Plastic-Degrading Clusters of Orthologous Groups in Prokaryotes

Shakira Mustari[1,&], Loan Tú Phạm[1,&], Kari Saikkonen[2], Miho Nakamura[3,4,5,*], Pere Puigbò[1,6,*]

[1] Department of Biology, University of Turku, 20500 Turku, Finland

[2] Biodiversity Unit, University of Turku, 20500 Turku, Finland

[3] Medicity Research Laboratory, Faculty of Medicine, University of Turku, Tykistökatu 6, 20520 Turku, Finland.

[4] Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University, 2-3-10 Kanda-Surugadai, Chiyoda, Tokyo 1010062 Japan

[5] Graduate School of Engineering, Tohoku University, 6-6 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi 9808579 Japan

[6] Department of Animal and Food Science, Faculty of Veterinary Medicine, Autonomous University of Barcelona, 08193 Bellaterra, Catalonia, Spain.

[&] Equal contribution of the authors

* Co-senior authors and for correspondence:

Pere Puigbò, pere.puigbo@uab.cat

Miho Nakamura, miho.nakamura@utu.fi

ABSTRACT

Microplastics and nanoplastics pollution (MNPP) is a major environmental issue due to its small size, long-lasting nature, and potentially harmful impacts on ecosystems and human health. Plastics can be degraded in nature by physical, chemical, and biological processes. Plastic biodegradation is a potential sustainable solution to MNPP, but understanding the fitness of potential plastic-degrading proteins (PPDPs) across environments requires comparative resources. Here, we introduce the Plastic-Degrading Clusters of Orthologous Groups (PDCOGs) database, available at https://phylobone.com/microworld/PDCOG. The resource includes 625,616 PPDPs from free-living prokaryotes classified in 51 PDCOGs. The PDCOG database is a robust resource that improves understanding of plastic biodegradation and supports research on global MNPP challenges. Overall, PDPPs represent 3.5% of proteins in prokaryotes and are widely distributed across the Earth, spanning across most prokaryotic species and environments. Nearly all prokaryotic species (>95%) have the potential to biodegrade at least one polymer type.

MAIN

Plastic pollution stands out as a major environmental concern because of its small size, persistence in the environment, and potential adverse effects on ecosystems and human health [1]. These tiny plastic particles, often measuring 0.001–1 µm (nanoplastics) and 1–1000 µm (microplastics) [2], are the result of the breakdown of larger plastic items or are intentionally manufactured on a small scale [3]. Microplastics and nanoplastics (MNP) can be found in various environments, including oceans, rivers, soil, and air, posing threats to marine life and wildlife, and potentially entering the food chain [4,5]. Moreover, the chemical composition of microplastics and their ability to be vectors of pollutants raise concerns about the indirect health impacts on humans when consuming contaminated food and water [6,7]. In addition, recent studies have shown that MNP may play an important role in the dissemination of antimicrobial resistant strains, due to their capacity to host microbial communities, chemical contaminants, and antimicrobial resistance genes through biofilm formations [8].

According to recent research estimates, over 14 million tons of microplastics have accumulated on the world's ocean floor [9]. Current approaches to tackling this problem are fragmented and offer only limited, short-term solutions to the pervasive issue of plastic discharge into the environment. However, this is a global phenomenon, affecting the most densely populated areas, but also remote areas of the earth. Particularly, the polar regions may become extremely sensitive during the next few years because of the effect of climate change [10,11]. The impact of plastic pollutants in the context of global climate warming is a matter of major concern [12]. Global warming will have an impact on the polar regions of the earth (the Arctic Ocean will be ice-free before the end of the century [13], whereas less than 1% of the Antarctic continent is ice-free, but it is estimated to increase to close to 25% by 2100 [14]). This leads to the accumulation of plastic debris from lower latitudes to the Arctic by ocean currents [10] and releases microplastic particles frozen in Arctic Sea ice [11].

Plastic pollutants from traditional plastics, such as polyethylene (PE), polystyrene (PS), polypropylene (PP), polyvinyl chloride (PVC), polyethylene terephthalate (PET), polyurethane (PUR), and other plastic polymers, gradually accumulate in the environment because of extremely low degradation rates [15]. In natural environments, the attention given to plastic biodegradation has surged because of its advantageous green and safe features [15]. Studies of the plastisphere, that is, the microbial biofilms that are formed on the surface of microplastics [16], allows for the identification of microbial strains involved in the biodegradation of plastic particles [17]. Until recently, it was thought that plastic biodegradation was exclusive to mesophilic species (only performed by microorganisms with an optimal temperature growth ranging from 20°C to 45°C [18]). A recent study focuses on sequences from thermophiles (i.e. microbial strains with optimal growth temperature above 45°C [18]), as thermostable enzymes are likely more suitable for industrial applications [19]. However, studies of the Arctic plastisphere have identified microbial strains with the ability to break down biodegradable plastic films at lower temperatures (<20°C) [20]. In addition, the capacity of plastic-degrading microbes to perform the biodegradation should vary across terrestrial, aquatic and air environments.

To understand the adaptation of plastic-degrading proteins in different environments (i.e. the enzymes that microbial species produce to perform the biodegradation polymers), it is essential to have resources that allow the comparison of proteins across environments and geographical locations. In this article, we introduce a dataset of Plastic-Degrading Clusters of Orthologous Groups (PDCOGs). This resource classifies each plastic-degrading protein into a set of 51 orthologous families, and each family has been annotated based on plastic-degrading properties, enzymatic reactions, functional domains, environment and geolocation. This is the most comprehensive resource of plastic-degrading enzymes and will allow a better understanding of the biodegradation process in different environments. Furthermore, it will allow further studies for the management of the MNP global problems.

RESULTS

A resource of plastic-degrading clusters of orthologous groups (PDCOGs)

The PDCOGs database is freely accessible at http://phylobone.com/microworld/PDCOG. The resource is organized into six sections and comprises 662,830 potential plastic-degrading proteins (PPDP) sequences compressed within 51 PDCOGs (Table 1, Supplementary Table S1). The PDCOGs resource is a protein-centric database that includes information on the 51 PDCOGs, along with functional and biological data, and offers a comprehensive repository of PPDPs across diverse environments on Earth. We have analyzed data from 11 natural and 28 synthetic (7 homochain and 21 heterochain) polymers. Furthermore, P(3HB-co-3MP), P(3HV) [21], P3HP [22] can be classified as both natural [23,24] and synthetic polymers. In general, these polymers are not produced in natural environments, but they are engineered productions, mainly petroleum based, which it is performed by microorganisms (e.g. *Escherichia coli*, *Clostridium butiricum*, and *Klebsiella pneumoniae*) [25]. The present PDCOG database classifies these plastics as natural. All plastic polymers described in the PDCOGs databases are potentially susceptible to be biodegraded by microbial species that have at least one copy of a PDCOGs (Supplementary Datasets SD1-2).

Associations of PDCOGs with prokaryotic species and polymer types

The database of PDCOGs was analyzed as a bipartite network that associates PDCOGs with polymer types and microbial species (Fig. 1 and Supplementary Fig. S1). This modular structure with variable node degrees shows diverse phylogenetic associations and a widespread distribution of enzymatic potential across bacterial phyla, highlighting a robust ecological basis for biodegradation (Supplementary Dataset SD3). Overall, PDCOGs are associated with a median number of three plastic polymers (minimum: 1 and maximum: 16). PDCOG026 and PDCOG041 are the ones that provide a larger number of associations with plastic polymers (12 and 16, respectively). Notably, 22 PDCOGs are associated exclusively with a single polymer type, suggesting a putative role in its degradation: PDCOG003 is linked to Nylon, PDCOG005 is linked to PET, PDCOG007 is linked to PET, PDCOG009 is linked to PET, PDCOG010 is linked to low density polyethylene (LDPE), PDCOG011 is linked to PET, PDCOG015 is linked to polyethylene glycol (PEG), PDCOG016 is linked to Polylactic acid (PLA), PDCOG017 is linked to LDPE, PDCOG017 is linked to PLA, PDCOG019 is linked to PET, PDCOG022 is linked to Natural Rubber (NR), PDCOG023 is linked to NR, PDCOG025 is linked to PLA, PDCOG028 is linked to PEG, PDCOG029 is linked to NR, PDCOG030 is linked to polyvinyl alcohol (PVA), PDCOG031 polyurethane (PU), PDCOG032 is linked to PS, PDCOG034 is linked to polycaprolactone (PCL), PDCOG035 is linked to Nylon, PDCOG042 is linked to Polylactic acid (PLA).

The distribution of microbial species, based on the five best BLAST hits to the seed dataset, varies across PDCOGs, ranging from 1 to 56 species. PDCOG041 is linked to 56 species, followed by PDCOG006 (34 species), PDCOG014 (33 species), PDCOG018 (25 species), PDCOG027 (21 species), PDCOG045 (19 species), and PDCOG001 (10 species). The remaining PDCOGs are each associated with fewer than 10 species (Supplementary Dataset SD3). Moreover, several bacterial groups are associated with a single PDCOG, such as Thermodesulfobacteria with PDCOG001, Dictyoglomota and Verrucomicrobiota with PDCOG018, Methylomirabilota with PDCOG014, and Methanobacteriota with PDCOG006.

General functional analysis of the PDCOGs

The PDCOGs includes a list of 51 orthologous groups that include a wide variety of biological functions. A functional characterization of the PDCOGs using the 26 functional categories (Supplementary Table S2) in the COG database shows that the functional categories of PPDP are I (Lipid transport and metabolism; n = 12), R (General function prediction only; n = 6), G (Carbohydrate transport and metabolism; n = 5) are the most abundant in plastic-degrading enzymes, followed by other categories in order: C=D=M=O=T>E=Q>P=V>H=J=L (Fig. 2A). Among the 51 PDCOGs, the enzymatic activities involved in plastic biodegradation include Oxidoreductases (EC1; 33.33%), transferases (EC2; 33.33%), hydrolases (EC3, 52.94%), lyases (EC4, 5.88%), isomerases (EC5, 1.96%), ligases (EC6, 9.8%) and translocases (EC7, 1.96%) (Supplementary Dataset SD4). Furthermore, the PDCOG database includes KEGG and gene ontology annotations, as well as Interpro functional domains (Supplementary Dataset SD4).

