

MULTI-OMIC INVESTIGATION OF PLASTIC-ASSOCIATED MICROBES:
BIOINFORMATIC INSIGHTS INTO PLASTIC BIODEGRADATION AND
NOVEL DEGRADING GENES ACROSS ENVIRONMENTS

A Dissertation
Presented to
The Academic Faculty

by

Rodney Sylvester Ridley, Jr.

In Partial Fulfillment
of the Requirements for the Master of Science in Bioengineering in the
School of Chemical and Biomolecular Engineering

Georgia Institute of Technology
December 2023

COPYRIGHT © 2023 BY Rodney Sylvester Ridley, Jr.

MULTI-OMIC INVESTIGATION OF PLASTIC-ASSOCIATED MICROBES:
BIOINFORMATIC INSIGHTS INTO PLASTIC BIODEGRADATION AND NOVEL
DEGRADING GENES ACROSS ENVIRONMENTS

Approved by:

Dr. Kostas Konstantinidis, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Blair Brettmann
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Thomas DiChristina
School of Biological Sciences
Georgia Institute of Technology

Date Approved: August 23, 2023

ACKNOWLEDGEMENTS

I would like to thank Dr. Kostas Konstantinidis for his support and direction of this overall project. Additionally, I would like to thank Seongwook Woo for his support in understanding microbial ecology and its relation to plastics. I thank Blake Lindner for his support in regards to metagenomic pipeline development as well as his input on various aspects of microbiology and data science. I additionally thank Roth Conrad for his notable support within metagenomic pipeline development, as well as support in developing the framework for integration and downstream analysis of the subsequent data. I would also would like to thank GGBC, UG Athens, GA, RRID:SCR_010994 for the library preparation and sequencing. The library preparation and sequencing process was supported by GGBC-PacBio SMRT grant awarded to Blake Lindner.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	III
LIST OF TABLES	V
LIST OF FIGURES	VI
LIST OF SYMBOLS AND ABBREVIATIONS	VII
SUMMARY	VIII
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: RESULTS	5
2.1 STUDY DATASET	5
2.2 PHYLOGENETIC DISTRIBUTION OF MICROBES IN PLASTIC ENVIRONMENT	9
2.3 OCEAN RESULTS	13
2.3.1 <i>Beta Diversity</i>	14
2.4 <i>IN-SITU</i> DISTRIBUTION OF KNOWN DEGRADING SPECIES	19
2.4.1 <i>KEGG Metabolism</i>	20
2.5 GENE LEVEL RESULTS	22
2.5.1 <i>In-Situ Distribution of Known Degrading Genes</i>	23
2.5.2 <i>Disease Related Genes</i>	30
2.6 DATABASE.....	33
CHAPTER 3: DISCUSSION.....	36
CHAPTER 4: MATERIALS AND METHODS.....	42
REFERENCES.....	52

LIST OF TABLES

Table 1 - Meta-omic studies included in this work.	8
Table 2 - ALDEx2 Species Level Results - Henricella Subset.....	17
Table 3 - ALDEx2 Genus Level results for degraded PE & PP samples versus all other oceanic samples	18
Table 4 - Differential Enrichment Results for Antibiotic Resistance Genes	32

LIST OF FIGURES

Figure 1 - Overview of the data used in this study.	5
Figure 2 - Phylogenetic Tree of Bacterial Genomes in Current Dataset	10
Figure 3 - Phylogenetic Tree of Plastic Associated Prokaryotes	10
Figure 4 - Number of Confirmed Degrading Genomes per Plastic Type	11
Figure 5 - Bray-Curtis Beta-Diversity of Environmental Ocean Samples	14
Figure 6 - Abundance of Known Plastic-Degrading Genomes Across the Plastisphere.	19
Figure 7 - Distribution of Known Plastic Degrading Genes in the Environment	23
Figure 8 - Jaccard Similarity of Known Plastic Degrading Genes across Environmental Samples	24
Figure 9 - UMAP Network Graph for Proteins Observed in Metagenomic 'Plastisphere'	26
Figure 10 – UMAP Network Graph colored by Metatranscriptomic Enrichment from Wu and colleagues	29
Figure 11 - UMAP Network Graph showing location of genes from other 'omics studies	30
Figure 12 - Plastics Meta-omic Database (PMDB) Preview	33

LIST OF SYMBOLS AND ABBREVIATIONS

ALDEx - ANOVA-Like Differential Expression
ANI – Average Nucleotide Identity
AAI – Average Amino Acid Identity
CAZy - Carbohydrate-Active enZymes Database
GTDB – Genome Taxonomy Database
LFC – Log fold change
KEGG - Kyoto Encyclopedia of Genes and Genomes
MAG – Metagenome Assembled Genome
NMDS - Non-metric Multidimensional Scaling
PS – polystyrene
PE – polyethylene
PVC – polyvinyl chloride
PET – polyethylene terephthalate
PBAT - polybutylene adipate terephthalate
PP – polypropylene
PHA – polyhydroxyalkanoate
PHB - polyhydroxybutyrate
PLA – polylactic acid
PA – polyamide
PCL – polycaprolactone
PF - ecovio® FT 2341
PU – polyurethane
PTFE - Polytetrafluoroethylene
PEEK - polyether ether ketone
PMMA – polymethyl methacrylate
PAA – PolyLack® Aqua Brillante
RNA – ribonucleic acid
DNA – deoxyribonucleic acid
VFDB – Virulence Factor Database
ARG – Antibiotic resistance gene
AMR – Antimicrobial resistance
iTOL – Interactive Tree of Life
PAZy – Plastic Active Enzymes Database
Plastizyme – Plastic degrading enzyme
PMDB – Plastics Meta-omic Database
TAD80 – 80% Truncated Average Depth
UMAP - Uniform Manifold Approximation and Projection for Dimension Reduction
UV - Ultraviolet

SUMMARY

Synthetic plastics and their resulting waste are ubiquitous across the planet, from the Arctic to the tropics. Despite increasing efforts to understand the fate and transport of these plastics, their fate and impact on the environment and public health are still not well understood. To better comprehend the microbial ecology associated with plastic waste and its potential for bioremediation, we conducted a large-scale analysis of all publicly available meta-omic studies investigating plastics in the environment.

Importantly, we observed low prevalence of previously reported plastic degrading populations throughout most environments, except for substantial enrichment in riverine systems. This indicates rivers may be the one of the most promising environments for sources of plastic bioremediation. Ocean samples associated with degrading plastics showed clear differentiation from non-degrading polymers, showing enrichment of novel putative biodegrading taxa in the degraded samples. In regards to plastisphere pathogenicity, we observe no association between virulence factors and plastics in any environment. Additionally, we report a co-occurrence network analysis of 10+ million proteins associated with the plastisphere. This analysis revealed a localized sub-region enriched with known and putative plastizymes. These novel putative plastizymes may be useful for deeper investigations of nature's ability to biodegrade man-made plastics. Finally, the combined data from our meta-analysis was used to construct a publicly available database. These data should allow for integrated exploration of the microbial plastisphere and aid the community in continued research efforts to better understand the fate of plastics in the environment.

CHAPTER 1: Introduction

The anthropogenic issue of plastic waste is widespread throughout the environment. 460 million tons of plastic were produced in 2019¹. In this same year, 353 MT of plastic were discarded, a quantity expected to triple by the year 2060². Of all plastics produced each year, approximately 9% are captured and recycled for reuse; the remaining plastics are either disposed of in landfills or in unknown locations in the environment³. Recent studies have observed microplastics in highly remote regions across the earth, from the polar ice caps^{4,5} to remote mountainous regions⁶. Plastic and its additives have been seen in various studies to have potentially deleterious effects on species across known life, including perturbation of the photosynthetic activity of *Prochlorococcus*⁷ as well as various diseases in birds⁸ and fish⁹. The predicted rates of plastics degradation vary widely in the literature, with studies such as Chamas et al. (2020) predicting rates up to 1000 years, based on 25 degradation studies available¹⁰. Additional works have looked to further describe the characteristics of marine plastic degradation based on physical and chemical polymer properties¹¹. Efforts to increase capture and recycling of plastics are underway, via mechanical, chemical, and biological means^{12,13}. As the production of plastics is also rapidly increasing, further understanding and utilization of the ecological response to and breakdown of plastics in the natural environment is also pertinent in our response to this growing issue.

Mechanochemical, UV, and biological degradation are the primary methods by which plastics degrade in the environment¹⁰. The in-situ microbial community is of specific interest, as these organisms are the primary means of reintroducing the compounds in

these materials into the global carbon cycle. Microbes also regularly perform this biotransformation for natural polymers including complex polysaccharides such as lignocellulose. Aside from the degradable class of 'bioplastics', synthetic polymers are generally considered highly recalcitrant materials, degrading at rates as aforementioned. Literature on the microbial degradation of plastics has nonetheless increased rapidly in recent years, due both to the increased efforts of scientists as well as microbes likely evolving to utilize the available manmade compounds in their habitat. Microbial isolates have been recently reported to break down plastics such as PE and PET within several months^{14,15}. Enzymes isolated from these species have been able to degrade these plastics within hours, as in the case of the engineered enzyme FastPETase¹⁶, deriving from PETase found in *Ideonella sankesis*. Database compilations of these isolate studies have been assembled in efforts such as PlasticDB¹⁷ and PAZy¹⁸. However, it is unclear whether these laboratory cultures and enzymes truly represent the microbial community *in-situ* which is acting on plastics.

Meta-omics has been a well-utilized tool towards understanding the microbial ecology of plastics. Much of this work has been performed at the 16S rRNA gene (or simply 16S) 'meta-barcoding' level, allowing for insights into the community structure present on plastics. These studies have confirmed that plastics do cause shifts in the local microbial community, which are distinct from other biofilms. A meta-analysis of these works can be found in the manuscript of Wright and colleagues¹⁹. The use of whole-genome meta-omic data, such as metagenomics, metatranscriptomics, and metaproteomics, has only recently begun to be utilized towards this task. These data

are particularly useful for studying the process of plastics breakdown, as pathways of breakdown are not well understood and, in general, are not represented by 16S data²⁰. Whole genome meta-omic data can provide a more accurate representation of the functional profile of a community, as it contains the entire enzymatic profile of a given community. Microbial species are known to vary widely in their accessory gene content based on their local environment²¹. Therefore, observing enzymatic potential and regulation of microorganism specifically on plastics is pertinent to understanding the issue of current interest.

Recent whole-genome meta-omic studies have been performed in several environments across a breadth of plastic types, giving snapshots into the world of microbial responses to anthropogenic plastic waste. In an effort to create a more comprehensive picture of the microbial 'plastisphere', we have compiled all publicly available meta-omic, isolate genome, and enzyme data relating to plastics in the environment. This dataset consists of over 6 terabases of sequence data and is the largest meta-omic analysis of plastics to date to the best of our knowledge. Through this, we aimed to gain a better understanding of how microbes are functionally responding to and degrading plastics in the environment. This analysis additionally provides novel means for direct enzymatic discovery from environmental plastic data, as well as insights for directed lab efforts in isolation of novel microbes to degrade plastics. In order to provide a means for researchers to more directly analyze and utilize the available protein sequences, genomes, and environmental metadata, we have

compiled the data from this study into a publicly available database, the Plastics Meta-omic DataBase (PMDB).

CHAPTER 2: RESULTS

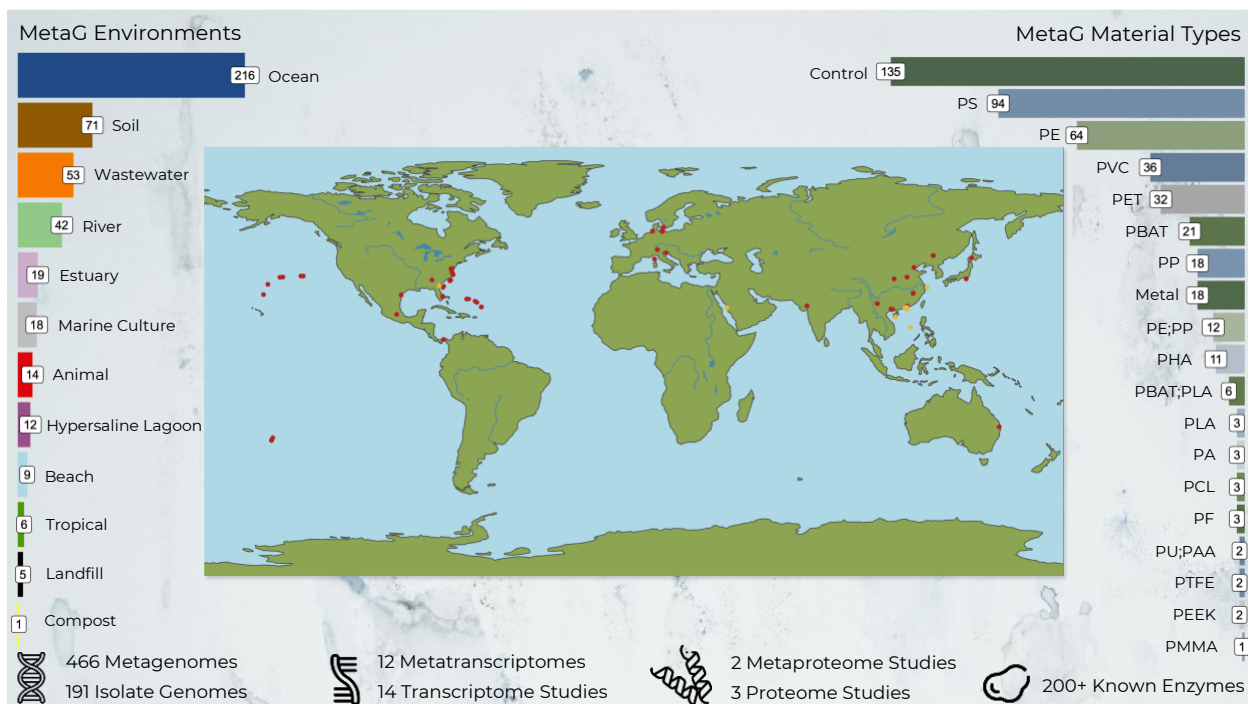


Figure 1 - Overview of the data used in this study.

Barplots and map show data for metagenomic sample locations and associated material types. Red dots on map indicate metagenomic sample was assembled via the custom metagenomic pipeline, yellow dots indicate samples from Zheng et al. biofilm study which were not assembled, but metagenome assembled genomes were instead taken from OceanDNA catalog for the sample. Bottom area shows information about the plastisphere data which was collected from the literature. Material type abbreviations: PS – polystyrene, PE – polyethylene, PVC – polyvinyl chloride, PET – polyethylene terephthalate, PBAT – polybutylene adipate terephthalate, PP – polypropylene, PHA – polyhydroxyalkanoate, PLA – polylactic acid, PA – polyamide, PCL – polycaprolactone, PF – ecovio® FT 2341, PU – polyurethane, PTFE – Polytetrafluoroethylene, PEEK – polyether ether ketone, PMMA – polymethyl methacrylate, PAA – PolyLack® Aqua Brillante, which is a polyether-polyurethane-acrylate (PE-PU-A) copolymer.

