Substrate promiscuity of xenobiotic-transforming hydrolases from stream biofilms impacted by treated wastewater

Yaochun Yu[1]*, Niklas Ferenc Trottmann[2]*, Milo R. Schärer[2], Kathrin Fenner[1,3], ✉Serina L. Robinson[2]

[1] Department of Environmental Chemistry, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf 8600, Switzerland

[2] Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf 8600, Switzerland

[3] Department of Chemistry, University of Zürich, 8057 Zürich, Switzerland

*co-first authors contributed equally to this work

✉corresponding author: serina.robinson@eawag.ch

Keywords: metagenomics, pollutant biotransformations, organic contaminants, enzyme promiscuity, biofilms, periphyton

Synopsis: We aimed to connect contaminant biotransformations to their responsible enzymes within complex microbiomes by testing substrate specificity of enzymes from stream metagenomes for their activity with 183 wastewater-relevant organic contaminants.

Abstract

Organic contaminants enter aquatic ecosystems from various sources, including through the effluents of wastewater treatment plants. Freshwater biofilms play a major role in the removal of organic contaminants from receiving water bodies, but knowledge of the molecular mechanisms driving contaminant biotransformations in complex biofilm communities remains limited. Previously, we demonstrated that biofilms in experimental flume systems grown at higher ratios of treated wastewater (WW) to stream water displayed an increased biotransformation potential for a number of organic contaminants. Specifically, we identified a positive correlation between WW percentage and biofilm biotransformation rates for the widely-used insect repellent, *N,N*-diethyl-meta-toluamide (DEET). Here, we conducted deep shotgun sequencing of flume biofilms and identified a positive correlation between WW percentage and metagenomic read abundances of DEET hydrolase (DH) homologs. To test the causality of this association, we constructed a targeted metagenomic library of DH homologs from flume biofilms. We screened our complete metagenomic library for activity with four different substrates and a subset thereof with 183 WW-related organic compounds. The majority of active hydrolases in our library preferred aliphatic and aromatic ester substrates while, remarkably, only a single reference enzyme was capable of DEET hydrolysis. Of the 626 total enzyme-substrate combinations tested, approximately 5% were active enzyme-substrate pairs. Metagenomic DH family homologs revealed a broad substrate promiscuity spanning 22 different compounds when summed across all enzymes tested. We biochemically characterized the most promiscuous and active enzymes identified based on metagenomic analysis from uncultivated *Rhodospirillaceae* and *Planctomycetaceae*. In addition to uncovering new DH family enzymes, we exemplified a framework for linking metagenome-guided hypothesis generation with experimental validation. Overall, this study expands the scope of known enzymatic contaminant biotransformations for metagenomic hydrolases from WW-receiving stream biofilm communities.

Introduction

Trace organic contaminants (TrOCs) released due to anthropogenic activities such as industrial production and agricultural practices are exerting increasing pressures on ecosystems and human health (Caliman and Gavrilescu, 2009; Daughton and Ternes, 1999; Evich et al., 2022; Fenner et al., 2013; Halden, 2011; Schwarzenbach et al., 2006). Many TrOCs and their transformation products exert toxic effects through bioaccumulation, mixture effects, and specific toxic interactions (Altenburger et al., 2018; Carles et al., 2021; Jin et al., 2023). In aquatic systems, microbes are demonstrated to be the major players in the biotransformation and removal of various TrOCs through metabolic and co-metabolic processes (Che et al., 2021; Fenner et al., 2013; Han et al., 2023, 2019; Xing et al., 2018; Yu et al., 2018). However, the specific enzymes involved in microbial biotransformation processes remain largely unknown although this knowledge is critical to improve TrOC removal by environmental microbiomes and to improve prediction of contaminant biotransformations (Fenner et al., 2021).

Freshwater biofilms are ideal systems for investigating TrOC biotransformation processes due to their complex microbial community composition and critical ecological function in aquatic environments (Battin et al., 2016; Desiante et al., 2021). Most biofilm studies have been conducted using marker gene sequencing to profile taxonomic composition and community shifts that correlate with TrOC biotransformations in treated wastewater (Carles et al., 2022, 2021; Desiante et al., 2022; Mansfeldt et al., 2020; Tien et al., 2013). Metagenomics and metatranscriptomics analyses have recently deepened the resolution of community analysis to identify functional shifts in TrOC-exposed engineered wetland biomats, yet the molecular mechanisms of TrOC biotransformations in freshwater biofilms or biomats remain poorly studied relative to other systems such as enriched soil consortia (Vasileiadis et al., 2022) or activated sludge (Rich and Helbling, 2023; Vasileiadis et al., 2022). Studies in activated sludge have additionally correlated microbial taxa, functional genes, and pollutant biotransformations using meta'omic datasets (Achermann et al., 2020; Athanasakoglou and Fenner, 2022; Kennes-Veiga et al., 2022; Stadler et al., 2018). However, proving correlations experimentally presents a 'needle-in-the-haystack' problem for the identification and validation of enzymes involved in TrOC biotransformations.

In our previous study, stream biofilms cultivated at different ratios of stream water and either ultrafiltered or unfiltered treated wastewater (WW) exhibited varying TrOCs biotransformation capacities (Desiante et al., 2022). Among 75 TrOCs tested, seven compounds (*N,N*-diethyl-meta-toluamide (DEET), acesulfame, caffeine, cyclamate, levetiracetam, saccharin, and valsartan, Figure S1) showed clear evidence for increased biotransformation rates in biofilms grown at higher WW percentages, as compared to controls grown with no or in ultrafiltered WW. This phenomenon, previously defined as the *downstream effect* (Desiante et al., 2022), strongly suggested that microorganisms present in the WW might affect the biotransformation capacity of downstream biofilms. Notably, all compounds that exhibited the *downstream effect* contain hydrolyzable moieties, primarily sulfonamide- and amide-functional groups (Figure S1). This motivated us to initiate a large-scale investigation into the potential roles of environmental hydrolases from wastewater effluent-impacted biofilms in degrading TrOCs.

To address this aim, we exemplify a metagenomic data-driven framework to investigate contaminant biotransformations and enzyme substrate specificity in complex stream biofilm communities. We observed increasing biotransformation rates of DEET, one of the seven *downstream effect* compounds, to align with higher metagenomic abundances of DEET hydrolase (DH) homologs in stream biofilms with increasing proportions of treated wastewater (WW). Because of the similarity in hydrolyzable moieties amongst all *downstream effect* compounds, we hypothesized that diverse metagenomic DH family homologs detected in WW-impacted biofilms have an expanded substrate specificity not only for DEET, but also for other WW-relevant contaminants. Through metagenome mining guided by protein structural modeling, we designed, constructed, and tested a targeted metagenomic library of 64 DH homologs. This work maps the sequence-structure-substrate relationships of DH family enzymes belonging to ten different bacterial phyla from complex stream biofilm communities. The results not only expand our fundamental understanding of how enzyme diversity relates to biotransformation processes, but also have implications for the identification of novel and functional TrOC-degrading enzymes from environmental microbiomes.

Materials and Methods

Biofilm sampling

The biofilm samples which in this study were used for metagenomic sequencing and hydrolase screening were previously analyzed by 16S and 18S rRNA amplicon sequencing (Carles et al., 2021) and organic contaminant biotransformation experiments (Desiante et al., 2022). Briefly, the experimental flume systems for biofilm growth consisted of two independent experiments described in detail by (Desiante et al., 2022), each with 16-20 water channels fed with different proportions of the following sources: stream water mixed in proportions with either 10%, 30%, or 80% treated wastewater (WW) or 30% or 80% ultrafiltered (pore size 0.4 μM) treated wastewater (Burdon et al., 2020). Biofilms were grown on glass slides placed in these flow-through channels for four weeks. Biofilms from different glass slides within the same channel were scraped independently to form biological replicates (n=4 for each treatment) and suspended in 400 mL of artificial river water (Carles et al., 2022, 2021; Desiante et al., 2022).

Metagenomic sequencing and analysis

DNA was extracted from biofilms and water samples using the DNeasy PowerBiofilm kit (QIAGEN) according to the manufacturer's instructions. The quality of DNA was evaluated by Nanodrop. DNA samples that passed the sequencing quality criteria were used for metagenomic library construction and then sequenced on an Illumina NovaSeq platform by Novogene (UK) with a paired-end (2 × 150) sequencing strategy to yield a total of 75 metagenomes with replicates for different sample types (Table S1). Adapters and low-quality reads (N > 10% and quality score ≤ 5) were removed by Novogene. *De novo* assembly of raw reads was performed using MEGAHIT v1.2.9 (Li et al., 2015). Gene calling was done using Prodigal (v2.6.3) (Hyatt et al., 2010).