Enzymatic characterization of the PDCOGs

We analyzed the associations between 212 diverse Enzyme Commission (EC) Functions (Supplementary Dataset SD4) and 39 plastic types (Supplementary Dataset SD2) using an Indicator Value (IV), and a permutation test to assess statistical significance of enzyme-plastic interactions. The interaction of 7 EC classes spanning up to the third hierarchical level (e.g., EC 1.x.x.-) and 39 plastics polymers (Fig. 3B) was analyzed by focusing on the constriction of diverse EC subclasses (for example 1.-.-.-, 1.1.-.-, 1.1.1.-, 1.1.1.1) into 85 aggregated broad categories of third hierarchical level. Based on the highest log IV value and significant p-values, the heatmap helps us understand the broad range of enzymatic activity. Log IV values range between 0 and 1, with a highly skewed distribution, with 75% of the values between 0.00 and 0.25. The enzymes carboxylic ester hydrolase and the enzyme hydrolyzing glycosidase exhibited a strong association with most polymers. However, both enzymes showed a poor association with synthetic-heterochain polymers. On the other hand, the log IV showed that synthetic-heterochain polymers are associated with more EC subclasses (n = 52) compared to the natural (n = 34) and synthetic-homochain (n = 26) polymers in the third hierarchical level composition.

Distribution of the PDCOGs across prokaryotes (archaea and bacteria)

Prokaryotic species have the potential to biodegrade synthetic-heterochain polymers (13.4%), followed by natural polymers (9.40%), and synthetic-homochain polymers (9.80%) (Fig. 2C). The relative abundance of the PDCOGs across prokaryotes in the swarm plot shows that bacterial species present a wider range of PDCOGs compared to archaeal species (Fig. 2B, Table 1). The median distribution of PDCOGs in prokaryotes is 2,296 species (35.37%), including 2,103 (38.52%) bacteria and 193 (8.29%) archaea, and interquartile ranges (IQR) of 23.52–65.00% for bacteria, and 3.37–38.60% for archaea, reflecting considerable variability among orthologs. Outliers, defined as data points exceeding 1.5 times the IQR, were identified across the bacterial and archaeal species. Furthermore, to assess the relative contributions of bacteria and archaea to the total extent of PPDP, the proportion was calculated for each PDCOG, based on their counts relative to the total prokaryotic counts. The median proportion extended in bacteria was 96.7% of the species, whereas in archaea the extent comprised 3.3% (Wilcoxon signed-rank test, p-value < 0.0001) (Fig. 2B). The enzyme-specific abundance pattern shows that only the PDCOG015 have a high presence of archaeal species (>90%), above the confidence intervals.

Phyletic distribution of the PDCOGs

A Venn's diagram of species overlap across plastic types, based on literature data (seed sequences), was used to infer shared enzymatic functions (Fig. 4A). To standardize seed sequences obtained from the literature, the top five BLAST hits were selected for the analysis. We analyzed the relative abundance of the resultant species to determine which plastic types host the most diverse biodegradability potential (synthetic-heterochain > natural > synthetic-homochain) and which taxa are most frequently associated with specific plastics (Pseudomonadata, Actinomycetota, Bacillota, and Bacteroidota) (Fig. 4B). Among them, synthetic-heterochain, natural, and synthetic-homochain plastics harbored 88%, 65% , and 32% of unique microbial populations respectively. Furthermore, 22% of the species were common across all plastic types. Notably, the highest overlap was observed between synthetic-heterochain and natural polymers (41%), whereas no species were shared between synthetic-homochain and the other two types (Fig. 3A).

Most prokaryotic species present enzymes with the capacity to break down plastic polymers (Fig. 4B). From the seed sequences of putative plastic-degrading proteins, a total of 51 PDCOGs were identified. These orthologous proteins with the potential capacity to break down plastic polymers are widely spread across prokaryotic species (Fig. 1, Supplementary Fig. S1, Supplementary Dataset SD3). Pseudomonadota is the bacterial group that presents a higher abundance of PPDP of all polymer types (natural and synthetic), followed by Actinomycetota, Bacteroidota and Bacillota (Fig. 4B). In archaea three phyla were identified as potential biodegraders of plastic polymers, including Methanobacteriota, Thermoproteota and Bathyarchaeota. Our analysis revealed that 65.68% of the cataloged species potentially biodegrade natural polymers, 32.47% species biodegrade synthetic homochain polymers and 87.82% biodegrade synthetic heterochain type. Notice that some of the strains have the potential capacity to biodegrade more than one polymer type.

Geographical distribution of the PDCOGs

The PDCOGs are widespread across environments (Fig. 5). Most of the PPDP in the environmental dataset include sequences from aquatic and continental environments (Table 1). As expected, these sequences were mostly found in aquatic, due to sampling overrepresentation. However, the PDCOG034, PDCOG035, PDCOG045, and PDCOG050 include a majority of continental sequences. In addition, A large portion of terrestrial sequences were collected in the USA, western Europe, east Asia and Australia, whereas marine sequences follow the pattern of sampling collection routes. Overall, across environments, the environmental dataset of PPDP contains fewer sequences than the expectation based on the results of the chi-squared test (Supplementary Table S3). Only, rock-dwelling (endoliths) has more PPDP than expected and soil has an expected number of PPDP. Furthermore, the distribution of each PDCOG compared to chance expectation (t-test) is highly variable across environments (Supplementary Dataset SD6).

DISCUSSION

General characterization of the PDCOGs

The PDCOGs resource provides a comprehensive database of PPDPs, organized by orthology and classified according to enzymatic function, biological occurrence, geographical distribution, and habitat variability. Interestingly, all enzymatic proteins in the database present known biological activity, with hydrolases (EC3) being the most abundant, followed by oxidoreductases (EC1) and transferases (EC2). In terms of functional classification, the most represented COG categories associated with plastic degradation are I (lipid transport and metabolism), M (cell wall/membrane/envelope biogenesis), and C (energy production and conversion). Nevertheless, we

speculate that future studies may reveal the evolution of PPDPs with minimal or no known biological function, resulting from the rapid adaptation of proteins to plastic-rich environments. Furthermore, distribution of orthologs across lineages also reflects horizontal gene transfer, lineage-specific gene loss, and rapid gene evolution linked to functional changes [26].

Differences in polymer structure are determinant to the formation of prokaryotic biofilms, and in consequence affecting presence and absences of PDCOGs. Overall, synthetic-heterochain (25%) polymers present a much larger number of unique species than natural (2%) and synthetic-homochain (10%) polymers. This analysis is based on seed sequences [27], derived from prokaryotic species, either isolated from natural environments or identified in laboratory experiments, that demonstrate the capacity to biodegrade polymers. However, all prokaryotic species that host PPDPs, classified here into PDCOGs, are potential plastic-biodegradators. Two main scenarios may explain why many are not detected in natural or experimental conditions: (1) sampling bias, as not all species have been tested, and (2) the presence of the enzyme does not guarantee its functionality in certain species. Further studies are needed to clarify these possibilities. Furthermore, the effects of PDCOGs on polymer structures do not account for the structural modifications that occur after plastics are exposed to natural environments.

Phyletic distribution of the PDCOGs

Since the term plastisphere was coined in 2013, to describe microbial communities with the capacity to colonize plastic debris in marine environments, several studies have advanced to the understanding of the biodiversity of microorganisms inhabiting plastic particles [28]. Microorganisms, among other life forms, have evolved strategies to colonize and biodegrade plastic particles [29]. The most striking result is that over 95% of prokaryotes (>98% of archaea and >96% of bacteria) have the potential to biodegrade at least one polymer type. In general, bacteria presentes a much larger number of plastic-degrading enzymes compared to archaea across the earth and environmental habitats. Bacterial dominance aligns with the numerical disparity between the two groups, and likely reflects a broader ecological adaptability and metabolic versatility, enabling them to occupy a wider range of plastisphere environments. However, archaea presents several species with rare, but functionally significant PDCOGs, possibly associated with environmental adaptations.

Pseudomonas is the bacterial genus with a larger number of PPDP and is the potentially dominant group for biodegradation across all polymer types, reflecting its well-known metabolic versatility and adaptability [30]. As an example, clinical isolates of *Pseudomonas aeruginosa* have shown the potential to biodegrade medically relevant plastics [31], while wild-type strains are effective at degrading crude oil in petrochemical wastewater [32]. Other prominent bacterial groups included Actinomycetota, Bacteroidota, and Bacillota. These results are consistent with previous studies on the phylogenetic distribution of 436 putative plastic-degrading species, where the five most represented phyla across prokaryotes, eukaryotes, and fungi were Proteobacteria (37.1%), Ascomycota (16.9%), Actinobacteria (14.3%), Firmicutes (13.8%), and Basidiomycota (8.1%). [33]. Interestingly, in our study we found that three archaea (Methanobacteriota, Thermoproteota, and Bathyarchaeota) are potential plastic degraders, suggesting that archaea may play an underexplored role in plastic biodegradation, particularly in extreme or anaerobic environments [33].

Geographical distribution of the PDCOGs

Although the PDCOGs database is largely represented in aquatic environments, only two terrestrial environments—soil and rock-dwelling—show an equal or higher-than-expected presence of PPDPs.

We hypothesize that microbial species from terrestrial environments with nutrient limitations [34] drive the abundance of PPDP in soil and rock-dwelling samples. Although nutrients are also scarce in marine environments, water columns accumulate organic carbon, nitrogen, phosphorus, and iron due to the chemical nature of the compounds that constitute marine dissolved organic matter. [35]. Therefore, the imbalanced accessibility to nutrients in marine and terrestrial environments may explain a lower abundance and lower proportion of catalytic enzymes. Nevertheless, we should not disregard a potential information gap between the COG [36] and the genome database [37] regarding the genome sequence information. Overall the results suggest a potential environmental-specificity of some PPDP with potential microbial adaptation to plastic pollution, but further quantitative studies are needed.

How to use the PDCOGs?

The PDCOGs database includes functional, biological and environmental information classified across PDCOGs. This database is complementary to other publicly available databases of putative plastic-degrading proteins and microbes [27,38–41]. Potential biodegradable plastics are compostable due to their chemical structure, which includes ester linkages that make them susceptible to breakdown by microbial enzyme systems, such as Poly-caprolactone (PCL), poly butylene succinate-co-adipate (PBSA), Poly butylene succinate (PBS), and PLA [42]. When compared to high polymers, like PET, which need highly efficient degrading enzymes, BPs are prone to degradation and overcome the environmental pollution and challenges associated with synthetic and natural plastics [43]. Addressing the problem of microplastic contamination requires comprehensive research projects to mitigate their environmental and health risks. Furthermore, microbial biodegradation is apparently highly dependent on environmental conditions and polymer structures. Therefore, the PDCOGs resources provide a robust and comprehensive tool for further studies aimed at understanding the potential and extent of microbial biodegradation.