2.1 Study Dataset

We found 27 available studies with publicly available meta-omic data which passed initial quality checks (Table 1). These studies covered were spread across 4 continents and included all major environments (Figure 1). The majority of samples were oceanic, with many samples also in the soil and wastewater environments. All available studies had metagenomic data, there were three studies which additionally had metatranscriptomic or metaproteomic data. We additionally included 2 long read Pacbio

metagenomes from wastewater and degraded wood environments, which were selected as these are common locations from which plastic degraders have previously been isolated. In general, studies primarily utilized incubations under natural environmental conditions lasting several weeks to a few months. Several studies also utilized laboratory mesocosms to better control for environmental factors such as UV or mechanical weathering.

Briefly, in this study, we utilized a custom pipeline to generate metagenome-assembled genomes (MAGs) from all available metagenomic sequence data, including rich annotations of these MAGs and assembled but not binned contigs from UniProt²², the Carbohydrate-Active enZYmes Database (CAZy)²³, and other relevant databases. As most of the available data was metagenomic data, we based our analysis primarily at this type of data. We additionally provided the same rich annotations for known plastic-degrading isolate genomes. After creating a non-redundant genomospecies set, we mapped the available metagenomic reads to these species representatives to gather abundance and distribution information on these genomes in the environment. We here define genomospecies as 95% average nucleotide identity (ANI) of the shared content between related genomes, following previous practice²⁴. We also performed this same mapping on a non-redundant 'plastisphere' (gene set to gain understanding on the functional level of this microbial community. Using big-data dimensionality reduction techniques, we subsequently integrate the enzymes from these metagenomic data with available non-ribosomal RNA and proteomic data in order to search the available

protein space for novel plastic degrading enzymes. Further details on data collection and methodology may be found in the methods section.

Table 1 - Meta-omic studies included in this work.

Study Title and Reference included in this study	Study Type	Internal Study ID	Environment	DOI
Whole community and functional gene changes of biofilms on marine plastic debris in response to ocean acidification ²⁵	Metagenome	S01	Ocean	https://doi.org/10.1007/s00248-022-01987-w
Plastic materials and water sources actively select and shape wastewater plastispheres over time ²⁶	Metagenome	S02	Wastewater	https://doi.org/10.1007/s11783-022-1580-1
New insights into the functioning and structure of the PE and PP plastispheres from the Mediterranean Sea ²⁷	Metagenome, Metaproteome	S03	Beach	https://doi.org/10.1016/j.envpol.2021.118678
Insights into plastic biodegradation: community composition and functional capabilities of the superworm (<i>Zophobas morio</i>) microbiome in styrofoam feeding trials ²⁸	Metagenome	S04	Mealworm	https://doi.org/10.1099/mgen.0.000842
Diversity and Activity of Communities Inhabiting Plastic Debris in the North Pacific Gyre ²⁹	Metagenome	S07	Ocean	https://doi.org/10.1128/mSystems.00024-16
Shotgun metagenomic data of microbiomes on plastic fabrics exposed to harsh tropical environments ³⁰	Metagenome	S08	Fiber	https://doi.org/10.1016/j.dib.2020.106226
Microbial Consortia of Putative Degradors of Low-Density Polyethylene-Associated Compounds in the Ocean ³¹	Metagenome	S09	Ocean	https://doi.org/10.1128/msystems.01415-21
Microplastics altered soil microbiome and nitrogen cycling: The role of phthalate plasticizer ³²	Metagenome	S10	Soil	https://doi.org/10.1016/j.ihazmat.2021.127944
Integrated metagenomic and metatranscriptomic analysis reveals actively expressed antibiotic resistomes in the plastisphere ³³	Metagenome, Metatranscriptome	S11	River	https://doi.org/10.1016/j.ihazmat.2022.128418
Selective enrichment of bacterial pathogens by microplastic biofilm ³⁴	Metagenome	S12	River	https://doi.org/10.1016/j.watres.2019.114979
Synergistic biodegradation of aromatic-aliphatic copolyester plastic by a marine microbial consortium ³⁵	Metagenome, Metatranscriptome, Metaproteome	S15	Marine Culture	https://doi.org/10.1038/s41467-020-19583-2
Shotgun Metagenomics Reveals the Benthic Microbial Community Response to Plastic and Bioplastic in a Coastal Marine Environment ³⁶	Metagenome	S16	Ocean	https://doi.org/10.3389/fmicb.2019.01252
Genomic and proteomic profiles of biofilms on microplastics are decoupled from artificial surface properties ³⁷	Metagenome, Metaproteome	S17	Ocean	https://doi.org/10.1111/1462-2920.15531
Plastics select for distinct early colonizing microbial populations with reproducible traits across environmental gradients ³⁸	Metagenome	S18	Ocean	https://doi.org/10.1111/1462-2920.16391
Plastisphere showing unique microbiome and resistome different from activated sludge ³⁹	Metagenome	S19	Wastewater, Mesocosm	https://doi.org/10.1016/j.scitotenv.2022.158330
Viral diversity and potential environmental risk in microplastic at watershed scale: Evidence from metagenomic analysis of plastisphere ⁴⁰	Metagenome	S20	River	https://doi.org/10.1016/j.envint.2022.107146
The plastisphere microbiome in alpine soils alters the microbial genetic potential for plastic degradation and biogeochemical cycling ⁴¹	Metagenome	S21	Soil	https://doi.org/10.1016/j.ihazmat.2022.129941
Exploring the Composition and Functions of Plastic Microbiome Using Whole-Genome Sequencing ⁴²	Metagenome	S22	Ocean	https://doi.org/10.1021/acs.est.0c07952
Soil Type Driven Change in Microbial Community Affects Poly(butylene adipate-co-terephthalate) Degradation Potential ⁴³	Metagenome	S25	Soil	https://doi.org/10.1021/acs.est.0c04850
Landfill microbiome harbour plastic degrading genes: A metagenomic study of solid waste dumping site of Gujarat, India ⁴⁴	Metagenome	S34	Landfill	https://doi.org/10.1016/j.scitotenv.2021.146184
Elucidation of the biodegradation pathways of bis(2-hydroxyethyl) terephthalate and dimethyl terephthalate under anaerobic conditions revealed by enrichment culture and microbiome analysis ⁴⁵	Metagenome	S42	Wastewater	https://doi.org/10.1016/j.cej.2022.137916
Degradation of Recalcitrant Polyurethane and Xenobiotic Additives by a Selected Landfill Microbial Community and Its Biodegradative Potential Revealed by Proximity Ligation-Based Metagenomic Analysis ⁴⁶	Metagenome	S43	Landfill	https://doi.org/10.3389/fmicb.2019.02986
Soil plastispheres as hotspots of antibiotic resistance genes and potential pathogens ⁴⁷	Metagenome	S53	Soil	https://doi.org/10.1038/s41396-021-01103-9
Deciphering the role of polyethylene microplastics on antibiotic resistance genes and mobile genetic elements fate in sludge thermophilic anaerobic digestion process ⁴⁸	Metagenome	S57	Wastewater	https://doi.org/10.1016/j.cej.2022.139520
Marine biofilms constitute a bank of hidden microbial diversity and functional potential ⁴⁹	Metagenome	S62	Ocean	https://doi.org/10.1038/s41467-019-08463-z
Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach ⁵⁰	Metagenome	S71	Wastewater	https://doi.org/10.1007/s00253-016-7655-9
Novel bacterial taxa in a minimal lignocellulolytic consortium and their potential for lignin and plastics transformation ⁵¹	Metagenome	S80	Culture	https://doi.org/10.1038/s43705-022-00176-7
Metagenomes from WWTP and Wood degrading environments	Metagenome	S82	Wastewater, Wood	Samples from our work

2.2 Phylogenetic Distribution of Microbes in Plastic Environment

In this study, we recovered 4,708 metagenome assembled genomes (MAGs) which are widely distributed across the tree of life (Figure 2). Of these MAGs, 3,392 were not previously classified at the species level in GTDB⁵² (r207). The average CheckM⁵³ completeness and contamination of these assembled MAGs were 82.1% and 3.4%, respectively. Within this study we have also included a subset of the OceanDNA MAG catalog⁵⁴, including all the MAGs collected from the ocean biofilms of Zhang and colleagues,⁴⁹ as well as any other genomospecies in the catalog which were not otherwise represented in this dataset.

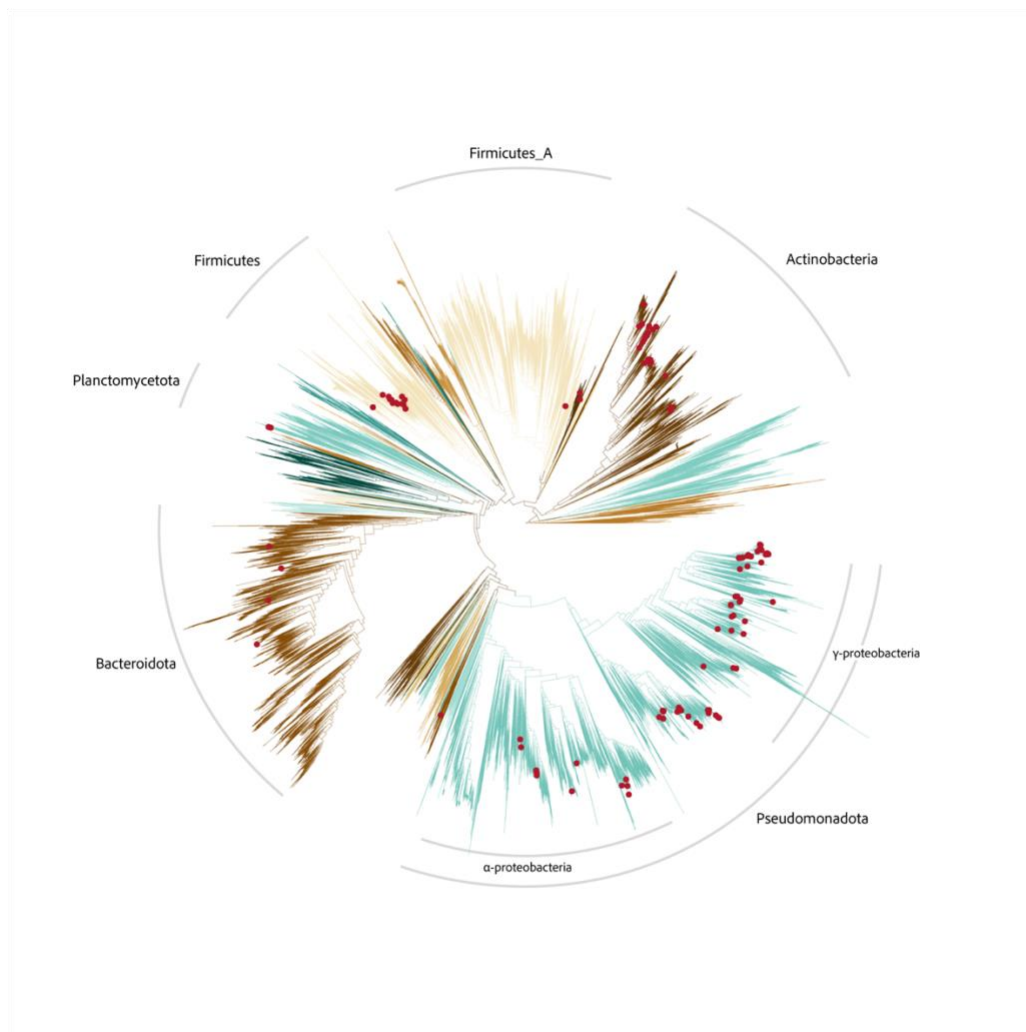


Figure 2 - Phylogenetic Tree of Bacterial Genomes in Current Dataset

The phylogenetic tree was built using GTDB-tk in de-novo mode; only branches corresponding to genomes in the current dataset are shown (as opposed to all genomes in the GTDB database). Branches are colored according to their assigned phylum. Selected phyla and classes of interest are annotated. Red dots indicate isolate genomes which either degraded plastic or were recovered from environmental plastic samples.

We additionally performed a literature search for microbes known to degrade or colonize plastics (isolated from environmental plastic samples, but without confirmed degradation activity). We collected a total of 166 prokaryotic genomes, 142 of which had confirmed degradation activity. Of these genomes, only 68 were previously reported in PlasticDB, reflecting the rapid discovery of plastic degraders in nature.

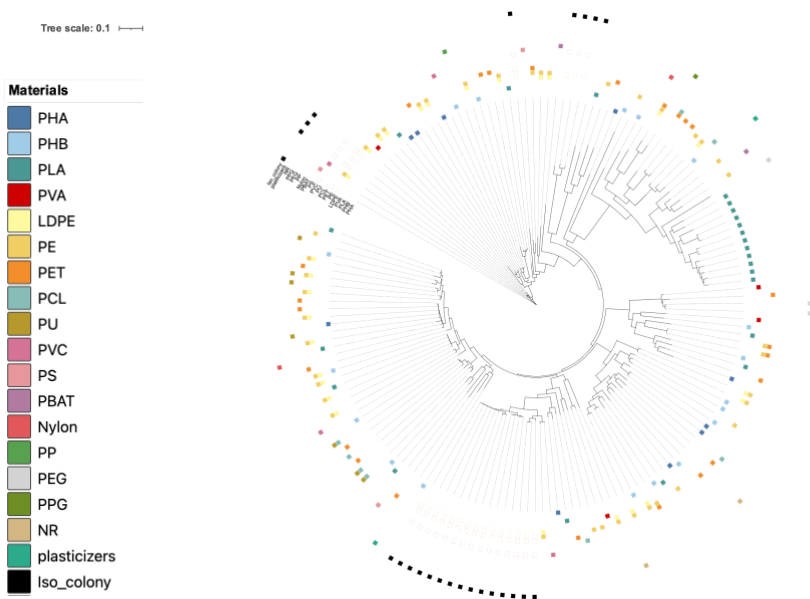


Figure 3 - Phylogenetic Tree of Plastic Associated Prokaryotes

Squares show the plastics reported to be degraded by the microbe. Black boxes represent isolates without confirmed degradation activity, the white boxes beneath these show the plastics from which this microbe was isolated.

