The phylogenetic and global distribution of bacterial polyhydroxyalkanoate 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.

Originality/Significance Statement

Biodegradable plastics like polyhydroxyalkanoates (PHAs) are a hot topic, following ubiquitous environmental plastic contamination, government bans on single-use plastics, and a growing need for sustainable alternatives to petroleum-based plastics. Understanding the microbial conversion of PHAs in the environment and finding biomolecular tools that can act on PHAs is increasingly important as PHAs grow in popularity. In this study, we screened thousands of metagenomes and metagenome-assembled genomes (MAGs) to substantially increase the recognized diversity of PHA depolymerases, the key enzymes mediating PHA biodegradation. We use datasets from seven continents to provide a global summary of the distribution of PHA depolymerase genes in natural environments and waste-management systems. In tandem, we increase the number of described phylum-level lineages with PHA biodegradation potential. This work contributes a new understanding of the phylogenetic and environmental distribution of PHA depolymerases.

Summary

Polyhydroxyalkanoates (PHAs) are a family of microbially-made polyesters commercialized as biodegradable plastics. PHA production rates are predicted to increase as concerns around environmental plastic contamination and limited fossil fuel resources have increased the importance of biodegradable and bio-based plastic alternatives. Microbially-produced PHA depolymerases are the key enzymes mediating PHA biodegradation, but only a few PHA depolymerases have been well-characterized and screens employing metagenomic sequence data are lacking. Here, we used 3,078 metagenomes to analyze the distribution of PHA depolymerases in microbial communities from diverse aquatic, terrestrial and waste management

systems. We significantly expand the recognized diversity of this protein family by screening 1,914 Gb of sequence data and identifying 13,869 putative PHA depolymerases in 1,295 metagenomes. Our results indicate that PHA depolymerases are unevenly distributed across environments. We predicted the highest frequency of PHA depolymerases in wastewater systems and the lowest in marine and thermal springs. In tandem, we screened 5,290 metagenome-assembled genomes to describe the phylogenetic distribution of PHA depolymerases, which is substantially broader compared to current cultured representatives. The Proteobacteria and Bacteroidota are key lineages encoding PHA depolymerases, but PHA depolymerases were predicted from members of the Bdellovibrionota, Methylomirabilota, Actinobacteriota, Firmicutes, Spirochaetota, Desulfobacterota, Myxococcota and Planctomycetota.

Introduction

Global concerns associated with environmental plastic contamination and fossil fuel use have led to a demand for sustainable alternatives to petroleum-based plastics[1]. Increased public awareness coupled with ongoing legislative changes aimed at reducing single-use plastics are promoting a pivot to bio-based and biodegradable alternatives like polyhydroxyalkanoates (PHAs)[2,3]. Conventional plastics rely on fossil resources and are extremely recalcitrant to degradation, causing them to persist upon disposal[4]. Annual plastic production now exceeds 380 million tons, with most plastics experiencing a short service lifespan before being discarded[5]. It is estimated that 79% of plastic waste accumulates in landfills or the environment[5] where it can cause ecological damage[6].

PHAs offer an exciting solution to curb environmental damage caused by liberal plastic use because they are made from renewable carbon resources (bio-based) and biodegrade into

70 non-harmful products in natural environments through the action of microbially-secreted PHA
71 depolymerases[7-12]. PHAs are a naturally produced family of polyesters synthesized by bacteria
72 and archaea as intracellular energy and carbon storage compounds[13,14]. Commercial PHAs
73 currently represent a small fraction (1.4%) of the bioplastic market[15] but show the highest growth
74 rates with production capacities forecasted to more than triple in the next five years[16].
75 Rates of environmental plastic degradation are influenced by polymer properties (*e.g.*,
76 available surface area, monomeric composition, additives), the extant microbial community, and
77 abiotic factors (*e.g.*, temperature, UV, pH)[17]. PHA biodegradation is largely reliant on the
78 abundance and diversity of microorganisms encoding extracellular PHA depolymerases (the key
79 PHA degrading enzymes)[18,19]. PHAs are broadly classified as either short-chain length (PHAscl)
80 with 3-5 carbon atoms per monomer or medium-chain length (PHAmcl) with 6-15 carbon atoms
81 per monomer, where the size of the monomer impacts enzyme substrate specificity[13].
82 PHA depolymerases are a diverse family of intracellular and extracellular
83 carboxylesterases and are members of the alpha/beta-hydrolase fold family[20]. To date,
84 approximately 35 PHA depolymerases have been isolated and biochemically characterized (Supp
85 Data 1 and references within), from organisms within the phyla Ascomycetes, Firmicutes,
86 Proteobacteria (most) and Actinobacteria. PHA depolymerases are classified based on
87 subcellular location (extracellular or intracellular), substrate specificity (mcl [EC 3.1.1.76] or scl
88 [EC 3.1.1.75]), and features of the catalytic domain[13,20]. Intracellular PHA (inPHA)
89 depolymerases mobilize amorphous PHA stores in intracellular, protein-associated granules
90 referred to as carbonosomes[21]. Extracellular PHA (ePHA) depolymerases are secreted from the
91 cell to scavenge exogenous semi-crystalline PHAs found in the environment[13], and are the focus
92 of this project.

93 Early culture-based studies have isolated ePHAscl-degrading microorganisms from
94 different environments including soil, compost, sewage and aquatic environments, but indicate
95 that ePHAmcl degraders appear to be relatively rare[7,13,19]. Reflecting these findings, relatively
96 few ePHAmcl depolymerases have been characterized compared to ePHAscl depolymerases[20,22].
97 All known ePHA depolymerases contain a catalytic triad composed of a serine (embedded in a
98 lipase box motif), histidine, and aspartic acid[20]. ePHAscl depolymerases typically consist of a
99 signal peptide, a catalytic domain, a linker domain and a C-terminal substrate binding domain[13].
100 There are two recognized ePHAscl depolymerase subtypes (denoted type 1 [T1] and type 2 [T2])
101 which differ in the arrangement of their catalytic domains. In ePHAsclT1 depolymerases, the
102 oxyanion hole is N-terminal to the lipase box, whereas in ePHAsclT2 depolymerases the
103 oxyanion hole is C-terminal to the lipase box[13]. Characterized ePHAmcl depolymerases do not
104 exhibit a C-terminal substrate binding domain; instead the N-terminal region is presumed to
105 mediate substrate binding[13]. ePHAmcl depolymerases share no significant homology with
106 ePHAscl depolymerases except for the catalytic triad[20] which is non-specific as it is shared by all
107 carboxylesterases[23].
108 With PHA production levels expected to increase, a strong understanding of the diversity
109 and distribution of PHA-degrading enzymes and organisms is critical for understanding the
110 impact of PHAs on the environment and to reveal biomolecular tools useful for processing these
111 plastics[19]. Culture-based research has developed a foundational understanding of the enzymes
112 and organisms mediating PHA biodegradation (reviewed by [13]). However, many microbial
113 lineages are currently unamenable to culturing and are only recognized from culture-independent
114 analyses[24]. With the advent of high-throughput sequencing and metagenomics, there has been a

115 proliferation of publicly available DNA sequence data from microbial communities inhabiting
116 diverse global environments[25,26], enabling broad surveys of gene distributions[27-29].
117 Here, we employed a culture-independent approach and leveraged large amounts of
118 publicly available environmental sequence data to characterize the phylogenetic and
119 environmental distribution of PHA depolymerases and predict novel PHA-degrading enzymes.
120 We screened 5,290 metagenome-assembled genomes (MAGs) and 3,078 metagenomes
121 sequenced from globally distributed natural environments and waste-management systems.
122 Biochemically characterized ePHA depolymerases were used to seed a homology-based search
123 to identity novel PHA depolymerase candidates from different environments and across the
124 bacterial tree of life. Our analyses significantly expand the recognized diversity of PHA
125 depolymerases and inform our understanding of the frequency of PHA depolymerases in the
126 environment, with implications for the microbial conversion of PHA litter. Furthermore, we
127 identify enzymes and organisms that may prove suitable for future bioremediation, chemical
128 processing or biotechnological applications.
129
130 **Methods**
131 *Identification of ePHA depolymerases*
132 A reference set of biochemically validated ePHA depolymerases were gathered from the
133 literature (Supp Data 1) and used as queries in a BLASTp search (threshold $e < 1x10^{-10}$) against
134 3,078 assembled and annotated metagenomes (Supp Data 2) and 5,290 MAGs (Supp Data 3) that
135 were publicly available on the Integrated Microbial Genomes and Microbiomes system
136 (IMG/M) https://img.jgi.doe.gov/m/[26]. Datasets screened were publicly available, but
137 permissions were requested from PIs associated with the datasets. Many metagenomes came

138 from datasets associated with publications, including[30-45]. A series of curation steps were taken to
139 increase PHA depolymerase prediction accuracy and to reduce or eliminate non-functional
140 homologs. Briefly, curation included eliminating severely truncated hits, ensuring key catalytic
141 residues were present using multiple sequence alignments (MSAs), and protein-tree based
142 curation using tree topology and branch length of hits placed among reference enzymes (for
143 detailed curation notes see Supp Data 4). BLAST output data for the reference query sequence
144 with the lowest e-value and a summary of the average, minimum, and maximum bitscores,
145 identities, e-values, and lengths of predicted PHA depolymerases can be found in Supp Data 5.
146 Signal peptides were predicted for predicted PHA depolymerases using Signal P 5.0
147 (https://services.healthtech.dtu.dk/service.php?SignalP) with default settings using both gram
148 negative and gram positive predictions[46].
149
150 *Dataset selection, ecosystem classification & normalization*
151 All 3,078 metagenomic datasets had a minimum of 0.5 Mb of assembled sequence data
152 and were selected to analyze the environmental distribution of PHA depolymerases. All
153 metagenomic datasets had environmental metadata available in accordance with the GOLD five-
154 level ecosystem classification system ([1] ecosystem [2] ecosystem category [3] ecosystem type
155 [4] ecosystem subtype and [5] specific ecosystem)[47]. All environmental data provided was used
156 without modification with the exception of 8 metagenomes with ecosystem subtype
157 classifications of loam (3) and peat (5) which were modified to soil to match other loam and peat
158 samples. Additionally, sediment samples were inconsistently classified at the [5] specific
159 ecosystem type or [3] ecosystem type levels, which was remedied by manually grouping all
160 aquatic sediment samples together as an ecosystem type (see Supp Data 2 for modifications).