METHODS

Seed of putative plastic-degrading proteins

A seed of putative plastic-degrading proteins (PPDP) sequences was obtained from the PlasticDB [27] and articles from the literature to expand the coverage of the dataset (Supplementary Table S4). The initial dataset (available from the PDCOGs webpage) contains a seed of PPDP sequences from bacterial species and potential plastic polymers that can be biodegraded (Supplementary Table S5).

Orthologous proteins

We collected all proteins from the COGs database [36]. This database contains 5,640,669 protein sequences classified based on orthology relationships into 5,050 COGs. Each protein is functionally annotated and the COGs are classified in 26 biological functions based on 4 main functional categories (Supplementary Table S2). As of the 2024 update, the COGDB includes genomic information of 2,296 (2,103 bacterial and 193 archaeal).

Putative plastic-degrading orthologous groups & phyletic pattern

The PDCOG dataset has been created on the basis of our BLASTp result from NCBI COG database (COGDB) [36] and PlasticDB [27] (Supplementary Datasets SD1-3). Each sequence from the PPDP datased was mapped onto the COGs dataset with the program BLASP for the identification of potential plastic-degrading orthologous groups. Briefly, a BLASTP search was done on the NCBI COG database and the PlasticDB database to have the best COG hits against the PlasticDB (BLASP parameters were

outfmt = 6, *e*-vaule = 0.001, max-target-seqs=5. The data of COG functional category and symbol were collected from the COGDB.

Functional annotation of the PDCOGs

Functional annotations of the PDCOGs (Supplementary Dataset SD4) were perfumed with the database Cluster of Orthologous Group Database (COGDB) [36], Kyoto Encyclopedia of Genes and Genomes (KEGG) [44], InterPro 104.0 [45], Expasy [46], NCBI Taxonomy [47]. The KEGG database was used to collect KO codes for each PDCOG, as well as the EC (enzymatic) and reaction codes. The categorical function of each EC class was characterized with Expasy. Protein functional domains and gene ontology (GO) information of each PDCOG was characterized with the InterPro database.

Reconstruction of the polymers dataset

The systematic workflow for the data curation of polymers included in the dataset is based on both cheminformatics analyses and literature reviews (Supplementary Table S5 and Supplementary Dataset SD2). We have used PU-REST API of Pubchem DB [48] for SMILEs, Polymer Genome DB, Smiles2Monomers [49] for monomer mapping, RDKit open source cheminformatics v2025.3.2 in Python for validation of SMILES and monomers, Beauti-fulSoup for scraping open access journals from PubMed Central [50]. Out of 39 polymers, 12 PSMILEs were obtained from PubChem, the rest were obtained from PI1M DB [51] for full repeat units vs. fragments, using filters '−Atomcount ≥ 10 to exclude fragments, -Molecular weight 500–5000 Da, atom count, functional groups' and excluding non-specific. Consequently after fetching 50 best batches per target polymer, the consensus based PSMILEs were obtained for the rest of the polymers. To evaluate recent advances in the potential biodegradation of additional polymers, including P3HO, PC, PEF, PMCL, PP and PPL, we surveyed 300 peer-reviewed studies published between 2000 and 2025. The selection specifically relied on prokaryotic relevance to those polymers and enzymes used in biodegradation.

Environmental annotation of the PDCOGs

Protein sequences of free-living species with geographical and environmental information were gathered from the Joint Genome Institute (DOE-JGI) [37]. We have analyzed 74,651,773 protein sequences with the Python *re* package to identify sequences with geographical and environmental annotations that belong to free-living microbial species. The filtered sequences were mapped on the PDCOGs with BLASTP. Thus, the final dataset included 662,830 protein sequences with environmental and functional annotations (Supplementary Table S3 and Supplementary Datasets SD5-6). These proteins are spread in 23 different environments across the Earth. The distribution map of the PDCOGs was built using pandas, matplotlib, cartopy.crs for handling of world map projections, and seaborn for the built-in color palettes. A chi-squared test was performed on the PDCOGs dataset to determine significant differences ($p<0.05$) from an expected distribution of the protein sequences based on protein number in the COGDB .

Plotting and statistical analysis

All data processing, statistical analyses and visualizations were performed with Python (version 3.9.6). Visualizations were generated using Matplotlib (v3.9.4), Plotly (v6.0.0.) and Seaborn (v0.13.2), whereas data processing with the Pandas (v2.2.3), SciPy (v1.13.1), and NumPy (v2.0.2) libraries, scipy.stats module (wilcoxon). The Wilcoxon signed-rank test (Figure 2B) was implemented to assess significant differences (p-value < 0.05) of PDCOG abundance distribution between the bacteria and archaea. The indicator value (IV) in the heatmap (Fig. 3B) is calculated based on .

Briefly, the method combines the specificity (Sp) and Fidelity (Fd), which was calculated for each enzyme-plastic pair [52], and downweighting was applied to reduce the influence of frequently occurring enzymes (Equation 1). Permutation test (10,000 iterations) assessed significance, yielding p-values ranging from 0.01 to 0.04 (Fig. 2B).

Equation 1.

$$IV = S \cdot F \cdot W$$

$$Sp = \frac{(P \cap e)}{(p \cup e)} \quad ; \quad Fd = \frac{(p \cap e)}{p} \quad ; \quad W = \frac{1}{log\,(1 + Np)}$$

Where, p is the total EC occurrences for the target plastic polymer; e is the total 4th-subclass EC occurrences under the 3rd-subclass EC across plastics; p∩e is the number of EC occurrences (4th EC subclass) for the target plastic under the 3rd EC subclass; p∪e is the union of the 3rd EC subclass annotation of the target plastic and across plastics; and W is a down-weighting factor.

AUTHOR CONTRIBUTIONS

Conception and design of the study: MN, PP; Funding acquisition: MN, KS, PP; Data collection: SM, LTF, PP; Data analysis: SM, LTF, PP; Manuscript drafting: SM, LTF, PP.; Manuscript revision for critical intellectual content: SM, LTF, MN, PP; Writing the final version of the manuscript: PP and MN. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

We thank members of the MicroWorld team, students and collaborators for their helpful discussions.

FUNDING

Scandinavian-Japan Sasakawa foundation (GA25-FUB-0132). PP is a Serra Húnter Professor.

REFERENCES

1. Kakar, F. L. *et al.* Climate change interaction with microplastics and nanoplastics pollution. in *Current developments in biotechnology and bioengineering* 387–403 (Elsevier, 2023). doi:10.1016/B978-0-323-99908-3.00003-8.
2. Bermúdez, J. R. & Swarzenski, P. W. A microplastic size classification scheme aligned with universal plankton survey methods. *MethodsX* **8**, 101516 (2021).
3. Tong, H. *et al.* Micro- and nanoplastics released from biodegradable and conventional plastics during degradation: Formation, aging factors, and toxicity. *Sci. Total Environ.* **833**, 155275 (2022).
4. Mamun, A. A., Prasetya, T. A. E., Dewi, I. R. & Ahmad, M. Microplastics in human food chains: Food becoming a threat to health safety. *Sci. Total Environ.* **858**, 159834 (2023).
5. Taipale, S. J. *et al.* Biodegradation of microplastic in freshwaters: A long-lasting process affected by the lake microbiome. *Environ. Microbiol.* **25**, 2669–2680 (2023).
6. Amelia, T. S. M. *et al.* Marine microplastics as vectors of major ocean pollutants and its hazards to the marine ecosystem and humans. *Prog. in Earth and Planet. Sci.* **8**, 12 (2021).
7. Shruti, V. C., Pérez-Guevara, F., Elizalde-Martínez, I. & Kutralam-Muniasamy, G. First study of its kind on the microplastic contamination of soft drinks, cold tea and energy drinks - Future research and environmental considerations. *Sci. Total Environ.* **726**, 138580 (2020).

