PNAS Nexus, 2025, 4, pgaf066
https://doi.org/10.1093/pnasnexus/pgaf066
Advance access publication 18 March 2025
Research Report

Natural-selected plastics biodegradation species and enzymes in landfills

Xiaoxing Lin[a], Rui Zhang[a], Feng-Qi Cui[b], Wenqing Hong[a], Shu Yang[c], Feng Ju[d], Chuanwu Xi[e], Xiao Sun[f] and Liyan Song[a,g,*]

[a]School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China
[b]Institute of Advanced Technology, University of Science and Technology of China, Hefei 230601, China
[c]Institute of Public Health Sciences, Division of Life Science and Medicine, University of Science and Technology of China, Hefei 230026, China
[d]Environmental Microbiome and Biotechnology Laboratory (EMBLab), Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310030, China
[e]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI 48109, USA
[f]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China
[g]Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Science, Chongqing 400714, China
*To whom correspondence should be addressed: Email: 20023@ahu.edu.cn
Edited By Ivet Bahar

Abstract

Biodegradation is a promising and environmentally friendly strategy for plastic pollution management. Landfills decompose municipal solid waste, including almost 50% of global plastic debris and even some of the oldest synthetic plastics, fostering naturally selected plastic biodegradation. Herein, we present a global collection of plastic biocatalytic enzymes from landfills using metagenomics and machine learning. Metagenomic analysis identified 117 plastic-degrading genes, with 39 incorporated in 22 prokaryotic metagenome-assembled genomes (MAGs). A machine-learning approach predicted 978,107 candidate plastic-degrading genes, 712 of which were encoded respectively by 150 MAGs. Our results highlight landfills as reservoirs of diverse, naturally selected plastic-degrading microbes and enzymes, serving as references and/or models for biocatalysis engineering and in situ bioremediation of plastic pollution.

Keywords: landfill, plastic biodegradation, bioresources, metagenomics, machine learning

Significance Statement

Plastic accumulation is a global threat. Although enzymatic and microbial degradation is an efficient method of plastic recycling, progress is hindered by culture-dependent isolation and a limited discovery of novel, powerful enzymes. Landfills, serving as active anaerobic refuse-decomposing bioreactors, receive nearly half of the global plastic waste. In this special system, utilizing publicly available metagenome data, we obtained 24 enzymes and 27 species for plastic biodegradation via metagenomics, and 712 enzymes and 150 metagenome-assembled genomes with plastic-degrading potential via machine learning. Our study highlights landfills as promising reservoirs for plastic-degrading bioresources and provides a foundation for advancements in biocatalysis engineering and bioremediation to combat plastic pollution.

Introduction

Plastic pollution is a serious threat to environmental sustainability and public health. By 2050, an estimated 11 billion metric tons of plastic are projected to accumulate in the environment (1). In 2019, ~353 million metric tons of plastic waste were produced worldwide, with 4% of plastic waste leaking into terrestrial environments and 2% accumulating in aquatic environments, including oceans, rivers, and coastlines (2). There are 300,000–500,000 landfills worldwide (3), which account for 80% of land-based plastic pollution entering aquatic systems (4), with almost half of the global plastic waste disposed of in landfills (5). Landfills serve as major reservoirs of plastic debris, spreading contamination to surrounding environments via leaching (6), airborne pathways (7), and landfill mining. During refuse decomposition, plastics fragment into microplastics, which bioaccumulate in the food chain (8), act as carriers for various contaminants, and are transported across long distances (9). Plastics also threaten public health by leaching endocrine-disrupting chemicals (10).

Addressing the threat of plastic pollution requires highly effective, safe, and economically feasible methods for plastic control and upcycling. Enzymatic degradation offers remarkable potential in this regard (11). For example, poly(3-hydroxyoctanoate) (PHA)

Competing Interest: The authors declare no competing interests.
Received: July 29, 2024. **Accepted:** February 6, 2025

depolymerase isolated from *Bacillus* achieved ~98% degradation of poly(3-hydroxybutyrate) (P3HB) (12). Similarly, an engineered variant of PETase demonstrated an average catalytic activity of 23.4 $mg_{(PET)}/h/mg_{(enzyme)}$, showing a 40-fold improvement in PET biodegradation compared with the wild type (13). Landfill microbiota has developed distinct functional potential in plastic biodegradation owing to selective effects from plastic accumulation and depolymerization (14). The landfill plastisphere, shaped by highly dynamic and activated physical, chemical, and biological conditions—such as highly organic matter (514.68 mg/g total organic matter, 234.86 g/kg total carbon) (15), a wide range of pH (4.5–9) (16), elevated temperature (up to 70 °C) (17), and diverse microbiota (18, 19)—develops unique characteristics during refuse decomposition. A previous study highlights the importance of plastic biodegradation in landfill carbon cycling (20). Over the past decade, microorganisms and biomolecules with plastic-degrading potential have been isolated and identified from landfills to expand available bioresources (21, 22). Given these strict selection pressures, we hypothesize that landfills harbor extensive microbiomes featuring promiscuous enzymes suited for plastic biodegradation (18).

Despite sporadic progress, the taxonomic, genetic, and functional diversity of microbial communities involved in plastic depolymerization remains poorly understood (23). Limited experimental evidence and the constraints of homology-based algorithms mean that our understanding of protein functions has not kept pace with the growing number of protein genes. Moreover, the low accuracy of protein function annotation hinders the effective application of biotechnology for environmental sustainability. Recently, a machine-learning (ML) model, contrastive learning–enabled enzyme annotation (CLEAN) (24), has well uncovered the functional diversity of proteins, surpassing other state-of-the-art ML tools. Combining bioinformatics with AI can help narrow the search for plastic-degrading enzymes and enhance the efficiency of known enzymes. Unlocking valuable but uncharacterized enzymes from microbial dark matter can advance biotechnology, for recycling and upcycling plastic wastes. Therefore, exploring landfill bioresources and associated habitats—termed the landfill system in this study—offers potential for energy-conserving biocatalytic approaches to manage and upcycle plastic waste.

Herein, we collected DNA samples from landfill systems on a global scale. Leveraging the machine-learning model CLEAN, we expanded the biocatalytic list within the landfill-associated plastisphere, surpassing classic database-dependent annotation by 35-fold. Furthermore, we inferred the catalytic activity and microbial hosts of 712 predicted proteins using tertiary structure modeling, superposition, and MAG reconstruction. This number was 6-fold higher than that obtained by searching against PlasticDB. The discovery of vast majority of landfills plastic biodegradation species and enzymes represents a vital step toward combating global plastic–microplastic–nanoplastic pollution, promoting sustainable practices, and advancing environmental and industrial sciences. It holds the potential to revolutionize waste management and create a cleaner, healthier planet.

Results

Preliminary discovery of plastic-degrading taxa and genes

The microbiomes of the investigated landfill system spanned 21 cities across six countries and six distinct environmental habitats, including refuse ($n = 45$), leachate ($n = 37$), bioreactor ($n = 4$), activated sludge from leachate treatment ($n = 6$), and landfill airborne particle ($n = 6$). These data are publicly available, and some of which were previously published. Detailed information on sampling geography and data deposition is provided in Fig. S1 and Dataset S1. Through annotation against PlasticDB (25) and the NCBI nonredundant (NR) database, we identified 24 enzymes involved in the biodegradation of 22 plastic chemicals and their taxonomical hosts (Fig. 1) across the global landfill microbiome (Dataset S2). Notably, among these enzymes, 3-hydroxyvalerate dehydrogenase (average reads per kilobase per million mapped reads [RPKM] value of 1.84 and maximum of 45.26), polyester hydrolase (1.60, 47.76), and cutinase (1.60, 108.49), were the most abundant enzymes across various landfill habitats, with the highest abundance found in refuse and leachate. Microbial hosts of these enzymes were mainly bacterial, including genera such as *Paracoccus*, *Halopseudomonas*, and *Saccharomonospora*.

Potential plastic-degrading protein prediction via machine learning

To further explore plastic-degrading bioresources, we employed two methods to annotate enzyme commission (EC) numbers for protein sequences: classic bioinformatics (a combination of homology-based annotation and Hiden Markov Model prediction) and machine-learning prediction (CLEAN (24) and ProteInfer (26) models).

Using the CLEAN model, we identified 978,107 sequences predicted with target EC numbers (Dataset S3), compared with 117 proteins annotated by PlasticDB (Dataset S2), 55,656 putative sequences (excluding the 117 annotated by PlasticDB) via KofamKOALA, 27,682 putative sequences via the Kyoto Encyclopedia of Genes and Genomes (KEGG) and 5,399 putative sequences via ProteInfer. Among these tools, CLEAN demonstrated superior predictive performance, identifying the largest number of putative plastic-degrading enzymes (Figs. 2, S2, and S3), suggesting a substantial reservoir of plastic-degrading genes yet to be unveiled.