161 Collectively, 104 unique GOLD ecosystems were described (Supp Data 2). Less than two
162 percent of metagenomes (40/3,078) lacked latitude and longitude locations, which were input
163 manually using data given in an associated study, or approximate geocoordinates were used
164 given samples' described geographical locations (Supp Data 2). To account for differing dataset
165 sizes, we calculated the frequency (hits/Gb) of PHA depolymerases occurring in metagenomes
166 and MAGs (Supp Fig 1; Supp Data 3). Frequencies for different GOLD level ecosystem
167 classifications were derived by dividing the total number of enzymes identified in metagenomes
168 from an ecosystem category or subtype by the sum total of the sequence data screened from
169 metagenomes occurring within a group. Frequencies in the MAGs for different phyla were
170 calculated by dividing the total number of enzymes identified from a phylum by the sum total of
171 the sequence data screened from MAGs of that phylum. Gene frequencies were calculated in this
172 manner to normalize based on clade size.
173
174 *Taxonomic affiliation of PHA depolymerases*
175 Phylogenetic affiliations of PHA depolymerases were predicted for the unbinned PHA
176 depolymerase-encoding scaffolds with a minimum length of 2,000 bps (7,260 scaffolds) using a
177 diamond-blast search[48] against NCBI's non-redundant database and subsequently conducting a
178 lowest common ancestor analysis (LCA) with MEGAN6 v6.19.5[49] using GTDB taxonomy[50]. For
179 MAGs, 143 MAGs encoding predicted PHA depolymerases were assessed using CheckM
180 v1.0.13[51], with MAGs kept for on-going analyses if they scored >50% complete and <10%
181 contaminated (Supp Data 3). MAGs were subsequently dereplicated using dRep v. 2.5.3[52].
182 GTDB-tk[50], as implemented on Kbase[53], was used to taxonomically identify dereplicated MAGs.
183 Finally, a concatenated ribosomal protein (rp) tree was constructed to phylogenetically place the

184 dereplicated MAGs that were >50% complete and <10% redundant (Supp Data 3). A dataset of
185 16 rps (rpL2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17 and S19) were
186 gathered from MAGs and independently aligned. Alignments were masked to remove columns
187 comprised of 95%+ gaps. MAGs with <50% of expected ungapped residues over the
188 concatenated alignment were excluded. Phyla containing predicted PHA depolymerase-encoding
189 MAGs were selected from a bacterial reference set[24] and included in the final concatenated
190 alignment. The final alignment was composed of 2,789 positions, 886 reference taxa and 98
191 MAGs. The final concatenated rp tree was inferred using FastTree[54] as implemented in Geneious
192 v.10.2.6. (https://www.geneious.com) and the tree was visualized using iTOL
193 (https://itol.embl.de/)[55].
194

195 **Results**

196 **Environmental distribution of predicted PHA depolymerases**

197 All IMG/M metagenomes included in this study had a minimum of 0.5 Mb of assembled
198 sequence data and had available GOLD environmental metadata (Supp Data 2). The 3,078
199 metagenomes were classified into 104 specific ecosystems from the GOLD hierarchy level 3
200 (Supp Data 2). Metagenomes were treated at the level of ecosystem category (level 1: aquatic,
201 terrestrial, solid waste or wastewater) and ecosystem type (level 2: 14 subcategories). We focus
202 on specific ecosystems or metagenomes only for select notable cases, however all data including
203 raw number of predicted PHA depolymerases, presence/absence, and frequency of predicted
204 enzymes identified from each metagenome and each specific ecosystem type can be found in
205 Supp Data 2.

206 A total of 13,869 PHA depolymerases (13,377 unique) were predicted in 1,295
207 metagenomes from aquatic (417/1,492 = 27.9%), solid waste (23/34 =67.6%), terrestrial
208 (764/1,323 = 57.7%) and wastewater (91/229 = 39.7%) environments (Fig 1). Of these, 43%
209 (5,949/13,869) had predicted signal peptides. Most metagenomes had a low frequency of
210 predicted PHA depolymerases (0.3-10 hits/Gb) (Supp Fig 1, Supp Data 2). Looking broadly at
211 different ecosystem categories, the highest frequency of predicted PHA depolymerases occurred
212 in wastewater systems (10.73 hits/Gb) and terrestrial environments (8.22 hits/Gb) whereas solid
213 waste and aquatic systems had averages of 5.53 and 4.09 hits/Gb respectively (Table 1, Supp
214 Data 2).

215 **1) Aquatic:** across all aquatic ecosystem types, around 30% (±5%) of metagenomes had
216 predicted PHA depolymerases, with the exception of thermal spring metagenomes (7/121=
217 5.8%) (Table 1). Despite this finding, marine ecosystem metagenomes had the lowest frequency
218 of predicted enzymes within aquatic environments (1.88 hits/Gb). Notably, no metagenomes
219 from marine samples nor thermal springs contained any predicted PHAmcl enzymes. The highest
220 frequency of predicted PHA depolymerases in this category occurred in aquatic sediments (6.05
221 hits/Gb). (Table 1).

222 **2) Solid waste:** 66.7% (20/30) of compost metagenomes and 75.0% (3/4) of landfill
223 metagenomes had predicted PHAscl depolymerases at a frequency of 9.37 hits/Gb and 3.04
224 hits/Gb respectively (Table 1).

225 3) **Terrestrial:** Soil microbial communities have been extremely well-sampled relative to
226 other environments. From our screen, 61.1% (740/1,211) of soil metagenomes had predicted
227 PHA depolymerases at a frequency of 8.22 hits /Gb. Other terrestrial environments with

228 predicted depolymerases were deep subsurface environments (21/104= 20.2%, 7.97 hits/Gb) and
229 rock-dwelling communities (3/3 = 100.0%, 10.81 hits/Gb) (Table 1).
230 **4): Wastewater:** Within wastewater environments, the majority of anaerobic digestors
231 (51/54 = 94.4%) and nutrient removal communities (27/35=77.1%) had predicted PHA
232 depolymerases, whereas only a small fraction of activated sludge communities had any
233 (13/140=9.3%). Despite this, activated sludge communities had a high frequency of predicted
234 PHA depolymerases (13.82 hits/Gb), second to nutrient removal systems (16.34 hits/Gb), while
235 anaerobic digestors were more similar to sediments and deep subsurface environments with 7.50
236 hits/Gb (Table 1; Supp Data 2).
237
238 **The environmental distribution of PHA depolymerase subtypes**
239 **1) PHAmcl depolymerases***:* PHAmcl depolymerases were sparsely predicted across
240 environments, with only 222 PHAmcl depolymerases identified from 153 metagenomes (Supp
241 Fig 2a) from aquatic (41/1,492 = 2.7%, 0.11 hits/Gb), terrestrial (94/1323 = 7.1%, 0.10 hits/Gb),
242 and wastewater (18/229 = 7.9%, 0.49 hits/Gb) environments (Table 1). The highest frequency of
243 PHAmcl depolymerases was detected in nutrient removal (1.15 hits/Gb) and activated sludge
244 (0.70 hits/Gb) wastewater systems. Solid waste systems had no predicted PHAmcl
245 depolymerases (Table 1). Of the 222 PHAmcl depolymerases predicted in this study, 208 placed
246 in maximum-likelihood trees with reference enzymes isolated from species of *Bdellovibrio*[56] and
247 *Pseudomonas*[57,58]. Another 14 predicted PHAmcl depolymerases formed a second clade with
248 enzymes recently isolated from *Streptomyces* species, which share limited sequence homology
249 with other ePHAmcl biomarker enzymes[22,59] (Supp Fig 2a).

250 **2) PHAscl depolymerases:** Predicted PHAscl depolymerases were more prevalent across
251 environments. A total of 10,245 PHAsclT1 depolymerases were predicted from 1,136
252 metagenomes (Supp Fig 2b), from aquatic (338/1,492 =22.7%, 2.81 hits/Gb), solid waste (13/34
253 = 38.2 %, 3.03 hits/Gb), terrestrial (713 /1,323 = 53.9%, 6.33 hits/Gb), and wastewater (72/229 =
254 31.4%, 5.15 hits/Gb) environments (Table 1). A large number of T1 depolymerases were
255 predicted in soil metagenomes (8,331 enzymes), but when scaled for dataset size this normalized
256 to a frequency of 6.32 hits/Gb. All three metagenomes sequenced from rock-dwelling microbial
257 communities encoded T1 enzymes and collectively had the highest frequency across ecosystem
258 types with 9.17 hits/Gb. Activated sludge wastewater systems also had a higher frequency of
259 predicted T1 enzymes with 8.47 hits/Gb.
260 For PHAsclT2 depolymerases, 3,402 were predicted from 908 metagenomes (Supp Fig
261 2c) from aquatic (264/1,492 =17.7%, 1.16 hits/Gb), solid waste (22/34 = 64.7%, 2.50 hits/Gb),
262 terrestrial (540/1,323 = 40.8%, 1.80 hits/Gb), and wastewater (82/229 = 35.8%, 5.10 hits/Gb)
263 environments (Table 1). Here too, a large number of hits (2,374) came from soil metagenomes,
264 but soil did not have a particularly high frequency of enzymes (1.80 hits/Gb) compared to other
265 ecosystem types. In fact, the frequency of predicted T2 enzymes was low (<2.5 hits/Gb) across
266 most ecosystems with the exceptions of compost (4.85 hits/Gb) and wastewater systems (5.10
267 hits/Gb).
268 Tree files, MSAs and fasta files for all predicted PHA depolymerases and information
269 linking hits to associated metagenome information including environmental metadata, size,
270 scaffold read depth etc., can be found in Supp Data 2 and https://github.com/vrviljak/-The-
271 phylogenetic-and-global-distribution-of-polyhydroxyalkanoate-bioplastic-degrading-genes.git.
272

273 **Taxonomic affiliation of unbinned PHA depolymerase-encoding scaffolds**

274 Only a few biochemically characterized ePHA depolymerases exist, which originate from
275 the Proteobacteria, Actinobacteriota, Firmicutes, Bdellovibrionota, and Ascomycetes (Supp Data
276 1). To assess the taxonomic distribution of unbinned PHA depolymerase-encoding scaffolds
277 >2000 bp in length (7,250 scaffolds), taxonomic affiliations were inferred using MEGAN[49] with
278 GTDB-tk[50] taxonomy (Supp Data 2). Most sequences were assigned to the Proteobacteria,
279 Actinobacteriota, Chloroflexota, Methylmirabilota, Bacteroidota, and Myxococcota (Supp Fig
280 3). Predicted PHAmcl depolymerase-encoding scaffolds were mainly affiliated with the
281 Myxococcota (26.0%), Proteobacteria (17.8% Gammaproteobacteria and 1.4%
282 Alphaproteobacteria), Actinobacteriota (16.4%), and Bdellovibrionota (4.1%). Scaffolds
283 encoding predicted PHAmcl depolymerases also originated from the Deinococcota (1),
284 Dormibacterota (1), and Thermoplasmatota (1) (Supp Data 2). 27.4% were unclassified.