8. Gross, N. *et al.* Effects of microplastic concentration, composition, and size on Escherichia coli biofilm-associated antimicrobial resistance. *Appl. Environ. Microbiol.* **91**, e0228224 (2025).
9. Barrett, J. *et al.* Microplastic Pollution in Deep-Sea Sediments From the Great Australian Bight. *Front. Mar. Sci.* **7**, (2020).
10. Bergmann, M. *et al.* Plastic pollution in the Arctic. *Nat. Rev. Earth Environ.* (2022) doi:10.1038/s43017-022-00279-8.
11. Obbard, R. W. *et al.* Global warming releases microplastic legacy frozen in Arctic Sea ice. *Earth's Future* **2**, 315–320 (2014).
12. Sharma, S., Sharma, V. & Chatterjee, S. Contribution of plastic and microplastic to global climate change and their conjoining impacts on the environment - A review. *Sci. Total Environ.* **875**, 162627 (2023).
13. Kelly, B. P., Whiteley, A. & Tallmon, D. The Arctic melting pot. *Nature* **468**, 891 (2010).
14. Lee, J. R. *et al.* Climate change drives expansion of Antarctic ice-free habitat. *Nature* **547**, 49–54 (2017).
15. Yang, X.-G. *et al.* Plastic biodegradation by in vitro environmental microorganisms and in vivo gut microorganisms of insects. *Front. Microbiol.* **13**, 1001750 (2022).
16. 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).
17. Amaral-Zettler, L. A., Zettler, E. R. & Mincer, T. J. Ecology of the plastisphere. *Nat. Rev. Microbiol.* **18**, 139–151 (2020).
18. Lehmann, M. *et al.* Adaptive laboratory evolution of a thermophile toward a reduced growth temperature optimum. *Front. Microbiol.* **14**, 1265216 (2023).
19. Hu, Z., Klupt, K., Zechel, D. L., Jia, Z. & Howe, G. Mining Thermophile Genomes for New PETases with Exceptional Thermostabilities Using Sequence Similarity Networks. *Chembiochem* e2500065 (2025) doi:10.1002/cbic.202500065.
20. Rüthi, J. *et al.* Discovery of plastic-degrading microbial strains isolated from the alpine and Arctic terrestrial plastisphere. *Front. Microbiol.* **14**, 1178474 (2023).
21. Watabe, Y., Shimomura, S., Ono, K. & Sasanuma, Y. Conformational characteristics of poly(3-hydroxyvalerate) (P3HV) and structure-property relationships of P3HV and poly(3-hydroxybutyrate). *Phys. Chem. Chem. Phys.* **25**, 27034–27044 (2023).
22. Meng, D.-C. *et al.* Production of poly(3-hydroxypropionate) and poly(3-hydroxybutyrate-co-3-hydroxypropionate) from glucose by engineering Escherichia coli. *Metab. Eng.* **29**, 189–195 (2015).
23. De Angelis, M. & Gobbetti, M. Lactobacillus sanfranciscensis CB1: manganese, oxygen, superoxide dismutase and metabolism. *Appl. Microbiol. Biotechnol.* **51**, 358–363 (1999).
24. García, B. *et al.* The chromomycin CmmA acetyltransferase: a membrane-bound enzyme as a tool for increasing structural diversity of the antitumour mithramycin. *Microb. Biotechnol.* **4**, 226–238 (2011).
25. Yang, B. *et al.* Metabolic engineering of Escherichia coli for 1,3-propanediol biosynthesis from glycerol. *Bioresour. Technol.* **267**, 599–607 (2018).
26. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
27. Gambarini, V. *et al.* PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database (Oxford)* **2022**, (2022).
28. Bocci, V. *et al.* Freshwater plastisphere: a review on biodiversity, risks, and biodegradation potential with implications for the aquatic ecosystem health. *Front. Microbiol.* **15**, 1395401 (2024).
29. Zeenat, Elahi, A., Bukhari, D. A., Shamim, S. & Rehman, A. Plastics degradation by microbes: A sustainable approach. *Journal of King Saud University - Science* **33**, 101538 (2021).
30. Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B. & Jackson, R. W. Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol. Rev.* **35**, 652–680 (2011).
31. Howard, S. A. *et al.* Pseudomonas aeruginosa clinical isolates can encode plastic-degrading enzymes that allow survival on plastic and augment biofilm formation. *Cell Rep.* **44**, 115650 (2025).
32. Samimi, M. & Shahriari-Moghadam, M. Efficient biodegradation of crude oil by novel

Pseudomonas aeruginosa strain isolated from petrochemical wastewater: identification and optimization. *Pet. Sci. Technol.* **43**, 1182–1199 (2025).
33. Gambarini, V. *et al.* Phylogenetic Distribution of Plastic-Degrading Microorganisms. *mSystems* **6**, (2021).
34. Kang, H. *et al.* Nutrient limitation mediates soil microbial community structure and stability in forest restoration. *Sci. Total Environ.* **935**, 173266 (2024).
35. Repeta, D. J. & Boiteau, R. M. Organic Nutrient Chemistry and the Marine Microbiome - The Chemistry of Microbiomes - NCBI Bookshelf. (2017).
36. Galperin, M. Y. *et al.* COG database update 2024. *Nucleic Acids Res.* **53**, D356–D363 (2025).
37. Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* **42**, D26-31 (2014).
38. Buchholz, P. C. F. *et al.* Plastics degradation by hydrolytic enzymes: The plastics-active enzymes database-PAZy. *Proteins* **90**, 1443–1456 (2022).
39. Ridley, R. S., Conrad, R. E., Lindner, B. G., Woo, S. & Konstantinidis, K. T. Potential routes of plastics biotransformation involving novel plastizymes revealed by global multi-omic analysis of plastic associated microbes. *Sci. Rep.* **14**, 8798 (2024).
40. Wicker, J. *et al.* enviPath--The environmental contaminant biotransformation pathway resource. *Nucleic Acids Res.* **44**, D502-8 (2016).
41. Ellis, L. B. & Wackett, L. P. Use of the University of Minnesota Biocatalysis/Biodegradation Database for study of microbial degradation. *Microb. Inform. Exp.* **2**, 1 (2012).
42. Al Hosni, A. S., Pittman, J. K. & Robson, G. D. Microbial degradation of four biodegradable polymers in soil and compost demonstrating polycaprolactone as an ideal compostable plastic. *Waste Manag.* **97**, 105–114 (2019).
43. Cai, Z. *et al.* Biological degradation of plastics and microplastics: A recent perspective on associated mechanisms and influencing factors. *Microorganisms* **11**, (2023).
44. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
45. Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* **53**, D444–D456 (2025).
46. Bairoch, A. The ENZYME database in 2000. *Nucleic Acids Res.* **28**, 304–305 (2000).
47. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)* **2020**, baaa062 (2020).
48. Team, Fair. PubChem. *FAIRsharing* (2015) doi:10.25504/fairsharing.qt3w7z.
49. Dufresne, Y., Noé, L., Leclère, V. & Pupin, M. Smiles2Monomers: a link between chemical and biological structures for polymers. *J. Cheminform.* **7**, 62 (2015).
50. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
51. Ma, R. & Luo, T. PI1M: A benchmark database for polymer informatics. *J. Chem. Inf. Model.* **60**, 4684–4690 (2020).
52. Wei, H., Therrien, C., Blanchard, A., Guan, S. & Zhu, Z. The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Res.* **36**, e50 (2008).

FIGURE LEGENDS

Fig. 1. Bipartite network of associations between the plastic-degrading clusters of orthologous groups (PDCOGs) with microbial species and enzymatic functions. The network was generated from the linkage data of organisms, COGs, and targeted plastics using Cytoscape v3.10.3. In the network, node colors represent: (i) the COG functional hierarchy based on connectivity to polymers, (ii) taxonomic clustering of species, and (iii) classification of plastic polymers. Edge colors indicate the strength of connections, weighted by interaction confidence.

Fig. 2. Functional, genomic and ecological diversity in the plastic-degrading clusters of orthologous groups (PDCOGs). A. Enzymatic roles and genomic functions of plastic biodegradation. The stacked bar plot illustrates the distribution of PDCOG across COG functional categories (CFC) within each EC class. The x-axis represents EC classes (EC1 to EC7), which categorizes enzymes based on their catalytic reactions. Each bar is divided into colored segments, with each color representing a distinct CFC (e.g., metabolism, cellular processes, information storage etc.). The count of correlated PDCOGs in each CFC is annotated within the corresponding segment of the bar, the sum of which in the y-axis indicates the quantification of COGs associated with each EC class. **B. Relative prokaryotic abundance in the PDCOGs.** Box plots show the percentage of Prokaryotes, Bacteria, and Archaea (primary y-axis) across 51 PDCOGs (secondary y-axis), with a swarm plot overlay indicating enzyme-specific contributions (enzymes color-coded). Statistical significance of Bacteria dominance over Archaea within Prokaryotes was determined by a Wilcoxon signed-rank test (***: $p < 0.001$). Outliers are denoted by black triangles. **C. Plastic polymer biodegradation profile.** Pie chart illustrates the proportional distribution of natural, synthetic-homochain, and synthetic-heterochain plastic polymers potentially biodegradable by PDCOG. Inner segment proportions indicate relative degradation capacity of different categories of polymers, whereas the outer layer represents average degradability with minimum and maximum counts. **D. Environmental distribution of PDCOGs.** Illustration of mean proportional incidence of plastic degraders in various environmental habitats based on genome count from Genome portal.

Fig. 3: Enzymatic characterization of plastic-degrading clusters of orthologous groups (PDCOGs) based on known biological enzymatic activities. A. Enzymatic activity of PDCOG across plastic types. Bars represent the percentage of total enzyme activity (0–100%) from 51 PDCOGs for each plastic type. The x-axis displays plastic types, while the y-axis shows enzyme activity (%). Due to the highest percentage being noticed under 50%, the y-axis depicts up to 50% for visualization. The legend (right) maps PDCOG domains to colors, with a purple-to-yellow gradient (bottom to top) indicating different protein clusters. **B. Heatmap of the Indicator Value (IV) for enzymes associated with plastic biodegradation.** This plot illustrates the Indicator Values (IV) for enzymes putatively involved in the biodegradation of 39 plastic types. With plastics on the x-axis and potential enzymes on the y-axis, the IV metric of relative abundance and specificity was calculated based on the annotations from the KEGG database [44]. Color intensity and log based value (10^{-4} lowest to 10^{-0} highest) highlights the strength of their association.

Fig. 4. Taxonomic distribution of plastic-degrading microbes and polymer types. A. Species-level distribution patterns among polymer classes. Venn's diagram illustrates the relative distribution of common species among plastic polymer categories natural, synthetic-heterochain, and synthetic-homochain for biodegradation**.** Each circle represents one plastic category with color grading, and the intersections denote shared association among multiple categories. Percent values are shown for major intersections, while low-frequency overlaps (<5%) are denoted in the legend for clarity. Circle positions and sizes vary based on their affinity with species. **B. Phylum-Specific Trends in Categorical Polymer Biodegradation.** The grouped bar plot compares the relative association of phyla with each plastic category. Each bar cluster on x-axis corresponds to all the unique prokaryotic

phyla in the dataset. The height of each bar along the y-axis denotes the phylum based proportion of unique species that is correlated to the plastics in each category (natural, synthetic-heterochain, synthetic-homochain), normalized by the total unique plastics within that category.

Fig. 5. World distribution of plastic degrading enzymes and their abundance. A. World distribution of plastic-degrading clusters of orthologous groups (PDCOGs) across environments. The geospatial distribution of PDCOGs highlight hotspots of functional diversity and frequency of plastic-degrading genes across sampled environments. Color gradients represent PDCOGs functional diversity, whereas their proportional variation denotes sequence abundance. Hotspots of enzymatic potential correlate with anthropogenic impact zones (e.g., soil, sediments, marine and aquatic etc.) and microbial niches adapted to synthetic polymer hydrolysis. **B. Percentage of Observed sequences in Relative to Expectation (O/E) across various environments.** The observed (O) and the expected values (E) of were calculated based on the data obtained from the JGI Genome Portal and the NCBI COG database respectively. Values represent the percentage of deviation from the expected baseline (100%). Bars to the left of the central axis indicate environments where observed values fall below expectations, while bars to the right show environments exceeding expectations. The "TOTAL" summary is displayed in orange for distinction.

TABLES

Table 1. List of PDCOGS and distribution across prokaryotic species, plastic types and environments.