123 **Metagenomic library design and construction**

124 The sole characterized reference DEET hydrolase sequence (DH) described in
125 *Pseudomonas putida* (A3R4E0_PSEPU) was used to query the metagenomes for homologs
126 using NCBI BLASTp v2.9.0 (Camacho et al., 2009). Default BLASTp parameters were amended
127 with a bit-score threshold of 50 and hits with a query cover of at least 20% and sharing at least
128 20% amino acid identity with the query sequence. The resulting 3,500 hits were clustered using
129 CD-HIT (Fu et al., 2012) at 60% amino acid sequence identity cut-off with a word size of 4 to
130 obtain 70 cluster representatives. Cluster representatives were filtered to include only full-length
131 open-reading frames with valid start and stop codons resulting in 64 remaining sequences. To
132 assign the likely taxonomic origin of the 64 DH homologs, the taxonomic classification tool Kaiju
133 was used (Menzel et al., 2016). Since Kaiju did not provide classification for many cases, a
134 reverse BLAST of the metagenomic sequences against the NCBI non-redundant protein database
135 and consensus taxonomic classification based on the identity of the top hit cross-referenced with
136 Kaiju classification (Table S2) The 64 metagenomic DH homolog sequences were codon-
137 optimized for *E. coli* expression using the build-optimization software, BOOST (Oberortner et al.,
138 2017), and cloned by the U.S. Department of Energy Joint Genome Institute with C-terminal 6x-
139 His tags, a tobacco etch virus cleavage cut-site, and additional flexible linker into the first multiple
140 cloning site of pCDF-Duet vectors retaining the NdeI and HindIII restriction sites. Sequence-
141 verified constructs were transformed into T7 Express Competent *E. coli* cells (NEB C2566I).

142 **Short read mapping and coverage calculations**

143 Read mapping and coverage calculations for the *P. putida* reference DH and full-length
144 DH homologs from the above-described metagenomic library were computed following the
145 methods previously described (Lee et al., 2021). Mapping of raw metagenomic reads to query
146 DHs was done using Bowtie2 (v2.4.4), depth was computed using samtools v1.12 and coverage
147 was calculated as described by Albertsen et al. (Albertsen et al., 2013; Langmead and Salzberg,
148 2012; Lee et al., 2021; Li et al., 2009). Bowtie2 parameters used are described in Table S3. Data
149 analysis, statistics and plotting were performed using R within the RStudio integrated
150 development environment (IDE) with the tidyverse library. (R core team, 2022; RStudio Team,
151 2020; Wickham et al., 2019).

152 **Phylogenetic analysis**

153 To analyze the phylogenetic relationship of DH family enzymes, 64 DHs in the
154 metagenomic library plus the reference *P. putida* DH and reference cocaine esterase (PDB ID:
155 3IDA) were aligned using Clustal-Omega v1.2.3 (Sievers and Higgins, 2018). Three outgroup
156 sequences were included as distantly related proteins with the alpha/beta hydrolase fold with PDB
157 IDs: 1MJ5, 3HI4 and 6QHV. A protein phylogenetic tree including 71 sequences was generated
158 from the multiple sequence alignment of DH homologs and reference sequences using IQ-TREE
159 2 v2.0.3 (Minh et al., 2020). Briefly, IQ-TREE's ModelFinder function was used to identify the top
160 model. ModelFinder identified the best model as LG+F+R5 which is based on an improved
161 general amino acid replacement matrix (Le and Gascuel, 2008). Non-parametric bootstrapping
162 was performed within IQ-TREE with 1,000 replicates.

163 **Structural modeling and comparison**

164 Full-length sequences of the 64 metagenomic DH proteins and reference enzymes were
165 used as inputs for structural modeling using ColabFold AlphaFold 2.0 (Mirdita et al., 2022) with
166 default parameters. Models were assessed for quality by predicted local distance difference test
167 (pLDDT) score and aligned pairwise with the reference *P. putida* structure (AF-A3R4E0-F1) using
168 the PyMOL 'super' command with default parameters. To construct a tree of structural
169 relationships, all-vs-all structural comparisons were made and the resulting matrix was converted
170 into a dendrogram using the neighbor-joining algorithm (Saitou and Nei, 1987) implemented in
171 the an 'nj' function in R ape package. Matching nodes between the sequence and structural
172 phylogenetic trees were paired and visualized using the treedataverse library (Yu, 2022).

173 **Colorimetric activity screening of *P. putida* DH and DH homologs**

174 Activity screening for the 64 metagenomic DH homologs and the *P. putida* reference
175 sequence was performed using an established whole-cell colorimetric screening protocol for
176 hydrolase activity (Robinson et al., 2020; Smith et al., 2020). Briefly, individual transformants were
177 inoculated in 5 mL lysogeny broth (LB, final concentration 50 mg/L) supplemented with
178 spectinomycin (final conc. 50 µg/mL). Cultures were incubated at 37°C with orbital shaking at 200
179 rpm until an optimal density of 600 nm (OD600) of 0.5–0.8 was reached. Incubator temperature
180 was then altered to 16°C before adding isopropyl-β-D-thiogalactopyranoside (IPTG, final conc.
181 0.1 mM) for induction of protein production. The samples were then incubated for 48 hours. Cell
182 pellets were collected by centrifugation of 0.5 mL cell suspensions at 5,000 x g for 10 min and
183 resuspended in 1 mL Tris-HCl buffer (0.05 M Tris-HCl, pH 8.0). OD600 was then measured to
184 calculate the cell suspension/Tris-HCL buffer ratio in order to normalize the samples to an OD600
185 of 0.1. Normalized samples were screened for esterase activity using two colorimetric substrates:
186 4-nitrophenyl butyrate and 4-nitrophenyl trimethylacetate (200 µM final concentration from 8 mM
187 stock solutions in ethanol (EtOH), for a 2% final EtOH in reactions below the limit which would
188 disturb enzyme activity). Individual reactions containing 195 µL normalized cell suspensions and
189 5 µL substrates were performed in 96-well clear-bottomed plates (Greiner). Standard curves of 4-
190 nitrophenol (8mM stock solutions in EtOH), a positive control of *P. putida* DH, and a negative
191 control of a heat-inactivated DH were included in each plate. Plates were measured using a
192 microplate reader (Synergy H1, BioTek) at 37°C using a continuous measurement program of 60-
193 time point measurements with a time interval of one minute at OD 410 nm.

194 **High-Performance Liquid Chromatography (HPLC) Analysis of *P. putida* DH and**
195 **metagenomic DH homologs hydrolysis activity**

196 The enzyme activity of *P. putida* DH and DH homologs with DEET and methylbenzoate
197 (MB) were performed using cell lysates. Protein expression was performed as described in the
198 above section but with 400 mL cultures. Cell pellets were collected and resuspended in Buffer A
199 (20 mM imidazole, 20 mM Tris-HCl, 500 mM NaCl, 10% (v/v) glycerol). Cell suspensions were
200 then lysed using a sonicator (SONOPULS HD 3200, BANDELIN) with a 3 mm probe (SONOPULS
201 MS73, BANDELIN) at a frequency of 20 kHz. The sample vials were kept in an ice-water bath to
202 prevent sample overheating caused by sonication. Cell lysates were then centrifuged for 30 min
203 at 17,000 x g at 4°C. Supernatants containing soluble proteins (here referred to as crude extract)

204 were used for downstream biotransformation experiments. DEET and MB (final conc. 250 μM)
205 were added to crude extracts separately. Subsequently, reaction progress was monitored by
206 sampling and quenching the reaction with an equal volume of acetonitrile at defined time intervals
207 at 0, 1, 3, 5 and 24h and measured by HPLC (Summit HPLC system, Dionex). Detailed
208 information regarding the HPLC Method is included in SI.

209 **Purification of *P. putida* DH and metagenomic DH homologs**

210 Induction and expression of DH homologs, cell lysis, and harvesting of cell pellets were
211 performed as described in the sections above but with 400 mL Terrific broth culture media (12 g/L
212 tryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol) supplemented with 10% (vol/vol) phosphate
213 buffer (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4). Protein purification was performed using benchtop
214 purification for initial tests and fast protein liquid chromatography (FPLC, ÄKTA Pure 25, Cytiva)
215 for data and figures reported in the manuscript. The detailed FPLC method can be found in the
216 SI. For benchtop purification, supernatants were applied to a column packed with Ni-NTA agarose
217 beads (Qiagen) pre-equilibrated in buffer A. The column was subsequently washed with buffer A
218 and then subjected to two washes with buffer B (comprising 40 mM Tris-HCl, 500 mM NaCl, and
219 10% (v/v) glycerol). Enzymes were eluted from the column using buffer C (containing 500 mM
220 imidazole, 500 mM NaCl, and 10% (v/v) glycerol). Fractions obtained from both benchtop and
221 FPLC purifications were later subjected to analysis by SDS-PAGE (Figure S2 – S4). Fractions
222 containing the enzymes of interest, consistent with their expected molecular weights, were eluted
223 by removing imidazole and then underwent buffer exchange into TNG buffer (50 mM Tris-HCl at
224 pH 7.4, 0.1 M NaCl, and 10% (v/v) glycerol). The enzymes were aliquoted, flash-frozen, and
225 stored at –80°C.