285 The majority of predicted PHAsclT1 enzymes were on scaffolds affiliated with the
286 Proteobacteria (64.1%), with some representation from the Actinobacteriota (5.0%),
287 Chloroflexota (4.7%), Methylomirabilota (3.4%), Bacteroidota (2.3%), Myxococcota (1.4%) and
288 Acidobacteriota (1.0%) (Supp Fig 3). Unclassified scaffolds represented a further 14.6%, and 17
289 other phyla were represented at <1% of sequences (Supp Data 2). PHAsclT2 enzymes were
290 mostly predicted to be from the Proteobacteria (80.4%), with some in the Actinobacteriota
291 (2.3%) (Supp Fig 3). T2 enzymes were predicted at <1% in 12 phyla, with 13.3% unclassified at
292 the phylum level (Supp Data 2). Six PHA depolymerase scaffolds were assigned to archaeal
293 phyla including the Halobacteriota (5), encoding predicted PHAsclT1 depolymerases, and one
294 Thermoplasmatota scaffold encoding a predicted PHAmcl depolymerase. 5,122 predicted PHA
295 depolymerase encoding scaffolds were classified down to the family level (2,138 or 29.5% of the

296 phylum-level classified scaffolds were unclassified at the family-level). Family classified
297 scaffolds were affiliated with 184 different families, including the Xanthobacteraceae (16.0%),
298 Burkholderiaceae (13.1%), Beijerinckiaceae (4.7%), Rhodocyclaceae (3.0%), Alteromonadaceae
299 (1.7%), Methyloligellaceae (1.6%) of the Proteobacteria, the Ktedonobacteraceae (3.09%) of the
300 Chloroflexota, and the GWA2-73-35 (2.4%) of the Methylomirabilota. A complete family-level
301 overview of predicted PHA depolymerases can be found in Supp Data 2.

302 Across ecosystem types the Proteobacteria (Alpha and Gamma) were the most abundant
303 predicted PHA depolymerase hosts (Fig 2). Soil environments harboured the highest diversity of
304 predicted PHA depolymerase hosts with 24 phyla predicted. Notably, several phyla were found
305 only in soil (at <1% abundance) - the Desulfobacterota, Dormibacterota, Elusimicrobiota,
306 Nitrospirota, Patescibacteria, Poribacteria, and Thermoplasmatota (Supp Data 2). The
307 Myxococcota-encoded PHA depolymerases appeared in relatively high levels in thermal springs
308 (19.1% of taxonomically assigned PHA depolymerase-encoding scaffolds) and lower levels in
309 some other ecosystem types (1.9% in freshwater, 1.2% in sediment, 1.5% in marine, 7.5% in
310 nutrient removal, 5.4% in activated sludge and 1.8% in anaerobic digestors). Chloroflexota-
311 associated PHA depolymerases were also relatively abundant in thermal springs (9.5%) and only
312 otherwise found in soil (4.5%). Five phyla were detected in all wastewater ecosystem types - the
313 Proteobacteria, Bacteroidota, Gemmatimonadota, Myxococcota, and Spirochaetota.

314 Predicted PHA depolymerases co-occurred in pairs on the same assembled scaffold 39
315 times (1.1%), including 28 pairs of predicted PHAsclT1 depolymerases predicted to be from the
316 Proteobacteria, Gemmatimonadota, Chloroflexota, Actinobacteriota, Myxococcota, and
317 Bacteroidota, six pairs of predicted PHAsclT2 depolymerases from the Proteobacteria and
318 Eremiobacterota, and one pair of predicted PHAmcl depolymerases from *Sandaracinus*

319 *amylolyticus* of the Myxococcota. Predicted PHAsclT1 and PHAsclT2 depolymerases co-
320 occurred in four cases, with three predicted as from the Gammaproteobacteria.
321
322 **Phylogenetic distribution of predicted PHA depolymerase-encoding MAGs**
323 To further describe the phylogenetic distribution of PHA depolymerases, we separately
324 screened 5,290 MAGs available on IMG/M for PHA depolymerases. A total of 4,927 bacterial
325 and 363 archaeal MAGs across over 70 phyla were included in our screen, which were
326 reconstructed from metagenomes originating from engineered (443), environmental (4,659),
327 host-associated (186), and unclassified (2) systems (Supp Data 3). We identified a total of 143
328 MAGs (2.7%) encoding a total of 231 predicted PHA depolymerases in MAGs from 10 different
329 bacterial phylum-level lineages. Of these, 71% (163/231) had predicted signal peptides. In our
330 screen, the highest frequency of predicted PHA depolymerases (hits/Gb) were identified in the
331 Proteobacteria (90.6 hits/Gb), followed by the Bdellovibrionota (52.5 hits/Gb), Myxococcota
332 (47.17 hits/Gb), Methylomirabilota (28.1 hits/Gb), Actinobacteriota (20.9 hits/Gb), Bacteroidota
333 (16.5 hits /Gb), Spirochaetota (16.1 hits /Gb), Firmicutes (12.8 hits /Gb), Desulfobacterota (5.8
334 hits /Gb), and Planctomycetota (3.4 hits/Gb). The remaining phyla represented in the MAG
335 dataset had no predicted hits (Supp Data 3). 110 dereplicated PHA depolymerase-encoding
336 MAGs were >50% complete and <10% redundant, and 98/110 encoded sufficient marker genes
337 for inclusion in a rp-based phylogenetic tree (Fig 3). GTDB-tk successfully classified all but
338 5/110 MAGs (Supp Data 3), with those 5 placing among the Proteobacteria in the concatenated
339 rp tree (Fig 3). Notably, we did not detect any MAGs with predicted PHA depolymerases from
340 MAGs associated with the Candidate Phyla Radiation/Patescibacteria.

341 **1) Proteobacteria:** The vast majority of PHA depolymerases were predicted in
342 dereplicated MAGs classified as Alphaproteobacteria (14/110) and Gammaproteobacteria
343 (63/110). From the Alphaproteobacteria we identified MAGs from 10 families including the
344 Hyphomicrobiaceae (3 MAGs) and the Sphingomonadaceae (4 MAGs). From the
345 Gammaproteobacteria we identified MAGs from 15 families with most from the
346 Burkholderiaceae (31 MAGs), the Rhodocyclaceae (10 MAGs), and the Alteromonadaceae (5
347 MAGs).Within the Proteobacteria, all three ePHA depolymerase types were detected (mcl = 4;
348 sclT1 = 83; sclT2 = 45).
349 **2) Bacteroidota:** Beyond the Proteobacteria, 11 MAGs from the Bacteroidota encoded
350 one (9 MAGs) or two (2 MAGs) predicted ePHAsclT1 depolymerases from seven families. An
351 2-12-FULL-35-15 and a GWF2-33-38 MAG each encoded two predicted ePHAsclT1 enzymes
352 that shared high sequence similarity. Bacteroidota with one predicted PHAsclT1 depolymerase
353 included members of the b-17BO (1), GWA2-32-17 (1), GWF2-33-38 (3), GWF2-35-48 (2),
354 Saprospiraceae (1), and UBA2798 (1).
355 **3) Myxococcota:** Seven Myxococcota (formally the deltaproteobacteria) encoded
356 predicted PHA depolymerases. Three PHAmcl depolymerases were predicted from two SG8-38
357 MAGs and a Polyangiaceae. A PHAsclT1 depolymerase was predicted from a Haliangiaceae,
358 Nannocystaceae, SG8-38 and three from a Polyangiaceae MAG.
359 **4) Actinobacteriota:** Five Actinobacteriota MAGs from the Cellulomonadaceae,
360 Pseudonocardiaceae and QHCA01 were predicted to encode ePHAsclT1 depolymerase(s). Three
361 were Pseudonocardiaceae, two of which were predicted to encode two ePHAsclT1
362 depolymerase.

363 **5) Firmicutes:** Two Clostridia (Firmicutes A) both from the Ruminococcaceae family
364 sequenced from an anerobic bioreactor were predicted to have a PHAmcl depolymerase. One
365 UBA5603 (Firmicutes) was predicted to encode a PHAsclT1 depolymerase.
366 **6) Others:** Three Bacteriovoracales from the Bdellovibrionota encoded one (1 MAG) or
367 two (2 MAGs) predicted PHAsclT1 depolymerases. A Spirochaetota from freshwater encoded
368 three predicted PHAsclT1 depolymerases, a Desulfobacterota, a Planctomycetota and a
369 Methylomirabilota (formally the Candidate phylum Rokubacteria) all encoded one. For a
370 detailed overview of the PHA depolymerase complement predicted for each MAG and
371 taxonomic affiliations, see Supp Data 3.
372
373 **MAGs encoding multiple predicted PHA depolymerases:** Of the 110 dereplicated MAGs with
374 predicted PHA depolymerases, 68 were predicted to encode one PHA depolymerase, 25 were
375 predicted to encode two, 16 were predicted to encode three, and one, a *Herminiimonas* species
376 (98.48% complete, 4.69% redundant) from the Burkholderiaceae family, was predicted to encode
377 nine PHA depolymerases (Supp Data 3). The *Herminiimonas* MAG was predicted to have two
378 closely related ePHAsclT1, four ePHAsclT2 and three closely related ePHAmcl depolymerases.
379 MAGs from the Burkholderiaceae (6), Rhodocyclaceae (6) and Halomonadaceae (1) from the
380 Proteobacteria all encoded three predicted PHAscl enzymes each encoding both T1 and T2
381 enzyme(s). One unclassified Proteobacteria that clustered closely with a Bradyrhizobiaceae
382 reference sequence in the rp tree encoded three closely related PHAsclT1 enzymes. MAGs from
383 the Polyangiaceae of the Myxococcota (1) and GWB1-36-13 of the Spirochaetota (1) each
384 encoded three predicted PHAscl T1 depolymerases. Twelve of the Proteobacteria with three