PDCOG	Number	COG	Fun	%Prok	Plastic types	Environmental distribution
PDCOG001	20,071	COG0154	J	88.81%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.56%), Aq (60.27%), Te (38.97%), Un (0.2%)
PDCOG002	21,153	COG0277	C	80.49%	N: 0/11; S_He: 2/21; S_Ho: 1/7	Ai (0.54%), Aq (56.41%), Te (42.84%), Un (0.21%)
PDCOG003	6,475	COG0400	R	52.44%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.59%), Aq (73.27%), Te (26.01%), Un (0.14%)
PDCOG004	12,730	COG0412	Q	57.01%	N: 2/11; S_He: 9/21; S_Ho: 1/7	Ai (0.46%), Aq (53.94%), Te (45.48%), Un (0.13%)
PDCOG005	16,725	COG0508	C	80.01%	N: 0/11; S_He: 1/21; S_Ho: 1/7	Ai (0.50%), Aq (68.55%), Te (30.75%), Un (0.2%)
PDCOG006	90,850	COG0596	HR	91.29%	N: 4/11; S_He: 8/21; S_Ho: 1/7	Ai (0.56%), Aq (54.26%), Te (45.01%), Un (0.16%)
PDCOG007	4,527	COG0627	V	43.07%	N: 1/11; S_He: 5/21; S_Ho: 1/7	Ai (0.46%), Aq (48.13%), Te (51.25%), Un (0.15%)
PDCOG008	15,526	COG0657	I	65.11%	N: 3/11; S_He: 6/21; S_Ho: 2/7	Ai (0.71%), Aq (51.13%), Te (48.04%), Un (0.12%)
PDCOG009	14,723	COG0726	MG	80.66%	N: 0/11; S_He: 0/21; S_Ho: 2/7	Ai (0.60%), Aq (51.69%), Te (47.48%), Un (0.22%)
PDCOG010	43,441	COG0834	TE	76.96%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.46%), Aq (54.29%), Te (45.04%), Un (0.22%)
PDCOG011	67,608	COG1012	I	89.46%	N: 0/11; S_He: 1/21; S_Ho: 1/7	Ai (0.44%), Aq (53.08%), Te (46.25%), Un (0.22%)
PDCOG012	139,666	COG1028	I	95.03%	N: 3/11; S_He: 0/21; S_Ho: 2/7	Ai (0.57%), Aq (55.03%), Te (44.21%), Un (0.19%)
PDCOG013	4,116	COG1073	T	30.31%	N: 1/11; S_He: 5/21; S_Ho: 1/7	Ai (0.56%), Aq (65.38%), Te (33.79%), Un (0.27%)
PDCOG014	2,434	COG1075	I	32.58%	N: 4/11; S_He: 8/21; S_Ho: 1/7	Ai (0.58%), Aq (49.92%), Te (49.30%), Un (0.21%)
PDCOG015	261	COG1382	O	8.89%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (1.53%), Aq (69.35%), Te (29.12%), Un (0.00%)
PDCOG016	4,483	COG1398	I	27.57%	N: 0/11; S_He: 0/21; S_Ho: 1/7	Ai (0.27%), Aq (61.25%), Te (38.39%), Un (0.09%)
PDCOG017	6,857	COG1404	O	62.37%	N: 1/11; S_He: 4/21; S_Ho: 0/7	Ai (0.61%), Aq (54.53%), Te (44.55%), Un (0.31%)
PDCOG018	9,776	COG1506	E	71.65%	N: 4/11; S_He: 4/21; S_Ho: 0/7	Ai (0.30%), Aq (71.69%), Te (27.91%), Un (0.11%)
PDCOG019	2,783	COG1647	Q	31.84%	N: 1/11; S_He: 5/21; S_Ho: 1/7	Ai (0.25%), Aq (63.71%), Te (35.79%), Un (0.25%)
PDCOG020	21,699	COG1680	V	66.77%	N: 1/11; S_He: 5/21; S_Ho: 0/7	Ai (0.41%), Aq (59.08%), Te (40.36%), Un (0.15%)
PDCOG021	4,746	COG1773	P	42.55%	N: 0/11; S_He: 0/21; S_Ho: 3/7	Ai (0.36%), Aq (62.45%), Te (37.04%), Un (0.15%)
PDCOG022	4,733	COG1858	O	35.76%	N: 0/11; S_He: 0/21; S_Ho: 2/7	Ai (0.42%), Aq (58.27%), Te (41.12%), Un (0.19%)
PDCOG023	18,460	COG2010	C	56.58%	N: 0/11; S_He: 0/21; S_Ho: 2/7	Ai (0.64%), Aq (56.21%), Te (43.01%), Un (0.14%)
PDCOG024	6,773	COG2132	DMP	51.09%	N: 0/11; S_He: 0/21; S_Ho: 4/7	Ai (0.75%), Aq (56.18%), Te (42.85%), Un (0.22%)
PDCOG025	590	COG2247	M	3.44%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.34%), Aq (68.98%), Te (30.68%), Un (0.00%)
PDCOG026	8,632	COG2267	I	70.86%	N: 3/11; S_He: 14/21; S_Ho: 0/7	Ai (0.47%), Aq (57.75%), Te (41.52%), Un (0.25%)
PDCOG027	2,972	COG2272	I	20.17%	N: 1/11; S_He: 9/21; S_Ho: 0/7	Ai (0.34%), Aq (63.36%), Te (36.07%), Un (0.24%)
PDCOG028	17,646	COG2303	IR	49.96%	N: 0/11; S_He: 1/21; S_Ho: 1/7	Ai (0.49%), Aq (56.58%), Te (42.76%), Un (0.17%)
PDCOG029	1,542	COG2703	T	20.82%	N: 0/11; S_He: 0/21; S_Ho: 1/7	Ai (0.71%), Aq (66.67%), Te (32.10%), Un (0.52%)
PDCOG030	10,074	COG2863	C	35.37%	N: 0/11; S_He: 1/21; S_Ho: 1/7	Ai (0.60%), Aq (59.90%), Te (39.40%), Un (0.11%)
PDCOG031	5,066	COG2931	Q	28.75%	N: 0/11; S_He: 2/21; S_Ho: 0/7	Ai (0.71%), Aq (61.94%), Te (36.89%), Un (0.45%)
PDCOG032	3,397	COG3145	L	33.97%	N: 0/11; S_He: 0/21; S_Ho: 3/7	Ai (0.71%), Aq (59.58%), Te (39.48%), Un (0.24%)
PDCOG033	2,814	COG3147	D	28.22%	N: 2/11; S_He: 0/21; S_Ho: 0/7	Ai (0.39%), Aq (66.24%), Te (33.08%), Un (0.28%)
PDCOG034	1,515	COG3179	G	11.80%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.20%), Aq (43.17%), Te (56.44%), Un (0.20%)
PDCOG035	2,522	COG3191	E	29.05%	N: 0/11; S_He: 2/21; S_Ho: 0/7	Ai (0.36%), Aq (42.27%), Te (57.14%), Un (0.24%)
PDCOG036	1,0447	COG3239	I	40.29%	N: 0/11; S_He: 0/21; S_Ho: 4/7	Ai (0.30%), Aq (61.02%), Te (38.57%), Un (0.11%)
PDCOG037	3,984	COG3243	I	28.83%	N: 2/11; S_He: 0/21; S_Ho: 0/7	Ai (0.85%), Aq (56.58%), Te (42.42%), Un (0.15%)
PDCOG038	693	COG3266	D	17.77%	N: 2/11; S_He: 0/21; S_Ho: 0/7	Ai (0.43%), Aq (82.40%), Te (16.74%), Un (0.43%)
PDCOG039	1,892	COG3325	G	16.33%	N: 0/11; S_He: 4/21; S_Ho: 0/7	Ai (0.16%), Aq (53.91%), Te (45.61%), Un (0.32%)
PDCOG040	278	COG3401	R	11.28%	N: 0/11; S_He: 4/21; S_Ho: 0/7	Ai (0.36%), Aq (57.91%), Te (41.73%), Un (0.00%)
PDCOG041	3,327	COG3509	G	25.04%	N: 8/11; S_He: 11/21; S_Ho: 2/7	Ai (0.75%), Aq (53.77%), Te (45.33%), Un (0.15%)
PDCOG042	1,500	COG3591	O	21.56%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.47%), Aq (49.87%), Te (48.87%), Un (0.80%)
PDCOG043	1,294	COG3979	G	11.67%	N: 0/11; S_He: 5/21; S_Ho: 0/7	Ai (0.31%), Aq (69.78%), Te (29.75%), Un (0.15%)
PDCOG044	2,311	COG4188	R	21.86%	N: 3/11; S_He: 8/21; S_Ho: 0/7	Ai (0.17%), Aq (65.73%), Te (34.01%), Un (0.09%)
PDCOG045	2,504	COG4553	I	14.94%	N: 3/11; S_He: 7/21; S_Ho: 0/7	Ai (0.72%), Aq (38.74%), Te (60.42%), Un (0.12%)
PDCOG046	9,970	COG1366	T	35.76%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.90%), Aq (57.92%), Te (40.52%), Un (0.65%)
PDCOG047	6,948	COG1520	M	53.53%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.39%), Aq (70.31%), Te (29.07%), Un (0.23%)
PDCOG048	15,226	COG2755	DI	73.78%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.63%), Aq (57.97%), Te (41.11%), Un (0.28%)
PDCOG049	2,710	COG2819	R	24.65%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.44%), Aq (66.94%), Te (32.55%), Un (0.07%)
PDCOG050	1,749	COG3240	I	12.46%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.80%), Aq (45.45%), Te (53.57%), Un (0.17%)
PDCOG051	611	COG4099	R	16.16%	N: 0/11; S_He: 1/21; S_Ho: 0/7	Ai (0.33%), Aq (73.65%), Te (26.02%), Un (0.00%)

Fun: Functional category in the COG database; N: Natural polymers (n = 10); S_Ho: Synthetic-homochain polymers (n = 8); S_He: Synthetic-heterochain polymers (n = 21); Ai: Air; Aq: Aquatic; Te: Terrestrial; Un: Unknown