226 **TrOCs Selection**

227 A total of 183 trace organic contaminants (TrOCs) were chosen for biotransformation
228 experiments with purified *P. putida* DH and the highly active DH homolog p055 selected as
229 described in the Results & Discussion. These TrOCs are commonly used pesticides,
230 pharmaceuticals, and artificial sweeteners and were selected to equally cover a number of
231 relevant enzymatic transformation reactions. Reference compounds are either purchased from
232 Sigma-Aldrich (St. Louis, MO) or generously provided by the Department of Environmental
233 Chemistry at Swiss Federal Institute of Aquatic Science and Technology (Eawag). Stock solutions
234 of TrOCs were individually prepared in dimethylsulfoxide (DMSO) at a concentration of 10 mM
235 and stored at −20°C until use. Detailed information regarding TrOCs can be found in the SI, Table
236 S4 and S5

237 **High-throughput Biotransformation Experiments of TrOCs in 96-well Plate**

238 Purified *P. putida* DH and the metagenomic DH homolog p055 were tested for their activity
239 with 14 sub-mixtures composed of the 183 TrOCs instead of testing individual compounds. To
240 avoid potential cumulative toxicities and competitive inhibition, we first conducted a hierarchical
241 clustering of all compounds based on their chemical fingerprints and reaction rules triggered by
242 enviPath (Wicker et al., 2016). Sub-mixtures were then designed to contain compounds with

243 different annotated reaction rules. Each of the 14 sub-mixtures (200 μM for each TrOC) was
244 prepared via appropriate dilution of the stock solutions in EtOH (Table S4 and S5) and stored at
245 −20°C for later biotransformation experiments. The capability of TrOCs biotransformation by
246 different enzymes was investigated in a clear-bottomed 96-well plate (Greiner). Individual TrOC
247 sub-mixtures were first added to an empty 96-well plate and the organic solvent EtOH was
248 completely evaporated before 245 μL 10 mM Tris-HCl buffer (pH=8) was added to each well. The
249 plate was shaken at 50 rpm for 15 min to redissolve the TrOCs. Individually purified enzymes
250 (final concentration of ~190 nM) were then added to a well plate and mixed thoroughly. Samples
251 for LC-HRMS/MS analysis were collected at 0, 1, 6, and 24h. At each sampling event, a 20 μL
252 sample was taken and mixed with the same amount of pure methanol to completely denature the
253 enzyme, 160 μL buffer was then added to the sample mixture for an approximately 10 times
254 dilution of the original sample. After centrifugation (13,000 × g, 4°C for 15 min), the supernatant
255 was collected and stored at 4°C for LC-HRMS/MS analysis. An enzyme-free abiotic fresh buffer
256 control and trichloroacetic acid (TCA) treated inactivated enzyme control were set up in the same
257 way as the biological experiment described above. All experiments were conducted in duplicate.
258 Samples were submitted to LC-HRMS/MS analysis within two days after collection. Matrix match
259 calibration standards were used for quantification. Quantification was based on concentrations
260 calculated from peak area integration using TraceFinder 5.1 (Thermo Fisher Scientific).

261 **Ultra-High-Performance Liquid Chromatography Coupled to High-Resolution Tandem**
262 **Mass Spectrometry (UHPLC-HRMS/MS) Analysis.**

263 The parent compounds and transformation products (TPs) were analyzed by UHPLC-
264 HRMS/MS (Q Exactive, Thermo Fisher Scientific) using a method as previously described but
265 with slight modifications (Yu et al., 2022). Briefly, for UHPLC analysis, 25 μL samples were loaded
266 onto a ACQUITY Premier BEH C18 Column (particle size 1.7 μm, 100 × 2.1 mm, Waters) and
267 eluted with nano-pure water (A) and methanol (B) (both amended with 0.1% formic acid) at a flow
268 rate of 300 μL/min, with a gradient as follows: 95% A: 0 – 1.5 min, 95% – 5% A:1 – 7.5 min, 5%
269 A: 7.5 – 9.5 min, and 95% A: .5 – 11.5 min. For HRMS, mass spectra were acquired in full scan
270 mode at a resolution of 70,000 at m/z 200 and a scan range of m/z 100 − 1000 in positive/negative
271 switching mode with electrospray ionization (ESI). Xcalibur 4.0 (Thermo Fisher Scientific) was
272 used for data acquisition and analysis. Suspect screening for DEET transformation products was
273 conducted using TraceFinder 5.1 (Thermo Fisher Scientific). ChemDraw Professional 20.0 and
274 MarvinSketch (v19.20.0, ChemAxon, http://www.chemaxon.com) were used for drawing,
275 displaying, and characterizing chemical structures. The minimum incubation time for different
276 control groups at which the maximum removal percentage of individual compounds was achieved
277 was calculated based on Equation 1. The results were included in Table S4 (for DEET hydrolase)
278 and Table S5 (for p055). Significant biological removal was defined as following criteria: (1) > 20%
279 removal in enzymatic control, and (2) at least 10% more removal compared to both abiotic buffer
280 control and inactivated enzyme control.

281

282

283 Equation 1.

284 $$Removal\ Percentage\ (\%) = \frac{C_0 - C_{final}}{C_0} \times 100\%$$

285 Where C_0 is the initial concentration measured by LC-HRMS/MS and C_{final} is the final
286 concentration after biotransformation experiments.

287 ## Results and Discussion

288 ### Metagenomic sequencing and analysis of WW-impacted biofilm communities

289 To investigate the functional genes involved in TrOC biotransformations, we conducted
290 deep shotgun metagenomic sequencing (up to 200 million reads per sample, Table S1) for 75
291 samples including WW-grown biofilms, WW, and stream water. In line with previous 16S rRNA
292 sequencing results (Carles et al., 2022), shotgun metagenomic analysis demonstrated the
293 microbial community structures of the biofilms, WW, and stream water samples were significantly
294 influenced by the ratio of treated WW relative to stream water (Figure S5). Specifically, in our
295 metagenomic analysis of the biofilm samples, we observed the relative abundance of
296 *Chloroflexota, Desulfobacterota, and Myxococcota* increased from 0% to 80% treated WW
297 (log_2Fold-change > 2; normalized mean baseline counts > 240, Figure S5). In agreement with the
298 16S rRNA analysis by (Carles et al., 2022), *Cyanobacteria* relative abundance decreased with
299 increasing treated WW%.

300 Previously, Desiante et al. identified 146 differentially-abundant amplicon sequence
301 variants (ASVs) through 16S rRNA analysis that mirrored the *downstream effect* in TrOC
302 biotransformations (Desiante et al., 2022). Since 16S rRNA studies do not capture functional
303 genes coding for enzymes, we aimed here to query the new assembled biofilm metagenomes to
304 identify responsible enzymes involved in TrOC biotransformations. Among the proposed
305 biotransformation pathways for the seven TrOCs showing a *downstream effect* (Desiante et al.,
306 2022), at the time of our analysis, only one compound (DEET) could be linked to an experimentally
307 validated hydrolase sequence (Rivera-Cancel et al., 2007). In the case of the *downstream effect*
308 compound cyclamate, the partial purification of an active hydrolase was reported but the protein
309 sequence remains undetermined (Nimura et al., 1974). Several other pathways were predicted
310 (Desiante et al., 2022) but only one additional hydrolytic biotransformation pathway for the
311 *downstream effect* compound, acesulfame K (Bonatelli et al., 2023), was experimentally validated
312 since we initiated this analysis. We detected no significant BLAST hits in our assembled
313 metagenomes for the acesulfame K hydrolase (Bonatelli et al., 2023). However, we detected more
314 than 3,500 DEET hydrolase (DH) homologs in our metagenomes, motivating us to pursue DH
315 enzymes as a use case.

316 Based on the high number and diversity of DH homologs, we next calculated the
317 abundances of DH homologs in assembled metagenomes of biofilms grown in different WW
318 percentages using the *Psuedomonas putida* DH as a query as it is the only characterized enzyme
319 known to have DH activity. We quantified sequence counts in the biofilms grown in 0%, 10%,
320 30%, and 80% unfiltered WW mixed with stream water for two independent flume experiments
321 (Figure 1A). Strikingly, in two separate and independent flume experiments, we similarly observed

322 a positive correlation between the metagenomic DH abundances in biofilms and the percentage
323 of unfiltered WW the biofilms were grown in (Figure 1A, R^2 = 0.77 for Exp. 2, R^2 = 0.42 for Exp.
324 1). We also calculated the coverage of metagenomic reads mapping to the *P. putida* DH reference
325 in the stream, filtered WW, and unfiltered WW samples and identified significantly more reads
326 mapping in WW relative to stream water (Figure S6, *p*-value < 2e-16).

327 The positive association between DH homolog abundances and treated wastewater
328 percentages suggested that DEET hydrolysis might be an important biotransformation reaction
329 occuring in WW and WW-impacted stream biofilms. A previous study on microbial degradation of
330 DEET using activated sludge exclusively reported transformation products of oxidative
331 transformation reactions i.e., amide *N*-deethylation and tolyl group oxidation (Helbling et al., 2010)
332 (Figure 1B). In contrast, direct amide hydrolysis in DEET was only previously observed in a pure
333 culture of *P. putida* isolated from activated sludge (Rivera-Cancel et al., 2007) (Figure 1B). Here
334 we detected the DEET hydrolysis transformation product (TP137) in biofilms grown at 80% WW
335 (Figure 1C). The transformation products were consistent with the higher abundance of DH
336 homologs found in these biofilms (Figure 1A). Besides the DEET hydrolysis product, we also
337 detected the product of tolyl group oxidation (i.e., TP222) in 80% WW-grown biofilms, indicating
338 that oxidative biotransformation of DEET also occurred. TP164 was not detected probably due to
339 its transient behavior as described previously (Helbling et al., 2010).