Notably, 58 out of the 117 putative sequences identified through PlasticDB were also detected as plastic-degrading genes through the CLEAN model, affirming the reliability of the model. The remaining 59 sequences missed by CLEAN may be attributed to limited information on the EC assignment. Similarly, KEGG-based annotation and ProteInfer also cannot fully cover these 59 sequences, indicating the need for additional EC information to enhance our EC-number list of plastic-degrading enzymes. Importantly, the large counts of putative sequences detected via CLEAN reflect the untapped potential of microbiome protein functional dark matter.

Moreover, all approaches—CLEAN, KEGG annotation, and HMM prediction—revealed that airborne particles harbored a higher abundance of plastic-degrading genes. Their average RPKM values represented 43, 38, and 71% of the total average, respectively, higher than that of PlasticDB annotation (10%). However, BLASTp annotation against PlasticDB showed that refuse was enriched with more plastic-degrading genes, representing 57% of the total average (Fig. 2). We compared the capabilities of multiple approaches in unveiling microbiome proteome dark matter and validated our choice of CLEAN, as a powerful tool for screening plastic-degrading enzymes in landfill systems. Specifically, 31,989 plastic-degrading genes belonging to 6 EC numbers were detected via CLEAN prediction for downstream analysis, namely EC 3.1.1.101 (poly(ethylene terephthalate) hydrolase, $n = 5{,}700$), EC 3.1.1.102 (mono(ethylene terephthalate) hydrolase, $n = 5{,}669$), EC 3.4.21.64 (peptidase K, $n = 352$, represents part of proteases), EC 3.5.1.13 (aryl-acylamidase,

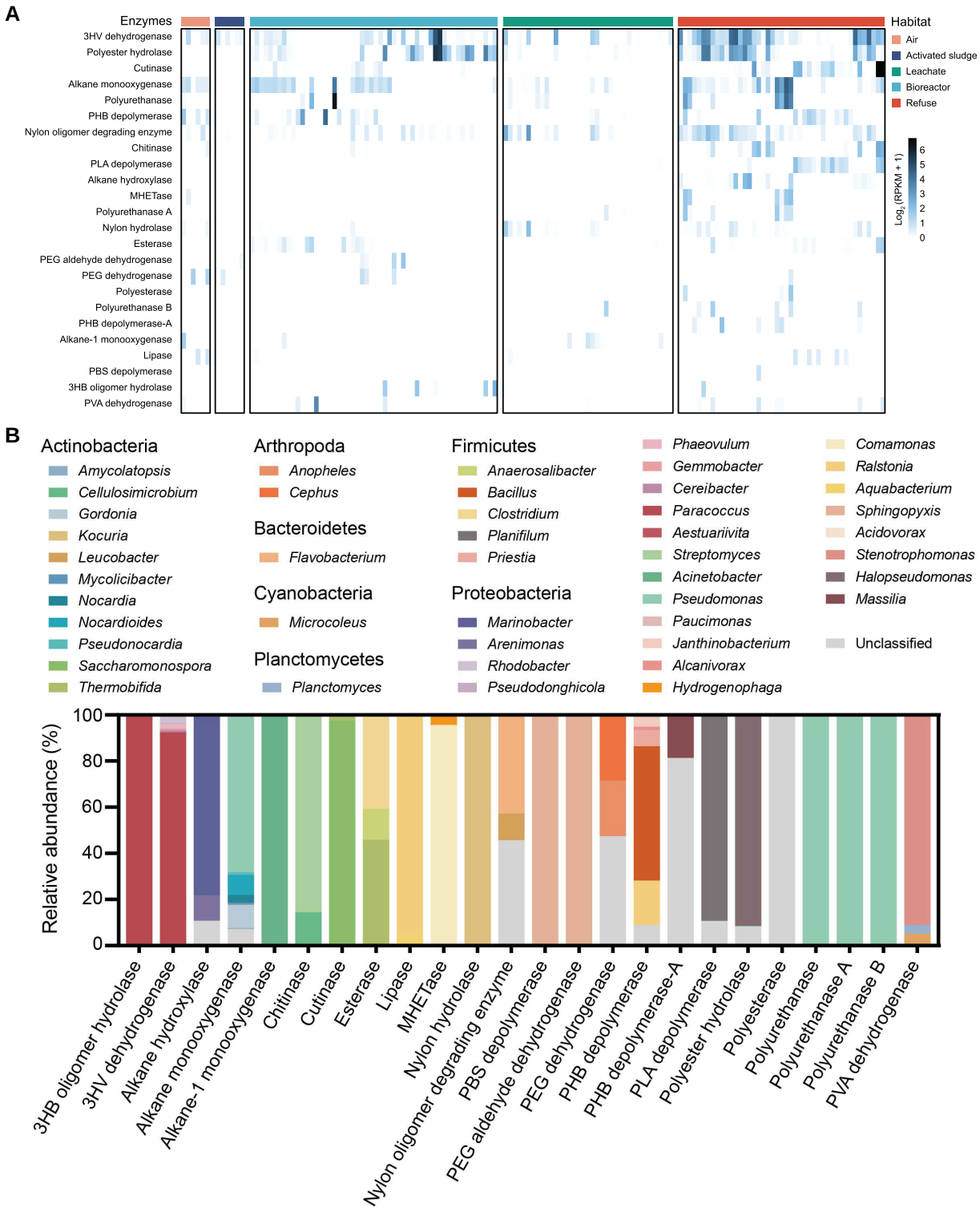

Fig. 1. Plastic-degrading genes and their microbial hosts across landfill-associated habitats. A) Relative abundance of plastic-degrading genes annotated against PlasticDB, and B) composition of taxonomical hosts of these genes. 3HV, 3-hydroxyvalerate; MHET, mono(2-hydroxyethyl) terephthalic acid; PBS, poly(butylene succinate); PEG, polyethylene glycol; P3HB, poly(3-hydroxybutyrate); PLA, polylactic acid; PVA, polyvinyl alcohol.

$n = 15{,}598$), EC 3.5.2.12 (6-aminohexanoate-cyclic-dimer hydrolase, $n = 1{,}027$—both EC 3.5.1.13 and EC 3.5.2.12 represent the amidase group responsible for hydrolyzing C–N bonds other than peptide bonds), and EC 3.7.1.7 (β-diketone hydrolase, $n = 3{,}643$, represents a part of carboxylesterases). All these distinct enzymes could be categorized as hydrolases.

Functional identification of putative plastic-degrading proteins

To discover as many diverse and novel plastic-degrading proteins as possible, we focused on those EC-number proteins detected solely via CLEAN (the six specific EC numbers mentioned above) for further analysis. To enhance the accuracy of isozyme

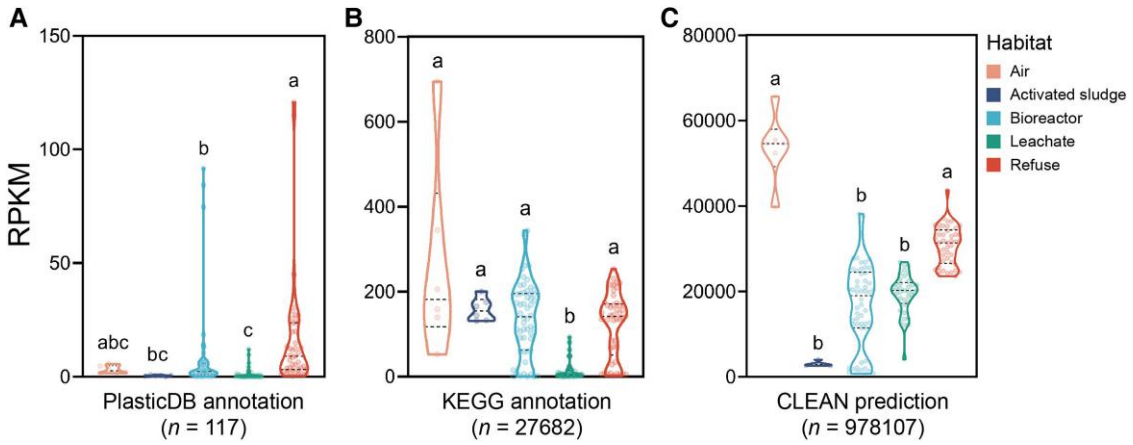

Fig. 2. Abundance comparison of putative plastic-degrading genes among different processes. The processes include A) BLASTp annotation against PlasticDB, B) BLASTp annotation against KEGG enzyme database, and C) machine-learning prediction via CLEAN model. Letters indicate the statistically significant difference of pairwise comparison between enrichments of plastic-degrading genes in each habitat ($P < 0.05$).

detection and minimize potential false positives, these CLEAN-extracted sequences were filtered by screening for conserved protein family domains. According to the functions of annotated domains, we divided these filtered sequences into four categories of enzymes: protease (Fig. 3), hydrolase (Fig. 4), amidase (Fig. 5), and carboxylesterase/lipase (Fig. 6). To functionally identify each category of enzymes, we analyzed their genetic diversity, conserved domains, motif arrangements, and representative tertiary structure. Quality assessment of the protein models confirmed the acceptable accuracy of 3D models for further superposition (Fig. S4 and Table S1). The consensus motif sequences are listed in Table S2. While most proteins within the family exhibited the motif arrangement pattern, many proteins in each phylogeny displayed close evolutionary correlations and similar structural folding despite belonging to different families.