385 predicted depolymerases were from the Burkholderiales and one from the Pseudomonadales,
386 each encoding both T1 and T2 enzyme(s).
387
388 **Phylogenetic distribution of MAG-encoded PHA depolymerase subtypes:**
389 **1) ePHAmcl depolymerases:** We predicted 8 unique ePHAmcl depolymerase genes from 6
390 dereplicated MAGs (Fig 4a). All PHAmcl depolymerases occurred as singletons with the
391 exception of the three predicted in the *Herminiimonas*-related MAG, which co-occurred with six
392 PHAscl depolymerases. Three PHAmcl depolymerases were predicted in members of the
393 Myxococcota and two were predicted in Firmicutes. Based on ML analyses, the predicted
394 Firmicutes ePHAmcl depolymerases were more closely related to the biochemically validated
395 *Streptomyces* ePHAmcl depolymerases, whereas the other predicted PHAmcl enzymes were
396 more closely related to those reference ePHA depolymerases isolated from *Pseudomonas* and
397 *Bdellovibrio* (Fig 4a).
398 **2) ePHAscl depolymerases:** In our MAG screen, we predicted 109 unique ePHAsclT1
399 depolymerases from 84 dereplicated MAGs from the Proteobacteria (59), Bacteroidota (11),
400 Myxococcota (4), Actinobacteriota (3), Bdellovibrionota (2) Desulfobacterota (1), Firmicutes
401 (1), Methylomirabilota (1), Planctomycetota (1) and Spirochaetota (1) (Fig 4b). Predicted T1
402 enzymes occurred as mostly as singletons (62 MAGs) but also in pairs (19 MAGs), or triplets (3
403 MAGs). T1 and T2 enzymes were predicted to co-occur in 17 Proteobacteria MAGs. We
404 predicted 44 unique ePHAsclT2 depolymerases in 33 dereplicated MAGs exclusively from
405 members of the Proteobacteria (Fig 4c). Three MAGs were not classified with GTDB-tk, but
406 placed among the Proteobacteria in the concatenated rp tree near Burkholderiales (1 MAG) and
407 Rhizobiales (2 MAGs) reference sequences. Most predicted T2 enzymes occurred as singletons

408 (24) or co-occurred with T1 enzymes (23). Notably, seven Burkholderiales from the
409 Rhodocyclaceae family encoded two predicted T2 enzymes as well as a predicted T1 enzyme.
410

411 **Discussion**

412 PHA depolymerases degrade PHA substrates in the environment and have potential
413 applications for bioremediation, composting, and chemical recycling. The objective of this study
414 was to increase the recognized diversity of PHA depolymerases and to elucidate their
415 phylogenetic affiliations and occurrence in the environment. We used a homology-based search
416 seeded with biochemically validated ePHA depolymerases followed by manual curation to
417 identify 13,869 predicted PHA depolymerases in 1,295 annotated and assembled metagenomes
418 from diverse global environments (from all seven continents) and affiliated these enzymes with a
419 broader-than-expected distribution of taxonomic lineages. Our results indicate that PHA
420 depolymerases are globally distributed and occur in most ecosystems. However, PHA
421 depolymerases were not universally detected in metagenomic datasets (*e.g.*, less than a third of
422 the 1,492 aquatic metagenomes screened had predicted hits) or occurred with low frequency
423 (Supp Fig 1). PHAmcl depolymerases were particularly poorly represented in the metagenomic
424 dataset, with only 3% of aquatic metagenomes and no solid waste metagenomes harbouring
425 predicted enzymes. There are very few reports of isolated PHAmcl degraders in the literature[22]
426 and our screen supports a sparse environmental distribution of these organisms. We allowed a
427 permissive size cut-off threshold of 0.5Mb of assembled sequence data to retain even small
428 datasets. Notably, the majority of thermal spring and volcanic metagenomes were small (81/121
429 and 4/5 metagenomes were <10Mb respectively), as well as a significant fraction of sediment,
430 freshwater and activated sludge (212/442, 213/481 and 60/140 were <10Mb respectively). Since

431 we were less likely to identify enzymes in smaller metagenomes, we performed normalization by
432 dataset size to calculate PHA depolymerase frequency. We note that PHA depolymerases may be
433 present in a greater proportion of environments than our screen predict. We explored the
434 relationship between metagenome size and the frequency of predicted PHA depolymerases and
435 did not find a strong linear relationship (R^2= 0.0282) (Supp Fig 4). Predicted PHA
436 depolymerases were placed in ML trees anchored with reference enzymes (Supp Fig 2). Based
437 on tree placement, only a handful of predicted enzymes were closely related to biomarker
438 enzymes; the majority of predicted enzymes formed more distantly related clades. If functional,
439 these enzymes significantly expand the known sequence diversity of PHA depolymerases.
440 PHA depolymerase subtypes were predicted at different frequencies across ecosystems.
441 PHAsclT1 depolymerases were the most prevalent followed by PHAsclT2 enzymes, with
442 PHAmcl depolymerases having substantially lower frequencies across all ecosystems.
443 Wastewater metagenomes, particularly those from activated sludge and nutrient removal
444 systems, had the highest frequency of predicted PHA depolymerases and may be ecological
445 hotspots for these enzymes. Interestingly, wastewater has been explored as a promising carbon
446 source for PHA production[60]. With the predicted presence of PHA degraders in the majority of
447 wastewater systems, there is the possibility of engineering a circular system where wastewater
448 acts as both a feedstock and a waste processing component in the bioplastic lifecycle. Notably, a
449 high frequency of PHAsclT1 depolymerases were detected in terrestrial metagenomes, with soil
450 metagenomes harboring a huge number of novel predicted T1 enzymes, though at average or low
451 relative frequencies.
452 Prior to our screen, bacterial ePHA depolymerases were characterized from the
453 Proteobacteria, Actinobacteriota, Bdellovibrionota (formerly within the Deltaproteobacteria), and

454 Firmicutes, as well as one fungal phylum, the Ascomycetes. Our screen infers PHA
455 depolymerases occur in 23 additional bacterial phyla and 2 archaeal phyla. An LCA analysis on
456 unbinned predicted PHA-depolymerase encoding scaffolds associated this predicted function
457 with 28 different phylum-level lineages, 10 of which were also identified in our MAG screen.
458 MAGs allowed us to more confidently taxonomically place putative PHA depolymerase hosts,
459 compare enzyme frequencies between lineages, and place predicted enzymes within a genomic
460 context whereas our work with metagenomic scaffolds provided a larger dataset originating from
461 a greater number of environmental samples. Together, our analyses affiliated predicted PHA
462 depolymerases predominantly with the Proteobacteria, Bdellovibrionota, Myxococcota,
463 Methylomirabilota, Actinobacteriota, and Chloroflexota (Fig 3, Supp Fig 3). Many
464 Proteobacterial PHA degraders have been identified through culture-based studies, particularly
465 from soils[13] and many of the PHA depolymerases used to seed our screen were from this phylum
466 (Supp Data 1 and references within). For this reason, we cannot be certain if this phylum is
467 genuinely enriched for this function compared to other phyla, meaning efforts to characterize
468 distantly related enzymes from non-Proteobacterial lineages may confirm a broader phylogenetic
469 distribution for this enzyme family.
470 A small number of predicted PHAscl depolymerase-encoding scaffolds identified from
471 metagenomes were classified as archaeal, from the Halobacteriota, and one predicted PHAmcl
472 depolymerase-encoding scaffold was classified as Thermoplasmatota. A number of distantly-
473 related genes were identified in archaeal MAGs, but none passed our curation steps. These
474 archaeal sequences may represent true PHA depolymerases that are distantly related from
475 reference enzymes and thus may be interesting targets for future characterization (Supp Data 6).

476 Reference ePHAmcl depolymerases originated from the Bdellovibrionota (formerly a
477 lineage within the Deltaproteobacteria), Gammaproteobacteria, and Actinobacteriota. We
478 identified PHAmcl depolymerase-encoding MAGs from the Proteobacteria, Firmicutes and
479 Myxococcota (formerly the Deltaproteobacteria), with the highest overall frequency occurring in
480 the Firmicutes. More phylum-level diversity was found in the LCA analyses of PHAmcl
481 depolymerase-encoding scaffolds, but interestingly no Firmicutes were identified. Most scaffolds
482 were affiliated with the Proteobacteria, Myxococcota, and Actinobacteriota, but many were
483 unclassified. Intriguingly, a few ePHAmcl-encoding scaffolds were classified as Acidobacteriota,
484 Deinococcota, and Dormibacterota, which are lineages that have not yet been associated with
485 PHAmcl depolymerases. T2 reference enzymes originated from the Proteobacteria (most), a
486 fungal species (1) and Actinobacteriota (1). All MAG-encoded PHAsclT2 depolymerases
487 occurred in members of the Proteobacteria (mainly the Burkholderiales) and often co-occurred
488 with PHAsclT1 depolymerases. The restriction of predicted T2 enzymes to only a few lineages is
489 an interesting result which was mirrored by our LCA analyses on T2-encoding scaffolds.
490 Two biochemically validated intracellular PHAsclT1 depolymerases (ACN29674.1 &
491 BAE44206.1) are homologous to our reference extracellular PHAsclT1 depolymerases. To
492 assess cellular localization of our predicted PHA depolymerases, we predicted signal peptides.
493 The majority of MAG-encoded PHA depolymerases are expected to localize extracellularly. In
494 contrast, fewer than half of the metagenome-derived PHA depolymerases had predicted signal
495 peptides. We predict that a small fraction of identified PHA depolymerases may represent
496 intracellular enzymes, but suspect that signal prediction algorithms may overlook valid signal
497 peptides, particularly for organisms not well-represented in current databases.