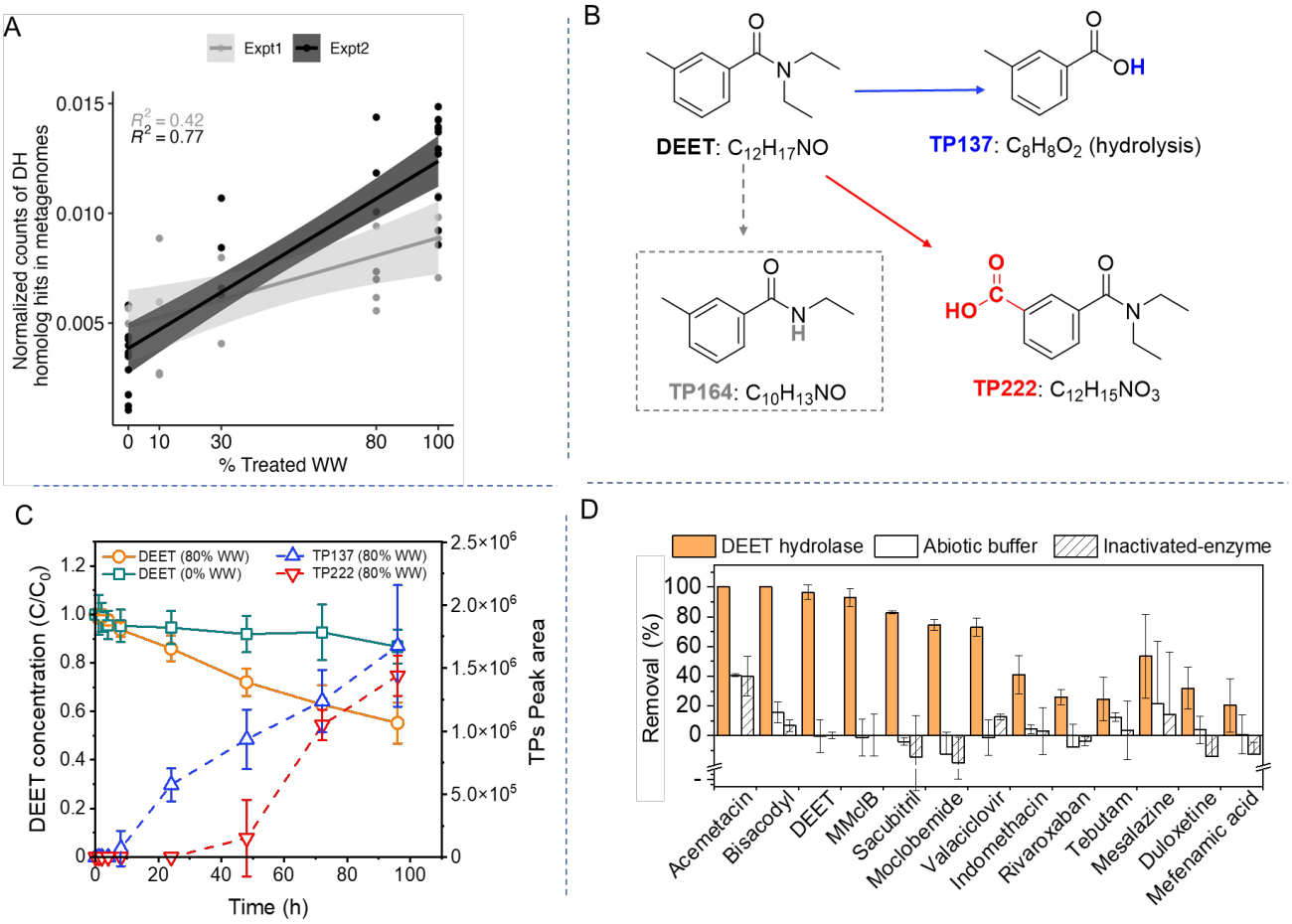
340 **DEET hydrolase biotransformation of TrOCs**

341 To further understand the substrate specificity of DEET hydrolases, we first cloned,
342 heterologously expressed, and purified the *P. putida* DH as a reference enzyme, and tested its
343 reactivity against a panel of 183 wastewater-related TrOCs. To avoid cumulative inhibitory effects,
344 the 183 WW-relevant compounds were pooled into 14 sub-mixtures (Table S4 and S5). Each sub-
345 mixture contains compounds that were expected to undergo different initial biotransformation
346 reactions predicted by enviPath (Wicker et al., 2016; Zhang and Fenner, 2023). Results showed
347 that *P. putida* DH significantly biotransformed 13 out of 183 compounds (Figure 1D). With the
348 exceptions of mesalazine, duloxetine, and mefenamic acid, all other biotransformed compounds
349 contained at least one hydrolyzable moiety (Figure S7). Among the chemicals screened, five
350 tertiary amide-containing compounds (acemetacin, DEET, MMclB, indomethacin, and tebutam)
351 exhibited significant enzymatic removal (Figure 1D, Table S4, and Figure S7). Other compounds
352 with tertiary amide functional groups, including chlorotoluron and diuron, were not transformed.
353 This observation suggests that the presence of other chemical functional groups may affect
354 enzyme binding such as the presence of bulky side chains as in the case of chlorotoluron and
355 diuron (Figure S7). In addition to the ability of the *P. putida* DH to catalyze hydrolysis of certain
356 tertiary amides, it also demonstrated promiscuous activity towards secondary amides and esters
357 (Figure 1D and Figure S7).

358 **Construction of a metagenomic library of DEET hydrolase family homologs.**

359 Previous studies on the *P. putida* DH placed it within the di-domain CocE/NonD family
360 encoding both alpha/beta hydrolase and Xaa-Pro dipeptidyl-peptidase protein domains (Rivera-
361 Cancel et al., 2012, 2007). Referred here for simplicity as the 'DH family', enzymes in this diverse
362 family were also shown to cleave cocaine (Bresler et al., 2000) and α-amino acid esters (Barends

et al., 2006). The broad substrate range of DH family enzymes beyond DEET hydrolysis along with our results from the *P. putida* DH (Figure 1D) suggests the potential for DH homologs to transform other TrOCs beside DEET. Therefore, we turned our attention to uncharacterized DH homologs detected in stream biofilm metagenomes. After metagenomic assembly and gene calling, we selected a library of 64 full-length DH homologs from our metagenomes based on taxonomic and sequence diversity (see Methods). Taxonomic annotations for the DH sequences were assigned (Methods) to total 10 different bacterial phyla: Acidobacteriota, Actinobacteriota, Bacteriodota, Chloroflexota, Firmicutes (Bacillota), Gemmatimonadota, Myxococcota, Planctomycetota, Proteobacteria (Pseudomonadota), and Verrucomicrobiota. Surprisingly, although we did recover DH homologs from *Pseudomonas* (Rivera-Cancel et al., 2007) and other closely-related genera in our biofilms, a full-length DH sequence with amino acid sequence identity greater than 40% to the *P. putida* DH reference could not be recovered from the assemblies. Full-length metagenomic DH homologs with start and stop codons intact recovered from biofilm metagenomes had an average of only 25% amino acid (AA) sequence identity with the *P. putida* DH (Table S2). The high level of sequence diversity and general conservation of DH homologs across phyla hints to a more central conserved role of DH homologs *e.g.*, in primary metabolism, beyond specialized DEET hydrolysis. In order to test this hypothesis, we next investigated DH family sequence-structure-function relationships (Figure 2).



382 **Figure 1**: A) Normalized numbers of hits of DH homologs recovered from assembled
383 metagenomic sequences fitted with a linear regression model (95% confidence interval) across
384 two independent experiments: experiment 1 (light gray) and experiment 2 (dark gray). B)
385 Documented DEET biotransformation pathways and transformation products (TPs) detected in
386 this study. The product TP164 in the dashed-gray box was reported previously (Helbling et al.,
387 2010) but not detected in this study. C) DEET degradation (solid line) and detected transformation
388 products (TPs) (dashed line) in 80% WW and 0% WW (no TP was detected in 0% WW samples).
389 D) Purified *P. putida* DH substrate specificity for active enzyme (in orange) relative to an
390 inactivated enzyme (hashed lines) and abiotic buffer control (white). Detailed selection criteria
391 and complete biotransformation data are available in the Methods and Table S4.

392 **Sequence-structure-function relationships among DH family enzymes**

393 We reconstructed a maximum-likelihood phylogeny of metagenomic DH homologs relative
394 to the reference *P. putida* DH (Figure 2A). Although enzymes from many folds are often sequence-
395 diverse, they still often share conserved protein structures (Illergård et al., 2009). Thus, to further
396 test structural similarity of our DH homologs, we used AlphaFold2 (Jumper and Hassabis, 2022)
397 to construct structural models for the 64 metagenomic DH homologs in our library. Structural
398 model comparisons revealed that enzymes shared an average of 1.73 ± 0.43 angstrom (Å) root-
399 mean squared deviation (RMSD) of atomic positions with the *P. putida* reference DH (Table S6).
400 This is on par with RMSD values measured within other contaminant-degrading enzyme families
401 sharing similar functions (Lakshminarasimhan et al., 2010; Perez-Garcia et al., 2021; Sidhu et al.,
402 2019). We then compared our sequence phylogeny (Figure 2B) with a neighbor-joining tree
403 generated from the RMSD matrix of all-vs-all structural alignments in our AlphaFold2 library
404 (Figure 2C). This head-to-head tree comparison enabled us to view how clades of sequence-
405 similar proteins might correlate with structure and identify structural outliers. DH homologs tend
406 to clade together with organisms from the same phylum both based on their sequence and
407 structure.