The phylogram of proteases consisted of 138 proteins, mainly affiliated with the peptidases S8 and S53 families, where we selected two proteins (named Pr1 and Pr2) for structural modeling (Fig. 3). Both of them had exhibited the typical triad of Ser-His-Asp, and their domain regions showed high similarity in superposition (Fig. S5aA and B) with their reference proteases (protein ID: 00040 and 00153). While both Pr1 and Pr2 contained conserved domains from the peptidases S8 and S53 superfamilies, Pr1 was predicted as peptidase K (EC 3.4.21.64) using the CLEAN model, consistent with predictions of high-order structure comparison, and Pr2 was predicted as aspartylglucosaminidase (EC 3.5.1.26), polyethylene terephthalate hydrolases (PETase, EC 3.1.1.101), and feruloyl esterase (EC 3.1.1.73), showing its functional diversity and potential for plastic-degrading activity.

The hydrolase phylogram comprised 144 proteins covering three conserved protein domain families, with the majority belonging to the alpha/beta hydrolase (ABH) superfamily (142 proteins). Three representative proteins (Hy1, Hy2, and Hy3) were selected for modeling. Proteins from the metallo-hydrolase-like superfamily were excluded from modeling owing to the lack of related domains in PlasticDB for reference. All representatives exhibited the ABH fold and Ser-His-Asp catalytic triad, showing structural similarity to their reference proteins with close phylogenetic correlations (Figs. 4, S3aC, S3bD and E). Notably, Hy2 is almost perfectly superimposed on a mono(ethylene terephthalate) hydrolase (MHETase) (protein ID: 00084). The superposition also suggested that these representatives may resemble polyurethane (PU) esterase (protein ID: 00013), MHETase, and PHA depolymerase (protein ID: 00016) in catalytic function. In addition, these enzymes appeared to perform diversely in catalyzing a range of substrates. According to the CLEAN model, Hy1 was predicted as MHETase, Hy2 as P3HB depolymerase (EC 3.1.1.75), PHA depolymerase (EC 3.1.1.76), and MHETase, while Hy3 was predicted as oxidized polyvinyl alcohol (PVA) hydrolase (EC 3.7.1.7).

The amidase phylogram contained 168 proteins across four conserved protein domain families, primarily dominated by the amidase superfamily and amidohydrolase family. Amidase/polyamidase enzymes specifically catalyze the hydrolysis of amide bonds in nonprotein substrates, such as nylon. We selected two representative proteins, Am1 and Am2. Am1 was predicted as 6-aminohexanoate-cyclic-dimer hydrolase (NylA, EC 3.5.2.12), and showed a strong alignment with polyamidase (numbered 3 in Fig. 5, protein ID: 00172), which is closely related to aryl–acyl amidases (27). Both of Am1 and Am2 contained the classical Lys-*cis*-Ser-Ser of the amidase signature family, along with the conserved GGSS(S/G)GS motif (Fig. S6). Additionally, Am2 was predicted as aryl–acyl amidase (EC 3.5.1.13), responsible for hydrolyzing amide bonds between aryl and acyl groups. The tertiary structure of Am2 displayed a barrel of seven beta sheets connected via alpha helices, similar to the well-characterized $(\beta/\alpha)_8$ barrel found within the amidohydrolase superfamily (28).

We identified 262 candidate carboxylesterase/lipase proteins across eight different conserved family domains and constructed a phylogram (Fig. 6). Six representative proteins (CL1 to CL6) were selected for homology modeling (Fig. S7). Three-dimensional superposition showed shared typical folding in these proteins, allowing their fundamental enzymatic activities. All of these representative proteins, except CL3 (Ser-His-Gly), possessed the catalytic triad of Ser-His-Asp. Additionally, all representatives possessed the GxSxG motif (Fig. S8), which is common in most carboxylesterase/lipase proteins. Structural comparison with reference proteins accessed from PlasticDB or the Protein Data Bank showed high similarity in conserved domains. For example, CL1 and CL2 were predicted as PETase via CLEAN, and CL4 was predicted as oxidized PVA hydrolase, closely aligned with PlasticDB reference proteins, PETase (protein ID: 00196), polyester hydrolase (protein ID: 00140), and carboxylesterase (protein ID: 00090), respectively (Fig. S5c and d). Protein CL5, with a P3HB depolymerase-relevant LpqC domain, showed a similar domain fold to 3-hydroxybutyrate oligomer hydrolase (Protein ID:

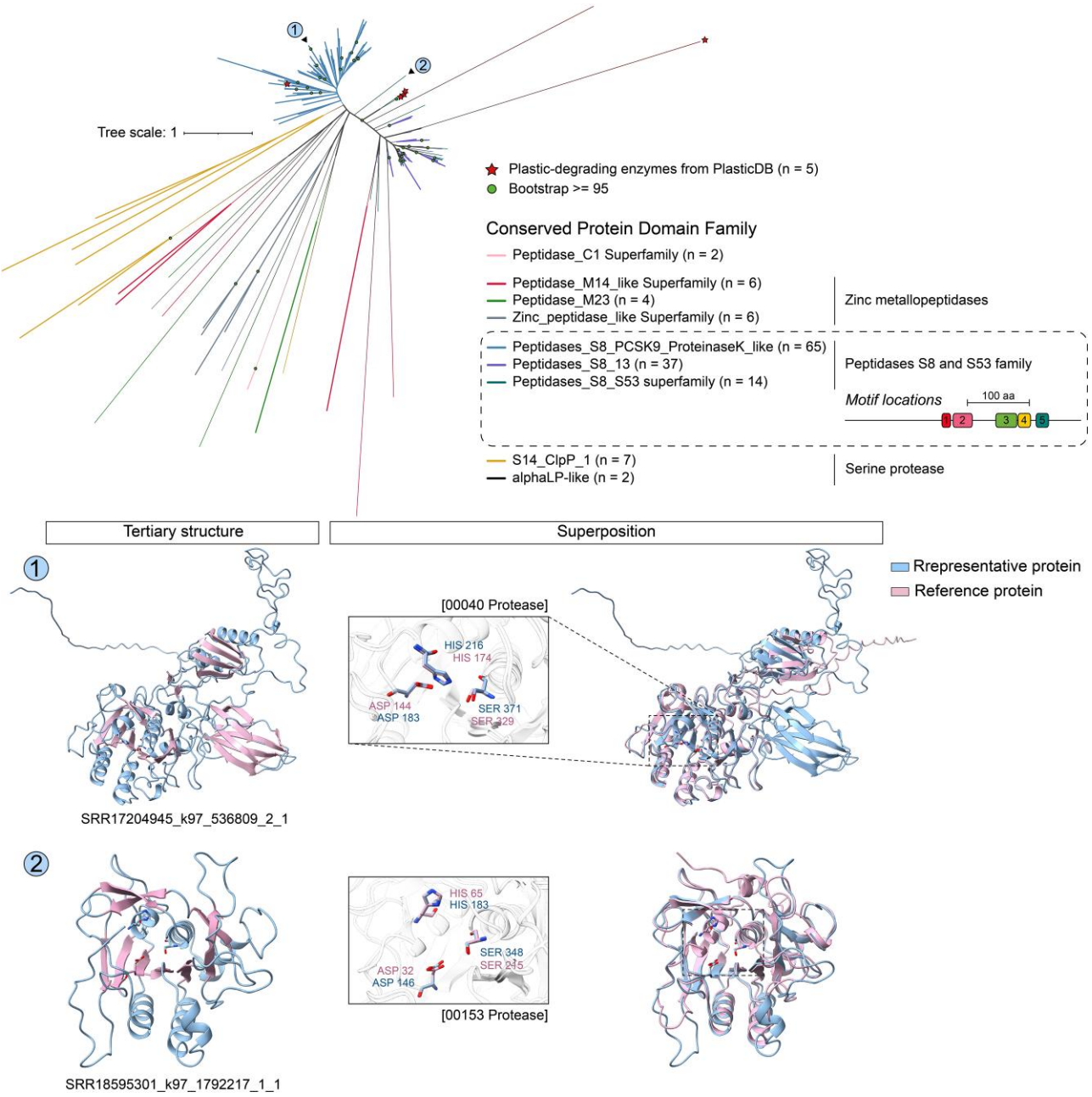

Fig. 3. Protein characteristics of putative protease sequences. The upper panel displays the phylogeny composed of putative proteases predicted by CLEAN and plastic-degrading proteases archived in PlasticDB. Each line represents a protein, colored by their conserved protein domain family. Legend includes the motif pattern of the major protein family (the motif sequences are listed in Table S2). The below panel shows the tertiary structure modeling and superposition analyses conducted on representative proteins.