Figure 1

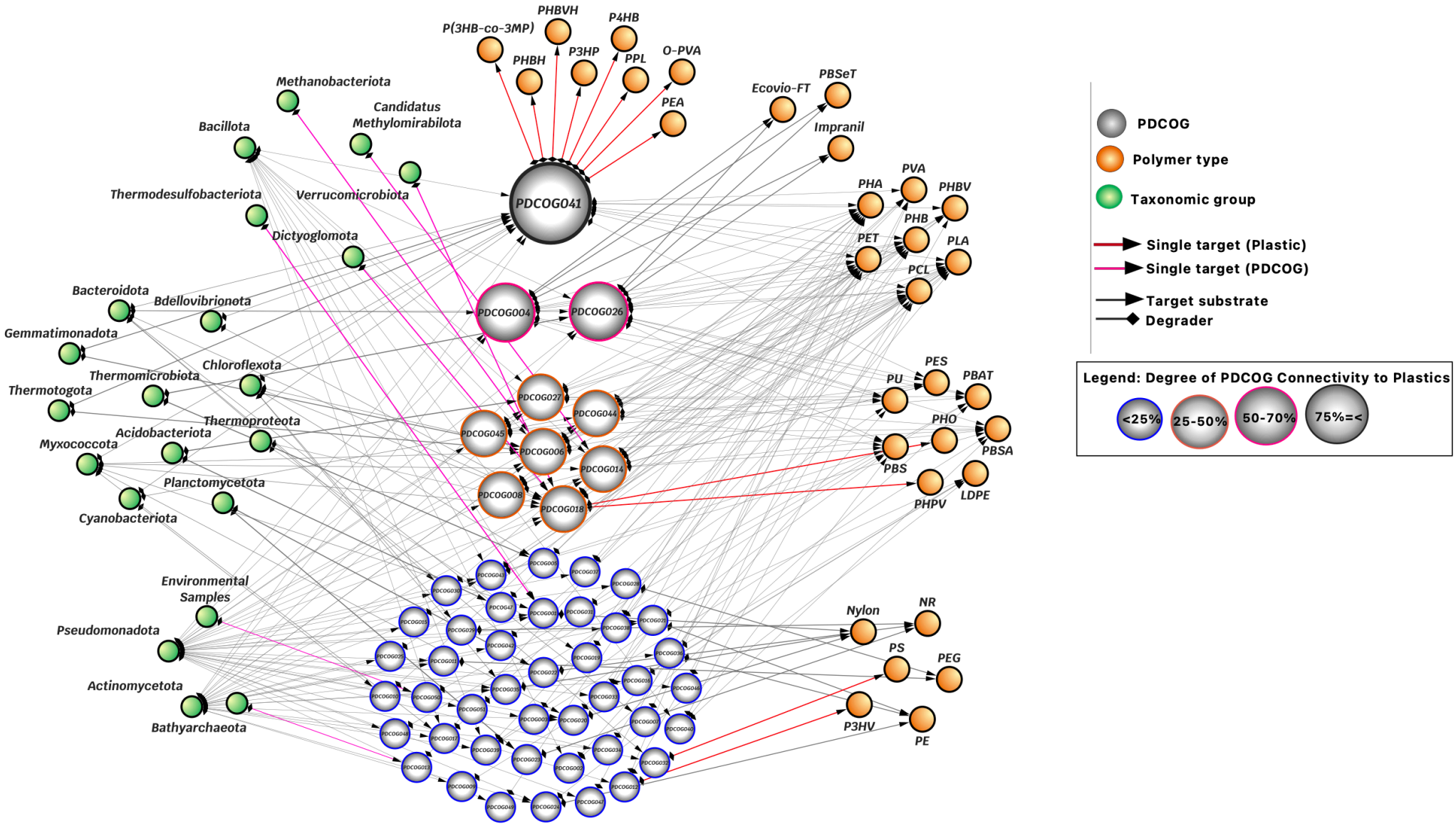

PDCOG041
PDCOG004
PDCOG026
PDCOG027
PDCOG044
PDCOG045
PDCOG006
PDCOG014
PDCOG008
PDCOG018
P(3HB-co-3MP)
PHBVH
P4HB
PHBH
P3HP
PPL
O-PVA
PEA
Ecovio-FT
PBSeT
Impranil
PHA
PVA
PHBV
PHB
PET
PLA
PCL
PU
PES
PBAT
PHO
PBSA
PBS
PHPV
LDPE
Nylon
NR
PS
PEG
P3HV
PE
Methanobacteriota
Candidatus Methylomirabilota
Bacillota
Thermodesulfobacteriota
Verrucomicrobiota
Dictyoglomota
Bacteroidota
Bdellovibrionota
Gemmatimonadota
Thermomicrobiota
Chloroflexota
Thermotogota
Thermoproteota
Acidobacteriota
Myxococcota
Planctomycetota
Cyanobacteriota
Environmental Samples
Pseudomonadota
Actinomycetota
Bathyarchaeota
PDCOG
Polymer type
Taxonomic group
Single target (Plastic)
Single target (PDCOG)
Target substrate
Degrader
Legend: Degree of PDCOG Connectivity to Plastics
<25%
25-50%
50-70%
75%=<

Figure 2

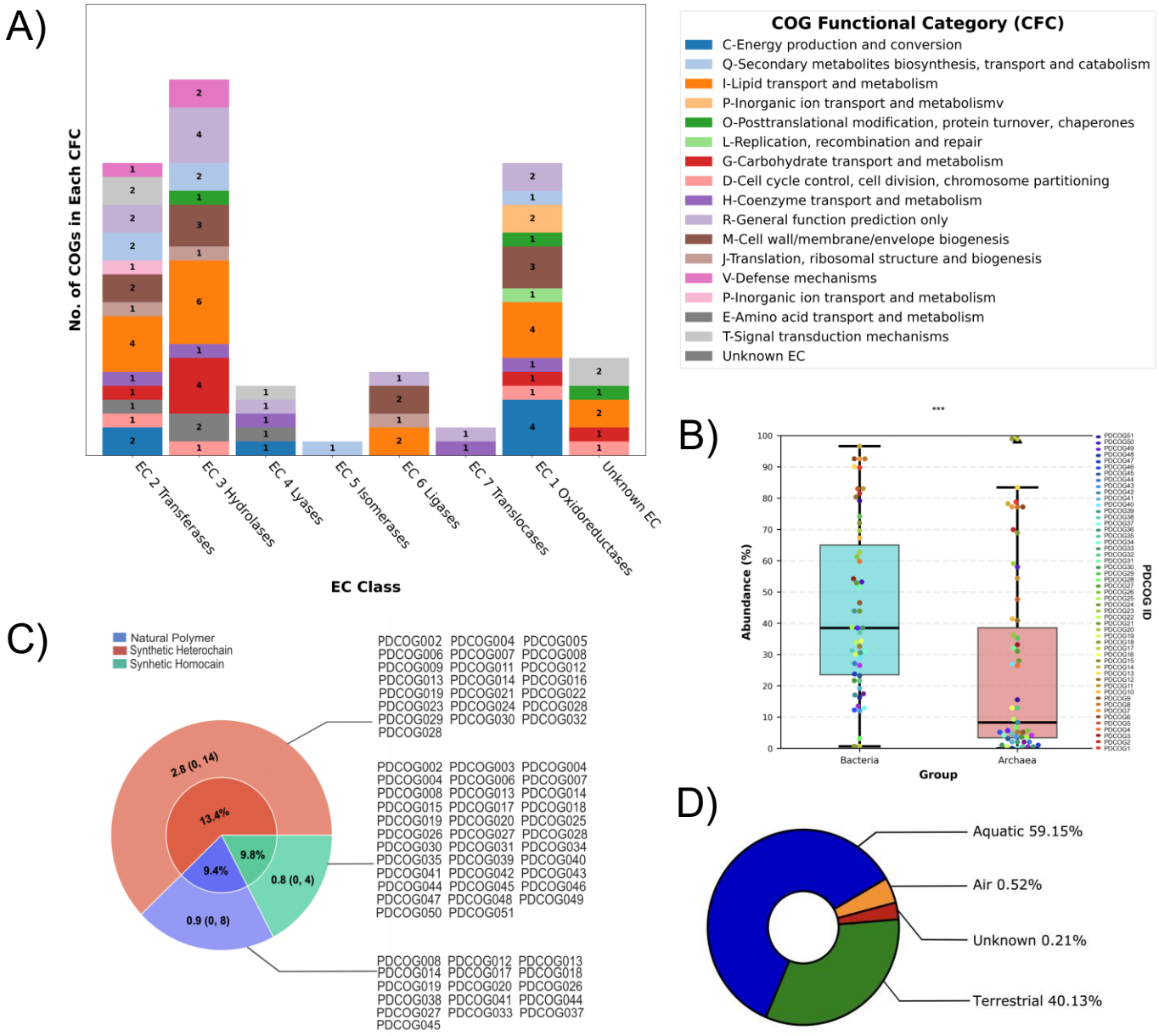

A)
No. of COGs in Each CFC
EC 2 Transferases
EC 3 Hydrolases
EC 4 Lyases
EC 5 Isomerases
EC 6 Ligases
EC 7 Translocases
EC 1 Oxidoreductases
Unknown EC
EC Class
COG Functional Category (CFC)
C-Energy production and conversion
Q-Secondary metabolites biosynthesis, transport and catabolism
I-Lipid transport and metabolism
P-Inorganic ion transport and metabolismv
O-Posttranslational modification, protein turnover, chaperones
L-Replication, recombination and repair
G-Carbohydrate transport and metabolism
D-Cell cycle control, cell division, chromosome partitioning
H-Coenzyme transport and metabolism
R-General function prediction only
M-Cell wall/membrane/envelope biogenesis
J-Translation, ribosomal structure and biogenesis
V-Defense mechanisms
P-Inorganic ion transport and metabolism
E-Amino acid transport and metabolism
T-Signal transduction mechanisms
Unknown EC
B)
***
Abundance (%)
Bacteria
Archaea
Group
PDCOG ID
C)
Natural Polymer
Synthetic Heterochain
Synthetic Homocain
2.8 (0, 14)
13.4%
9.8%
9.4%
0.8 (0, 4)
0.9 (0, 8)
PDCOG002 PDCOG004 PDCOG005 PDCOG006 PDCOG007 PDCOG008 PDCOG009 PDCOG011 PDCOG012 PDCOG013 PDCOG014 PDCOG016 PDCOG019 PDCOG021 PDCOG022 PDCOG023 PDCOG024 PDCOG028 PDCOG029 PDCOG030 PDCOG032 PDCOG028
PDCOG002 PDCOG003 PDCOG004 PDCOG004 PDCOG006 PDCOG007 PDCOG008 PDCOG013 PDCOG014 PDCOG015 PDCOG017 PDCOG018 PDCOG019 PDCOG020 PDCOG025 PDCOG026 PDCOG027 PDCOG028 PDCOG030 PDCOG031 PDCOG034 PDCOG035 PDCOG039 PDCOG040 PDCOG041 PDCOG042 PDCOG043 PDCOG044 PDCOG045 PDCOG046 PDCOG047 PDCOG048 PDCOG049 PDCOG050 PDCOG051
PDCOG008 PDCOG012 PDCOG013 PDCOG014 PDCOG017 PDCOG018 PDCOG019 PDCOG020 PDCOG026 PDCOG038 PDCOG041 PDCOG044 PDCOG027 PDCOG033 PDCOG037 PDCOG045
D)
Aquatic 59.15%
Air 0.52%
Unknown 0.21%
Terrestrial 40.13%