408 Important for this comparison, we verified that our DH homolog structural models were of
409 high confidence with an average predicted local distance difference test (pLDDT) scores >80
410 (Table S6). Structural alignments of catalytic residues in our library with the *P. putida* DH active
411 site revealed 100% conservation of the catalytic serine and greater than 90% conservation of a
412 complete Ser-His-Asp catalytic triad (Table S6). Our metagenomic library contained more
413 variation (Table S6) with respect to oxyanion hole residues (*P. putida* DH: Y84 and W167) and
414 two other known substrate interacting residues M170 and W214 (Rivera-Cancel et al., 2012).
415 Moving forward, we hypothesized that these variants occurring in nature might correspond to
416 altered substrate specificity of the homologs in our metagenomic library.

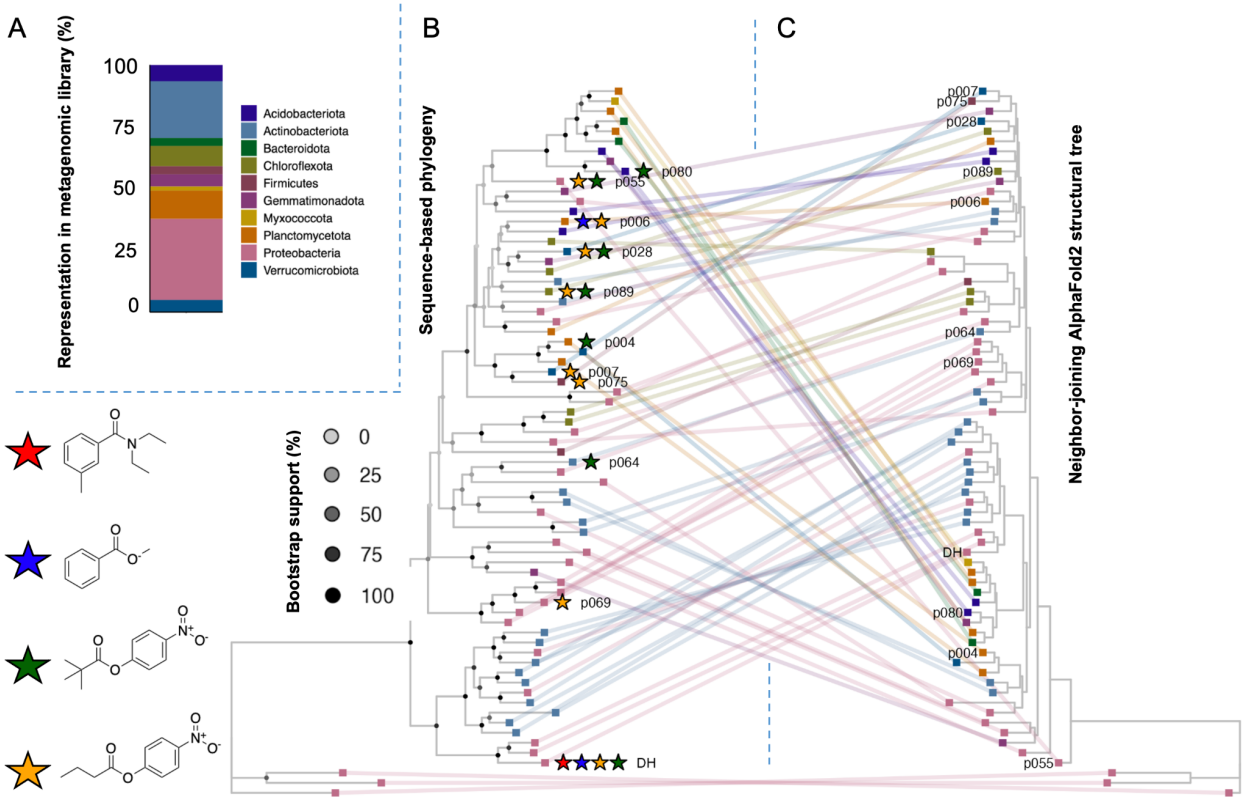


Figure 2. A) Representation of phyla as a percentage in the metagenomic DH homolog library (n = 64). B) Maximum-likelihood sequence-based phylogeny and C) neighbor-joining structural tree of DH family hydrolases. Tips with the same identity are linked between trees. Bootstrapping support for internal nodes is indicated by shading. Tips are colored based on phylum of origin of hydrolases. Star colors correspond to activity measured with distinct substrates: Red = DEET, Blue = methylbenzoate, Green = 4-nitrophenyl trimethylacetate, Gold = 4-nitrophenyl butyrate. Functionally-divergent alpha/beta hydrolases (PDB IDs: 1MJ5, 3HI4 and 6QHV) are used as an outgroup.

Functional screening of the metagenomic DH homolog library

In order to test the substrate specificity of our metagenomic DH homolog library, we codon optimized, cloned and and expressed each of the 64 metagenomic sequences in *E. coli*. We screened for hydrolase activity with four different ester or amide-containing substrates (Figure 2): 4-nitrophenyl butyrate, 4-nitrophenyl trimethylacetate, methylbenzoate, and DEET using a previously optimized whole-cell protocol (Robinson et al., 2020). Eleven out of 64 metagenomic library enzymes tested were measured as active (Methods in Supporting Information) with at least one substrate, primarily with the linear and branched 4-nitrophenyl aliphatic esters (Figure 2). Substrates with ester linkages were preferred by metagenomic DH homologs likely because esters are generally more susceptible to nucleophilic attack than amides, e.g., by the catalytic serine of the α/β-fold hydrolase triad. Out of the 64 enzymes screened, only one additional metagenomic hydrolase (denoted p006) also hydrolyzed methylbenzoate (Figure 4A). As reported previously (Rivera-Cancel et al., 2012), the *P. putida* DH exhibited significantly higher specific activity for the hydrolysis of methylbenzoate over DEET (Figure 4A).

In contrast, compounds containing tertiary amide groups, such as DEET, are more sterically hindered than carboxylic acid esters. To our surprise, out of the 64 DH homologs tested, only the reference *P. putida* DH was capable of DEET hydrolysis (Figure 4). This may be due to the fact that the *P. putida* DH contains several substrate-interacting residues in a unique combination relative to the metagenomic DH homolog library. Specifically, Y84, W167 and M170 exclusively co-occur in the *P. putida* DH. Although each of these residues occurred individually or in pairs in different members of our metagenomic DH homolog library, all three residues together were not in any of our metagenomic DH homologs (Table S6). This suggests that while altering individual residues near the active site affects DEET hydrolysis kinetics (Rivera-Cancel et al., 2012), the co-occurrence of multiple residues may be essential to enable an enzyme to catalyze DEET hydrolysis. The co-occurrence of various combinations of these residues within metagenomic sequences creates a reservoir of naturally occurring sequences, which could potentially have combined to give rise to DEET hydrolase functionality.

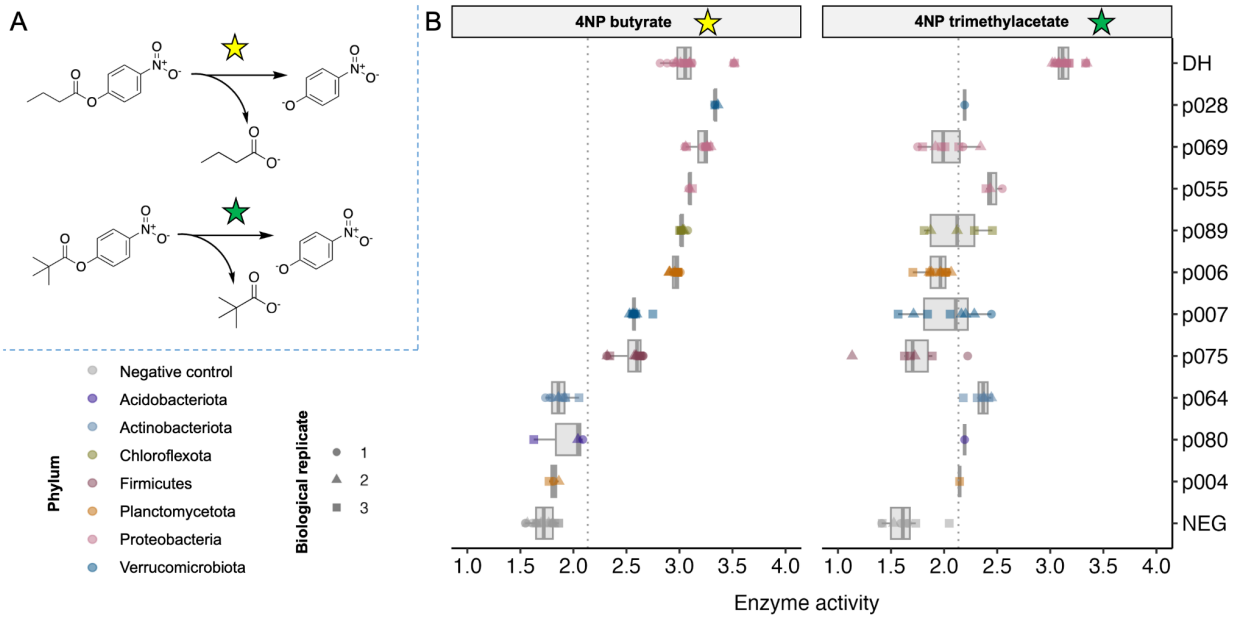


Figure 3. A) Linear and branched 4-nitrophenyl (4NP) ester substrates tested: 4NP butyrate (yellow star) and 4NP trimethylacetate (green star) both release 4NP as a chromophore. B) The ten most active metagenomic enzymes in our library tested in a whole-cell *E. coli* screening assay for 4NP hydrolysis activity. Units for enzyme activity are reported as $\log_{10}$ of nanomoles of 4NP released per hour by an *E. coli* cell culture with an optical density unit (OD) of 1.0, enabling direct comparison with enzymatic 4NP production rates previous studies (Robinson et al., 2020; Smith et al., 2020). Enzyme activity is measured compared to negative control empty vector (NEG) and reference *P. putida* DEET hydrolase (DH). The average activity level calculated across all enzymes and substrates is drawn as a dotted gray line.