00070), despite an imperfect ABH fold (composed of seven β-strands instead of eight; Figs. S5dK and S7E).

ML produced a more functional proteome than classic metagenomic annotation, suggesting that landfill microbes have evolved a more diverse and versatile plastic-depolymerizing ability. Overall, the detailed analysis of these enzyme categories provides valuable insights into their structural and functional characteristics, aiding in the identification of potential candidates for plastic degradation.

Taxonomical hosts of candidate plastic-degrading proteins

A total of 1,537 metagenome-assembled genomes (MAGs) were recovered, meeting the criteria of >90% completeness and <5% contamination. Among these, 149 MAGs were assigned putative plastic-degrading genes based on CLEAN model predictions and included in the phylogeny (Fig. 7).

These MAGs contained at least one of the investigated genes for plastic degradation and were abundant in refuse (maximum RPKM of 43.45) and bioreactors (18.03) inoculated with municipal solid waste (MSW) and landfill leachate recirculation. They were mainly affiliated with the phyla Pseudomonadota and Actinomycetota (Fig. 7). Carboxylesterase/lipases or ABH-encoding genes were mainly enriched in the phylum Pseudomonadota, while amidase and protease genes were detected in both Actinomycetota and Pseudomonadota, the major carriers of putative plastic-degrading genes. Notably, 49 MAGs carried at least one putative

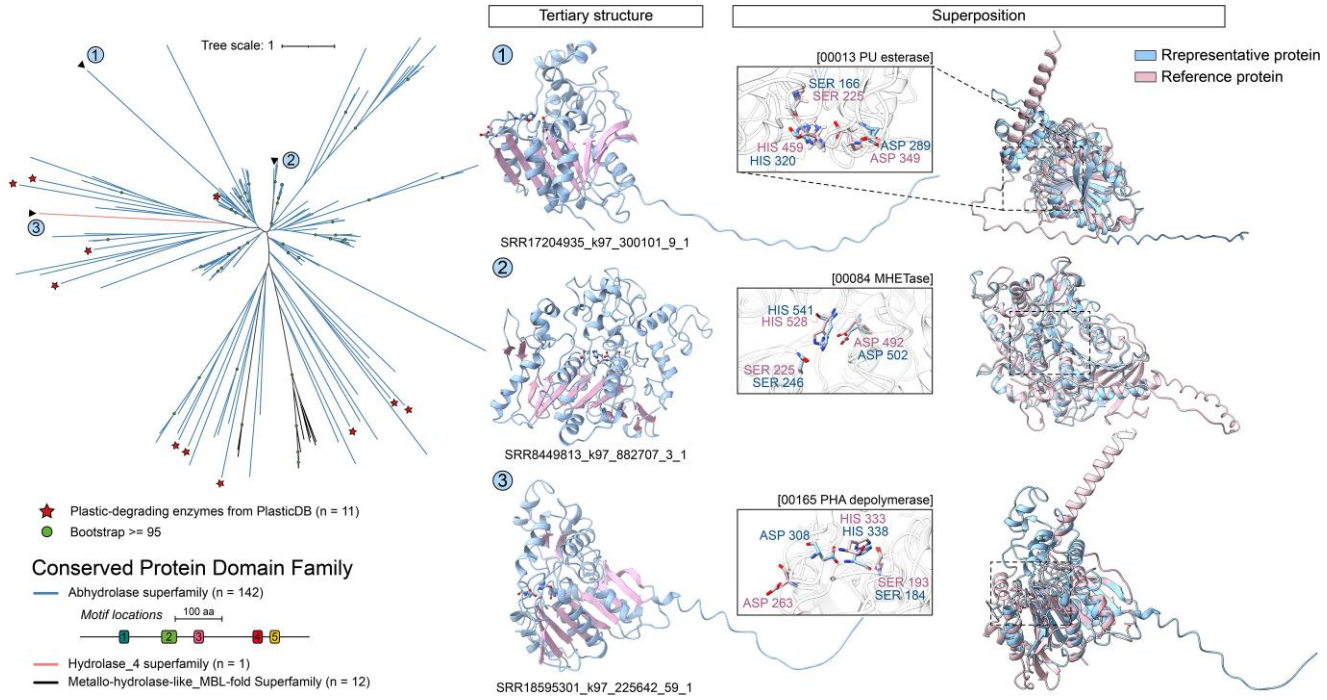

Fig. 4. Protein characteristics of putative hydrolase sequences. The left panel displays the phylogeny composed of putative hydrolase predicted by CLEAN and plastic-degrading hydrolase archived in PlasticDB. Each line represents a protein, colored by their conserved protein domain family. Legend includes the motif pattern of the major protein family (the motif sequences are listed in Table S2). The right panel shows the tertiary structure modeling and superposition analyses conducted on representative proteins.

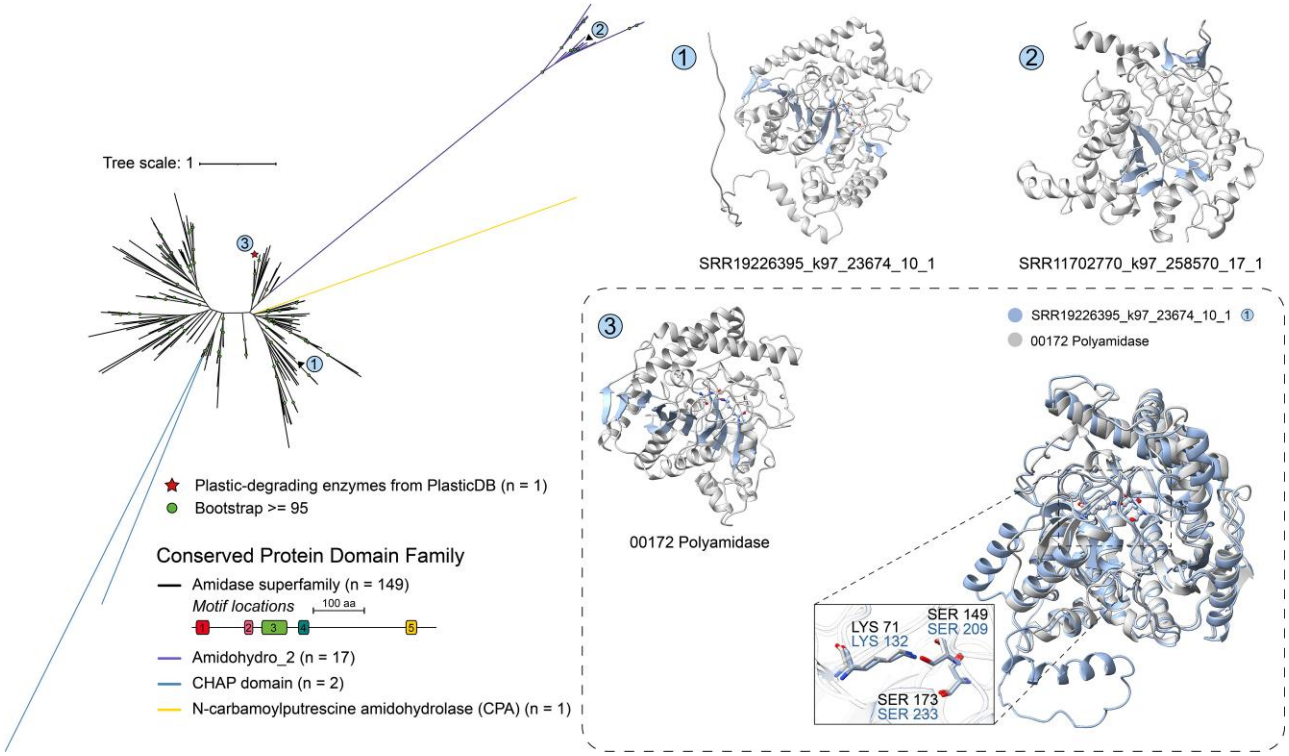

Fig. 5. Protein characteristics of putative amidase sequences. The left panel displays the phylogeny composed of putative amidase predicted by CLEAN and plastic-degrading amidase archived in PlasticDB. Each line represents a protein, colored by their conserved protein domain family. Legend includes the motif pattern of the major protein family (the motif sequences are listed in Table S2). The right panel shows the tertiary structure modeling and superposition analyses conducted on representative proteins.