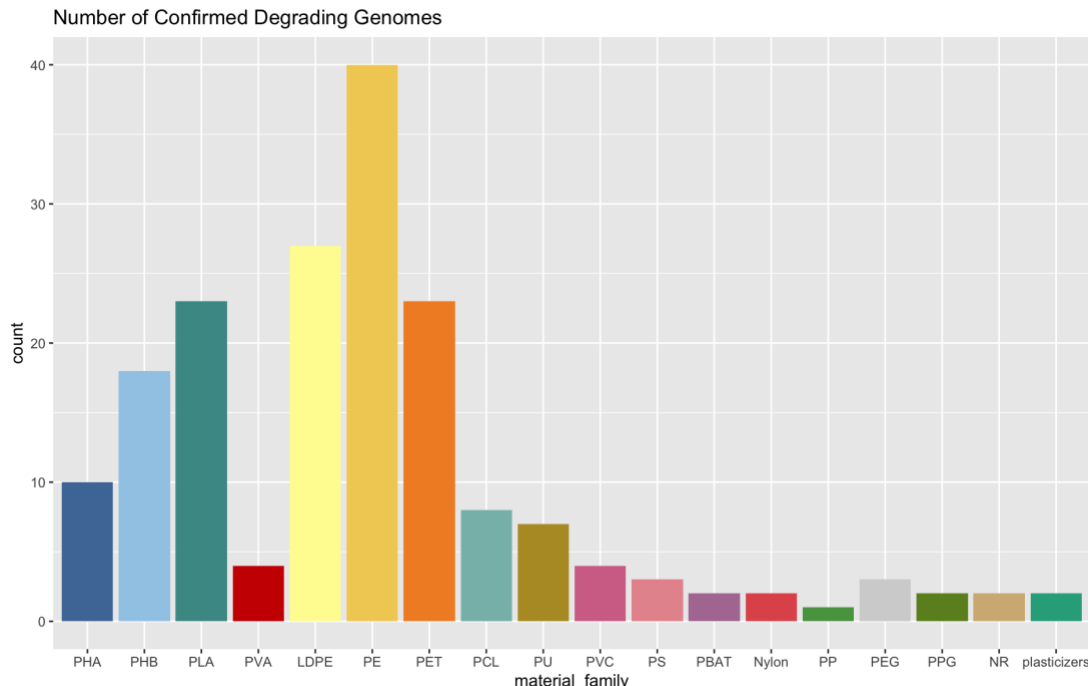


Figure 4 - Number of Confirmed Degrading Genomes per Plastic Type

The plastic with the most reported degradation activity was polyethylene, followed by poly(lactic acid) (PLA) and poly(ethylene terephthalate) (PET) (Figure 3). Polyethylene (PE) is largely considered a recalcitrant polymer⁵⁵; however this plastic had the widest phylogenetic distribution of reported degradation activity of any other plastic. Primarily reported were degraders of low-density polyethylene (LDPE), though there were some reports of degradation of high-density polyethylene (HDPE). There are still very few known enzymes with rapid degradation activity on polyethylene, the most rapidly degrading enzyme being from the recent waxworm study by Sanluis-Verdes and colleagues⁵⁶. Homologs to these two enzymes were not seen in any of these available bacterial genomes.

To note, this does not lead us to conclude that polyethylene is more biodegradable than more traditional biopolymers such as polyhydroxybutyrate (PHB) and polycaprolactone

(PCL), as the latter likely have had lower isolation efforts due to their known biodegradability⁵⁷. Biopolymers such as PHB and PCL also generally biodegraded rapidly in the isolation studies we observed.

Of the isolates with reported biodegradation of plastics, 70 were assigned to class *Gammaproteobacteria*. *Pseudomonas* (class *Gammaproteobacteria*) was by far the most observed genus of bacterial degrading species, containing 31 genomes with a wide variety of plastic degrading activity. *Pseudomonas* have long been well utilized for remediation of various xenobiotics⁵⁸, and are well represented in plastic degradation as well. The next most observed group with 21 degrading genomospecies was the *Burkholderiaceae* family (class *Betaproteobacteria*) – this group includes the notable *Ideonella sankesis*¹⁵ capable of PET degradation, as well as many isolates capable of PHA and PE degradation. *Actinomycetes* and *Bacilli* were also well represented within the degrading genomes. These groups are also well represented in the degradation of natural complex polymers, such as lignin⁵⁹.

Of the 16 plastic types which had multiple known degraders (polypropylene only had one), there were no material types which we observed to have monophyletic degradation activity. This suggests plastic degradation is not a lineage specific function, but likely evolves due to ecological selection in the environment and possibly horizontal gene transfer (HGT) of the selected degradation genes. Synthetic polymers have a wide variety of natural counterparts from a variety of environments, thus it is not surprising that species from many various areas are adapting to utilize plastics as a carbon

source. As the cleavage of plastics into short chain hydrocarbons can often occur by only a single or few genes, it is possible that HGT of these genes is likely frequent *in-situ*.

2.3 Ocean Results

The ocean is one of the major areas of interest for the effects of plastic, as it is a primary final location for mismanaged plastic waste⁶⁰, causing major environmental perturbations such as the Great Pacific Garbage Patch. This has also been one of the most deeply sequenced areas of the plastic environment, comprising over half of the available sequence data. Thus, we highlight some of the findings we had specifically for the ocean portion of the study below.

2.3.1 Beta Diversity

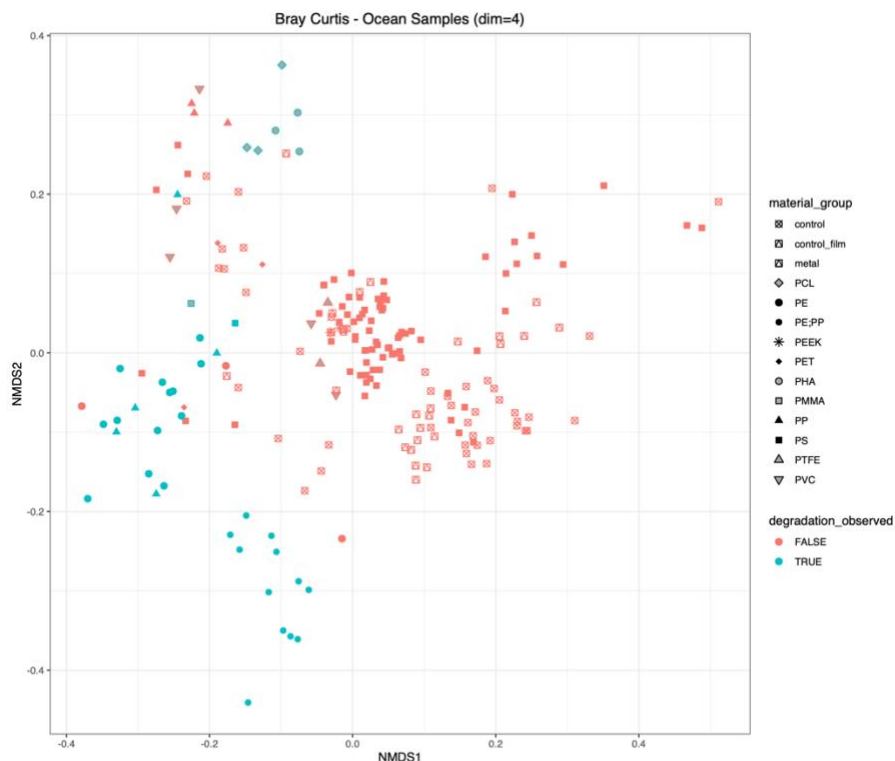


Figure 5 - Bray-Curtis Beta-Diversity of Environmental Ocean Samples

Blue points represent samples with degradation, while red points did not show degradation.

We calculated nucleotide level beta-diversity (i.e. how similar overall the communities are among each other) estimates among the ocean samples via Simka⁶¹, visualizing the resulting data using NMDS. We observed clear patterns of clustering for environmentally degraded plastics vs. non-degraded samples (Figure 5). Microbial communities on marine plastics otherwise largely clustered by material family and individual study. Metal biofilms and seawater control samples also clearly separated from plastic biofilms. Natural biofilms often were indistinguishable from polystyrene and other highly crystalline plastics, possibly connoting that microenvironments on non-biodegradable plastics may appear similar to other inert substrate materials in-situ.

Degraded plastics formed two clear groups, relating to polyolefins (PE, PP) and traditionally biodegradable polyesters (PCL, PHA). In the case of the PE and PP data, quantification of the microbial effect on, or rate of, the biodegradation was challenging to infer, as most of these samples were environmentally degraded without previous knowledge of the specifically microbial effect on biodegradation.

For the cluster of degraded PE and PP samples in figure 2, we assessed differential enrichment relative to non-degraded samples using ALDEx2⁶². We observed the differential enrichment of several genera; namely, *Rhodobacteraceae* genera *Roseovarius*, *Tateyamaria*, and *JABSSA01* along with several cyanobacterial taxa. *Roseovarius* is a group of chemoheteroorganotrophs⁶³ previously observed to be enriched on plastics both *in-situ* and under *laboratory* settings as a putative degrader⁶⁴. Genus *JABSSA01* is an uncultivated group seen previously during elevated carbon fluxes in the ocean⁶⁵. *Tateyamaria* is previously known in the context of breaking down compounds used as plastic additives⁶⁶. There are four *Rhodobacteraceae* species with complete genomes currently known to degrade plastics, including a PET degrader, as well as PHA and PLA degraders. These specific genomospecies were not differentially enriched in the present samples, however.

The most enriched genus within degraded PE and PP samples was the genus *Henriciella* (effect size 0.68, wilcoxon adj. p-value of 1.79E-04). This genus is a member of the family *Hyphomonadaceae* (class *Alphaproteobacteria*), and we note this as a putative plastic degrader due to its known ability to degrade hydrocarbons⁶⁷. We also observed that MAGs assigned to this genus across the study contained a significant

number of homologs of known plastic degrading genes. The *Henriciella* group contained the most homologs of any genus without confirmed plastic degrading isolates. Most genomes from this genus had a common pattern of sharing four homologs to known plastic degrading enzymes. There commonly was a gene showing 53% amino acid sequence identity to both PET esterase WP_085690612.1⁶⁸ and PLA degrading esterase AHG30919.1⁶⁹ in *Henriciella* MAGs; these two enzymes have 21% identity to each other. There also was commonly a homolog around 58% identity to PEG aldehyde dehydrogenase BAF98449.1, a gene characterized by Ohta and colleagues⁷⁰. Additionally common were homologs to PHA dehydrogenase from Lu and colleagues⁷¹ at 52% identity and PUR degrading protein ACD16728.1⁷² at 55% identity. Regions encoding these genes were generally spread out in different regions across the genome.

At the species level, there were 6 MAGs assigned to the *Henriciella* genus which were enriched (effect size > 0) at the species level, the most enriched being S07_18r1486_mtb_spa_t.17. All of these MAGs were unclassified at the species level in GTDB, and contained the genes described above. These MAGs and genes are available in the database connected to this study.

Table 2 - ALDEx2 Species Level Results - Henriciella Subset

ALDEx2 Species Level results for degraded PE & PP samples vs. all other ocean samples. Subset of only genomes from *Henriciella* genus are shown. Diff.btw – median centered-log ratio (clr) difference between groups, Diff.win – median of largest difference in clr values within groups, wi.eBH- Expected Benjamini-Hochberg corrected p-value of wilcoxon test.

species	diff.btw	diff.win	effect	genus	wi.eBH
s__unclassified S07_18r1486_mtb_spa_t.17	2.124	5.442	0.347	g__Henriciella	0.231
s__unclassified S07_18r1486_ros_spa_n.1	1.280	5.143	0.223	g__Henriciella	0.387
s__unclassified S18_41ao8643_mtb_idb_t.15	0.893	4.899	0.159	g__Henriciella	0.488
s__unclassified S18_44ar8646_mtb_spa_t.12	0.739	4.745	0.138	g__Henriciella	0.576
s__unclassified S18_30ad8632_mtb_spa_n.1	0.762	4.939	0.133	g__Henriciella	0.580
s__unclassified S07_23w1491_mtb_idb_n.8	0.117	4.497	0.023	g__Henriciella	0.767
s__unclassified S18_40an8642_mtb_idb_n.6_sub	0.001	4.537	0.000	g__Henriciella	0.764
s__unclassified S07_21u1489_mtb_idb_n.17	-0.018	4.595	-0.003	g__Henriciella	0.756
s__unclassified S18_44ar8646_mtb_spa_t.10	-0.152	4.457	-0.031	g__Henriciella	0.750
s__unclassified S03_8h9866_mtb_spa_t.8	-0.204	4.284	-0.042	g__Henriciella	0.746
s__unclassified S22_4d1151_mtb_idb_t.3	-0.205	4.273	-0.043	g__Henriciella	0.729
s__Henriciella sp013213825	-0.236	4.306	-0.047	g__Henriciella	0.734
s__Henriciella algicola	-0.281	4.255	-0.055	g__Henriciella	0.702
s__unclassified S09_2b3130_mtb_idb_t.12_sub	-0.308	4.306	-0.064	g__Henriciella	0.692
s__unclassified OceanDNA-b23505	-0.328	4.232	-0.070	g__Henriciella	0.718
s__unclassified OceanDNA-b23499	-0.360	4.272	-0.071	g__Henriciella	0.691
s__Henriciella sp002172915	-0.363	4.271	-0.076	g__Henriciella	0.682
s__unclassified OceanDNA-b23503	-0.580	4.349	-0.119	g__Henriciella	0.629
s__unclassified OceanDNA-b23517	-0.647	4.462	-0.127	g__Henriciella	0.623
s__unclassified OceanDNA-b23491	-0.684	4.488	-0.131	g__Henriciella	0.591

Table 3 - ALDEx2 Genus Level results for degraded PE & PP samples versus all other oceanic samples

Only significantly enriched groups (adjusted p-value less than 0.1) are shown. Abbreviations are identical to those in table 2.