498 We identified several hosts with multiple predicted PHA depolymerases, which may
499 indicate metabolic flexibility on PHA substrates. We consider these organisms and their PHA
500 depolymerases important targets for future characterization. Most notably, we identified a
501 *Herminiimonas* species from the Burkholderiales which encoded a complement of nine predicted
502 PHA depolymerases, suggesting it may be an efficient and/or highly versatile PHA degrader. We
503 also identified a scaffold encoding two PHAmcl enzymes that was classified as from a member
504 of the Myxococcota, indicating this family may be enriched for this rarer enzyme subtype.
505 Biodegradable plastics like PHA offer a promising solution to the devastating
506 environmental effects caused by conventional plastics, however it is important to understand the
507 potential impacts of adding a novel substrate into the environment. Our screen identified low
508 frequencies or non-detection of PHA depolymerases in many environments. The presence of
509 PHA waste may enrich low-abundance community members with PHA degradation capacity and
510 significantly alter microbial community structure. In a controlled trial, biofilms associated with
511 PHAs were dominated by sulfate-reducing microorganisms in comparison to polyethylene
512 terephthalate controls[61]. The potential for low abundance organisms to respond to PHAs in an
513 environment may allow more robust PHA degradation than our screen predicts, but also may
514 have unexpected impacts on microbial community dynamics and biogeochemical cycling.
515 In conclusion, our work presents a carefully curated set of predicted PHA depolymerases
516 to conservatively catalogue the existing environmental sequence diversity and the associated
517 taxonomic affiliations of this enzyme family. We dramatically expand the inventory of PHA
518 depolymerases available to target for future biochemical characterization, and affiliate many of
519 these enzymes with novel taxonomic lineages for this function. We also provide an overview of
520 the occurrence and frequency of PHA depolymerases detected in the environment. Future

research should employ a combination of metagenomic predictions, possibly leveraging hidden markov models (HMMs) for broader diversity capture, burial degradation trials with functional profiling, and biochemical validation of predicted enzymes (particularly for enzymes distantly related to biomarker enzymes) in targeted environments to confirm predictions developed here.

Data Availability

All predicted PHA depolymerase faa files, MSAs, and tree files are available at: https://github.com/vrviljak/-The-phylogenetic-and-global-distribution-of-polyhydroxyalkanoate-bioplastic-degrading-genes.git. All trees visualized with iTOL for this project can be viewed at: https://itol.embl.de/shared/2DHP8UFMQVeYO

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.

542 **Table 1:** Environmental distribution of 13,869 predicted PHA depolymerases (222 ePHAmcl, 10,245 ePHAsclT1 and 3,402

543 ePHAsclT2 depolymerases) identified by screening 3,078 metagenomes. Number of metagenomes screened, percent of metagenomes

544 with at least one or more predicted PHA depolymerase, the number of predicted PHA depolymerases identified, and the frequency of

545 predicted PHA depolymerases (hits/Gb) per ecosystem type are included.

Metagenomes Screened (n)	**# of predicted PHA depolymerases**				**% of metagenomes with predicted PHA depolymerases**				**Frequency of predicted PHA depolymerases (hits/Gb)**			
Ecosystem	**sclT1**	**sclT2**	**mcl**	**Total**	**sclT1**	**sclT2**	**mcl**	**Total**	**sclT1**	**sclT2**	**mcl**	**Total**
Aquatic (1,492)												
Freshwater (481)	648	140	23	811	26.0	13.5	3.3	27.4	3.94	0.85	0.14	4.93
Marine (373)	187	126	20	333	19.3	21.4	4	31.4	1.06	0.71	0.11	1.88
Saline and Alkaline (75)	41	18	0	59	24.0	20.0	0	34.7	2.91	1.28	0.00	4.18
Sediment (442)	488	286	10	784	26.2	23.3	2.3	30.5	3.76	2.21	0.08	6.05
Thermal springs (121)	15	1	0	16	5.8	0.8	0	5.8	3.14	0.21	0.00	3.34
	1,379	571	53	2,003	22.7	17.7	2.7	27.9	2.81	1.16	0.11	4.09
Solid Waste (34)												
Compost (30)	27	29	0	56	33.3	63.3	0	66.7	4.52	4.85	0.00	9.37
Landfill (4)	19	9	0	28	75.0	75.0	0	75	2.06	0.98	0.00	3.04
	46	38	0	84	38.2	64.7	0	67.6	1.78	1.91	0.00	5.53
Terrestrial (1,323)												
Deep subsurface (104)	59	16	1	76	18.3	7.7	1	20.2	6.19	1.68	0.10	7.97
Rock-dwelling (3)	28	5	0	33	100	66.7	0	100	9.17	1.64	0.00	10.81
Soil (1,211)	8,331	2,374	130	10,835	57.1	43.8	7.7	61.1	6.32	1.80	0.10	8.22
Volcanic (5)	0	0	0	0	0	0	0	0	0.00	0.00	0.00	0.00
	8,418	2,395	131	10,944	53.9	40.8	7.1	57.7	6.33	1.80	0.10	8.22
Wastewater (229)												
Activated Sludge (140)	133	73	11	217	9.3	5.7	5	9.3	8.47	4.65	0.70	13.82
Anaerobic digestor (54)	152	179	7	338	70.4	87	13	94.4	3.37	3.97	0.16	7.50
Nutrient removal (35)	117	146	20	283	60	77.1	11.4	77.1	6.76	8.43	1.15	16.34
	402	398	38	838	31.4	35.8	7.9	39.7	5.15	5.10	0.49	10.73
Total: 3,078	10,245	3,402	222	13,869	36.9	29.5	5	42.1	5.35	1.78	0.12	7.24

546

547

548 **Figures Legends**
549
550 **Figure 1:** A total of 13,869 PHA depolymerases were predicted in 1,295 metagenomes from
551 aquatic (417), terrestrial (764), solid waste (23) and wastewater (91) environments. **A)** World
552 map showing geocoordinates of samples with metagenomes containing one or more predicted
553 PHA depolymerase, colored by GOLD ecosystem type. Metagenomes sequenced from all seven
554 continents were included in our dataset but were unequally represented, with certain regions
555 having more sampled locations and sample types than others. **B)** Metagenomes from North
556 America were the most heavily represented in our dataset and nearly every ecosystem type
557 showed some evidence for PHA depolymerase genes within this region. Several samples share
558 geocoordinates and represent samples taken at different depths, samples taken within close
559 proximity to each other, or at different time points.
560
561 **Figure 2:** Phylogenetic affiliation of scaffolds >2000 bps in length encoding at least one
562 predicted PHA depolymerase across different ecosystems, presented as relative abundances.
563 "Unassigned" proteins could not be not classified at the phylum-level. Phyla that contributed less
564 than 1% of proteins (per ecosystem type) were combined into the "other" category. Number of
565 PHA depolymerase-encoding scaffolds for each environmental category is listed (n).
566
567 **Figure 3:** The predicted phylogenetic distribution of MAGs encoding predicted PHA
568 depolymerases. A concatenated rp tree inferred using FastTree shows the placement of 98 high-
569 quality, dereplicated MAGs encoding predicted PHA depolymerases placed among reference
570 sequences selected from the bacterial domain. Tree has been rooted between the Proteobacteria
571 and the other phyla to enhance readability. Tree scale bar indicates number of substitutions per

572 site. The colored ring labelled "PHA-encoding MAGs" is colored to indicate the phylum-level
573 classification for each MAG as predicted by GTDB-tk. Above this, the "References" colored
574 ring indicates the phylum-level classification of the bacterial reference sequences. The outer ring
575 bar plot shows the number of PHA depolymerases predicted in each MAG. The type(s) of
576 predicted PHA depolymerase(s) (mcl, sclT1, or sclT2) found in that MAG are indicated with
577 colored squares on the innermost rings.
578
579 **Figure 4:** Environmental origin and phylogenetic relationships of the 231 PHA depolymerases
580 predicted from all 143 MAGs. Tree scale bar indicates number of substitutions per site. Outer
581 rings are colored by environmental origin of predicted PHA depolymerases. Inner rings are
582 colored by phylum of the MAG encoding the predicted PHA depolymerase. ML trees showing
583 **A)** 10 (8 unique) PHAmcl depolymerases **B)** 156 (133 unique) PHAscl T1 depolymerases and **C)**
584 65 (56 unique) PHAscl T2 depolymerases.
585
586 **Supplementary Figure Legends**
587 **Supplementary Figure 1:** The distribution of predicted PHA depolymerases showing
588 geocoordinates of metagenomes containing predicted PHA depolymerases. Metagenomes are
589 colored by the frequency of predicted PHA depolymerases in hits/Gb. For metagenomes sharing
590 geocoordinates, the metagenome with the highest frequency of predicted enzymes is shown.
591
592 **Supplementary Figure 2:** Environmental origin and phylogenetic relationships of the 13,869
593 predicted PHA depolymerases identified in 1,295 metagenomes. Colored ring indicates the
594 environmental origin of the metagenome encoding each predicted PHA depolymerase.

595 Biomarker reference enzymes are indicated with black arrows. Protein trees were inferred with
596 FastTree and show the sequence diversity of A) 222 predicted PHAmcl depolymerases B) 10,245
597 predicted ePHAsclT1 depolymerases and C) 3,402 predicted PHAsclT2 depolymerases, each
598 with biochemically confirmed biomarker proteins included as references. Tree scale bar indicates
599 number of substitutions per site.
600
601 **Supplementary Figure 3:** The phylogenetic affiliation of predicted PHA depolymerase enzymes
602 on scaffolds >2,000 bps in length from 1,295 metagenomes using an LCA approach (as
603 implemented in MEGAN). Shown is the relative abundance of various phyla and classes (for
604 Proteobacteria only) of predicted PHA depolymerases for each enzyme subtype. "Unassigned"
605 proteins could not be not classified at the phylum-level and all phyla that comprised less than 1%
606 of proteins were combined into the "other" category. Number of scaffolds (n) for each enzyme
607 category is provided.
608
609 **Supplementary Figure 4:** Plot showing the relationship between assembled metagenome size
610 (Gb) and predicted PHA depolymerase frequency (hits/Gb). Plot shows a very weak positive
611 linear relationship (R^2 value = 0.0282) when Y-intercept was set to 0. Metagenomes are colored
612 by environment of origin. Area captured in gray box is expanded in figure inset.
613
614 **Supplementary Data Files**
615 All supplementary data files can be accessed through the Open Science Framework at doi:
616 10.17605/OSF.IO/VQTM6. Supplemental Data 5 is only available via OSF due to its large size.
617 **Supplementary Data 1:** Supp_Data_1_PHA_depolymerases_references.xlsx

Supplementary Data 2: Supp_Data_2_PHA_depolymerases_from_global_metagenomes.xlsx

Supplementary Data 3: Supp_Data_3_PHA_depolymerase_MAGs.xlsx

Supplementary Data 4: Supp_Data_4_Detailed_PHA_depol_curation_methods.txt

Supplementary Data 5:

Supp_Data_5_BLAST_output_of_predicted_PHA_depolymerase_summary.xlsx can be accessed via the Open Science Framework at doi: 10.17605/OSF.IO/VQTM6

Supplementary Data 6: Supp_Data_6_unique_archaeal_IMG_MAG_hits_e10.xlsx

References

1. Shah AA, Hasan F, Hameed A, Ahmed S. Biological degradation of plastics: A comprehensive review. *Biotechnology Advances*. 2008;26(3):246-265.