plastic-degrading gene. For example, *Saccharomonospora azurea* contained two peptidase genes and one amidase gene, with a high relative abundance of 43.47 in refuse. Furthermore, MAG1485 (assigned to Pseudomonadota phylum, *FEN-1219* genus), MAG613 (Myxococcota phylum, Polyangiaceae family), and MAG1354 (Verrucomicrobiota phylum, *Rariglobus* genus) carried

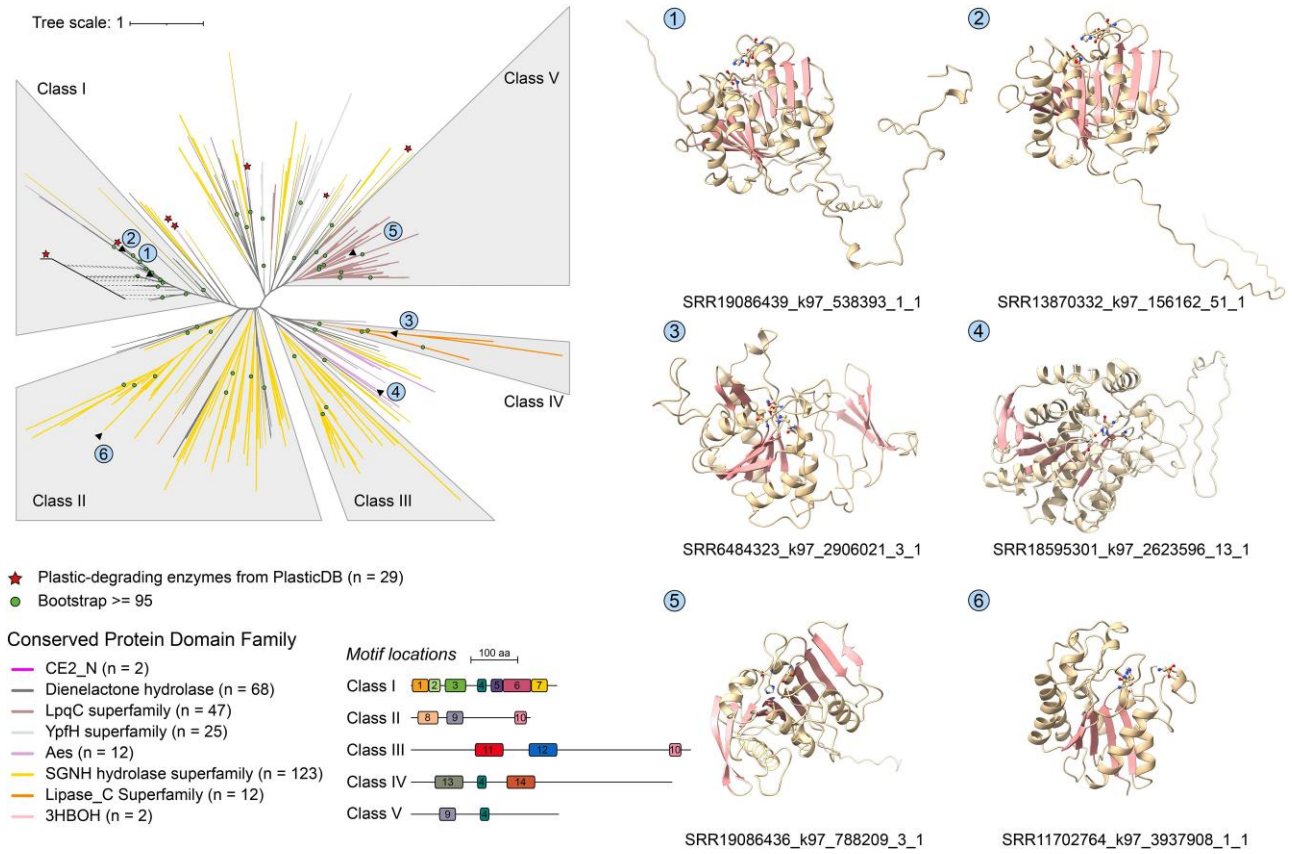

Fig. 6. Protein characteristics of putative carboxylesterase and lipase sequences. The left panel displays the phylogeny composed of putative carboxylesterase/lipase predicted by CLEAN and the plastic-degrading one archived in PlasticDB. Each line represents a protein, colored by their conserved protein domain family. Legend includes the motif pattern of the major protein family (the motif sequences are listed in Table S2). The right panel shows the tertiary structure modeling and superposition analyses conducted on representative proteins.

six plastic-degrading genes. Among these, MAG1485 had five hydrolase genes related to plastic catalysis, while MAG613 and MAG1354 mainly carried carboxylesterase genes. In addition, three archaeal MAGs (MAG527, MAG900, and MAG920) possessing plastic-degrading genes were discovered, assigned to the genera *SM1-50* (Thermoplasmatota phylum), *Methanosarcina* (Halobacteriota phylum), and *Methanobacterium* (Methanobacteriota phylum), respectively.

Discussion

Plastic pollution demands multifaceted strategies for effective management of this urgent issue, such as reducing the demand and usage of plastics, designing better materials, and transforming plastics into value-added products (29). However, the management of long-existing plastic debris in the environment is a serious challenge owing to its widespread and random distribution and the difficulty of collective engineering treatment. Understanding plastic biodegradation in specific environments provides promising avenues for in situ bioremediation of plastic pollution. Learning from nature is a useful practice for advancing science and technology. Given that landfills and activated bioreactors for MSW decomposition serve as the reservoir of plastic debris collection and biodegradation, it is reasonable to discover diverse and powerful plastic-degrading bioresources. Moreover, the varying pH (4.5–9) (30) and temperature exceeding 70 °C (31) from hydrolysis and acetogenesis to the methanogenesis phase during refuse decomposition, suggest the presence of thermophilic and acid–base-resistant enzymes in landfills, which may be suitable for industrial applications under high-temperature or fluctuating pH conditions (32).

Despite our research highlighting the potential of the landfill microbiome for plastic biodegradation as revealed via ML tools, it was limited by the absence of experimental validation to identify the enzymes from uncultured biomes. This limitation underscores the critical need for subsequent studies to provide empirical evidence to confirm the predicted protein functions. Given the huge collection of putative enzyme-coding genes, validation will require a long-term investment and joint efforts. Herein, we identified 712 proteins and 150 MAGs using the CLEAN model, an earlier published model by Yu et al. (24), from global landfill-associated systems as novel plastic-degrading microbial bioresources. According to the substrates that reference enzymes target, these identified 712 proteins are possibly responsible for the decomposition of polylactic acid, PU, PET, PHA, P3HB, poly(butylene adipate-co-terephthalate) (PBAT), and nylon. Omura et al. (33) reported that bioplastics, including P3HB, poly(3-hydroxybutyrate-co-3-hydroxyvalerate), poly[(R)-3-hydroxybutyrate-co-4-hydroxybutyrate], polycaprolactone, poly (butylene succinate), and PBAT, can be decomposed at the deep-sea floor, while poly(L-lactic acid), polyethylene, polypropylene, polystyrene, and PET cannot. Metagenomics showed that the QZLD01 order carries >10 genes for PHA depolymerase, while in our study, MAG806, also assigned to the QZLD01 order, was identified to possess a putative gene encoding MHETase. Additionally, in contrast with soil and marine, where there were more oxidoreductases predicted as plastic-degrading enzymes (34), our predictive result indicated that landfills harbor a significant number of hydrolases for