Figure 3

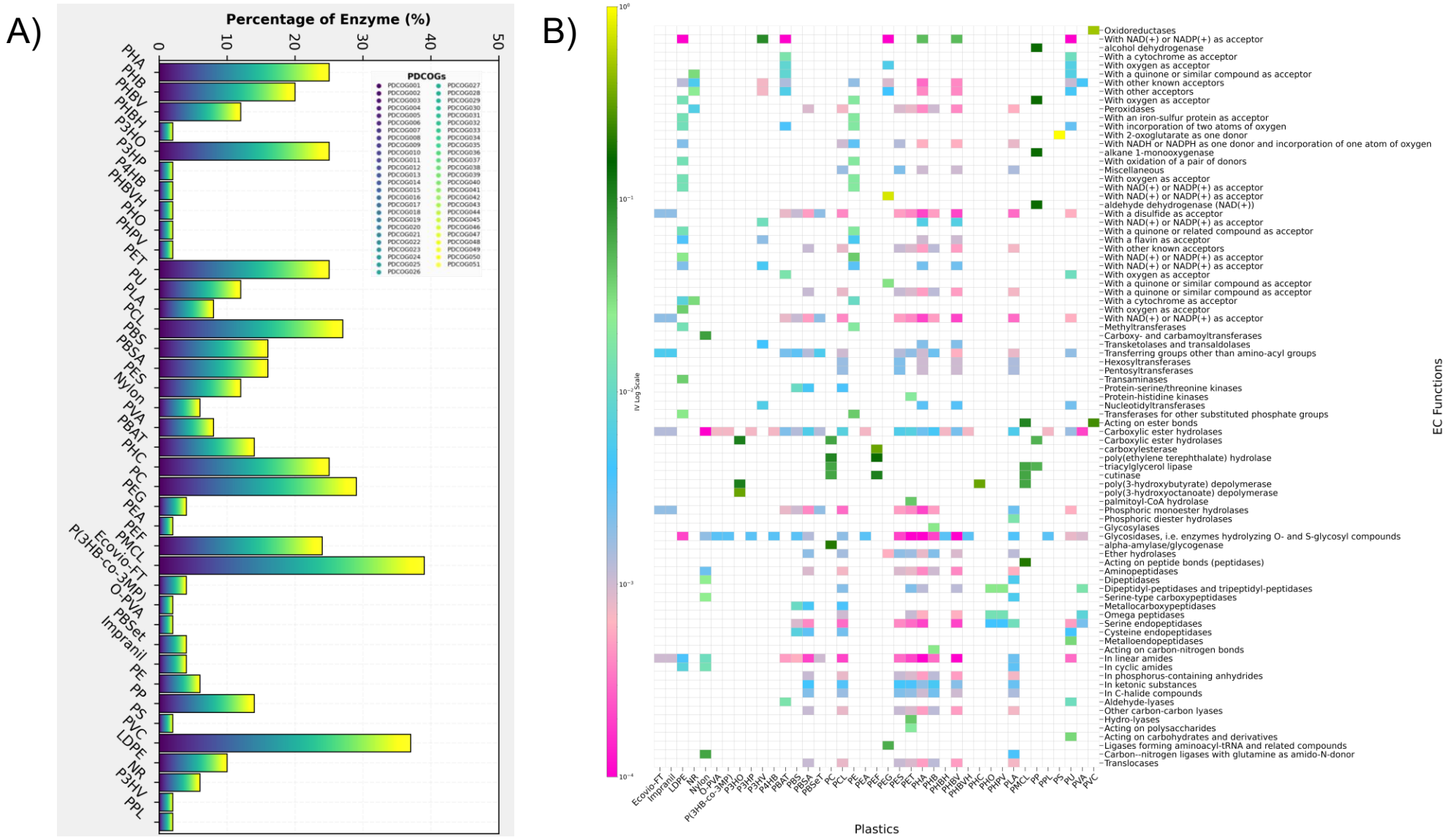
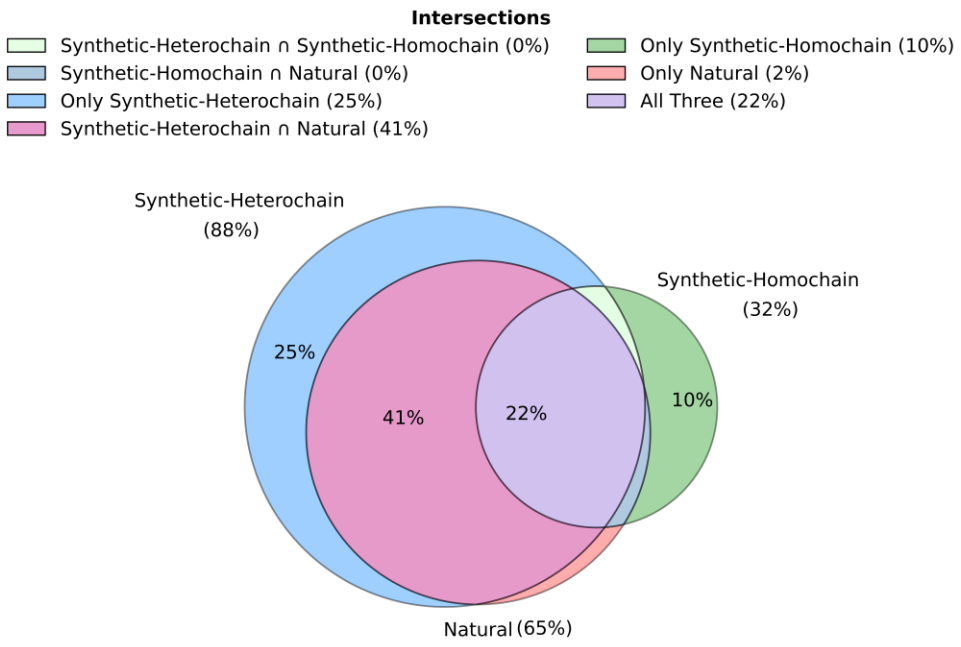
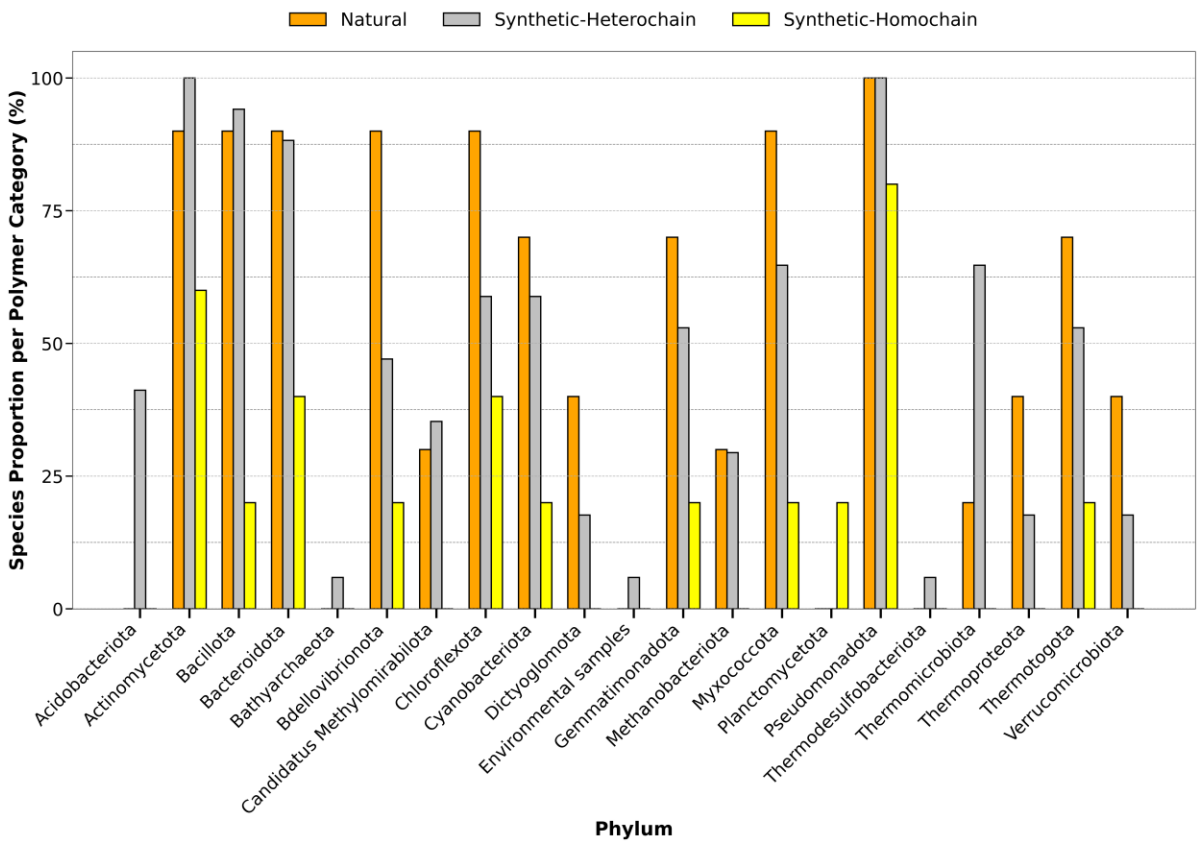