2. United Nations Environment Programme. Legal limits on single-use plastics and microplastics: A global review of national laws and regulations. *World Resources Institute*. 2018:1-114.

3. Aeschelmann F, Carus M. Biobased building blocks and polymers in the world: Capacities, production, and Applications–Status quo and trends towards 2020. *Industrial biotechnology*. 2015;11(3):154-159.

4. Danso D, Chow J, Streit WR. Plastics: Environmental and biotechnological perspectives on microbial degradation. *Applied and environmental microbiology*. 2019;85(19):e01095-19.

5. Geyer R, Jambeck JR, Law KL. Production, use, and fate of all plastics ever made. *Science advances*. 2017;3(7):e1700782.

640 6. Li WC, Tse HF, Fok L. Plastic waste in the marine environment: A review of sources,
641 occurrence and effects. *Science of the Total Environment*. 2016;566-567:333-349.

642 7. Nishida H, Tokiwa Y. Distribution of poly(β-hydroxybutyrate) and poly(ε-caprolactone)
643 aerobic degrading microorganisms in different environments. *Journal of Environmental Polymer*
644 *Degradation*. 1993;1(3):227-233.

645 8. Mergaert J, Swings J. Biodiversity of microorganisms that degrade bacterial and synthetic
646 polyesters. *Journal of Industrial Microbiology & Biotechnology*. 1996;17(5-6):463-469.

647 9. Boyandin AN, Prudnikova SV, Karpov VA, et al. Microbial degradation of
648 polyhydroxyalkanoates in tropical soils. *International Biodeterioration & Biodegradation*.
649 2013;83:77-84.

650 10. Altaee N, El-Hiti G, Fahdil A, Sudesh K, Yousif E. Biodegradation of different formulations
651 of polyhydroxybutyrate films in soil. *SpringerPlus*. 2016;5(1):1-12.

652 11. Dilkes-Hoffman LS, Lant PA, Laycock B, Pratt S. The rate of biodegradation of PHA
653 bioplastics in the marine environment: A meta-study. *Marine Pollution Bulletin*. 2019;142:15-
654 24.

655 12. Volova TG, Boyandin AN, Vasil'ev AD, et al. Biodegradation of polyhydroxyalkanoates
656 (PHAs) in the south China sea and identification of PHA-degrading bacteria. *Microbiology*.
657 2011;80(2):252-260.

658 13. Jendrossek, Handrick. Microbial degradation of polyhydroxyalkanoates. *Annual Review of*
659 *Microbiology*. 2002;56(1):403-432.

660 14. Fernandez-Castillo R, Rodriguez-Valera F, Gonzalez-Ramos J, Ruiz-Berraquero F.
661 Accumulation of poly(β-hydroxybutyrate) by Halobacteria. *Applied and Environmental*
662 *Microbiology*. 1985;51(1):214-216.

663 15. European Bioplastics. Bioplastics market data 2018: Global production capacities of
664 bioplastics 2018-2023. *Nova-Institute*. 2018:1-4.

665 16. European Bioplastic. Bioplastic market data 2019: Global production capacities of
666 bioplastics 2018-2023. *Nova-Institute*. 2019.

667 17. Tokiwa Y, Calabia BP, Ugwu CU, Aiba S. Biodegradability of plastics. *International journal*
668 *of molecular sciences*. 2009;10(9):3722-3742.

669 18. Vigneswari S, Lee TS, Bhubalan K, Amirul AA. Extracellular polyhydroxyalkanoate
670 depolymerase by *Acidovorax* sp. DP5. *Enzyme research*. 2015;2015:212159.

671 19. Tokiwa Y, Calabia B. Review degradation of microbial polyesters. *Biotechnology Letters*.
672 2004;26(15):1181-1189.

673 20. Knoll M, Hamm TM, Wagner F, Martinez V, Pleiss J. The PHA depolymerase engineering
674 database: A systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA)
675 depolymerases. *BMC bioinformatics*. 2009;10(1):89.

676 21. Bresan S, Sznajder A, Hauf W, Forchhammer K, Pfeiffer D, Jendrossek D.
677 Polyhydroxyalkanoate (PHA) granules have no phospholipids. *Scientific reports*.
678 2016;6(1):26612.

679 22. Martínez, de Santos, García-Hidalgo, et al. Novel extracellular medium-chain-length
680 polyhydroxyalkanoate depolymerase from *Streptomyces exfoliatus* K10 DSMZ
681 41693: A promising biocatalyst for the efficient degradation
682 of natural and functionalized mcl-PHAs. *Applied Microbiology and Biotechnology*.
683 2015;99:9605-9615.

684 23. Sood S, Sharma A, Sharma N, Kanwar SS. Carboxylesterases: Sources, characterization and
685 broader applications. *Insights in Enzyme Research*. 2018;1(1):1-11.

686 24. Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. *Nature*
687 *microbiology*. 2016;1(5):16048.

688 25. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server – a public resource
689 for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*.
690 2008;9(1):386.

691 26. Markowitz VM, Chen IA, Chu K, et al. IMG/M: The integrated metagenome data
692 management and comparative analysis system. *Nucleic Acids Research*. 2012;40(D1):123-129.

693 27. Podar M, Gilmour CC, Brandt CC, et al. Global prevalence and distribution of genes and
694 microorganisms involved in mercury methylation. *Science advances*. 2015;1(9):e1500675.

695 28. Holert J, Cardenas E, Bergstrand LH, et al. Metagenomes reveal global distribution of
696 bacterial steroid catabolism in natural, engineered, and host environments. *mBio*.
697 2018;9(1):2345.

698 29. Danso D, Schmeisser C, Chow J, et al. New insights into the function and global distribution
699 of polyethylene terephthalate (PET)-degrading bacteria and enzymes in marine and terrestrial
700 metagenomes. *Applied and Environmental Microbiology*. 2018;84(8):2773.

701 30. He S, Malfatti SA, McFarland JW, et al. Patterns in wetland microbial community
702 composition and functional gene repertoire associated with methane emissions. *mBio*.
703 2015;6(3):e00066-e00015.

704 31. Anantharaman K, Brown CT, Hug LA, et al. Thousands of microbial genomes shed light on
705 interconnected biogeochemical processes in an aquifer system. *Nature communications*.
706 2016;7(1):13219.

707 32. Butterfield C, Li Z, Andeer P, et al. Proteogenomic analyses indicate bacterial methylotrophy
708 and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ*. 2016;4:e2687.

709 33. Hug LA, Thomas BC, Sharon I, et al. Critical biogeochemical functions in the subsurface are
710 associated with bacteria from new phyla and little studied lineages. *Environmental microbiology*.
711 2016;18(1):159-173.

712 34. Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. Metatranscriptomic
713 insights on gene expression and regulatory controls in candidatus accumulibacter phosphatis. *The*
714 *ISME Journal*. 2015;10(4):810-822.

715 35. Hartman WH, Ye R, Horwath WR, Tringe SG. A genomic perspective on stoichiometric
716 regulation of soil carbon cycling. *The ISME Journal*. 2017;11(12):2652-2665.

717 36. Kimbrel JA, Ballor N, Wu Y, et al. Microbial community structure and functional potential
718 along a hypersaline gradient. *Frontiers in microbiology*. 2018;9:1492.

719 37. Schulz F, Alteio L, Goudeau D, et al. Hidden diversity of soil giant viruses. *Nature*
720 *communications*. 2018;9(1):4881-9.

721 38. Yao Q, Li Z, Song Y, et al. Community proteogenomics reveals the systemic impact of
722 phosphorus availability on microbial functions in tropical soil. *Nature ecology & evolution*.
723 2018;3(3):1-28.

724 39. Acinas S, Sanchez P, Salazar G, et al. Metabolic architecture of the deep ocean microbiome.
725 *bioRxiv*. 2019;635680.

726 40. Diamond S, Andeer P, Li Z, et al. Mediterranean grassland soil C–N compound turnover is
727 dependent on rainfall and depth, and is mediated by genomically divergent microorganisms.
728 *Nature microbiology*. 2019;4:1356-1367.

729 41. Lavy A, McGrath D, Carnevali P, et al. Microbial communities across a hillslope-riparian
730 transect shaped by proximity to the stream, groundwater table, and weathered bedrock. *Ecology*
731 *and Evolution*. 2019;9:6869-6900.

732 42. Sharrar A, Crits-Christoph A, Meheust R, Diamond S, Starr E, Banfield J. Bacterial
733 secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation
734 type. *bioRxiv*. 2019:818815.

735 43. Alteio LV, Schulz F, Seshadri R, et al. Complementary metagenomic approaches improve
736 reconstruction of microbial diversity in a forest soil. *mSystems*. 2020;5(2):768.

737 44. Flores-Nunez V, Fonseca-Garcia C, Desgarennes D, Eloe-Fadrosh E, Woyke T, Partida-
738 Martinez L. Functional signatures of the epiphytic prokaryotic microbiome of agaves and cacti.
739 *Frontiers in microbiology*. 2020;10:1-13.

740 45. Bouma-Gregson K, Olm MR, Probst AJ, Anantharaman K, Power ME, Banfield JF. Impacts
741 of microbial assemblage and environmental conditions on the distribution of anatoxin-a
742 producing cyanobacteria within a river network. *The ISME Journal*. 2019;13(6):1618-1634.

743 46. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal
744 peptide predictions using deep neural networks. *Nature biotechnology*. 2019;37(4):420-423.

745 47. Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine database (GOLD) v.7: Updates
746 and new features. *Nucleic acids research*. 2018;47(D1):649-659.

747 48. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND.
748 *Nature methods*. 2014;12(1):59-60.