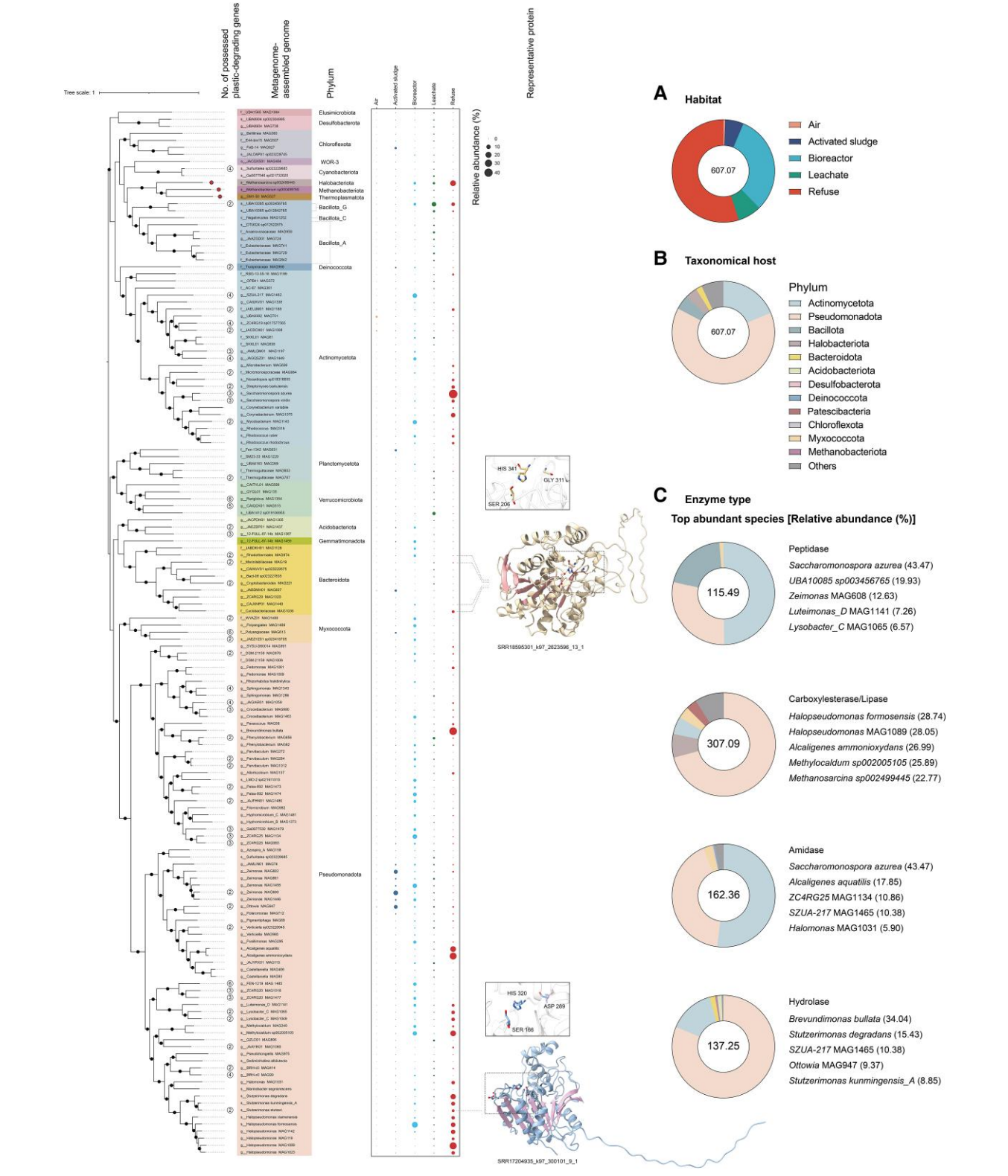

Fig. 7. Composition and distribution of species that host plastic-degrading genes/putative enzymes. The phylogeny notes the archaeal genomes with colored circles, the number of possessed plastic-degrading genes (>1) by each MAG and shows the abundance of MAGs in each habitat. Five of these MAGs may encode two aforementioned enzymes. Bootstrap values indicated by black dots, only bootstrap >95 showed. Pie charts display A) the habitat assignment for plastic-degrading enzymes, B) phyla composition of each enzymatic category, and C) the five most popular genomes from each enzymatic category. The total RPKM value is shown at the center of the corresponding pie chart.

plastic biodegradation. These results suggested the expandability of catalytic function within landfill microorganisms under natural selection.

However, the number of plastic-degrading enzymes likely exceeds our predictions. First, only a minor fraction (i.e. <1%) of the total microorganisms on earth have been cultured (14), meaning that

the plastic-degrading microbes, and their biodegradation proteins we have unraveled in this study represent just the tip of the iceberg. Second, given the structural similarity between plastics and biomass, enzymes responsible for biomass degradation in landfill microbiomes may also contribute to plastic biodegradation (35). In addition, cellulose-degrading functions exist in more taxa than currently known, including Firmicutes, Bacteroidetes, and Fibrobacteres (36), implying the potential for other enzymes to aid in plastic degradation. Besides enzymes that catalyze substrates structurally similar to plastics, promiscuous enzymes with multiple functions should also be included to expand the range of available enzymes, whether evolved naturally or genetically engineered. In this study, abundant genes were assigned to multiple EC numbers, accounting for 18.60% of total nonredundant protein sequences, suggesting the presence of promiscuous enzymes. Multifunctional enzymes are widespread in nature, adapting to diverse environments (37), and offering valuable evolutionary insights for engineering plastic-degrading proteins. Additionally, we need an in-depth understanding of the molecular phenotype of plastic-degrading proteins, including their structural and functional traits, as well as the genotype-to-phenotype correlation in plastic-degrading microorganisms.

In summary, these bioresources compile a promising pool for the biodegradation of landfill plastics and their derivatives. This study aims to provide valuable insights for promoting plastic pollution management and upcycling through the following: (i) the use of synthetic biology and protein engineering to expand the biocatalyst list, (ii) efficient enzyme prediction based on AI, and (iii) the development of accurate bioindicators for monitoring plastic pollution. Compared with studies on plastic pollution in aquatic and terrestrial ecosystems, research on landfill systems is limited, largely owing to the high cost of landfill sampling. Herein, we sampled three large-scale landfills and combined them with global landfill system samples. Although the sample size is 148, it provides valuable insights into plastic biodegradation in landfill systems. We anticipate that increased sampling from old landfills, where microbiota have been undergoing long-term plastic biodegradation, will offer further insights into the microbial biodegradation of even the first-generation plastics on earth.

Methods

Data collection

As of December 3, 2022, we screened 148 metagenomic samples from landfills and landfill-associated environments (named landfill systems) available in NCBI to assess the global occurrence and distribution of plastic-degrading proteins. Detailed information on metagenomic samples is listed in Dataset S1. Briefly, these metagenomic samples were collected from 21 cities across six countries (including China, CN; Italy, IT; Canada, CA; Great Britain, GB; Jamaica, JM; India, IN; Fig. S1), and from different habitats, including landfill refuse ($n = 45$, abbreviated as R), landfill leachate ($n = 37$, L), bioreactors operated by MSW decomposition and leachate recirculation ($n = 54$, B), activated sludge from leachate treatment plants ($n = 6$, AS), and landfill airborne particle (PM 2.5, $n = 6$, A). Although our dataset includes samples from multiple countries, a large proportion (74%) is from China, implying limited global coverage in our study. Despite the high cost of landfill sampling, which has resulted in limited attention to landfill systems, we believe that these diverse samples provide valuable insights into landfills as powerful plastic-degrading reactors and bioresource reservoirs for plastic-degrading microorganisms.

Curation of EC number of plastic-degrading enzymes

We searched each protein name in PlasticDB against the BRENDA Enzyme Database (https://www.brenda-enzymes.org/index.php) and KEGG to obtain the corresponding EC numbers and then verified them according to the descriptions of EC numbers in these databases or relevant publications to filter out unrelated EC numbers. This process allowed us to preliminarily summarize the currently known EC numbers of plastic-degrading enzymes based on PlasticDB inclusion. Additionally, we added six additional plastic-degrading enzymes and their corresponding EC numbers based on previous publications. These EC numbers basically cover the catalytic reactions related to plastic biodegradation (see Dataset S3 for details). While other databases, such as PMBD and PAZy, also contain relevant data on plastic biodegradation, this study adopted PlasticDB as the primary data source owing to its similar collection size of confirmed enzymes. However, we recommended considering multiple comprehensive and representative databases to fully uncover the functional potential of the uncultured microbiome.

Taxonomical and functional annotation

Raw reads were subjected to quality control, filtering, and trimming using Fastp (v0.20.0; https://github.com/OpenGene/fastp). Then, BWA (v0.7.9a) was used to obtain host decontamination and clean data, followed by assembly of contigs using MEGAHIT (v1.1.2, parameter: -min-contig-len 300). Open reading frames (ORFs) from the assembled contigs were predicted using Prodigal (v2.6.3; https://github.com/hyattpd/Prodigal). These ORFs were clustered at 99% sequence similarity using CD-HIT (v4.6.1, http://www.bioinformatics.org/cd-hit/) to generate an NR gene catalog, where the longest sequence in each cluster served as the representative sequence of each gene. Annotation of these assemblies was performed with DIAMOND (v2.0.13, parameter: blastp, $e \leq 10^{-5}$, identity $\geq 80\%$) against the following databases: (i) the NCBI NR protein database, for taxonomical assignment; (ii) the KEGG for preliminary catalyzing functional profiling of each gene via EC-number assignment, with genes annotated with these EC numbers being recruited; and (iii) PlasticDB (https://plasticdb.org/) for annotation of plastic-degrading proteins, which, as of the analyses, cataloged 573 microorganisms and 174 proteins for plastic degradation.