465 **Biochemical characterization of a new metagenomic hydrolase p055 and its reactivity with**
466 **TrOCs**

467 Among the 64 enzymes in our functional metagenomic library, p055 had the highest
468 summed average enzyme activity for 4-nitrophenyl trimethylacetate and 4-nitrophenyl butyrate.
469 Taxonomic analysis could only resolve the origin of the metagenomic p055 sequence at the
470 family-level as belonging to photoheterotrophic purple nonsulfur bacteria assigned to the family
471 *Rhodospirillaceae.* Interestingly, p055 metagenomic abundances also showed a clear increasing
472 trend with WW percentage and it was the only active enzyme in our library which showed this
473 trend (Figure S8). Furthermore, the p055 model also formed a 'structural outlier' in the
474 comparative phylogenies (Figure 2). With respect to sequence identity, p055 claded with other
475 metagenomic hydrolases but it clearly separated from the rest of the DH homologs on the basis
476 of its structure. Closer analysis revealed it lacked two elongated alpha helices characteristic of
477 the reference *P. putida* DH homologs (Figure 4). Of note, the helical regions are in a high-
478 confidence region of the AlphaFold2 models (pLDDT > 95) suggesting the structural deviation
479 was not a modeling artifact. Since these helices are absent from p055 but present in the *P. putida*
480 DH are located in the vicinity of the active site, we hypothesized p055 structural variation might
481 correspond to substrate variation.

482 To test p055 substrate specificity, we expressed and purified p055 to homogeneity as a
483 66 kDa protein with a yield of 0.640 mg/mL (Figure S4). We calculated the specific enzymatic
484 activity in buffers over a range of different pHs and determined p055 had a pH optimum between
485 8.5 – 9 (Figure S9). The enzyme p055 had an optimal temperature of 35°C (Figure S10) and
486 displayed activity up to 45°C which corresponds well with the optimal growth temperatures of the
487 *Rhodospirillaceae* as the predicted source (Baldani et al., 2014). However, the enzyme was not
488 thermostable. Following incubation for one hour at 40°C, it lost 50% of its activity (Figure S11).
489 Loss of activity at 40°C is not unexpected since this enzyme originates from an environmental
490 biofilm grown in WW and stream water where temperatures ranged between 12°C to 14°C
491 (Desiante et al., 2022).

492 A screening of the substrate promiscuity of this environmental hydrolase p055 with the
493 183 TrOCs revealed reactivity towards 8 substrates (Figure 5C). Notably, when comparing the
494 substrate specificity to the *P. putida* DH reference, we identified 20 substrates which were
495 different between the enzymes and two which were shared. Namely, both p055 and *P. putida* DH
496 shared similar reactivity to catalyze the hydrolysis of bisacodyl and moclobemide. Bisacodyl is
497 known as a prodrug that can be metabolized by intestinal and bacterial enzymes to its
498 deacetylated active metabolite. As active esterases, *P. putida* DH and p055 are both capable of
499 facilitating this hydrolytic process.

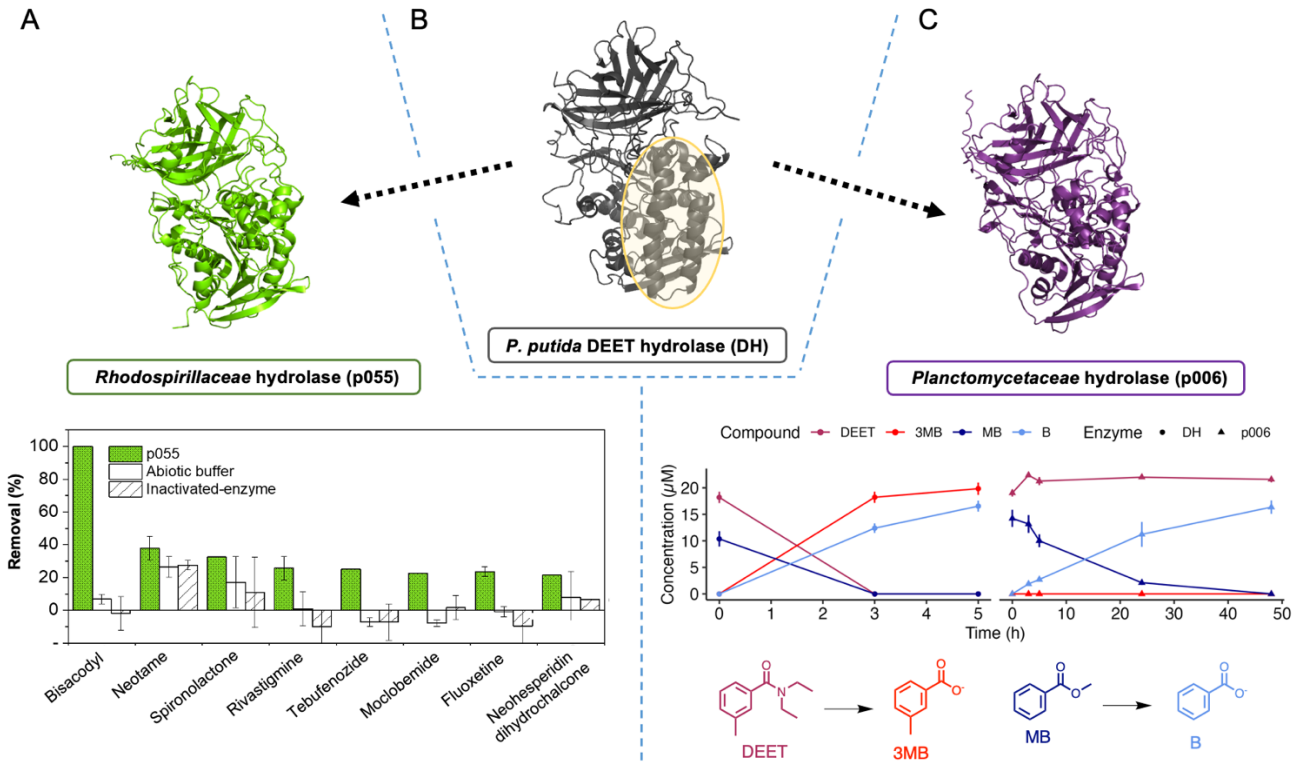


Figure 4. A) AlphaFold2 model of DH homolog p055 and substrate specificity as measured by LC-MS/MS, B) AlphaFold2 model of *P. putida* DH and C) AlphaFold2 model of p006 and HPLC measurements of DEET and methylbenzoate (MB) and product formation of 3-methylbenzoate (red, denoted as "3MB") and benzoate (light blue, denoted as "B"), respectively, by the *P. putida* DH and p006 (C). Detailed selection criteria and complete biotransformation data are available in Methods and Table S5.

Environmental Relevance and Implications

Through combined metagenome mining and experimental characterization of diverse DH homologs from ten different phyla, our study yielded new insights into the involvement of DH family enzymes in TrOC biotransformations. Overall, we detected activity in approximately 5% of the 626 enzyme-substrate combinations tested. The success rate with our targeted approach was higher than untargeted functional metagenomics approaches where on average less than 0.08% of colonies screened are active, depending on the screening method (Ferrer et al., 2016). This highlights the benefits of *in silico* analysis to prioritize candidates within metagenomic libraries and map sequence-structure-function relationships.

This work provided lessons for the computational prediction of responsible enzymes for TrOC removal including the importance of the threshold used for AA sequence similarity when selecting protein candidates to test. A relatively low AA sequence similarity of 30% or less enters into the protein 'twilight zone' where it is difficult to assign function based on homology (Rost, 1999). We specifically targeted this twilight zone in our study as it represents undersampled sequence space and enabled our discovery of metagenomic 'structural outliers' such as p055

522 with altered substrate specificity. However, our results also underpin that correlations in 'omic
523 data sets do not equal causation and predictions must be tested. The fact that our metagenomic
524 DH homologs were active with other substrates but not with the predicted substrate (DEET)
525 revealed functional differences even among closely-related enzymes. This suggests that besides
526 gene sequence similarity, further knowledge on individual amino acids, secondary structural
527 features (e.g., alpha helices), as well as experimental or environmental conditions (e.g., pH,
528 temperature, matrix) must be tested to improve the prediction of the substrate specificity of
529 enzymes for TrOC removal in environmental samples.