genus	rab.all	diff.btw	diff.win	effect	wi.ep	wi.eBH	class	order	Family
g_Henriciella	2.012	4.343	5.480	0.712	8.50E-07	3.06E-04	c_Alphaproteobacteria	o_Caulobacteriales	f_Hyphomonadaceae
g_JABSSA01	0.845	3.208	5.722	0.531	2.35E-04	9.63E-03	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_Leptothoe	0.686	3.215	6.327	0.483	3.54E-03	4.96E-02	c_Cyanobacteria	o_Phormidesmiales	f_Phormidesmiaceae
g_Tateyamaria	0.777	2.783	6.163	0.418	9.36E-03	9.05E-02	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_Rivularia	0.615	2.777	6.563	0.415	1.02E-02	8.67E-02	c_Cyanobacteria	o_Cyanobacteriales	f_Nostocaceae
g_unclassified S03_3c9861_ros_idb_n.25_sub	0.528	2.368	5.241	0.404	6.05E-03	6.58E-02	c_Cyanobacteria	o_PCC-6307	f_unclassified S03_3c9861_ros_idb_n.25_sub
g_Phormidesmis	0.765	2.419	6.133	0.362	1.01E-02	9.31E-02	c_Cyanobacteria	o_Phormidesmiales	f_Phormidesmiaceae
g_MED-G52	1.358	-1.945	5.308	-0.342	1.63E-02	1.08E-01	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_UBA9145	1.366	-2.022	4.974	-0.358	1.49E-02	9.98E-02	c_Gammaproteobacteria	o_Pseudomonadales	f_Pseudohongiellaceae
g_UBA8309	1.412	-2.139	5.249	-0.367	1.00E-02	8.41E-02	c_Alphaproteobacteria	o_Puniceispirillales	f_Puniceispirillaceae
g_UBA724	1.378	-2.007	5.031	-0.367	1.25E-02	9.51E-02	c_Bacteroidia	o_Flavobacteriales	f_Flavobacteriaceae
g_R2A130	1.562	-2.179	5.281	-0.371	1.11E-02	9.49E-02	c_Alphaproteobacteria	o_Rhizobiales	f_Rhizobiaceae
g_Boseongicola	1.561	-2.213	5.154	-0.383	8.27E-03	7.71E-02	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_JABDJO01	1.515	-2.273	5.059	-0.394	6.91E-03	7.38E-02	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_UBA10364	1.497	-2.211	5.165	-0.396	7.15E-03	6.92E-02	c_Bacteroidia	o_Flavobacteriales	f_Schleiferiaceae
g_UBA1268	1.724	-2.301	4.928	-0.414	3.19E-03	4.23E-02	c_Planctomycetia	o_Pirellulales	f_UBA1268
g_Luminiphilus	1.578	-2.255	4.890	-0.416	5.37E-03	5.89E-02	c_Gammaproteobacteria	o_Pseudomonadales	f_Haliaceae
g_UBA11606	1.796	-2.264	4.670	-0.435	2.12E-03	3.31E-02	c_Acidimicrobiia	o_Acidimicrobiales	f_UBA11606
g_HIMB30	1.706	-2.497	5.127	-0.436	3.09E-03	4.05E-02	c_Gammaproteobacteria	o_Pseudomonadales	f_Litoricolaceae
g_Synechococcus_C	1.859	-2.779	5.679	-0.457	1.10E-03	2.37E-02	c_Cyanobacteria	o_PCC-6307	f_Cyanobiaceae
g_HIMB11	1.952	-2.673	4.840	-0.484	7.00E-04	1.71E-02	c_Alphaproteobacteria	o_Rhodobacterales	f_Rhodobacteraceae
g_UBA4421	2.234	-3.089	4.988	-0.554	1.14E-04	5.49E-03	c_Gammaproteobacteria	o_Pseudomonadales	f_HTCC2089
g_Synechococcus_E	2.878	-4.375	5.365	-0.730	3.10E-06	5.16E-04	c_Cyanobacteria	o_PCC-6307	f_Cyanobiaceae

For the datasets with observed plastic degradation, there was also a clear abundance of sunlight, seen in the location and depth of samples in the corresponding articles, as well as the abundance of *Cyanobacteria* in the datasets. It is challenging to ascertain the difference between UV (or other abiotic processes) degradation and microbial degradation, although the species reported here appear to have strong potential to do so. Further culturing and experimentation would be necessary to confirm these species as plastic degraders in the ocean environment.

2.4 *In-Situ* Distribution of Known Degrading Species

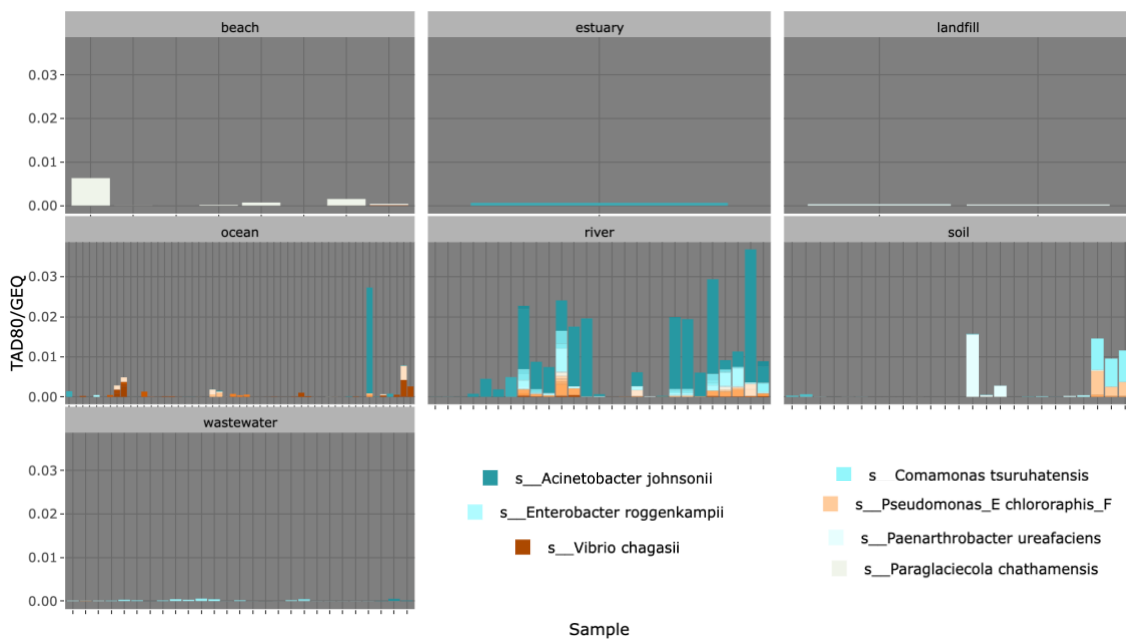


Figure 6 - Abundance of Known Plastic-Degrading Genomes Across the Plastisphere.

Abundances shown represent the 80% truncated average depth (TAD80) normalized by genomic equivalents (GEQ) in the sample. Only samples with a detect of a degrading species are shown. Colors corresponding to the most abundant species are shown in the figure key.

When looking across the environment for the genomes of species known to cause biodegradation, we observed a somewhat surprisingly low prevalence of these known degraders across the environmental metagenomes. Further, there was no correlation

between the available metadata on degraded plastics and the presence or relative abundance of these degraders in situ (Adonis2, method='bray'). This may connote that environmental biodegradation of plastics takes place via novel microbial lineages not yet cultured via laboratory methods. Surprisingly, however, we found an abundance of these known degraders in the only environmental river study available, produced by Li and colleagues⁴⁰. These samples had a strong enrichment of known degraders such as *Acinetobacter johnsonii* and *Comamonas testosteroni*. Despite the stark enrichment based on our analysis, the study by Li and colleagues did not include any metadata connoting that these plastics were biodegraded. The other two major river studies available, both by Wu and colleagues^{33,34} did show the presence of a few known degraders, such as *Azotobacter vinelandii* and *A. johnsonii*. However, they did not show enrichment of the wide variety of known degraders at the species or any higher taxonomic level as the study of Li and colleagues. These studies were produced using continuous flow bioreactors of river water, rather than in-situ incubations, which could have selected for different taxa.

2.4.1 KEGG Metabolism

We also observed the enrichment of many KEGG pathways in the river related to the degradation of various compounds associated with the degradation of plastic derivative compounds, such as styrene (map00643), caprolactam (map00930), and PAH degradation (map00624). This along with the genome level enrichment data may suggest that the river is primed for the utilization of many of the chemical compounds found in plastics.

The available plastic samples from wastewater, were primarily composed of polystyrene samples from Li and colleagues³⁹. These samples did not show as strong enrichment patterns as the riverine samples when compared to control wastewater samples. Instead, wastewater samples only showed the enrichment of a few potentially plastic-compound related pathways such as degradation of glycans and vitamin B6 production. The few wastewater samples for which degradation was reported did show additional enrichment for degradation of PAH and other aromatic pathways, as well as pathways for biofilm formation.

As the available soil samples primarily consisted of more readily degradable plastics (PBAT, PLA, PE), we expected to observe a greater number of KEGG pathways relating to biodegradation of related compounds in this habitat. PAH degradation and biofilm formation pathways were again enriched in plastic degradation associated soil samples. Increased in presence across all plastic samples in the soil vs. controls were also the degradation of steroids, caprolactam, and terpenoids.

We did not observe general enrichment patterns at the KEGG metabolism level for putative degradation pathways in the ocean environment, likely due to the wide variety of controls within this sample set, as well as the inadequate functional annotation present for many genes in this group. Across all other environments in the plastisphere vs. controls, we observed the enrichment in degradation pathways for various polymer related aromatics, such as caprolactam, toluene, and styrene. Additionally enriched were many amino acid production pathways, as well as radical generating pathways

such as cytochrome P450. Beta oxidation pathways and synthesis of polysaccharides was also commonly observed as abundant. These results suggest that in the environment, plastic may be utilized similarly to natural polymeric biofilms, selecting for microbes that are able to utilize their chemical components and integrate them into cellular biopolymers. Radical oxidation appears to be a widely used mechanism for utilization of and survival on these recalcitrant compounds. Notably, there is a lack of anaerobic samples within this dataset, thus these findings would obviously apply only to aerobic environments.

2.5 Gene Level Results

The known protein space for plastics in terms of available genes is relatively small - there are approximately 200 genes which are reported to degrade plastics in the PlasticDB¹⁷ and PAZy¹⁸ databases. We did find a number of recent studies claiming degradation via genes not present in either of these databases; genes from these studies will be included in a later version of our database after further screening.

The majority of known degrading genes are capable of degrading bioplastics such as PBAT, PLA, and PHB. Recent studies such as by Erickson and colleagues⁷³ have notably increased the number of enzymes and protein folds confirmed to degrade PET. Several studies have already reported specifically on the distribution of PET degrading plastics in the oceanic environment, which give further detail in the models and potential specific to the methods involved in this plastic⁷⁴. Plastics such as PE, PVC, and PP have relatively few known genes, and most of these are characterized by fairly low rates

of degradation. This is likely due to the high activation energy of the polyolefinic backbones of these plastics.

In order to observe patterns for these known degrading genes across the in-situ plastic environment, we opted to assess their relative abundance and enrichment in plastic-associated samples via an amino-acid identity clustering approach, detailed in the Materials and Methods section. This analysis allowed us to gather high confidence matches for these genes across the environment, and to observe whether these genes or their potential homologs were associated with biodegradation in the plastic environment.

2.5.1 In-Situ Distribution of Known Degrading Genes

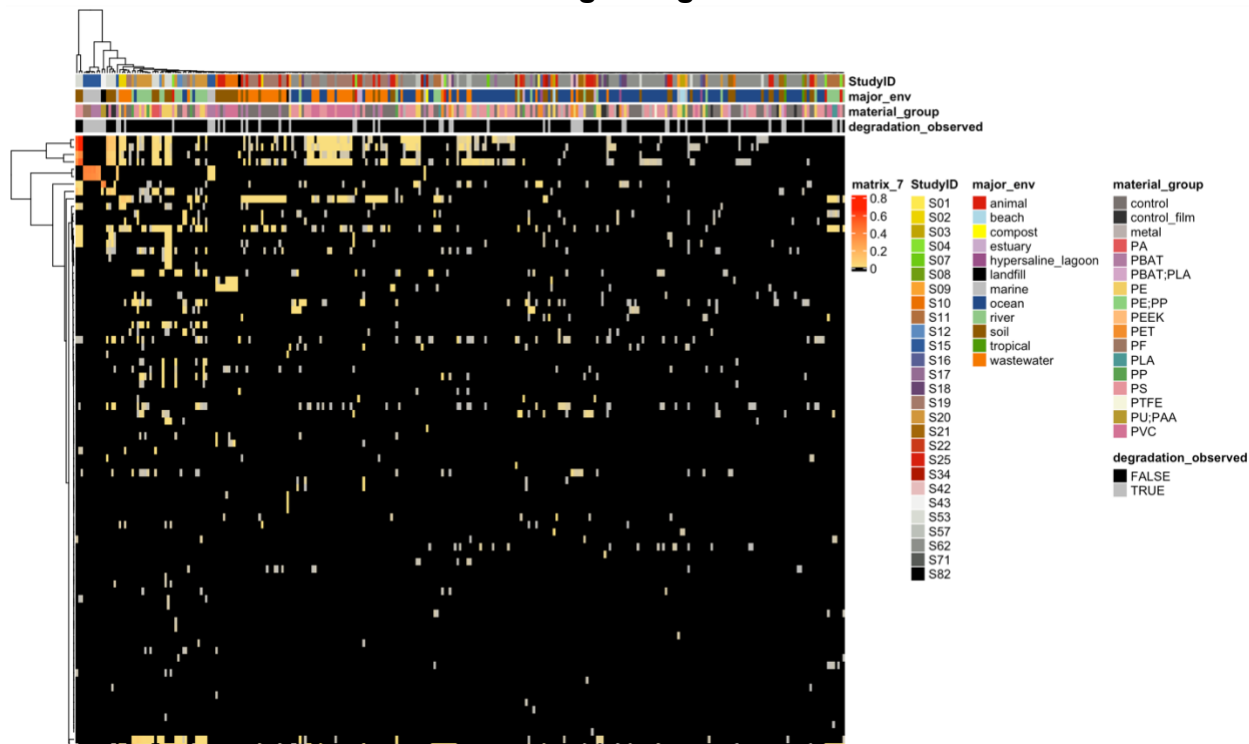


Figure 7 - Distribution of Known Plastic Degrading Genes in the Environment

Rows correspond to known genes, and columns correspond to samples. Gene presence was determined by truncated average depth values of metagenomes mapped to the plastisphere gene set, normalized by genomic equivalents. Only genes known to degrade plastic are shown. Heatmap bars contain metadata relating to the specific sample.

Although the known degrading genes did not follow the patterns of the available metadata, we expected that these genes would often appear in similar samples *in-situ*, as the corresponding populations which carry these genes were likely fulfilling a functional niche in the locations that they were observed. This was indeed the case, and we observed that these enzymes often had clusters of high Jaccard similarity, shown in figure 8. The clusters observed were not limited to enzymes discovered from the same microbial species, but often clustered by similar chemical features such as polyurethane (PUR) and nylon oligomers.

As the enzymes in the environment did not follow the available metadata on degraded samples, differential abundance was not considered as the primary strategy for discovery of enzymes which could be putative degraders of plastics. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)⁷⁵ has been a useful technique in dimensionality reduction and clustering of biological features in recent literature, having notable applications in spatial transcriptomics⁷⁶ as well as meta-omic data studies⁷⁷. UMAP is also well adapted for our specific dataset as many of the current correlation-based network methods are not capable of handling datasets as large as the current study, which consists of 10+ million unique protein sequences across over 400 samples. We therefore used UMAP to conduct a graph-network based analysis of the proteins across our study, selecting Jaccard similarity as the 'distance' metric. Our underlying hypothesis was that as the known degrading proteins are often observed in similar samples across the dataset, other proteins with comparable functionality would have similar observance patterns. This framework additionally

allowed us to integrate the data from other meta-omic datasets into our downstream analysis. Further details of the methodology and parameters we selected for UMAP can be found in the methods section.