We employed PlasticDB annotation and KEGG annotation (both of them were using BLASTp mode), and Hidden Markov Model prediction (using KofamKOALA), representing the traditional method, to compare their predictive power with that of the state-of-the-art method. Briefly, protein sequences with low homology to those in PlasticDB (identity < 80%) were further annotated for EC numbers using KofamKOALA (38) (default option, $e \leq 0.01$), followed by recruitment of proteins with the target EC.

The abundance of genes was normalized based on the RPKM calculated using SOAPaligner with default parameters. The relative abundances of genes, i.e. percentages of the total number, were calculated based on the RPKM values.

Enzyme prediction based on machine-learning model

To accurately discover the vast majority of plastic-degrading proteins in landfills, we employed two machine-learning models, ProteInfer (26) and CLEAN (24), utilizing their official open-source pretrained weights to predict the EC number of protein sequences.

ProteInfer predicts protein functions (such as EC numbers) by learning the mapping between full-length protein sequences and functional annotations based on deep convolutional neural networks. We split a total of 47,951,567 NR gene sequences into 480 files, processed on 8 Nvidia A6000 GPUs with a batch size of 128. CLEAN was developed based on a supervised contrastive learning framework and trained with a contrastive loss function, extracting vector representations from query protein sequences using ESM-1b and multilayer perceptron. CLEAN characterizes the function of understudied proteins by calculating the pairwise Euclidean distance between query sequence embeddings and all enzymes in each EC number cluster. The EC number of enzymes with notably close distances is assigned to the query protein. Similarly, we split the same 47,951,567 NR gene sequences into 6,000 files, which were processed across 15 Nvidia 3090Ti GPUs with a batch size of 3,000. The total computation time costs 1,800 GPUhours, with an average processing time of 0.24 h per file. Proteins assigned to the target EC numbers were recruited and counted for comparison with results obtained via classic bioinformatics methods (as described under *Taxonomical and Functional Annotation*). The predictive power of these tools was assessed by the number of recruited sequences labeled with the target EC.

Functional identification of plastic-degrading genes

CLEAN detected numerous proteins with the abovementioned EC numbers related to plastic biodegradation, including those undetected by BLAST-based annotation against KEGG. To discover as many diverse and novel plastic-degrading proteins as possible, we collected those proteins detected only by CLEAN for further analysis. To comprehensively understand the function of these protein sequences, we first compared protein sequences longer than 70% of the average length to the Conserved Domain Database using Batch CD-Search (default parameters, https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) to search their protein domains. Protein sequences were discarded if they possessed a domain that (i) was incomplete or (ii) had a function or catalytic reaction unrelated to plastic degradation. Moreover, motif detection on these protein sequences was performed on MEME (v 5.5.4, https://meme-suite.org/meme/tools/meme) with default parameters. Similarly, plastic-degrading proteins accessed from PlasticDB were processed in the same way as previously studied protein sequences but without length-based filtering. In sum, according to their domain functions, all of these proteins (obtained via CLEAN model) were reclassified into four categories: (i) proteases, $n = 138$; (ii) amidases, $n = 168$; (iii) carboxylesterases/lipases, $n = 262$; and (iv) hydrolases, $n = 144$, with the majority belonging to the α/β hydrolase superfamily. Accordingly, these four sets of putative plastic-degrading genes were then used for subsequent sequence comparisons and phylogeny generation, as described in the "Genome reconstruction" and "Phylogenetic analysis" sections.

Protein tertiary structure modeling

We aimed to cover each main protein family when selecting representative protein sequences for tertiary and superposition analysis. These representative proteins possessed a close evolutionary correlation with PlasticDB enzymes identified in the same family. Tertiary structures of proteins from PlasticDB could be accessed directly (except for protein ID 00196, which needed modeling owing to the lack of a corresponding entry in the database). SWISS-MODEL (https://swissmodel.expasy.org/) was used to construct the 3D structures. First, query sequence of the representative protein was submitted to SWISS-MODEL to search templates, which were prioritized based on the following criteria: (i) high sequence identity (>40% in this study); (ii) higher coverage; (iii) higher expected global model quality estimation (GMQE) score (ranging from 0 to 1, with higher values indicating higher accuracy of the model). After modeling, quality assessment was performed using SWISS-MODEL to obtain Ramachandran plots and GMQE scores. The Ramachandran plot provides a simple view of the predicted protein conformation. The structural quality of all models was further estimated using (i) SAVESv6.0 (https://saves.mbi.ucla.edu/) for ERRAT, Verify3D and Ramachandran plot analysis; and (ii) SWISS-MODEL tool QMEAN (https://swissmodel.expasy.org/qmean/) to obtain the qualitative model energy analysis distance constraint (QMEANDisCo) global score, which ranges from 0 to 1, with higher values indicating higher expected quality. Model quality assessment details are provided in Table S1. In addition, the validated structures were viewed and superimposed using ChimeraX (https://www.cgl.ucsf.edu/chimerax/) with default parameters. The similarity between the two protein structures was assessed by pairwise sequence alignments and Cα rmsd.

Genome reconstruction

MAGs were generated through single-sample binning. First, an index was built for efficient alignment by mapping clean sequence data against the assembled contigs using Bowtie2 (v2.3.4.1, https://github.com/BenLangmead/bowtie2) (39). Then, MAGs were constructed by binning assembled contigs (length > 1,000 bp) individually using two different algorithms, Semibin2 (v1.5.1) and MetaBAT2 (v2.15). Refinement of MAGs was performed using DAStool (v1.1.6) (40) to combine results from both binning tools. CheckM2 (v1.0.12) (41) was used to estimate the completeness and contamination of MAGs. MAGs with completeness > 90% and contamination < 5% were considered high-quality genomes. This process resulted in 1,099 high-quality genomes from MetaBin2 and 1,220 from SemiBin2. These high-quality genomes were dereplicated using dRep (v3.4.4) (42) with default parameters (i.e. the MAGs were dereplicated at the species level). Finally, 1,537 NR high-quality MAGs were obtained for downstream analyses.

MAG taxonomic and functional analysis

The 1,537 high-quality MAGs were taxonomically classified using GTDB-Tk (v2.3.2, https://github.com/Ecogenomics/GTDBTk) (43) against the Genome Taxonomy Database (GTDB, https://gtdb.ecogenomic.org/). Subsequently, the relative abundance of each MAG was calculated using CoverM (v0.6.1, https://github.com/wwood/CoverM) with the "Relative Abundance" method. Briefly, the relative abundance of each MAG was calculated by the percentage of mapped reads based on the ratios of its mean coverage in the sample (see the equation below, where A means the relative abundance of each MAG):

$$A_{\%} = \frac{\text{The mean coverage of each MAG}}{\text{The sum of each MAGs' mean coverage}} \times \frac{\text{Number of mapped reads}}{\text{Number of total reads}}$$

Identification of plastic-degrading genes hosted in each MAG was performed using DIAMOND (with the same operation parameters as described in section 2.2.1). Afterward, we compared ORFs within each MAG against PlasticDB and the putative plastic-degrading

gene sets (the four categories of plastic-degrading protein genes) to retrieve the plastic-degrading microbial species. As result, 22 MAGs carried plastic-degrading genes annotated from PlasticDB, while 150 MAGs carried plastic-degrading genes predicted via the CLEAN model.

Phylogenetic analysis

We generated four phylograms corresponding to the four sets of proteins previously categorized. Sequences were aligned with MUSCLE v5.1 (44), and the alignment was curated to remove poorly aligned positions using the trimAL wrapper in TBtools v2.027 (45). Before tree construction, the best-fit model was selected using IQ-tree v1.6.12 (parameters: -m MFP -bb 1000 -bnni) (46), and the maximum-likelihood phylogenetic tree for each category of protein was calculated; the adjusted bootstrap values were calculated to validate the certainty of each tree. Additionally, we analyzed the phylogeny of 150 MAGs possessing putative plastic-degrading genes using PhyloPhlAn 3.0 (47) (parameters: -d phylophlan -diversity low -fast -f 02_tol.cfg) to generate an alignment file for 149 species proteomes (1 was discarded owing to <100 mapped markers). The tree file was generated with a bootstrap value via IQ-tree (parameters: -m LG -bb 1000 -bnni). Phylogeny was conducted using the BIONJ algorithm and the LG model on the 149 MAGs based on their proteome alignment. Bootstrap values were calculated using the optimized ultrafast bootstrap approximation with 1,000 replicates. Herein, phylogenetic trees were visualized using iTOL (v 6.8, https://itol.embl.de/).