A)
Percentage of Enzyme (%)
0
10
20
30
40
50
PDCOGs
PHA
PHB
PHBV
PHBH
P3HO
P3HP
P4HB
PHBVH
PHO
PHPV
PET
PU
PLA
PCL
PBS
PBSA
PES
Nylon
PVA
PBAT
PHC
PC
PEG
PEA
PEF
PMCL
Ecovio-FT
P(3HB-co-3MP)
O-PVA
PBSet
Impranil
PE
PP
PS
PVC
LDPE
NR
P3HV
PPL
B)
IV Log Scale
Plastics
EC Functions
Oxidoreductases
With NAD(+) or NADP(+) as acceptor
alcohol dehydrogenase
With a cytochrome as acceptor
With oxygen as acceptor
With a quinone or similar compound as acceptor
With other known acceptors
With other acceptors
With oxygen as acceptor
Peroxidases
With an iron-sulfur protein as acceptor
With incorporation of two atoms of oxygen
With 2-oxoglutarate as one donor
With NADH or NADPH as one donor and incorporation of one atom of oxygen
alkane 1-monooxygenase
With oxidation of a pair of donors
Miscellaneous
With oxygen as acceptor
With NAD(+) or NADP(+) as acceptor
With NAD(+) or NADP(+) as acceptor
aldehyde dehydrogenase (NAD(+))
With a disulfide as acceptor
With NAD(+) or NADP(+) as acceptor
With a quinone or related compound as acceptor
With a flavin as acceptor
With other known acceptors
With NAD(+) or NADP(+) as acceptor
With NAD(+) or NADP(+) as acceptor
With oxygen as acceptor
With a quinone or similar compound as acceptor
With a quinone or similar compound as acceptor
With a cytochrome as acceptor
With oxygen as acceptor
With NAD(+) or NADP(+) as acceptor
Methyltransferases
Carboxy- and carbamoyltransferases
Transketolases and transaldolases
Transferring groups other than amino-acyl groups
Hexosyltransferases
Pentosyltransferases
Transaminases
Protein-serine/threonine kinases
Protein-histidine kinases
Nucleotidyltransferases
Transferases for other substituted phosphate groups
Acting on ester bonds
Carboxylic ester hydrolases
Carboxylic ester hydrolases
carboxylesterase
poly(ethylene terephthalate) hydrolase
triacylglycerol lipase
cutinase
poly(3-hydroxybutyrate) depolymerase
poly(3-hydroxyoctanoate) depolymerase
palmitoyl-CoA hydrolase
Phosphoric monoester hydrolases
Phosphoric diester hydrolases
Glycosylases
Glycosidases, i.e. enzymes hydrolyzing O- and S-glycosyl compounds
alpha-amylase/glycogenase
Ether hydrolases
Acting on peptide bonds (peptidases)
Aminopeptidases
Dipeptidases
Dipeptidyl-peptidases and tripeptidyl-peptidases
Serine-type carboxypeptidases
Metallocarboxypeptidases
Omega peptidases
Serine endopeptidases
Cysteine endopeptidases
Metalloendopeptidases
Acting on carbon-nitrogen bonds
In linear amides
In cyclic amides
In phosphorus-containing anhydrides
In ketonic substances
In C-halide compounds
Aldehyde-lyases
Other carbon-carbon lyases
Hydro-lyases
Acting on polysaccharides
Acting on carbohydrates and derivatives
Ligases forming aminoacyl-tRNA and related compounds
Carbon--nitrogen ligases with glutamine as amido-N-donor
Translocases

Figure 4


Intersections
Synthetic-Heterochain ∩ Synthetic-Homochain (0%)
Synthetic-Homochain ∩ Natural (0%)
Only Synthetic-Heterochain (25%)
Synthetic-Heterochain ∩ Natural (41%)
Only Synthetic-Homochain (10%)
Only Natural (2%)
All Three (22%)
Synthetic-Heterochain
(88%)
Synthetic-Homochain
(32%)
25%
41%
22%
10%
Natural (65%)

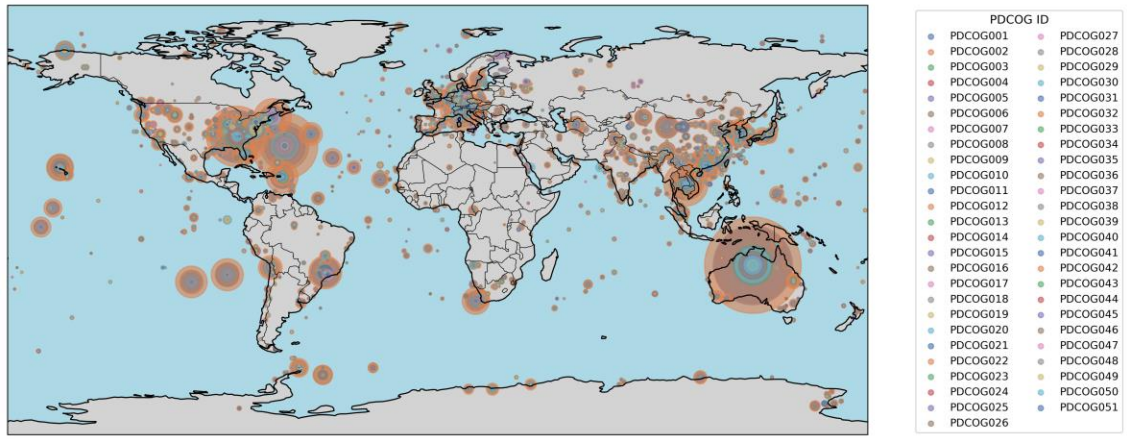
A)


Natural
Synthetic-Heterochain
Synthetic-Homochain
Species Proportion per Polymer Category (%)
100
75
50
25
0
Acidobacteriota
Actinomycetota
Bacillota
Bacteroidota
Bathyarchaeota
Bdellovibrionota
Candidatus Methylomirabilota
Chloroflexota
Cyanobacteriota
Dictyoglomota
Environmental samples
Gemmatimonadota
Methanobacteriota
Myxococcota
Planctomycetota
Pseudomonadota
Thermodesulfobacteriota
Thermomicrobiota
Thermoproteota
Thermotogota
Verrucomicrobiota
Phylum

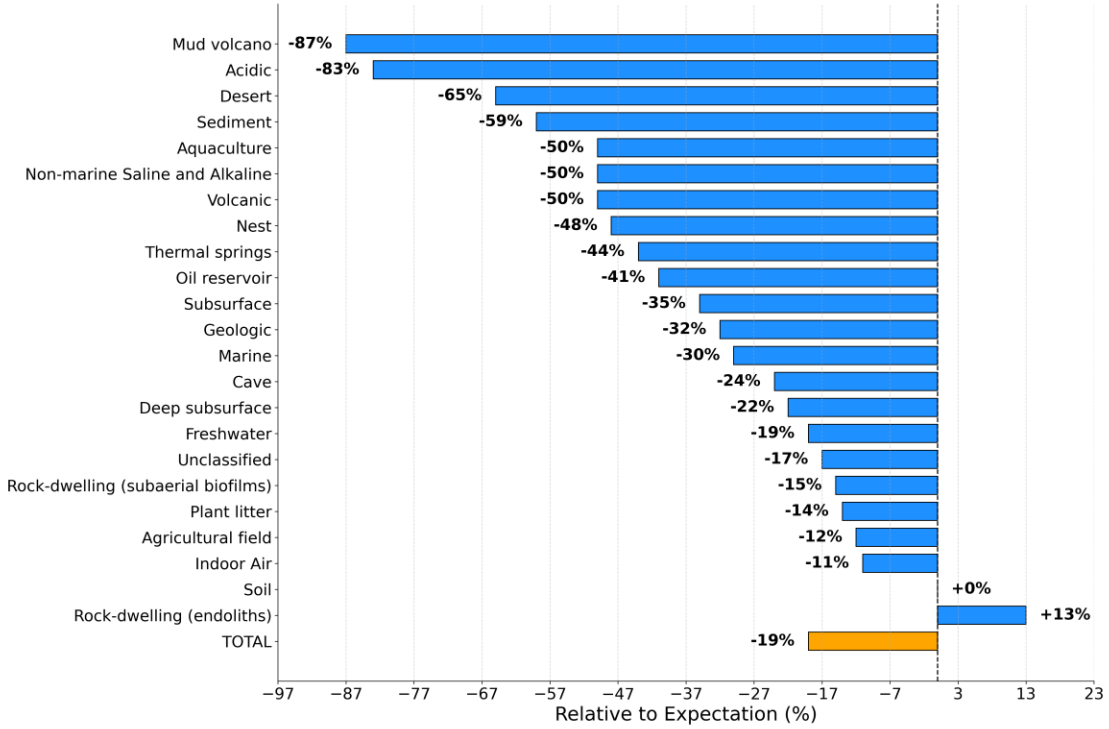
B)

Figure 5

A)


PDCOG ID
PDCOG001
PDCOG002
PDCOG003
PDCOG004
PDCOG005
PDCOG006
PDCOG007
PDCOG008
PDCOG009
PDCOG010
PDCOG011
PDCOG012
PDCOG013
PDCOG014
PDCOG015
PDCOG016
PDCOG017
PDCOG018
PDCOG019
PDCOG020
PDCOG021
PDCOG022
PDCOG023
PDCOG024
PDCOG025
PDCOG026
PDCOG027
PDCOG028
PDCOG029
PDCOG030
PDCOG031
PDCOG032
PDCOG033
PDCOG034
PDCOG035
PDCOG036
PDCOG037
PDCOG038
PDCOG039
PDCOG040
PDCOG041
PDCOG042
PDCOG043
PDCOG044
PDCOG045
PDCOG046
PDCOG047
PDCOG048
PDCOG049
PDCOG050
PDCOG051

B)


Mud volcano -87%
Acidic -83%
Desert -65%
Sediment -59%
Aquaculture -50%
Non-marine Saline and Alkaline -50%
Volcanic -50%
Nest -48%
Thermal springs -44%
Oil reservoir -41%
Subsurface -35%
Geologic -32%
Marine -30%
Cave -24%
Deep subsurface -22%
Freshwater -19%
Unclassified -17%
Rock-dwelling (subaerial biofilms) -15%
Plant litter -14%
Agricultural field -12%
Indoor Air -11%
Soil +0%
Rock-dwelling (endoliths) +13%
TOTAL -19%
−97 −87 −77 −67 −57 −47 −37 −27 −17 −7 3 13 23
Relative to Expectation (%)