530 The fact that many metagenomic DH homologs were hydrolases active with other
531 substrates, yet not active with DEET, also raises the question of the responsible enzyme(s) for
532 DEET hydrolysis in stream biofilms. While low numbers of metagenomic short reads mapped to
533 the *P. putida* DH (Figure S6), DH homologs with high similarity to the reference DH were not
534 recovered in the metagenomic assembly process and appear to be low abundance. This
535 emphasizes the potential limitations of assembly-based approaches for capturing 'rare', low-
536 abundance genes involved in TrOC biotransformations. Activity-based rather than sequence-
537 based screening approaches to uncover potential additional DH enzymes represent a clear future
538 direction. Complementary strategies such as (meta)transcriptomics to identify candidates with
539 differential expression patterns (Achermann et al., 2018; Athanasakoglou and Fenner, 2022;
540 Maini Rekdal et al., 2019) may point toward additional enzymes to screen not detected by shifts
541 in metagenomic abundances.

542 Here, we examined the only enzyme family for which we could match an observed TrOC
543 biotransformation to a validated reference sequence detected in our metagenomes. Given the
544 specificity of these enzymes, we expect diverse hydrolase families beyond DH and DH homologs
545 contribute to hydrolysis of diverse TrOCs (Hübner et al., 2021), including *downstream effect*
546 compounds. It is also worth noting that, within microbial communities, enzymes and their
547 producers exist in a mixture and certain transformation products of a TrOC may be further
548 transformed by enzymes produced by different community members (Wang et al., 2022). All the
549 evidence points to the importance of maintaining (bio)diversity to effectively cope with challenges
550 caused by anthropogenic activities, which we here sampled through characterizing new DH
551 homologs spanning ten different microbial phyla from complex biofilm communities. Future
552 studies investigating diverse enzyme families are necessary to expand reference sets for TrOC
553 biodegradation, which will not only deepen our understanding of the fate of TrOCs in
554 environments, but also facilitate the development of more accurate prediction models for
555 assessing TrOC reactivity.

556 **Data Availability**

557 All sequencing data are publicly available from NCBI under BioProject accession number
558 PRJNA1008123. Scripts for data analysis and figures in this manuscript are available on
559 www.github.com/MSM-group/DEET-hydrolase.

560

561

Acknowledgments

We acknowledge Thomas Fleischmann, René Gall, and Max Hofland for their technical support. We also acknowledge Yasuo Yoshikuni, Ian Blaby, and Miranda Harmon-Smith for support with metagenomic library constructs. The work (proposal: 10.46936/10.25585/60008420) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. We thank Elia Ceppi and Philipp Longree from the Department of Environmental Chemistry at Eawag for assisting with the development of UHPLC-HRMS/MS methods, and Martina Kalt and Miguel Heussi for preparing chemical stock solutions. We thank Kunyang Zhang for assisting with cheminformatics analyses. We thank Giulia Gionchetta, Robert Niederdorfer, and Helmut Bürgmann for their support in the metagenomic analysis. This work was supported by Eawag Discretionary Funding (for S.L.R. and K.F.) and Swiss National Science Foundation (project no. 200021L_201006, for Y.Y. and K.F.; PZPGP2_209124 for S.L.R.).

References

Achermann, S., Falås, P., Joss, A., Mansfeldt, C.B., Men, Y., Vogler, B., Fenner, K., 2018. Trends in micropollutant biotransformation along a solids retention time dradient. Environ. Sci. Technol. 52, 11601–11611.

Achermann, S., Mansfeldt, C.B., Müller, M., Johnson, D.R., Fenner, K., 2020. Relating metatranscriptomic profiles to the micropollutant biotransformation potential of complex microbial communities. Environ. Sci. Technol. 54, 235–244.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538.

Altenburger, R., Scholze, M., Busch, W., Escher, B.I., Jakobs, G., Krauss, M., Krüger, J., Neale, P.A., Ait-Aissa, S., Almeida, A.C., Seiler, T.-B., Brion, F., Hilscherová, K., Hollert, H., Novák, J., Schlichting, R., Serra, H., Shao, Y., Tindall, A., Tollefsen, K.E., Umbuzeiro, G., Williams, T.D., Kortenkamp, A., 2018. Mixture effects in samples of multiple contaminants - An inter-laboratory study with manifold bioassays. Environ. Int. 114, 95–106.

Athanasakoglou, A., Fenner, K., 2022. Toward characterizing the genetic basis of trace organic contaminant biotransformation in activated sludge: The role of multicopper oxidases as a case study. Environ. Sci. Technol. 56, 313–324.

Baldani, J.I., Videira, S.S., dos Santos Teixeira, K.R., Reis, V.M., de Oliveira, A.L.M., Schwab, S., de Souza, E.M., Pedraza, R.O., Baldani, V.L.D., Hartmann, A., 2014. The family *Rhodospirillaceae*, in the Prokaryotes. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 533–618.

Barends, T.R.M., Polderman-Tijmes, J.J., Jekel, P.A., Williams, C., Wybenga, G., Janssen, D.B., Dijkstra, B.W., 2006. *Acetobacter turbidans* alpha-amino acid ester hydrolase: how a single mutation improves an antibiotic-producing enzyme. J. Biol. Chem. 281, 5804–5810.

Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., Packmann, A.I., 2016. The ecology and biogeochemistry of stream biofilms. Nat. Rev. Microbiol. 14, 251–263.

Bonatelli, M.L., Rohwerder, T., Popp, D., Liu, Y., Akay, C., Schultz, C., Liao, K.-P., Ding, C., Reemtsma, T., Adrian, L., Kleinsteuber, S., 2023. Recently evolved combination of unique sulfatase and amidase genes enables bacterial degradation of the wastewater

607 micropollutant acesulfame worldwide. Front. Microbiol. 14, 1223838.
608 Bresler, M.M., Rosser, S.J., Basran, A., Bruce, N.C., 2000. Gene cloning and nucleotide
609 sequencing and properties of a cocaine esterase from *Rhodococcus* sp. strain MB1. Appl.
610 Environ. Microbiol. 66, 904–908.
611 Burdon, F.J., Bai, Y., Reyes, M., Tamminen, M., Staudacher, P., Mangold, S., Singer, H.,
612 Räsänen, K., Joss, A., Tiegs, S.D., Jokela, J., Eggen, R.I.L., Stamm, C., 2020. Stream
613 microbial communities and ecosystem functioning show complex responses to multiple
614 stressors in wastewater. Glob. Chang. Biol. 26, 6363–6382.
615 Caliman, F.A., Gavrilescu, M., 2009. Pharmaceuticals, Personal Care Products and Endocrine
616 Disrupting Agents in the Environment – A Review. Clean Soil Air Water 37, 277–303.
617 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L.,
618 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.
619 Carles, L., Wullschleger, S., Joss, A., Eggen, R.I.L., Schirmer, K., Schuwirth, N., Stamm, C.,
620 Tlili, A., 2022. Wastewater microorganisms impact microbial diversity and important
621 ecological functions of stream periphyton. Water Res. 225, 119119.
622 Carles, L., Wullschleger, S., Joss, A., Eggen, R.I.L., Schirmer, K., Schuwirth, N., Stamm, C.,
623 Tlili, A., 2021. Impact of wastewater on the microbial diversity of periphyton and its
624 tolerance to micropollutants in an engineered flow-through channel system. Water Res.
625 203, 117486.
626 Che, S., Jin, B., Liu, Z., Yu, Y., Liu, J., Men, Y., 2021. Structure-specific aerobic defluorination of
627 short-Chain fluorinated carboxylic acids by activated sludge communities. Environ. Sci.
628 Technol. Lett. 8, 668–674.
629 Daughton, C.G., Ternes, T.A., 1999. Pharmaceuticals and personal care products in the
630 environment: agents of subtle change? Environ. Health Perspect. 107 Suppl 6, 907–938.
631 Desiante, W.L., Carles, L., Wullschleger, S., Joss, A., Stamm, C., Fenner, K., 2022. Wastewater
632 microorganisms impact the micropollutant biotransformation potential of natural stream
633 biofilms. Water Res. 217, 118413.
634 Desiante, W.L., Minas, N.S., Fenner, K., 2021. Micropollutant biotransformation and
635 bioaccumulation in natural stream biofilms. Water Res. 193, 116846.
636 Evich, M.G., Davis, M.J.B., McCord, J.P., Acrey, B., Awkerman, J.A., Knappe, D.R.U.,
637 Lindstrom, A.B., Speth, T.F., Tebes-Stevens, C., Strynar, M.J., Wang, Z., Weber, E.J.,
638 Henderson, W.M., Washington, J.W., 2022. Per- and polyfluoroalkyl substances in the
639 environment. Science 375, eabg9065.
640 Fenner, K., Canonica, S., Wackett, L.P., Elsner, M., 2013. Evaluating pesticide degradation in
641 the environment: blind spots and emerging opportunities. Science 341, 752–758.
642 Fenner, K., Elsner, M., Lueders, T., McLachlan, M.S., Wackett, L.P., Zimmermann, M., Drewes,
643 J.E., 2021. Methodological advances to study contaminant biotransformation: new
644 prospects for understanding and reducing environmental persistence? ACS EST Water 1,
645 1541–1554.
646 Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O.V., Golyshin, P.N.,
647 2016. Estimating the success of enzyme bioprospecting through metagenomics: current
648 status and future trends. Microb. Biotechnol. 9, 22–34.
649 Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-
650 generation sequencing data. Bioinformatics 28, 3150–3152.
651 Halden, R., 2011. Contaminants of emerging concern in the environment: ecological and human
652 health considerations. OUP USA.
653 Han, P., Rios-Miguel, A.B., Tang, X., Yu, Y., Zhou, L.-J., Hou, L., Liu, M., Sun, D., Jetten,
654 M.S.M., Welte, C.U., Men, Y., Lücker, S., 2023. Benzimidazole fungicide biotransformation
655 by comammox *Nitrospira* bacteria: transformation pathways and associated proteomic
656 responses. J. Hazard. Mater. 445, 130558.
657 Han, P., Yu, Y., Zhou, L., Tian, Z., Li, Z., Hou, L., Liu, M., Wu, Q., Wagner, M., Men, Y., 2019.

