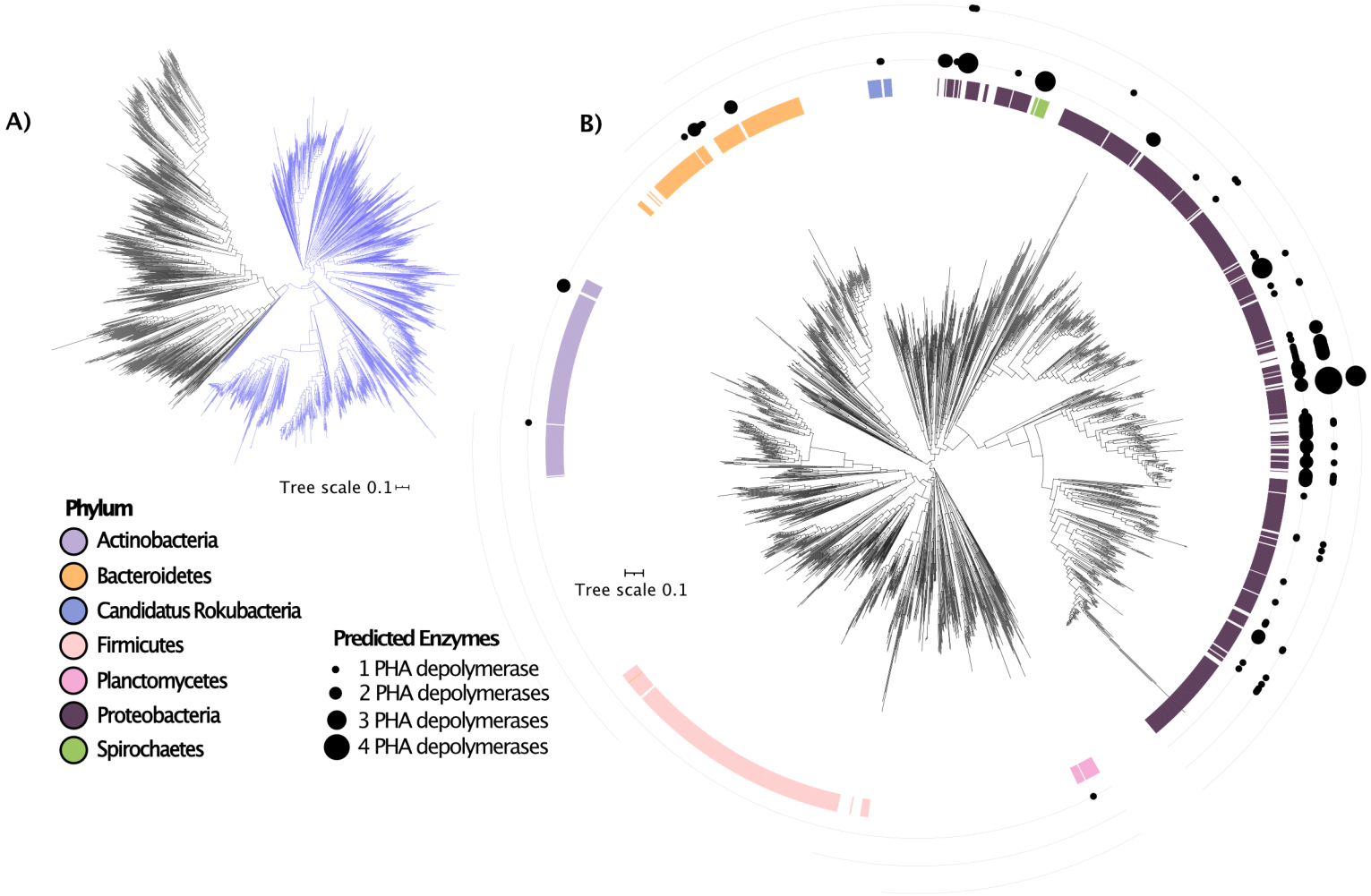
B)