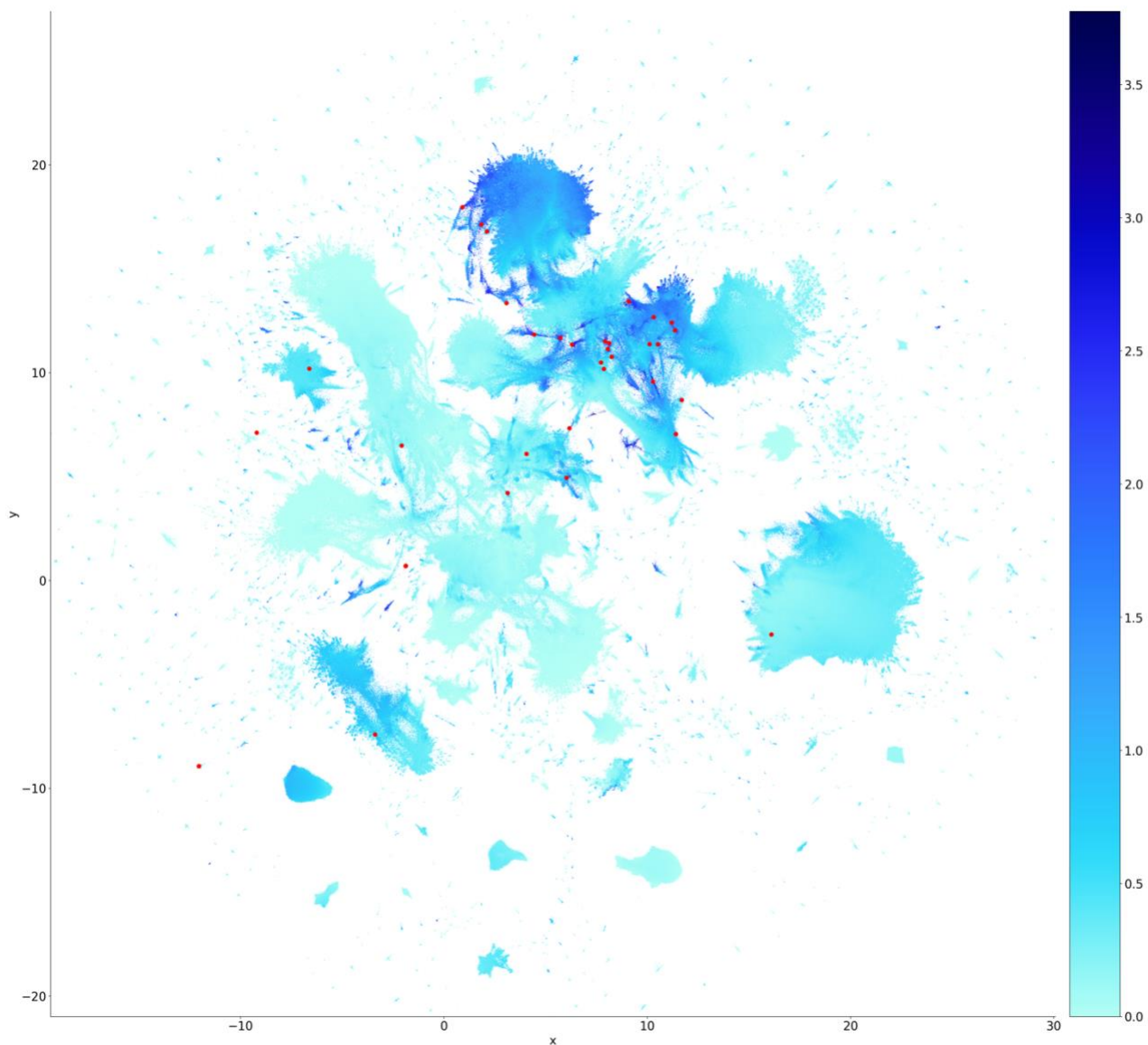


Figure 9 - UMAP Network Graph for Proteins Observed in Metagenomic 'Plastisphere'

Points are colored based on the summation of each gene's Bray-Curtis similarity to known plastic degrading genes in terms of presence/absence in the same samples (figure key on the right). Red dots are specific locations of proteins known to degrade plastic. Only genes present in at least 9 samples are present in the plot.

The known degrading proteins from the current datasets co-located into a relatively small region of the UMAP graph, connoting their strong similarity to each other in

observances across samples relative to the rest of the gene set present in the plastisphere. When observing the Bray-Curtis similarity of other genes in the dataset to the known degrading genes (figure 9), we observed that these genes were also commonly found in this same local space, confirming the successful embedding of these genes into this dimensionally reduced space, and indicating their putative association with plastic biodegradation. We additionally observed that the genes from the available proteomic studies consisting of proteins which were present on plastic samples were located in this same region. When analyzing the differentially expressed metatranscriptomes between PVC and PLA plastics, a notable portion of the most differentially expressed genes were also located in this same space. When we searched the annotations of the genes in this region, we found it was enriched with genes relating to oxidation and hydrolysis functionalities. Some of the genes of primary interest we found were WP_129973456.1 – a peroxidase found in *Pseudomonas* sp. B10 (GCF_004153525) – a species previously isolated for PET degradation⁷⁸, though its degradation pathway has not previously been described. This gene is a dye decolorizing type peroxidase (DyP) – this family of genes has been previously linked to promiscuous lignin degradation⁷⁹. This protein family, however, has to our knowledge never been tested for plastic degradation. 13 of 14 sample observances of this gene were on plastics samples, potentially connoting the environmental selection for this gene in putative degradation environments. An additional gene of interest from this space was S80_1a7091_mxb_fly_p.002~BEPAJJ_15120, carried by a MAG classified as *Ochrobactrum_B sp014138095*. This MAG was abundant in the minimal lignocellulolytic consortium constructed by Rodríguez and colleagues for lignin and

plastics transformation⁵¹. *Ochrobactrum* has previously been reported to degrade UV-treated low-density polyethylene; however, this study identified the corresponding organism(s) only at the 16S level⁸⁰. This gene is a superoxide dismutase, a gene previously observed to be abundant in proteomes of other plastic degraders⁸¹, however this gene has also not yet been specifically confirmed to degrade plastics. A final gene of interest we report is WP_021472099.1, a multicopper oxidase from *Paenarthrobacter ureafaciens* (GCF_002049485), a species previously reported to degrade nylon oligomers⁸². This gene was observed in 9 samples across the environment, primarily in the river and soil samples. This gene was also specifically associated with plastic samples, being only seen on samples containing plastic in the environment. These genes may be tested in the future to confirm their degradation activity on plastics. We expect that this UMAP space is highly enriched with novel genes which may perform plastic degradation activity, thus we have included this UMAP information in the database which we make publicly available through this study.

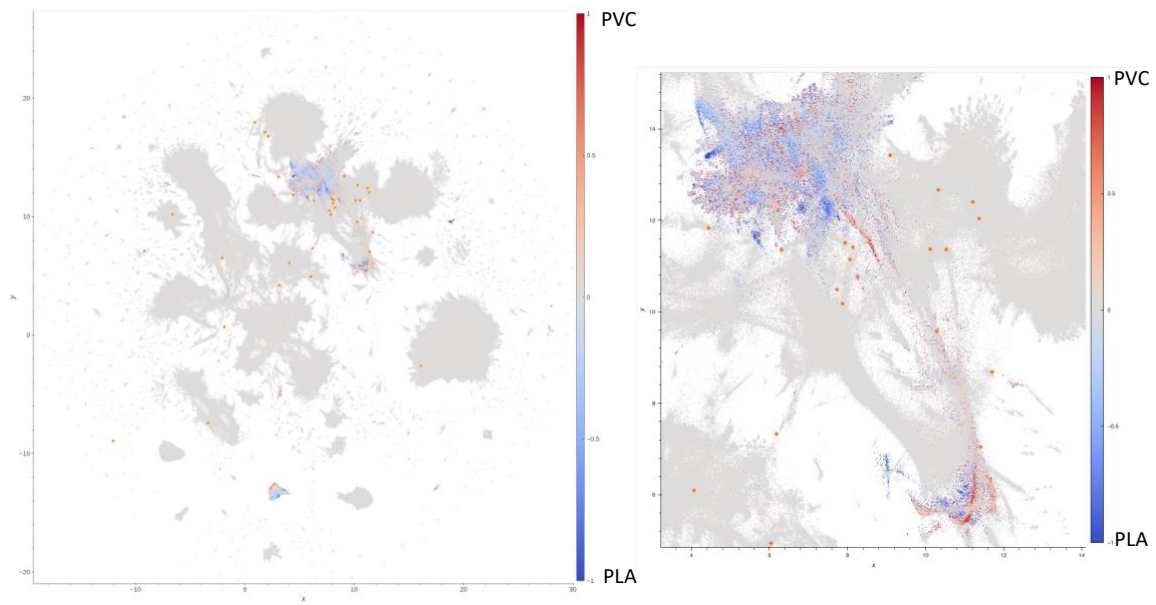


Figure 10 – UMAP Network Graph colored by Metatranscriptomic Enrichment from Wu and colleagues

Orange dots are specific locations of proteins known to degrade plastic. Colors are pseudo-log fold change (difference between) in polyvinyl chloride (PVC) versus polylactic acid (PLA) samples calculated via ALDEx2.

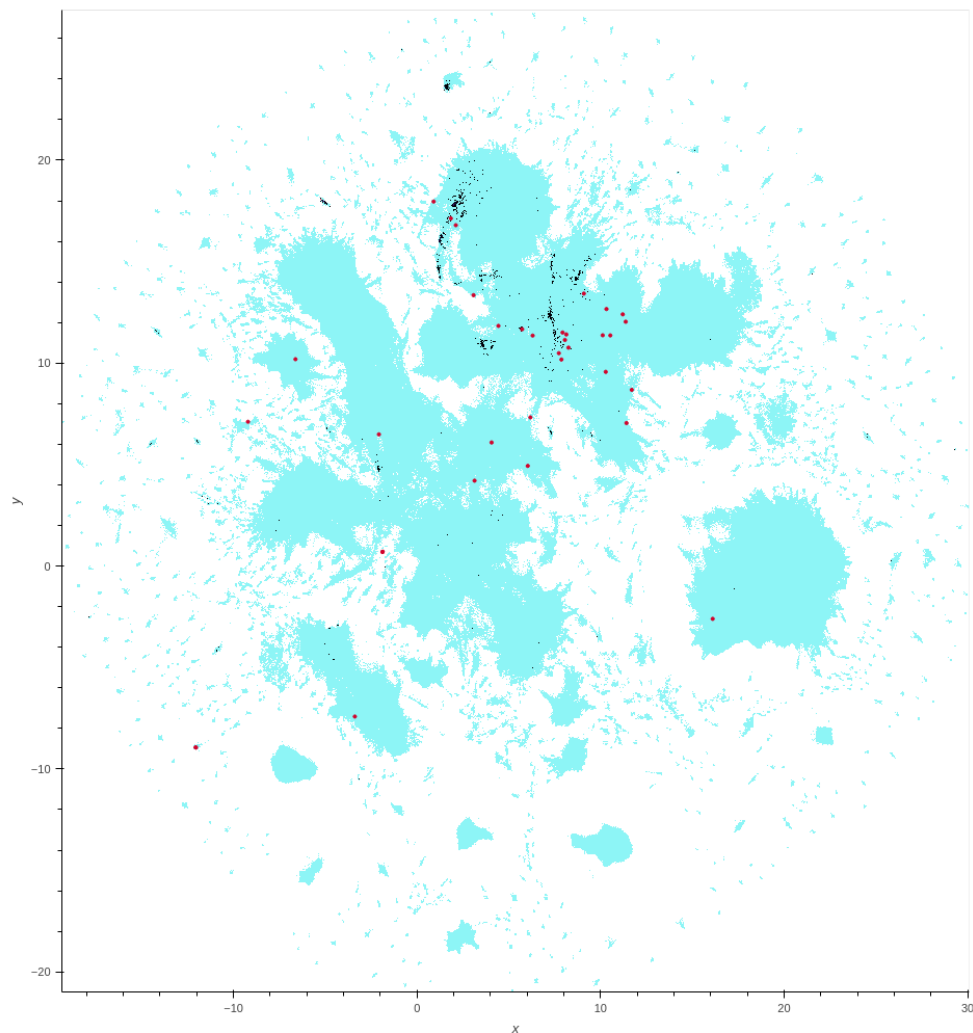


Figure 11 - UMAP Network Graph showing location of genes from other 'omics studies

Red dots are specific locations of proteins known to degrade plastic. Black dots are the locations of genes which were observed in transcriptomic, proteomic, or metaproteomic datasets. Of note is the dense regions of black points within the same region as many genes known to perform plastic biodegradation. Metatranscriptomic genes from Wu and colleagues (figure 10) are not specifically selected in this plot.

2.5.2 Disease Related Genes

Several recent studies have reported potential disease relatedness between plastics and pathogenic microbial species. The concern of plastic being associated with disease causing microbes has been previously connected to the ability of plastics to adsorb small molecules such as pharmaceuticals, including antimicrobials to its surface⁸³. As microplastics are released into the environment by a broad host of sources which are

not yet well managed⁸⁴, there is also a concern that microbial pathogens could also adhere to these materials and thus escape into the environment⁸⁵. Many of these recent studies focus on antibiotic resistance as a means of studying this area of concern. Antibiotic resistance shows the capacity of a microbe to resist common antibiotics known in the literature. An additional feature of interest would be virulence factors, which would connote the potential of a microbe to infect and spread within host cells. We therefore opted to utilize genes annotated via AMRFinderPlus⁸⁶ for antibiotic resistance gene (ARG) abundance, and the Virulence Factor Database⁸⁷ (VFDB) for genes relating to pathogenicity. For groupwide comparisons, we used permutational multivariate analysis of variance (PERMANOVA) via adonis2, and differential enrichment of specific genes was assessed using ALDEx2. The underlying expectation would be a greater relative abundance of these genes on metagenomic samples which contained plastic than in control samples. We did find small but significant differential enrichment of ARGs across the environment in plastic samples vs. controls (adonis2, $R^2=0.004$, $p=0.004$). Most notably, chloramphenicol resistance was specifically differentially enriched in plastic samples across the environment, beta-lactamase resistance genes were enriched in soils, as well as other enriched ARGs in wastewater and rivers (Table 4). The presence of plastics connotes the presence of human pollution, thus it is difficult to confirm plastics as a causative factor in this increased resistance rather than the outstanding increase in resistance due to common causes such as antimicrobial use by humans. In terms of pathogenic genes via VFDB, we observed no increase in the presence of these genes between plastic samples and controls in any of the major environments of the study. We note that these data do not

specifically deny the possibility of a public health risk, as metagenomic abundance data does not truly test the infectivity of a microbial population.

Table 4 - Differential Enrichment Results for Antibiotic Resistance Genes

Values reported are based on tests from either all samples (All) or the subset of samples from the specific environment noted. Only genes with significant positive enrichment ($p < 0.1$, effect > 0) are shown.