749 49. Huson DH, Beier S, Flade I, et al. MEGAN community edition - interactive exploration and
750 analysis of large-scale microbiome sequencing data. *PLoS computational biology*.
751 2016;12(6):e1004957.

752 50. Chaumeil P, Mussig AJ, Hugenholtz P, Parks DH. GTDB-tk: A toolkit to classify genomes
753 with the genome taxonomy database. *Bioinformatics*. 2019;36(6):1925-1927.

754 51. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. CheckM:
755 Assessing the quality of microbial genomes recovered from isolates, single cells, and
756 metagenomes. *Genome research*. 2015;25(7):1043-1055.

757 52. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: A tool for fast and accurate genomic
758 comparisons that enables improved genome recovery from metagenomes through de-replication.
759 *The ISME Journal*. 2017;11(12):2864-2868.

760 53. Arkin AP, Cottingham RW, Henry CS, et al. KBase: The united states department of energy
761 systems biology knowledgebase. *Nature biotechnology*. 2018;36(7):566-569.

762 54. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for
763 large alignments. *PloS one*. 2010;5(3):e9490.

764 55. Letunic I, Bork P. Interactive tree of life (iTOL) v3: An online tool for the display and
765 annotation of phylogenetic and other trees. *Nucleic acids research*. 2016;44(W1):242-245.

766 56. Martínez V, de la Peña F, García-Hidalgo J, de la Mata I, García JL, Prieto MA.
767 Identification and biochemical evidence of a medium-chain-length polyhydroxyalkanoate
768 depolymerase in the *Bdellovibrio bacteriovorus* predatory hydrolytic arsenal. *Applied and*
769 *Environmental Microbiology*. 2012;78(17):6017-6026.

770 57. Schirmer A, Jendrossek D. Molecular characterization of the extracellular poly(3-
771 hydroxyoctanoic acid) [P(3HO)] depolymerase gene of *Pseudomonas fluorescens* GK13 and of
772 its gene product. *Journal of Bacteriology*. 1994;176(22):7065-7073.

773 58. Kim D, Chul Kim H, Kim S, Rhee Y. Molecular characterization of extracellular medium-
774 chain-length poly(3-hydroxyalkanoate) depolymerase genes from *Pseudomonas alcaligenes*
775 strains. *The Journal of Microbiology*. 2005;43(3):285-294.

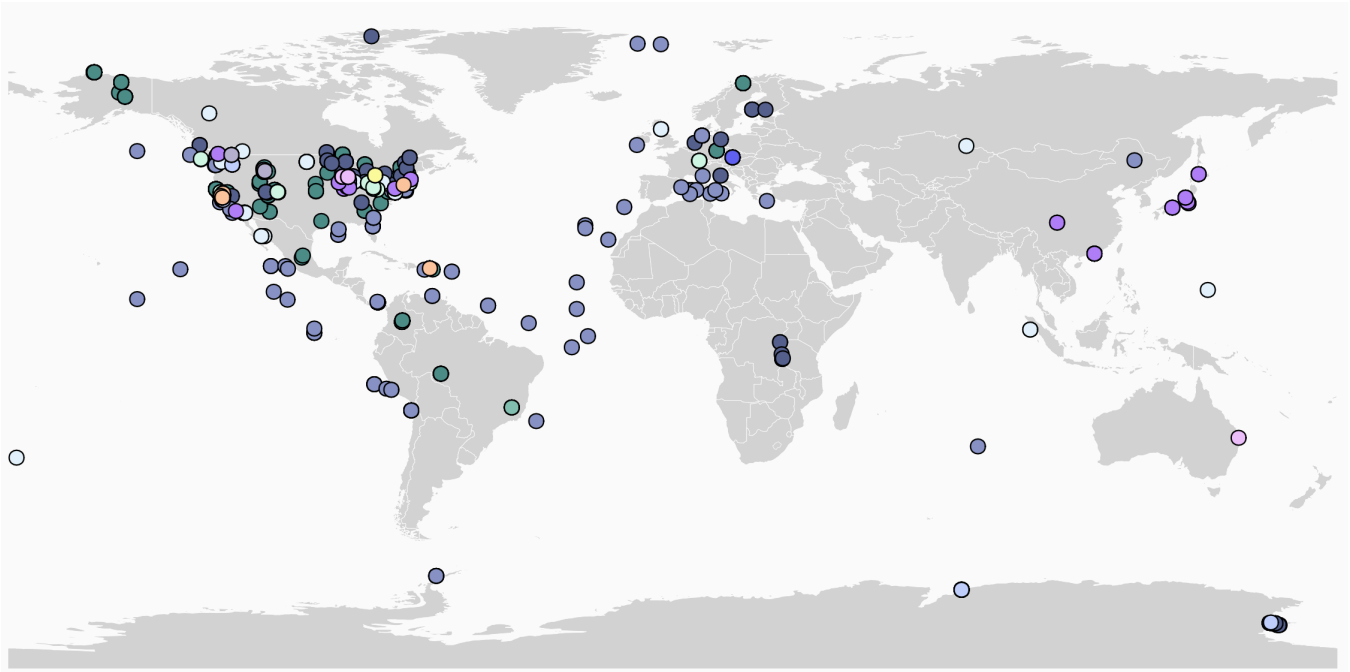
776 59. Gangoiti J, Santos M, Prieto M, de la Mata I, Serra J, Llama M. Characterization of a novel
777 subgroup of extracellular medium-chain-length polyhydroxyalkanoate depolymerases from
778 Actinobacteria. *Applied and environmental microbiology*. 2012;78(20):7229-7237.

779 60. Pittmann T, Steinmetz H. Polyhydroxyalkanoate production on wastewater treatment plants:
780 Process scheme, operating conditions and potential analysis for German and European municipal
781 waste water treatment plants. *Bioengineering*. 2017;4(54):1-24.

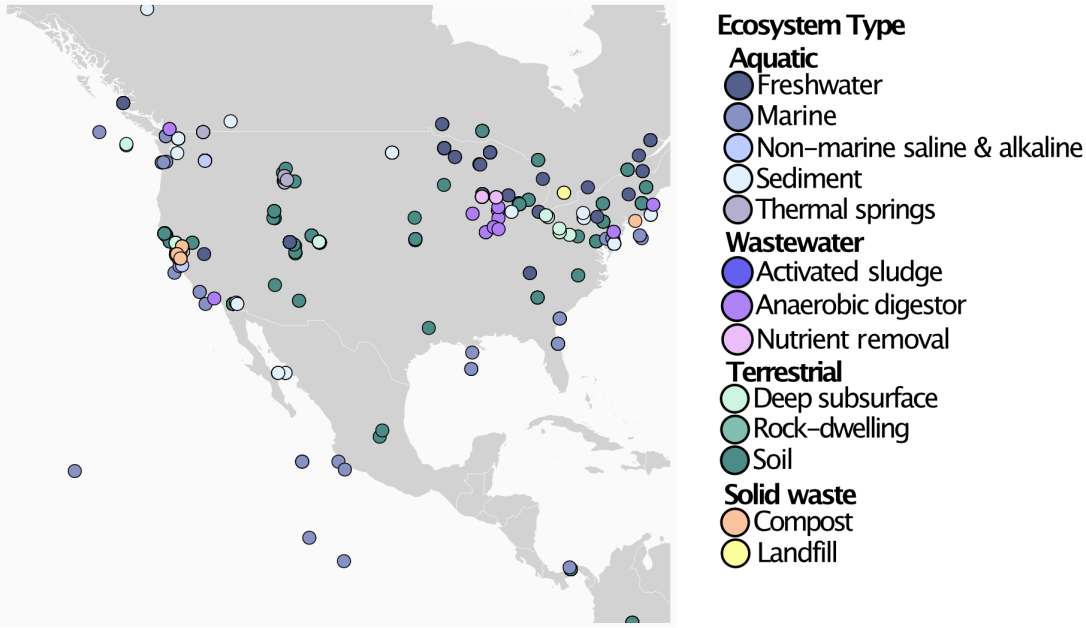
782 61. Pinnell LJ, Turner JW. Shotgun metagenomics reveals the benthic microbial community
783 response to plastic and bioplastic in a coastal marine environment. *Frontiers in microbiology*.
784 2019;10:1252.