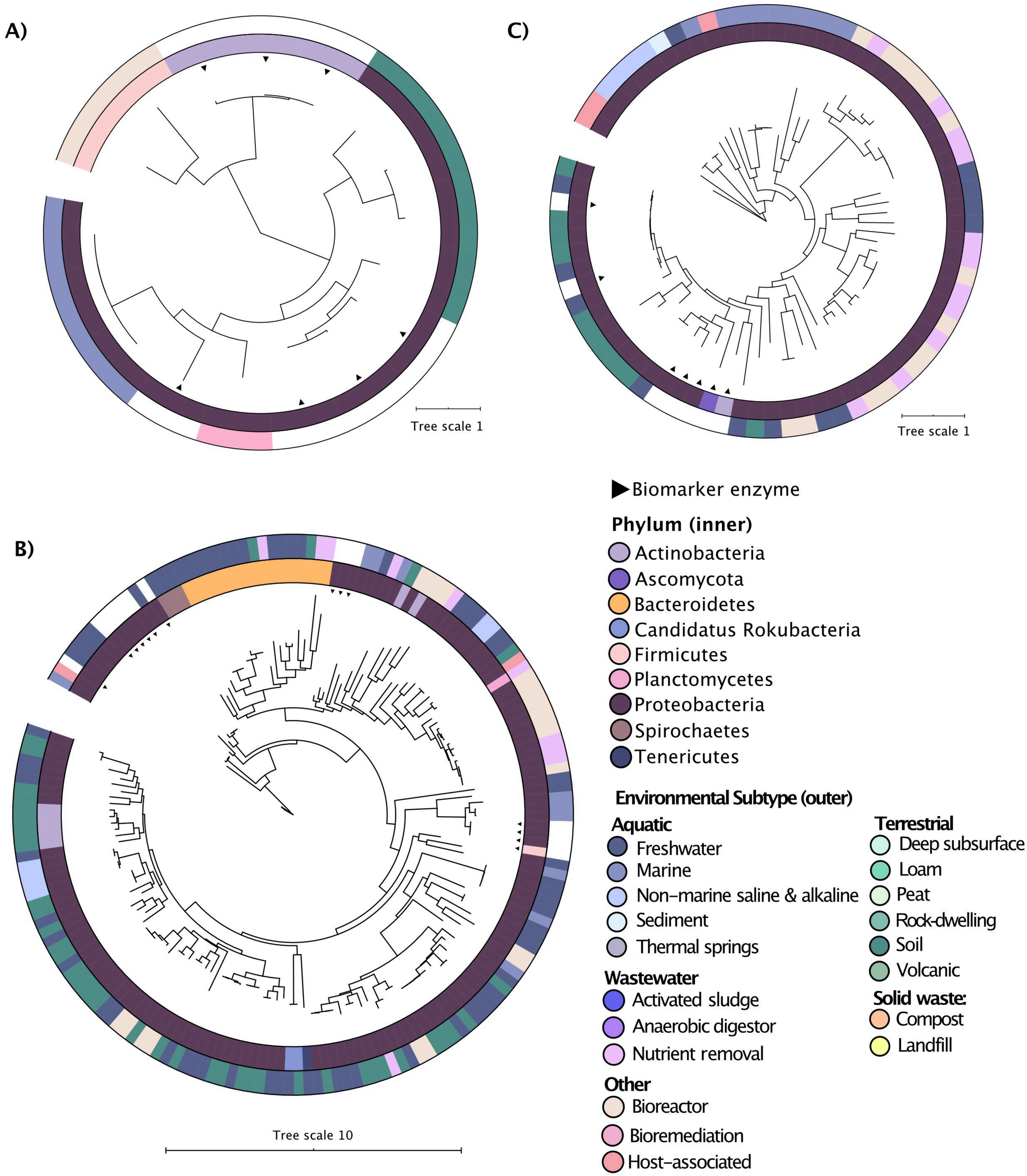

Environmental Subtype
Aquatic
Freshwater
Marine
Non–marine saline & alkaline
Sediment
Thermal springs
Wastewater
Activated sludge
Anaerobic digestor
Nutrient removal
Terrestrial
Deep subsurface
Loam
Peat
Rock–dwelling
Soil
Volcanic
Solid waste
Compost
Landfill

Metagenomes with predicted PHA depolymerase(s) (%)

Environment	PHAsclT1	PHAsclT2	PHAmcl	Total	Metagenomes (n)
Aquatic					
Freshwater	18%	11%	2%	19%	946
Marine	20%	22%	4%	31%	452
Saline and Alkaline	24%	21%	2%	36%	87
Sediment	44%	39%	4%	48%	114
Thermal springs	2%	1%	0%	2%	383
	17%	**13%**	**2%**	**21%**	**1,982**
Solid Waste					
Compost	33%	63%	0%	67%	30
Landfill	75%	75%	0%	75%	4
	38%	**65%**	**0%**	**68%**	**34**
Terrestrial					
Deep subsurface	13%	6%	1%	15%	143
Loam	0%	0%	0%	0%	3
Peat	0%	0%	0%	0%	5
Rock-dwelling	100%	67%	0%	100%	3
Soil	48%	37%	6%	51%	1,437
Volcanic	0%	0%	0%	0%	5
	45%	**34%**	**6%**	**48%**	**1,596**
Wastewater					
Activated Sludge	9%	6%	5%	9%	141
Anaerobic digestor	70%	87%	13%	94%	54
Nutrient removal	60%	77%	11%	77%	35
	31%	**36%**	**8%**	**40%**	**230**
Total	**30%**	**24%**	**4%**	**34%**	**3,842**




A)
Tree scale 0.1
Phylum
Actinobacteria
Bacteroidetes
Candidatus Rokubacteria
Firmicutes
Planctomycetes
Proteobacteria
Spirochaetes
Predicted Enzymes
1 PHA depolymerase
2 PHA depolymerases
3 PHA depolymerases
4 PHA depolymerases
B)
Tree scale 0.1


A)
Tree scale 1
C)
Tree scale 1
B)
Tree scale 10
Biomarker enzyme
Phylum (inner)
Actinobacteria
Ascomycota
Bacteroidetes
Candidatus Rokubacteria
Firmicutes
Planctomycetes
Proteobacteria
Spirochaetes
Tenericutes
Environmental Subtype (outer)
Aquatic
Freshwater
Marine
Non-marine saline & alkaline
Sediment
Thermal springs
Wastewater
Activated sludge
Anaerobic digestor
Nutrient removal
Other
Bioreactor
Bioremediation
Host-associated
Terrestrial
Deep subsurface
Loam
Peat
Rock-dwelling
Soil
Volcanic
Solid waste:
Compost
Landfill