Gene	Difference between (pseudo-lfc)	wilcox adj. p value
All		
chloramphenicol efflux MFS transporter - cml	1.28	0.07
Soil		
β -lactamase oxacillinase - blaOXA	2.86	0.03
chloramphenicol efflux MFS transporter - cml	3.70	0.06
Wastewater		
organomercurial lyase - MerB	1.26	0.08
River		
tetracycline efflux MFS transporter - Tet(G)	4.12	0.08

2.6 Database

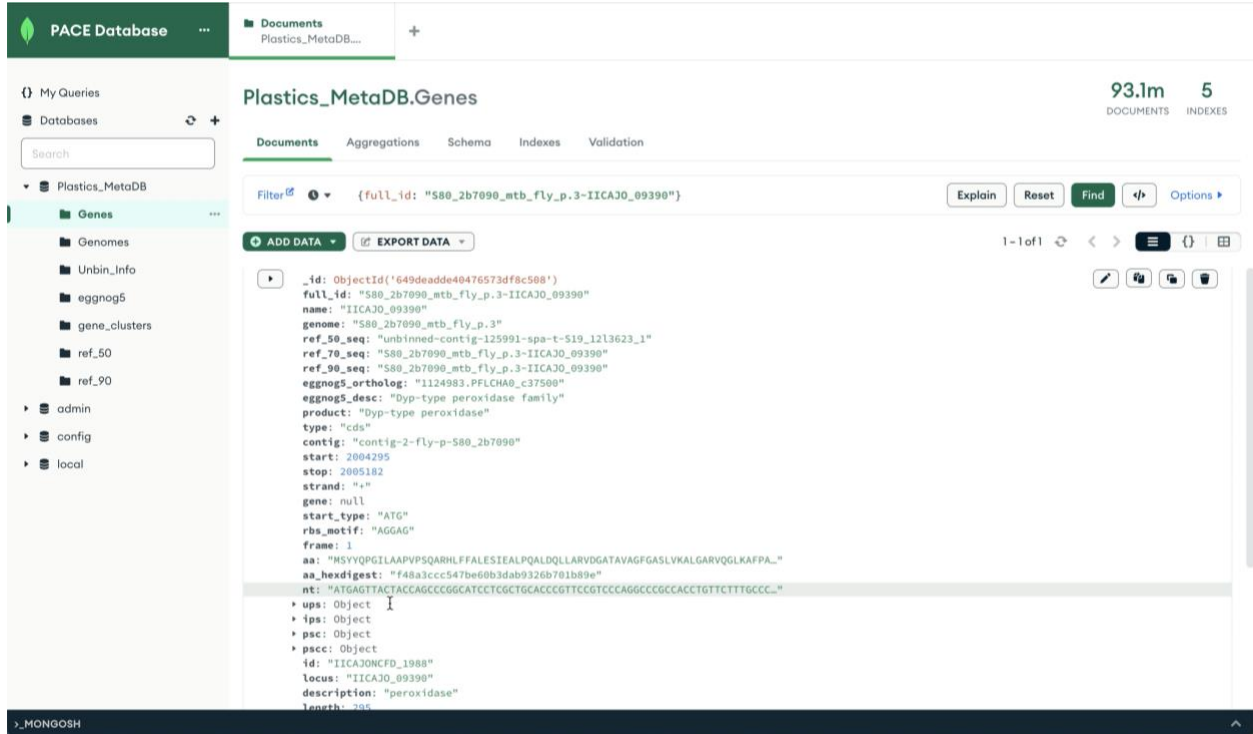


Figure 12 - Plastics Meta-omic Database (PMDB) Preview

Database preview of MongoDB based database of the gene and genome sequences, sample metadata, and other associated metadata produced by this study.

In order to make the data of this study accessible to researchers, we have compiled the datasets within an online database, the Plastics Meta-omic Database (PMDB). This database contains the sequences of all proteins and genomes used in the study, including rich annotations as well as environmental distributions.

Metadata for all metagenomic and metatranscriptomic samples contained within the dataset is included in the *Metadata* section of the database. Each record in this section contains the associated metadata manually compiled from the manuscripts themselves, as well as associated data in the sequence archives from which sequences were originally downloaded. This metadata includes details on the environment, material type, and whether degradation was observed in a particular sample.

Each genome within the *Genomes* section of PMDB is uniquely named based on its sample of origin using the study identifiers within Table 1. Genomes from external sources such as NCBI or the OceanDNA catalog retain their names from these original sources. Each genome record includes taxonomic classification and quality information from CheckM, in addition to whether the genomospecies is a known degrader of plastic. To search for genomes which were enriched in any of the subsets mentioned in this manuscript, we will also include a section *MetaG_Genomes*. This section includes abundance information for all genomes observed in the dataset, as well as differential enrichment information from ALDEx2 for selected environmental subsets of interest. The *Genes* section of the database contains the over 91 million unique proteins observed in the plastisphere. All proteins contained within a genome are linked by name to this genome, while unbinned proteins are linked to the sample from which they originate. Additionally included in the *Genes* section is all protein confirmed in the current dataset to degrade plastic – this subset is easily accessed by searching *is_known = True*. These known degrading proteins additionally contain citation information and details on the plastic type they degrade. Every protein within PMDB is fully text searchable by name as well as annotation information, allowing researchers to easily find all plastic-associated proteins which perform a specific function, such as hydrolases or polysaccharide lyases. Proteins are also searchable by enzyme families, such as Enzyme Commission (EC) numbers or Pfam family. Protein groups by 90%, 70%, and 50% sequence identity are also identifiable, including whether there is a protein known to degrade plastic within these limits of identity to any protein of interest.

In order to more directly search for proteins that have a high probability of degrading plastics, we additionally make available the UMAP embeddings of all proteins observed in at least 9 samples across the environment. These will be available in a separate section, *Protein - UMAP Network*. These embeddings are searchable by range of the graph, as well as by jaccard and bray-curtis similarity to known degrading proteins, in order to easily find proteins which were in the local protein space that was observed to be enriched as described previously. We additionally will provide access to an interactive graph of this UMAP space, so researchers may also browse this space directly. Metagenomic abundance information for each gene in the *MetaG_Proteins* also will be included, with details on the metadata associated with the samples in which the protein was observed. Metatranscriptomic abundance information for genes from Wu and colleagues will be available in the *MetaT_Proteins* section.

We additionally plan to make available the 90% amino acid identity dereplicated plastisphere protein set searchable by Diamond or BLAST. Researchers will be able query their own sequences against PMDB to ascertain where a protein has been observed *in-situ*, as well its location in the protein UMAP network space. This will allow scientists to quickly gather detailed information on their sequence of interest, as well as whether a protein has high likelihood of being capable of degrading plastic based on the meta-omic datasets available in the literature.

The data within PMDB is downloadable in JSON or tabular format for entire sections or search results of a particular query, through the MongoDB framework backend by which the database is built.

CHAPTER 3: Discussion

Plastics are xenobiotic materials which have been dispersed across the environment by mismanaged waste streams and uncaptured micro and nanoplastics. In this study, we have conducted the largest analysis to date of meta-omic data associated with plastics, comprising the in-situ microbial response to this environmental pollutant.

Microbes capable of utilizing these compounds are phylogenetically diverse, however *Pseudomonas* and *Burkholderiaceae* groups notably include many plastic degrading species. These known degraders were found to be very sparse in the environmental datasets analyzed here, however. The river was the main location we found known degrading species and enzymes to be relatively abundant across a wide variety of plastic types. Why would riverine systems be enriched with these known degrading species? Rivers are naturally eutrophic environments, containing many lignocellulosic compounds from plant sources. These compounds are chemically similar to plastics, thus adaptation to these new synthetic carbon sources with natural homologs would not be evolutionarily difficult. Additionally, rivers are a major acceptor of land runoff, as well as anthropogenic pollution such as microplastics when located near densely populated areas. Thus, one of the major places which plastics have likely been long available for microbes to adapt to is in riverine systems. There is also generally bioavailable oxygen in this habitat, allowing for ready incorporation of oxygen onto plastics by radical or enzymatic oxygenation. Somewhat surprisingly, there were few isolation studies which looked to specifically isolate plastic degraders from the river. Thus, we recommend this

as a habitat/type of environment from which further exploration for plastic degraders may be undertaken.

The ocean environment has been a deeply sequenced habitat of the plastisphere. We observed clear distinctions between plastics which had degradation and other samples that did not report plastic degradation but only plastic presence via NMDS, and further distinction between those plastics and free-living seawater samples. Natural biofilms did not strongly separate from non-degraded hard plastics, possibly indicating that these substrates may be treated similarly in-situ. We observed the genus *Henriciella* to be a group of high interest for future isolation and plastic degradation studies in the marine environment, based on its differential enrichment in degraded samples along with many homologs to known plastic degrading genes present in the *Henriciella* MAGs recovered here. *Henriciella* and these various taxa previously related to the flux of carbon in the environment may be essential pieces to understanding the biotransformation of these synthetic polymers. Isolation and characterization of these microbes may open new insights into the ability of microbes to incorporate synthetic xenobiotics back into the global carbon cycle. Omics-based studies that target the activity (not only presence and relative abundance) may also be useful to elucidate these mechanisms.

Metatranscriptomic and proteomic comparisons of these microbial communities between microplastics, cellulosic compounds such as laminarin, and controls may better elucidate how these microbes specifically respond to plastics. Additionally, metabolomic and lipidomic studies may allow us to see the transformation and incorporation of these compounds into microbial systems.

Microbial data on plastics in other environments consistently showed an enrichment for KEGG pathways corresponding to the utilization of plastic compounds. Present pathways allowed for the overall putative mechanism of radical oxidation, followed by degradation into compounds capable of entering beta-oxidation and subsequent transformation via common microbial pathways into amino acids, polysaccharides, and other forms of carbon. This schema is common across the plastic biodegradation literature and appears to be generally applicable across the observed metagenomic community as well. We would encourage follow up studies specifically observing for the generation of reactive oxygen species (ROS) by microbes, in order to ascertain whether this is a mechanism utilized *in-situ*. Studies have previously used UV pre-treated plastics to introduce radicals and subsequently oxygen onto the surface of plastics; this has been seen to increase degradation rates, but not to levels desirable for 'biodegradable' plastics⁸⁸. The ability of microbes to control the use of ROS as a method of biotransformation of synthetic polymers is an area of interest for further study. This may be completed by direct probing of ROS concentration in plastic-containing cultures, and comparing this to concentrations in the presence of natural lignocellulosic biopolymers, or benign substrates such as rocks.

As noted previously, there is also very few anaerobic samples within this dataset, limiting the application of these data to aerobic environments. The response of microbes to lignocellulosic compounds is largely different based on the presence of oxygen, with single enzymes being deployed to degrade these compounds when oxygen is available. However, in anaerobic environments, large multi-enzyme complexes such as the

cellulosome have been observed⁸⁹. It is possible these differences in deconstruction methods may be present for plastics as well.

Enzymes capable of degrading plastics present themselves as one of the most promising options for bioremediation and valorization of these materials. Enzymes known to degrade plastics appeared sparsely in the available metagenomic dataset, with no correlation to samples metadata relating to plastic or biodegradation. These enzymes however did appear to correlate with one another in observances relating generally to material families. The known space of plastic degrading genes is fairly small, comprising approximately 200 enzymes. As neither degrading genes or genomes *in-situ* corresponded to the available metadata, we expect that the microbial populations and enzymes performing biotransformation of plastics *in-situ* have largely not yet been reported and may be distinct from laboratory methods.

In order to search the available protein space for putative plastic degraders, we do expect that previously described enzymes are the best 'hook' for finding novel genes capable of activating the carbon-carbon and carbon-heteroatom backbones of synthetic polymers. UMAP embeddings showed the known degrading genes to localize to a dimensionally reduced protein space, connoting their strong similarity in terms of observance among all the genes in the plastisphere protein space. Proteins which were enriched in other forms of meta-omics also were frequently observed in this same region. We therefore predict this space to be enriched with novel degrading genes capable of performing biodegradation of plastics, and encourage researchers to utilize proteins from this dataset for further testing and study.

Within the available data, a common limitation for the studies was the lack of additional metadata, including additive information, environmental metadata such as temperature and humidity, and degradation information via robust testing methods, such as XPS and GPC. Additionally, data such as microbial loads for true quantification of microbial absolute abundances limit the power of the statistical methods utilized by this study. We additionally note the lack of meta-transcriptomic and meta-proteomic studies available, which would further expand our understanding of expression resolved in-situ responses to plastic at the enzymatic level. We encourage further datasets that include these forms of data, to allow greater granularity on biological responses to plastics.

We observed enrichment for known degraders in riverine environments, which are known to be locations containing many human pollutants and similar polymeric compounds. 16S data has been gathered for other locations which may also contain many of these compounds, such as landfills or insect guts. This data however is fairly sparse at the whole-genome level. Increasing sequencing from these locations will add further granularity into the ability of plastics to be biodegraded in these habitats. This may also allow for more in-depth comparison of the metabolic responses to plastic in locations of enriched degradation-associated microbial communities. Polymeric materials additionally vary in degradability based on crystallinity, molecular weight, and chemical composition, alongside other factors. Studying these factors plays an important role in understanding degradation mechanisms for specific polymers. We also again note the difficulty of differentiating biodegradation from UV or mechanical

degradation. Efforts to perform in-situ experimentation controlling for these effects may give greater detail into the mechanism by which microbes respond to plastics in our environment.

The current manuscript has covered the major trends and observances we see across the plastic environment. Additionally of interest are changes in the community across various polymer types and local habitats. Further in-depth study may be performed in terms of how the physicochemical characteristics of polymers effect the microbial community, as well as the various within-environment factors which may shape the ability of microbes in these habitats to degrade plastic. This data is available in the current dataset and connected database, and we encourage researchers to access these for further exploration.

CHAPTER 4: Materials and Methods

Data Curation

In order to collect all currently available meta-omic studies relating to plastics, terms involving various major plastics and meta-omics (metagenomics, metatranscriptomics, etc.) were searched on Web of Knowledge and Google Scholar. Only papers containing publicly available data were retained. Similar search terms were used for isolate studies. Previously curated databases PlasticDB¹⁷ and PAZy¹⁸ were also parsed for studies containing complete genomes, as well as known enzymes to perform biodegradation. 16S rRNA gene (16S) data was not considered for this study, as previous papers have described this data in detail¹⁹.

For each paper retained, metadata was collected in regards to geolocation, environment, and degradation information. Material information including material type, size, and general polymer family was also collected. Degradation was considered to be present based on the available information within the corresponding manuscripts. If degradation was not specifically reported in the paper, plastics were considered to be degraded if the collected samples showed strong signs of degradation via oxidation or physical breakdown, or were otherwise known biodegradable plastics kept for long incubation periods. Degradation levels were categorized by the information in supplemental table 4. Samples without degradation specifically reported were noted as degradation level 0. Each study and sample was given a unique identifier, which were utilized for subsequent analysis. Study and sample metadata is available at the github link provided, as well as the database provided by this paper.