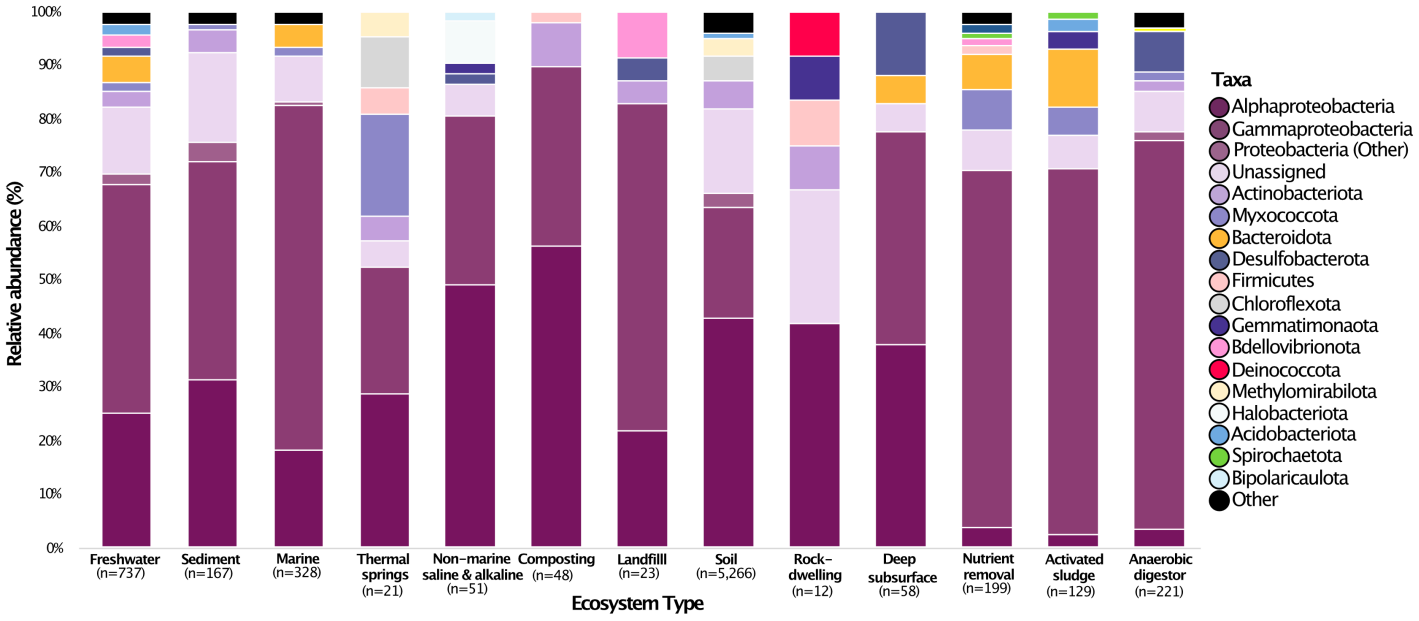
785

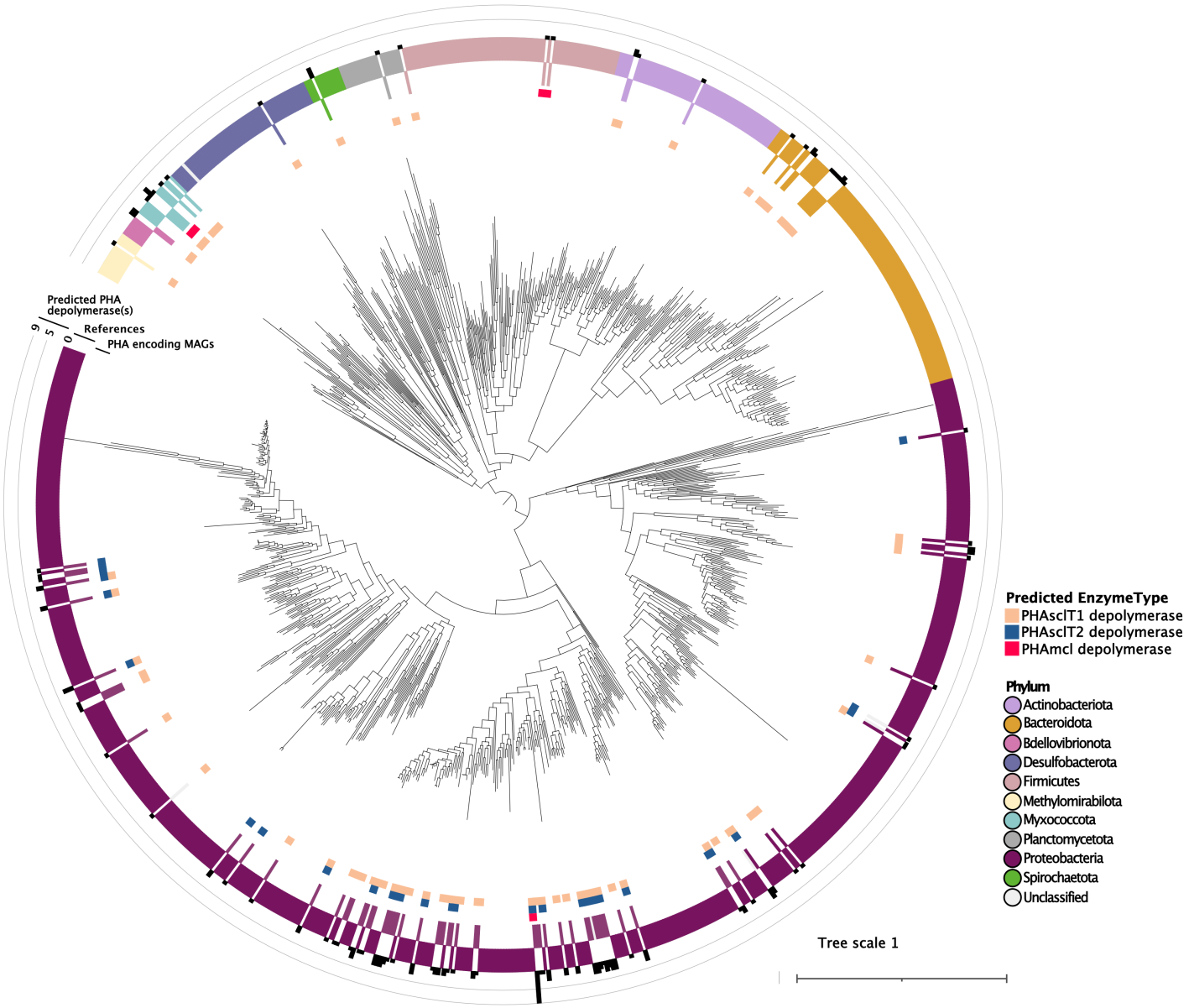
A)

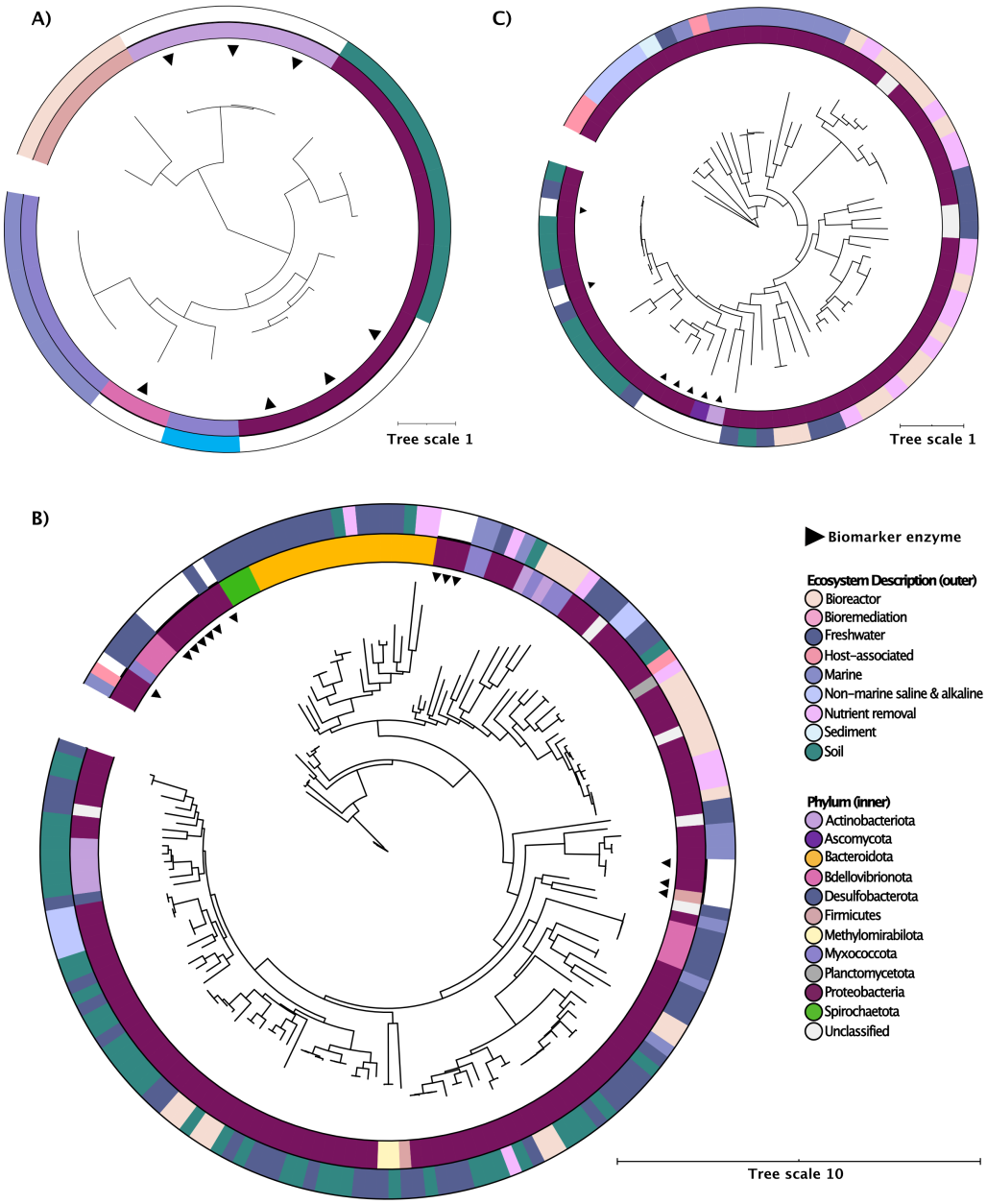


B)


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


Relative abundance (%)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Freshwater (n=737)
Sediment (n=167)
Marine (n=328)
Thermal springs (n=21)
Non-marine saline & alkaline (n=51)
Composting (n=48)
Landfilll (n=23)
Soil (n=5,266)
Rock-dwelling (n=12)
Deep subsurface (n=58)
Nutrient removal (n=199)
Activated sludge (n=129)
Anaerobic digestor (n=221)
Ecosystem Type
Taxa
Alphaproteobacteria
Gammaproteobacteria
Proteobacteria (Other)
Unassigned
Actinobacteriota
Myxococcota
Bacteroidota
Desulfobacterota
Firmicutes
Chloroflexota
Gemmatimonaota
Bdellovibrionota
Deinococcota
Methylomirabilota
Halobacteriota
Acidobacteriota
Spirochaetota
Bipolaricaulota
Other


Predicted PHA depolymerase(s)
References
PHA encoding MAGs
9
5
0
Predicted EnzymeType
PHAsclT1 depolymerase
PHAsclT2 depolymerase
PHAmcl depolymerase
Phylum
Actinobacteriota
Bacteroidota
Bdellovibrionota
Desulfobacterota
Firmicutes
Methylomirabilota
Myxococcota
Planctomycetota
Proteobacteria
Spirochaetota
Unclassified
Tree scale 1


A)
Tree scale 1
C)
Tree scale 1
B)
Biomarker enzyme
Ecosystem Description (outer)
Bioreactor
Bioremediation
Freshwater
Host-associated
Marine
Non-marine saline & alkaline
Nutrient removal
Sediment
Soil
Phylum (inner)
Actinobacteriota
Ascomycota
Bacteroidota
Bdellovibrionota
Desulfobacterota
Firmicutes
Methylomirabilota
Myxococcota
Planctomycetota
Proteobacteria
Spirochaetota
Unclassified
Tree scale 10