Statistical analyses

Data statistics and visualization were performed using GraphPad Prism 9.0 (https://www.graphpad.com/) and R (v 4.2.3) (https://www.r-project.org/). The relative abundance of plastic-degrading proteins was normalized using logarithmic normalization and visualized with the R package "pheatmap." A bubble plot based on MAG relative abundance was generated with R package "ggplot2." A combination of the Kruskal–Wallis test and Dunn's multiple comparison test was used for the statistical analysis in GraphPad Prism 9.0 (https://www.graphpad.com/).

Acknowledgments

The CLEAN software used in this study was developed by the Department of Chemical and Biomolecular Engineering at the University of Illinois Urbana-Champaign. The authors thank Y. Zhou for valuable guidance on proteomic structure and function. They thank the reviewers for their suggestions and constructive comments on this manuscript.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by the National Nature Science Foundation of China (grant no. 52000016 for S.Y. and grant no. 51878642 for L.S.).

Author Contributions

Conceptualization: X.L., F.J., C.X., and L.S. Methodology: X.L., R.Z., F.-Q.C., W.H., X.S., and L.S. Data curation: X.L. Software: F.-Q.C. Investigation: W.H., S.Y., and X.S. Visualization: X.L., W.H., and S.Y. Project administration: S.Y. and L.S. Supervision: S.Y. and L.S. Writing—original draft: X.L., R.Z., F.-Q.C., W.H., and L.S. Writing—review & editing: S.Y., F.J., C.X., and L.S.

Data Availability

All data are available in the main text and Supplementary Material. Other datasets are available in a public repository (https://doi.org/10.17605/OSF.IO/ABH38), including the following: (i) NR protein-coding gene set from Prodigal, (ii) prediction results of protein with target EC number by KofamKOALA, ProteInfer, and CLEAN, (iii) KEGG annotation result, (iv) taxonomical annotation results of the identities between MAGs and their closest taxonomies, and (v) RPKM value of each NR gene.

References

1 Brahney J, Hallerud M, Heim E, Hahnenberger M, Sukumaran S. 2020. Plastic rain in protected areas of the United States. *Science*. 368(6496):1257–1260.
2 OECD. *Global plastics outlook: economic drivers, environmental impacts and policy options*. OECD Publishing, Paris, 2022.
3 Wang Y, *et al.* 2024. Methane emissions from landfills differentially underestimated worldwide. *Nat Sustain*. 7(4):496–507.
4 Imhof HK, Ivleva NP, Schmid J, Niessner R, Laforsch C. 2013. Contamination of beach sediments of a subalpine lake with microplastic particles. *Curr Biol*. 23(19):R867–R868.
5 Bruna A. 2022. Data from "Global plastic waste flow 2019." https://www.statista.com/statistics/1357641/plastic-waste-lifecycle-worldwide/
6 He PJ, Chen LY, Shao LM, Zhang H, Lü F. 2019. Municipal solid waste (MSW) landfill: a source of microplastics?—evidence of microplastics in landfill leachate. *Water Res*. 159:38–45.
7 Rillig MC. 2012. Microplastic in terrestrial ecosystems and the soil? *Environ Sci Technol*. 46(12):6453–6454.
8 McCormick A, Hoellein TJ, Mason SA, Schluep J, Kelly JJ. 2014. Microplastic is an abundant and distinct microbial habitat in an urban river. *Environ Sci Technol*. 48(20):11863–11871.
9 Su Y, *et al.* 2019. Occurrence of microplastics in landfill systems and their fate with landfill age. *Water Res*. 164:114968.
10 Flaws J, Damdimopoulou P, Patisaul HB, Gore A, Raetzman L, Vandenberg LN. *Plastics, EDCs & health. A guide for public interest organizations and policymakers on endocrine disrupting chemicals & plastics*. Endocrine Society, 2020.
11 Kaushal J, Khatri M, Arya SK. 2021. Recent insight into enzymatic degradation of plastics prevalent in the environment: a mini-review. *Clean Eng Technol*. 2:100083.
12 Cho JY, *et al.* 2021. Polyhydroxyalkanoates (PHAs) degradation by the newly isolated marine Bacillus sp. JY14. *Chemosphere*. 283: 131172.
13 Meng X, *et al.* 2021. Protein engineering of stable IsPETase for PET plastic degradation by Premuse. *Int J Biol Macromol*. 180:667–676.
14 Li C, *et al.* 2024. Ecology and risks of the global plastisphere as a newly expanding microbial habitat. *Innovation (Camb)*. 5(1):100543.
15 An Y, *et al.* 2022. Microbial mercury methylation potential in a large-scale municipal solid waste landfill, China. *Waste Manag*. 145:102–111.
16 Kjeldsen P, *et al.* 2002. Present and long-term composition of MSW landfill leachate: a review. *Crit Rev Environ Sci Technol*. 32(4):297–336.
17 Hanson JL, Yesiller N, Kendall LA. 2005. Integrated temperature and gas analysis at a municipal solid waste landfill. Proceedings of the 16th International Conference on Soil

Mechanics and Geotechnical Engineering (Osaka). Millpress Science Publishers/IOS Press. p. 2265–2268.
18 Yang S, Li L, Peng X, Song LY. 2021. Leachate microbiome profile reveals bacteria, archaea and eukaryote dynamics and methanogenic function during solid waste decomposition. *Bioresour Technol*. 320(Pt A):124359.
19 Yang S, Song L. 2019. Succession of bacterial community structure and metabolic function during solid waste decomposition. *Bioresour Technol*. 291:121865.
20 Lin X, *et al*. 2023. A landfill serves as a critical source of microplastic pollution and harbors diverse plastic biodegradation microbial species and enzymes: study in large-scale landfills, China. *J Hazard Mater*. 457:131676.
21 Skariyachan S, Taskeen N, Kishore AP, Krishna BV. 2022. Recent advances in plastic degradation—from microbial consortia-based methods to data sciences and computational biology driven approaches. *J Hazard Mater*. 426:128086.
22 Gaytán I, *et al*. 2020. 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.
23 Meyer-Dombard DR, Bogner JE, Malas J. 2020. A review of landfill microbiology and ecology: a call for modernization with 'Next Generation' technology. *Front Microbiol*. 11:1127.
24 Yu TH, *et al*. 2023. Enzyme function prediction using contrastive learning. *Science*. 379(6639):1358–1363.
25 Gambarini V, *et al*. 2022. PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database (Oxford)*. 2022:baac008.
26 Sanderson T, Bileschi ML, Belanger D, Colwell LJ. 2023. ProteInfer, deep neural networks for protein functional inference. *Elife*. 12: e80942.
27 Heumann S, *et al*. 2009. A novel aryl acylamidase from *Nocardia farcinica* hydrolyses polyamide. *Biotechnol Bioeng*. 102(4):1003–1011.
28 Roodveldt C, Tawfik DS. 2005. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry*. 44(38):12728–12736.
29 2023. On the plastics crisis. *Nat Sustain*. 6(10):1137–1137.
30 Liu X, Yang S, Wang Y, Zhao H-P, Song L. 2018. Metagenomic analysis of antibiotic resistance genes (ARGs) during refuse decomposition. *Sci Total Environ*. 634:1231–1237.
31 De la Cruz FB, Cheng Q, Call DF, Barlaz MA. 2021. Evidence of thermophilic waste decomposition at a landfill exhibiting elevated temperature regions. *Waste Manag*. 124:26–35.
32 Hong HS, *et al*. 2023. Discovery and rational engineering of PET hydrolase with both mesophilic and thermophilic PET hydrolase properties. *Nat Commun*. 14(1):4556.
33 Omura T, *et al*. 2024. Microbial decomposition of biodegradable plastics on the deep-sea floor. *Nat Commun*. 15(1):568.
34 Zrimec J, Kokina M, Jonasson S, Zorrilla F, Zelezniak A. 2021. Plastic-degrading potential across the global microbiome correlates with recent pollution trends. *mBio*. 12(5):e0215521.
35 Lee K, Jing Y, Wang Y, Yan N. 2022. A unified view on catalytic conversion of biomass and waste plastics. *Nat Rev Chem*. 6(9):635–652.
36 Ransom-Jones E, McCarthy AJ, Haldenby S, Doonan J, McDonald JE. 2017. Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. *mSphere*. 2(4):e00300-17.
37 Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem*. 79:471–505.
38 Aramaki T, *et al*. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 36(7):2251–2252.
39 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–354.
40 Sieber CMK, *et al*. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol*. 3(7):836–843.
41 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 25(7):1043–1055.
42 Olm MR, Brown CT, Brooks B, Banfield JF. 2017. Drep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 11(12):2864–2868.
43 Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 36(6):1925–1927.
44 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.
45 Chen CJ, *et al*. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 13(8):1194–1202.
46 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
47 Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun*. 4:2304.