Bioinformatics Pipeline for Data Processing

Metagenomic samples were processed using a custom pipeline developed using snakemake. Briefly, samples were trimmed using fastp with default settings, and additionally normalized using BBNorm. For studies containing paired end reads, both trimmed and normalized libraries were assembled using metaSPAdes⁹⁰ and IDBA-UD⁹¹. For large and high complexity samples, MEGAHIT⁹² was used as a second assembler instead of IDBA-UD. For single-end samples, IDBA-UD and MEGAHIT were used as assemblers. For long-read samples, metaFlye⁹³ was used as an assembler. Resulting assembled contigs were filtered with a minimum length of 1 kb, and assembly statistics were collected using MetaQuast⁹⁴. The trimmed reads were mapped to the assemblies using bwa-mem2⁹⁵ and contig depths were collected using CoverM with method 'metabat', and otherwise default settings. Each of the four resultant assemblies was binned using MaxBin2⁹⁶, MetaBAT2⁹⁷, and Rosella. For long-read studies, GraphMB⁹⁸ was used as a fourth binner. Bin qualities were assessed using CheckM⁵³ and CheckM2⁹⁹. Bins were dereplicated at the sample level using dRep¹⁰⁰ with ANI level 95.0, S_algorithm method fastANI¹⁰¹ minimum completeness of 50 and maximum contamination of 10. Genome statistics were collected using SeqKit¹⁰². Dereplicated bins were annotated using Bakta (version 1.8.1, database version 5.0)¹⁰³, eggNOG-mapper¹⁰⁴, CAZy²³ via dbCAN2¹⁰⁵, and KEGG¹⁰⁶ via KofamScan¹⁰⁷. MAG taxonomy was assessed using GTDB-tk v2.2.1¹⁰⁸.

Diversity information for each sample was also estimated: alpha-diversity was assessed using Nonpareil¹⁰⁹, and taxonomic information was collected using Kraken¹¹⁰ and

Bracken¹¹¹. Genomic equivalents in each sample were estimated using MicrobeCensus¹¹² with parameters $n = 100,000,000$ and $q = 10$.

Pipelines used and other supplemental data aforementioned are available on github at the link https://github.com/Rridley7/Plastic_assc_info.

For samples from the biofilm study by Zhang and colleagues⁴⁹, MAGs from the OceanDNA⁵⁴ database were collected. Additionally included in the dataset were species representative genomes from OceanDNA which did not already have a same-species-representative in the dataset, based on a 95% nucleotide sequence identity threshold as computed by FastANI. For genomic equivalents in these metagenomic samples, reads were trimmed and assessed with the same tools and parameters as in the pipeline.

For dereplication of MAGs and isolate genomes of the entire study, dRep was used with the same parameters as previously described, and sample quality information from CheckM. MAGs which passed CheckM2 but did not pass CheckM were also retained after manual checks and dereplication using skani¹¹³.

Beta-diversity was assessed for metagenomic reads using Simka⁶¹ with default settings, which calculates beta-diversity using nucleotide kmer diversity. Non-metric multidimensional scaling (NMDS) of the resulting bray-curtis distances was assessed using metaMDS from the vegan package and visualized using ggplot2¹¹⁴.

All vs. all genome comparisons across the study were completed at the ANI and AAI level using fastANI¹⁰¹ and fastAAI¹¹⁵ respectively.

Genome and Gene Mapping

Genomic and gene level abundances in the environment were assessed by mapping metagenomic reads back to non-redundant genome and gene sets. The dereplicated genome set described previously was used for the genome level abundances.

Additionally, the species representatives from the OceanDNA set were utilized.

For gene level abundances, both genes carried by genomes and assembled but unbinned genes were considered. Briefly, 95% ANI genomospecies clusters from dRep¹⁰⁰ were clustered using Roary¹¹⁶ with default settings. Additionally, known genes from PAZy and PlasticDB with nucleotide sequences available were added to this database of binned genes. This database was clustered using MMSeqs2¹¹⁷ at 99.9% nucleotide identity and coverage to remove duplicate genes.

To collect unbinned genes, contigs were taken from the trimmed metaSPAdes, IDBA-UD, or metaFlye assembly for each sample. Genes from all contigs were predicted using Pyrodigal¹¹⁸, a python library binding to Prodigal¹¹⁹. Genes were subsequently dereplicated using MMSeqs2 at 99.9% nucleotide identity and coverage. This gene set was mapped to the binned genes using minimap2¹²⁰ using setting “—for-only”. Genes mapping to the binned set with nucleotide identity >95% and coverage >98% were removed from further analysis as redundant with the binned gene list. The remaining

unmapped genes were subsequently clustered at 95% nucleotide identity and 98% alignment length using mmseqs2, to produce a non-redundant unbinned gene set. Unbinned genes were annotated using the same tools as described in the pipeline.

Reads from all samples were mapped to genomes and genes using bwa-mem2 and CoverM using nucleotide identity >95 %, read alignment >70%, and covered fraction >10% as mapping thresholds. Minimap2 was used for long reads with the same settings. For the gene mapping, an iterative subtractive mapping approach was used. Briefly, reads were first mapped to binned genes, then unbinned reads were collected using SAMtools¹²¹. These unmapped reads were then mapped to the unbinned genes using the same parameters as previously described. The mapping pipeline is also available on github at the same link as above.

Abundance at the genome level was assessed using the truncated average depth at 80% (TAD80) metric, normalized by the genomic equivalents (GEQ) estimate from MicrobeCensus within a given sample. Gene abundances were assessed using a less stringent TAD of 90% normalized by GEQ. This metric gives similar results to transcripts per million (TPM), however GEQ accounts for more directly for the number of genomes in a metagenomic sample, when external microbial loads are not available.

Genome Relative Abundance and Phylogenetic Analysis

For genomic abundance, ALDEx2¹²² was used on samples from individual environmental subsets, to assess differential enrichment accounting for data

compositionality. Heatmaps of relative abundance were produced using ComplexHeatmap¹²³, and barplots were produced using ggplot2.

Phylogenetic trees were produced via two methods. A tree comparing the genomes from the current study to the current prokaryotic tree of life was produced using GTDB-tk *in de novo* mode. Trees involving only genomes from the current dataset were produced using PhyloPhlan¹²⁴. Trees were annotated and visualized using ggtree¹²⁵ and iTOL¹²⁶.

Gene Analysis

For analysis at the gene level, genes were clustered into high identity gene families subsequent to mapping. The gene set containing all genes from genomes in the plastic environment, known genes, and unbinned genes was collected, consisting of 92,930,684 protein sequences. These genes were clustered using MMseqs2 at the 90%, 70% and 50% amino-acid identity levels. 90% and 70% identity clustered required 80% coverage, while 50% was reduced to 70% coverage for lower stringency. Parameters `--cov-mode 1 --cluster-mode 2 --cluster-reassign` were used for all clustering. The procedure was completed in a cascading fashion, with sequences unique at higher levels being given as input to subsequent clusterings. Gene cluster statistics were then collected, giving note to which clusters contained known degrading genes or genes from higher meta-omic (transcriptomic, proteomic) datasets. Gene statistics subsequent processing of the dataset were completed primarily using Dask¹²⁷ for multi-threaded and larger-than-memory processing. Abundances for each gene

cluster were considered by using the sum of abundances of the gene assigned to the cluster. Individual gene abundances were calculated by read mapping, using 10% trimmed mean average depth (TAD90), which accounts for gene length and edge effects when mapping to short genes, as described above.

KEGG Pathway Analysis

For KEGG analysis, the top match from KofamScan annotations were used for the non-clustered dataset, with a minimum e-value threshold of $1e-5$. For genes without a KofamScan annotation, KEGG annotations were collected using KEGG modules provided through UniProt annotations via Bakta. Genes still without annotation following the latter step were not included in subsequent analysis. Genes were summed in KEGG modules using TAD90 for abundances. These abundances were subsequently passed to ALDEx2 within subsets based on environment, using 'lvha' as the denominator. Results from ALDEx2 were sorted by effect size, and passed to clusterProfiler¹²⁸ using method gseKEGG with parameters $pvalueCutoff = 0.05$, $nPermSimple = 10000$, and $eps=0$. KEGG results are also available at the Github link provided.

UMAP

Uniform Manifold Approximation and Projection⁷⁵ (UMAP) embeddings were produced via use of the 90% level clustered gene set, using all gene clusters observed in a minimum of 9 samples, to avoid spurious correlations. The jaccard distance metric was used for all runs; other metrics were selected via manual tuning. The main graphs presented in this manuscript used metrics $n_neighbors=20$ and $min_dist=0.3$. UMAP data was visualized using hvplot and datashader¹²⁹, using bokeh as a backend

framework¹³⁰. Bray-curtis similarities of metagenomic abundances to known plastic degrading enzymes were calculated using Dask and scipy¹³¹. Values reported within figures are summed cumulatively across all known plastizymes for each gene.

Virulence and AMR Genes

For antibiotic resistance and virulence related genes, gene annotations for the non-redundant gene dataset were collected using AMRFinderPlus⁸⁶ and VFDB⁸⁷. Genes from across the dataset were annotated by Bakta as previously described – the Bakta program reports annotations from each of the above programs within its output. Gene abundances from all genes matching a specific annotation were summed as previously described. Group-wise comparisons were assessed using adonis2, and differential enrichment of specific genes was tested using ALDEx2. Comparisons which were for plastics across all environments used denom='zero', while individual environments used denom='iqlr'. Heatmaps were generated using ComplexHeatmap in R.

Other Meta-omic Datasets

Data from transcriptomic and metaproteomic datasets were included in the non-redundant protein set. Genes from these sets were given numeric rankings based on the study type (proteome given higher ranking than transcriptome), and whether differential enrichment was observed in the dataset. Genes from these studies were annotated using the same methods as described previously.

For data from the meta-transcriptomic dataset published by Wu and colleagues³³, reads were first trimmed using fastp. Trimmed reads were then sorted using SortMeRNA v.

2.1¹³², to retain non-ribosomal reads. Non-ribosomal reads were mapped to a non-redundant metagenomic gene set using coverm and bwa-mem2 using the same parameters as previously described. The non-redundant gene set was produced using metaSPAdes assembled contigs from the pipeline, dereplicated at 95% nucleotide identity and 98% coverage using MMSeqs2. Differential abundance of mapped transcripts was assessed using ALDEx2¹²², and plots were produced using ggplot2.

Database

Study metadata, genomes, genes, and UMAP data were collected into a database using the MongoDB framework. This database will be soon made available to the public for browsing and searching user's datasets against the known meta-omic data from environmental studies. There are several sections available in the database, corresponding to gene annotations, genomes, gene clusters, and study metadata previously described (Supplemental Figure 5). Sections in the database are fully searchable by annotation, name, and sequence information. Sample metadata and genome sequences will also be separately made available for access.

Metagenome Sampling and DNA Extraction

The return-activated sludge was obtained from a wastewater treatment plant in the Atlanta Metropolitan area, Georgia, USA. Around 1 L of sludge was put into a sterile 1 L glass bottle and was transported in ice to the lab. 50 mL of sludge was aliquoted to a 50 mL conical tube and centrifuged at 4 °C, 5000 g for 10 minutes. The resulting supernatant was thrown away and the pellet was used for extraction immediately. The

degraded wood was sampled from a dead and decaying broad-leaved tree in the Atlanta Metropolitan area, Georgia, USA in the winter. The inside of the tree was scooped with a sterile 50 mL conical tube and the lid was closed immediately. The sample was stored at 4 °C until the extraction. The DNA was extracted with the Qiagen DNeasy PowerMax Soil kit following the manufacturer's protocol. The quality of the extracted DNA was analyzed with Thermo Fisher Scientific NanoDrop 1000 Spectrophotometer and the quantity was accessed with Invitrogen Qubit 1X dsDNA HS assay kit and Invitrogen Qubit 3.0 Fluorometer. The checked DNA was stored at -20 °C until the sequencing.

The library preparation and sequencing were performed by the Georgia Genomics and Bioinformatics Core (GGBC) (Athens, Georgia, USA). The DNA extracts were sent to GGBC where Pacific Biosciences (PacBio) Single Molecule, Real-Time (SMRT) bell multiplex library was constructed without the shearing step and was sequenced with a single PacBio SMRT Cell on the Sequel II system.

REFERENCES

1. OECD. Plastics use in 2019. <https://www.oecd-ilibrary.org/content/data/efff24eb-en> (2022).
2. OECD. *Global Plastics Outlook: Policy Scenarios to 2060*. (OECD, 2022). doi:10.1787/aa1edf33-en.
3. OECD. *Global Plastics Outlook: Economic Drivers, Environmental Impacts and Policy Options*. (OECD, 2022). doi:10.1787/de747aef-en.
4. Aves, A. R. *et al.* First evidence of microplastics in Antarctic snow. *The Cryosphere* **16**, 2127–2145 (2022).
5. Bergmann, M. *et al.* Plastic pollution in the Arctic. *Nat. Rev. Earth Environ.* **3**, 323–337 (2022).
6. Allen, S. *et al.* Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nat. Geosci.* **12**, 339–344 (2019).
7. Tetu, S. G. *et al.* Plastic leachates impair growth and oxygen production in *Prochlorococcus*, the ocean's most abundant photosynthetic bacteria. *Commun. Biol.* **2**, 184 (2019).
8. Charlton-Howard, H. S., Bond, A. L., Rivers-Auty, J. & Lavers, J. L. 'Plasticosis': Characterising macro- and microplastic-associated fibrosis in seabird tissues. *J. Hazard. Mater.* **450**, 131090 (2023).
9. Wang, W., Ge, J. & Yu, X. Bioavailability and toxicity of microplastics to fish species: A review. *Ecotoxicol. Environ. Saf.* **189**, 109913 (2020).

10. Chamas, A. *et al.* Degradation Rates of Plastics in the Environment. *ACS Sustain. Chem. Eng.* **8**, 3494–3511 (2020).
11. Min, K., Cuiffi, J. D. & Mathers, R. T. Ranking environmental degradation trends of plastic marine debris based on physical properties and molecular structure. *Nat. Commun.* **11**, 727 (2020).
12. Ragaert, K., Delva, L. & Van Geem, K. Mechanical and chemical recycling of solid plastic waste. *Waste Manag.* **69**, 24–58 (2017).
13. Danso, D., Chow, J. & Streit, W. R. Plastics: Environmental and Biotechnological Perspectives on Microbial Degradation. *Appl. Environ. Microbiol.* **85**, e01095-19 (2019).
14. Ojha, N. *et al.* Evaluation of HDPE and LDPE degradation by fungus, implemented by statistical optimization. *Sci. Rep.* **7**, 39515 (2017).
15. Yoshida, S., Hiraga, K., Taniguchi, I. & Oda, K. Ideonella sakaiensis, PETase, and MHETase: From identification of microbial PET degradation to enzyme characterization. in *Methods in Enzymology* vol. 648 187–205 (Elsevier, 2021).
16. Lu, H. *et al.* Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* **604**, 662–667 (2022).
17. Gambarini, V. *et al.* PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database* **2022**, baac008 (2022).
18. Buchholz, P. C. F. *et al.* Plastics degradation by hydrolytic enzymes: The PLASTICS-ACTIVE enzymes database— PAZY. *Proteins Struct. Funct. Bioinforma.* prot.26325 (2022) doi:10.1002/prot.26325.
19. Wright, R. J., Langille, M. G. I. & Walker, T. R. Food or just a free ride? A meta-analysis reveals the global diversity of the Plastisphere. *ISME J.* **15**, 789–806 (2021).