Specific micropollutant biotransformation pattern by the comammox bacterium. Environ. Sci. Technol. 53, 8695–8705.
Helbling, D.E., Hollender, J., Kohler, H.-P.E., Singer, H., Fenner, K., 2010. High-throughput identification of microbial transformation products of organic micropollutants. Environ. Sci. Technol. 44, 6621–6627.
Hübner, U., Wolff, D., Achermann, S., Drewes, J.E., Wick, A., Fenner, K., 2021. Analyzing (Initial) biotransformation reactions as an organizing principle for unraveling the extent of trace organic chemical biotransformation in biofiltration systems. ACS EST Water. https://doi.org/10.1021/acsestwater.1c00145
Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.
Illergård, K., Ardell, D.H., Elofsson, A., 2009. Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. Proteins 77, 499–508.
Jin, B., Liu, H., Che, S., Gao, J., Yu, Y., Liu, J., Men, Y., 2023. Substantial defluorination of polychlorofluorocarboxylic acids triggered by anaerobic microbial hydrolytic dechlorination. Nature Water 1, 451–461.
Jumper, J., Hassabis, D., 2022. Protein structure predictions to atomic accuracy with AlphaFold. Nat. Methods 19, 11–12.
Kennes-Veiga, D.M., Trueba-Santiso, A., Gallardo-Garay, V., Balboa, S., Carballa, M., Lema, J.M., 2022. Sulfamethoxazole enhances specific enzymatic activities under aerobic heterotrophic conditions: a metaproteomic approach. Environ. Sci. Technol. 56, 13152–13159.
Lakshminarasimhan, M., Madzelan, P., Nan, R., Milkovic, N.M., Wilson, M.A., 2010. Evolution of new enzymatic function by structural modulation of cysteine reactivity in *Pseudomonas fluorescens* isocyanide hydratase. J. Biol. Chem. 285, 29651–29661.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Lee, J., Ju, F., Maile-Moskowitz, A., Beck, K., Maccagnan, A., McArdell, C.S., Dal Molin, M., Fenicia, F., Vikesland, P.J., Pruden, A., Stamm, C., Bürgmann, H., 2021. Unraveling the riverine antibiotic resistome: The downstream fate of anthropogenic inputs. Water Res. 197, 117050.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome project data processing subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Maini Rekdal, V., Bess, E.N., Bisanz, J.E., Turnbaugh, P.J., Balskus, E.P., 2019. Discovery and inhibition of an interspecies gut bacterial pathway for levodopa metabolism. Science 364, eaau6323.
Mansfeldt, C., Deiner, K., Mächler, E., Fenner, K., Eggen, R.I.L., Stamm, C., Schönenberger, U., Walser, J.-C., Altermatt, F., 2020. Microbial community shifts in streams receiving treated wastewater effluent. Sci. Total Environ. 709, 135727.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7, 11257.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. Mol. Biol. Evol. 37, 1530–1534.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682.

Nimura, T., Tokieda, T., Yamaha, T., 1974. Partial purification and some properties of cyclamate sulfamatase. J. Biochem. 75, 407–417.

Oberortner, E., Cheng, J.-F., Hillson, N.J., Deutsch, S., 2017. Streamlining the design-to-build transition with build-optimization software tools. ACS Synth. Biol. 6, 485–496.

Perez-Garcia, P., Kobus, S., Gertzen, C.G.W., Hoeppner, A., Holzscheck, N., Strunk, C.H., Huber, H., Jaeger, K.-E., Gohlke, H., Kovacic, F., Smits, S.H.J., Streit, W.R., Chow, J., 2021. A promiscuous ancestral enzyme´s structure unveils protein variable regions of the highly diverse metallo-β-lactamase family. Commun Biol 4, 132.

R core team, 2022. R: A language and environment for statistical [WWW Document]. The R Project for Statistical Computing. URL https://www.R-project.org/ (accessed 22).

Rich, S.L., Helbling, D.E., 2023. Daily monitoring at a full-scale wastewater treatment plant reveals temporally variable micropollutant biotransformations. Environ. Sci. Technol. 57, 10404–10414.

Rivera-Cancel, G., Bocioaga, D., Hay, A.G., 2007. Bacterial degradation of *N,N*-diethyl-m-toluamide (DEET): cloning and heterologous expression of DEET hydrolase. Appl. Environ. Microbiol. 73, 3105–3108.

Rivera-Cancel, G., Sanders, J.M., Hay, A.G., 2012. Kinetics of hydrolysis and mutational analysis of *N,N*-diethyl-m-toluamide hydrolase from *Pseudomonas putida* DTB. FEBS J. 279, 1044–1053.

Robinson, S.L., Smith, M.D., Richman, J.E., Aukema, K.G., Wackett, L.P., 2020. Machine learning-based prediction of activity and substrate specificity for OleA enzymes in the thiolase superfamily. Synth. Biol. 5, ysaa004.

Rost, B., 1999. Twilight zone of protein sequence alignments. Protein Eng. 12, 85–94.

RStudio Team, 2020. RStudio: Integrated Development Environment for R [WWW Document]. URL http://www.rstudio.com/ (accessed 22).

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science 313, 1072–1077.

Sidhu, C., Solanki, V., Pinnaka, A.K., Thakur, K.G., 2019. Structure elucidation and biochemical characterization of environmentally relevant novel extradiol dioxygenases discovered by a functional metagenomics approach. mSystems 4, e00316-19.

Sievers, F., Higgins, D.G., 2018. Clustal omega for making accurate alignments of many protein sequences. Protein Sci. 27, 135–145.

Smith, M.D., Robinson, S.L., Molomjamts, M., Wackett, L.P., 2020. Assay reveals microbial OleA thiolases initiating hydrocarbon and β-Lactone biosynthesis. MBio 11, 00111-20.

Stadler, L.B., Delgado Vela, J., Jain, S., Dick, G.J., Love, N.G., 2018. Elucidating the impact of microbial community biodiversity on pharmaceutical biotransformation during wastewater treatment. Microb. Biotechnol. 11, 995–1007.

Tien, C.-J., Lin, M.-C., Chiu, W.-H., Chen, C.S., 2013. Biodegradation of carbamate pesticides by natural river biofilms in different seasons and their effects on biofilm community structure. Environ. Pollut. 179, 95–104.

Vasileiadis, S., Perruchon, C., Scheer, B., Adrian, L., Steinbach, N., Trevisan, M., Plaza-Bolaños, P., Agüera, A., Chatzinotas, A., Karpouzas, D.G., 2022. Nutritional inter-dependencies and a carbazole-dioxygenase are key elements of a bacterial consortium relying on a *Sphingomonas* for the degradation of the fungicide thiabendazole. Environ. Microbiol. 24, 5105–5122.

Wang, M., Chen, X., Liu, X., Fang, Y., Zheng, X., Huang, T., Tang, Y.-Q., Ackermann, M., Nie,

Y., Wu, X.-L., 2022. Even allocation of benefits stabilizes microbial community engaged in metabolic division of labor. Cell Rep. 40, 111410.

Wicker, J., Lorsbach, T., Gütlein, M., Schmid, E., Latino, D., Kramer, S., Fenner, K., 2016. enviPath-the environmental contaminant biotransformation pathway resource. Nucleic Acids Res. 44, D502–8.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4, 1686.

Xing, Y., Yu, Y., Men, Y., 2018. Emerging investigators series: occurrence and fate of emerging organic contaminants in wastewater treatment plants with an enhanced nitrification step. Environ. Sci.: Water Res. Technol. 4, 1412–1426.

Yu, G. (2022). Data Integration, Manipulation and Visualization of Phylogenetic Trees (1st ed.). Chapman and Hall/CRC.

Yu, Y., Che, S., Ren, C., Jin, B., Tian, Z., Liu, J., Men, Y., 2022. Microbial defluorination of unsaturated per- and polyfluorinated carboxylic acids under anaerobic and aerobic conditions: a structure specificity study. Environ. Sci. Technol. 56, 4894–4904.

Yu, Y., Han, P., Zhou, L.-J., Li, Z., Wagner, M., Men, Y., 2018. Ammonia monooxygenase-mediated cometabolic biotransformation and hydroxylamine-mediated abiotic transformation of micropollutants in an AOB/NOB coculture. Environ. Sci. Technol. 52, 9196–9205

Zhang, K., Fenner, K., 2023. enviRule: An End-to-end System for Automatic Extraction of Reaction Patterns from Environmental Contaminant Biotransformation Pathways. Bioinformatics. 39, btad407.