20. Ru, J., Huo, Y. & Yang, Y. Microbial Degradation and Valorization of Plastic Wastes. *Front. Microbiol.* **11**, 442 (2020).
21. Conrad, R. E. *et al.* Toward quantifying the adaptive role of bacterial pangenomes during environmental perturbations. *ISME J.* **16**, 1222–1234 (2022).
22. The UniProt Consortium *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
23. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
24. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
25. Kerfahi, D. *et al.* Whole community and functional gene changes of biofilms on marine plastic debris in response to ocean acidification. *Microb. Ecol.* **85**, 1202–1214 (2023).
26. Xia, Y. *et al.* Plastic materials and water sources actively select and shape wastewater plastispheres over time. *Front. Environ. Sci. Eng.* **16**, 145 (2022).
27. Delacuvellerie, A., Géron, A., Gobert, S. & Wattiez, R. New insights into the functioning and structure of the PE and PP plastispheres from the Mediterranean Sea. *Environ. Pollut.* **295**, 118678 (2022).
28. Sun, J., Prabhu, A., Aroney, S. T. N. & Rinke, C. Insights into plastic biodegradation: community composition and functional capabilities of the superworm (*Zophobas morio*) microbiome in styrofoam feeding trials. *Microb. Genomics* **8**, (2022).
29. Bryant, J. A. *et al.* Diversity and Activity of Communities Inhabiting Plastic Debris in the North Pacific Gyre. *mSystems* **1**, e00024-16 (2016).

30. Radwan, O. & Ruiz, O. N. Shotgun metagenomic data of microbiomes on plastic fabrics exposed to harsh tropical environments. *Data Brief* **32**, 106226 (2020).
31. Pinto, M., Zhao, Z., Klun, K., Libowitzky, E. & Herndl, G. J. Microbial Consortia of Putative Degradable Low-Density Polyethylene-Associated Compounds in the Ocean. *mSystems* **7**, e01415-21 (2022).
32. Zhu, F. *et al.* Microplastics altered soil microbiome and nitrogen cycling: The role of phthalate plasticizer. *J. Hazard. Mater.* **427**, 127944 (2022).
33. Wu, X., Liu, Z., Li, M., Bartlam, M. & Wang, Y. Integrated metagenomic and metatranscriptomic analysis reveals actively expressed antibiotic resistomes in the plastisphere. *J. Hazard. Mater.* **430**, 128418 (2022).
34. Wu, X. *et al.* Selective enrichment of bacterial pathogens by microplastic biofilm. *Water Res.* **165**, 114979 (2019).
35. Meyer-Cifuentes, I. E. *et al.* Synergistic biodegradation of aromatic-aliphatic copolyester plastic by a marine microbial consortium. *Nat. Commun.* **11**, 5790 (2020).
36. Pinnell, L. J. & Turner, J. W. Shotgun Metagenomics Reveals the Benthic Microbial Community Response to Plastic and Bioplastic in a Coastal Marine Environment. *Front. Microbiol.* **10**, 1252 (2019).
37. Oberbeckmann, S. *et al.* Genomic and proteomic profiles of biofilms on microplastics are decoupled from artificial surface properties. *Environ. Microbiol.* **23**, 3099–3115 (2021).
38. Bos, R. P. *et al.* Plastics select for distinct early colonizing microbial populations with reproducible traits across environmental gradients. *Environ. Microbiol.* 1462-2920.16391 (2023) doi:10.1111/1462-2920.16391.

39. Li, Q., Tian, L., Cai, X., Wang, Y. & Mao, Y. Plastisphere showing unique microbiome and resistome different from activated sludge. *Sci. Total Environ.* **851**, 158330 (2022).
40. Li, R., Zhu, L., Cui, L. & Zhu, Y.-G. Viral diversity and potential environmental risk in microplastic at watershed scale: Evidence from metagenomic analysis of plastisphere. *Environ. Int.* **161**, 107146 (2022).
41. Rüthi, J. *et al.* The plastisphere microbiome in alpine soils alters the microbial genetic potential for plastic degradation and biogeochemical cycling. *J. Hazard. Mater.* **441**, 129941 (2023).
42. Bhagwat, G. *et al.* Exploring the Composition and Functions of Plastic Microbiome Using Whole-Genome Sequencing. *Environ. Sci. Technol.* **55**, 4899–4913 (2021).
43. Han, Y. *et al.* Soil Type Driven Change in Microbial Community Affects Poly(butylene adipate- co -terephthalate) Degradation Potential. *Environ. Sci. Technol.* **55**, 4648–4657 (2021).
44. Kumar, R. *et al.* Landfill microbiome harbour plastic degrading genes: A metagenomic study of solid waste dumping site of Gujarat, India. *Sci. Total Environ.* **779**, 146184 (2021).
45. Kuroda, K. *et al.* Elucidation of the biodegradation pathways of bis(2-hydroxyethyl) terephthalate and dimethyl terephthalate under anaerobic conditions revealed by enrichment culture and microbiome analysis. *Chem. Eng. J.* **450**, 137916 (2022).

46. Gaytán, I. *et al.* Degradation of Recalcitrant Polyurethane and Xenobiotic Additives by a Selected Landfill Microbial Community and Its Biodegradative Potential Revealed by Proximity Ligation-Based Metagenomic Analysis. *Front. Microbiol.* **10**, 2986 (2020).
47. Zhu, D., Ma, J., Li, G., Rillig, M. C. & Zhu, Y.-G. Soil plastispheres as hotspots of antibiotic resistance genes and potential pathogens. *ISME J.* **16**, 521–532 (2022).
48. Wang, S., Zeng, D., Jin, B., Su, Y. & Zhang, Y. Deciphering the role of polyethylene microplastics on antibiotic resistance genes and mobile genetic elements fate in sludge thermophilic anaerobic digestion process. *Chem. Eng. J.* **452**, 139520 (2023).
49. Zhang, W. *et al.* Marine biofilms constitute a bank of hidden microbial diversity and functional potential. *Nat. Commun.* **10**, 517 (2019).
50. Chao, Y., Mao, Y., Yu, K. & Zhang, T. Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach. *Appl. Microbiol. Biotechnol.* **100**, 8225–8237 (2016).
51. Díaz Rodríguez, C. A. *et al.* Novel bacterial taxa in a minimal lignocellulolytic consortium and their potential for lignin and plastics transformation. *ISME Commun.* **2**, 89 (2022).
52. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).

53. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
54. Nishimura, Y. & Yoshizawa, S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Sci. Data* **9**, 305 (2022).
55. Ghatge, S., Yang, Y., Ahn, J.-H. & Hur, H.-G. Biodegradation of polyethylene: a brief review. *Appl. Biol. Chem.* **63**, 27 (2020).
56. Sanluis-Verdes, A. *et al.* Wax worm saliva and the enzymes therein are the key to polyethylene degradation by *Galleria mellonella*. *Nat. Commun.* **13**, 5568 (2022).
57. Suzuki, M., Tachibana, Y. & Kasuya, K. Biodegradability of poly(3-hydroxyalkanoate) and poly(ϵ -caprolactone) via biological carbon cycles in marine environments. *Polym. J.* **53**, 47–66 (2021).
58. Palleroni*, N. J., Pieper, D. H. & Moore, E. R. B. Microbiology of Hydrocarbon-Degrading *Pseudomonas*. in *Handbook of Hydrocarbon and Lipid Microbiology* (ed. Timmis, K. N.) 1787–1798 (Springer Berlin Heidelberg, 2010). doi:10.1007/978-3-540-77587-4_129.
59. Janusz, G. *et al.* Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol. Rev.* **41**, 941–962 (2017).
60. Jambeck, J. R. *et al.* Plastic waste inputs from land into the ocean. *Science* **347**, 768–771 (2015).
61. Benoit, G. *et al.* Multiple comparative metagenomics using multiset k-mer counting. *PeerJ Comput. Sci.* **2**, e94 (2016).

62. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
63. Slobodkina, G. *et al.* Lithoautotrophic lifestyle of the widespread genus *Roseovarius* revealed by physiological and genomic characterization of *Roseovarius autotrophicus* sp. nov. *FEMS Microbiol. Ecol.* **98**, fiac113 (2022).
64. Pinto, M. *et al.* Putative degraders of low-density polyethylene-derived compounds are ubiquitous members of plastic-associated bacterial communities in the marine environment. *Environ. Microbiol.* **22**, 4779–4793 (2020).
65. Poff, K. E., Leu, A. O., Eppley, J. M., Karl, D. M. & DeLong, E. F. Microbial dynamics of elevated carbon flux in the open ocean’s abyss. *Proc. Natl. Acad. Sci.* **118**, e2018269118 (2021).
66. Despotović, D. *et al.* Utilization of diverse organophosphorus pollutants by marine bacteria. *Proc. Natl. Acad. Sci.* **119**, e2203604119 (2022).
67. Zettler, E. R., Mincer, T. J. & Amaral-Zettler, L. A. Life in the “Plastisphere”: Microbial Communities on Plastic Marine Debris. *Environ. Sci. Technol.* **47**, 7137–7146 (2013).
68. Edwards, S. *et al.* Microbial Consortia and Mixed Plastic Waste: Pangenomic Analysis Reveals Potential for Degradation of Multiple Plastic Types via Previously Identified PET Degrading Bacteria. *Int. J. Mol. Sci.* **23**, 5612 (2022).
69. Tchigvintsev, A. *et al.* The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Appl. Microbiol. Biotechnol.* **99**, 2165–2178 (2015).

70. Ohta, T., Tani, A., Kimbara, K. & Kawai, F. A novel nicotinoprotein aldehyde dehydrogenase involved in polyethylene glycol degradation. *Appl. Microbiol. Biotechnol.* **68**, 639–646 (2005).
71. Lu, J., Takahashi, A. & Ueda, S. 3-Hydroxybutyrate Oligomer Hydrolase and 3-Hydroxybutyrate Dehydrogenase Participate in Intracellular Polyhydroxybutyrate and Polyhydroxyvalerate Degradation in *Paracoccus denitrificans*. *Appl. Environ. Microbiol.* **80**, 986–993 (2014).
72. Kang, T., Lin, J., Yang, L. & Wu, M. Expression, isolation, and identification of an ethanol-resistant ethyl carbamate-degrading amidase from *Agrobacterium tumefaciens* d3. *J. Biosci. Bioeng.* **132**, 220–225 (2021).
73. Erickson, E. *et al.* Sourcing thermotolerant poly(ethylene terephthalate) hydrolase scaffolds from natural diversity. *Nat. Commun.* **13**, 7850 (2022).
74. Wright, R. J., Bosch, R., Langille, M. G. I., Gibson, M. I. & Christie-Oleza, J. A. A multi-OMIC characterisation of biodegradation and microbial community succession within the PET plastisphere. *Microbiome* **9**, 141 (2021).
75. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018) doi:10.48550/ARXIV.1802.03426.
76. Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S. & Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat. Commun.* **11**, 1537 (2020).
77. Danko, D. *et al.* A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell* **184**, 3376–3393.e17 (2021).

78. León-Zayas, R., Roberts, C., Vague, M. & Mellies, J. L. Draft Genome Sequences of Five Environmental Bacterial Isolates That Degrade Polyethylene Terephthalate Plastic. *Microbiol. Resour. Announc.* **8**, e00237-19 (2019).
79. Brissos, V., Tavares, D., Sousa, A. C., Robalo, M. P. & Martins, L. O. Engineering a Bacterial DyP-Type Peroxidase for Enhanced Oxidation of Lignin-Related Phenolics at Alkaline pH. *ACS Catal.* **7**, 3454–3465 (2017).
80. Montazer, Z., Habibi-Najafi, M. B., Mohebbi, M. & Oromiehei, A. Microbial Degradation of UV-Pretreated Low-Density Polyethylene Films by Novel Polyethylene-Degrading Bacteria Isolated from Plastic-Dump Soil. *J. Polym. Environ.* **26**, 3613–3625 (2018).
81. Zadjelovic, V. *et al.* A mechanistic understanding of polyethylene biodegradation by the marine bacterium *Alcanivorax*. *J. Hazard. Mater.* **436**, 129278 (2022).
82. Takehara, I., Kato, D.-I., Takeo, M. & Negoro, S. Draft Genome Sequence of the Nylon Oligomer-Degrading Bacterium *Arthrobacter* sp. Strain KI72. *Genome Announc.* **5**, e00217-17 (2017).
83. Li, J., Zhang, K. & Zhang, H. Adsorption of antibiotics on microplastics. *Environ. Pollut.* **237**, 460–467 (2018).
84. Xu, C. *et al.* Are we underestimating the sources of microplastic pollution in terrestrial environment? *J. Hazard. Mater.* **400**, 123228 (2020).
85. Kirstein, I. V. *et al.* Dangerous hitchhikers? Evidence for potentially pathogenic *Vibrio* spp. on microplastic particles. *Mar. Environ. Res.* **120**, 1–8 (2016).

86. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
87. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917 (2022).
88. Gravouil, K. *et al.* Transcriptomics and Lipidomics of the Environmental Strain *Rhodococcus ruber* Point out Consumption Pathways and Potential Metabolic Bottlenecks for Polyethylene Degradation. *Environ. Sci. Technol.* **51**, 5172–5181 (2017).
89. Schwarz, W. H. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* **56**, 634–649 (2001).
90. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
91. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
92. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
93. Kolmogorov, M. *et al.* metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **17**, 1103–1110 (2020).
94. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).

95. Vasimuddin, Md., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (IEEE, 2019).
doi:10.1109/IPDPS.2019.00041.
96. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
97. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
98. Lamurias, A., Sereika, M., Albertsen, M., Hose, K. & Nielsen, T. D. Metagenomic binning with assembly graph embeddings. *Bioinformatics* **38**, 4481–4487 (2022).
99. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. *CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning*. <http://biorxiv.org/lookup/doi/10.1101/2022.07.11.499243> (2022)
doi:10.1101/2022.07.11.499243.
100. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
101. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
102. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**, e0163962 (2016).

103. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genomics* **7**, (2021).
104. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
105. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
106. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
107. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
108. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
109. Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems* **3**, e00039-18 (2018).
110. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
111. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

112. Nayfach, S. & Pollard, K. S. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**, 51 (2015).
113. Shaw, J. & Yu, Y. W. *Fast and robust metagenomic sequence comparison through sparse chaining with skani*.
<http://biorxiv.org/lookup/doi/10.1101/2023.01.18.524587> (2023)
doi:10.1101/2023.01.18.524587.
114. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016). doi:10.1007/978-3-319-24277-4.
115. Konstantinidis, K. *et al.* *FastAAI: Efficient Estimation of Genome Average Amino Acid Identity and Phylum-level relationships using Tetramers of Universal Proteins*.
<https://www.researchsquare.com/article/rs-1459378/v1> (2022) doi:10.21203/rs.3.rs-1459378/v1.
116. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
117. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
118. Larralde, M. Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. *J. Open Source Softw.* **7**, 4296 (2022).
119. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
120. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

121. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
122. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS ONE* **8**, e67019 (2013).
123. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
124. Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
125. Xu, S. *et al.* Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* **1**, (2022).
126. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
127. Dask Development Team. Dask: Library for dynamic task scheduling. (2016).
128. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
129. Bednar, J. A. *et al.* holoviz/datashader: Version 0.15.1. (2023)
doi:10.5281/ZENODO.8118007.
130. Bokeh Development Team. Bokeh: Python library for interactive visualization. (2018).
131. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

132. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).